Simulation Based Estimation of Discrete Choice Models

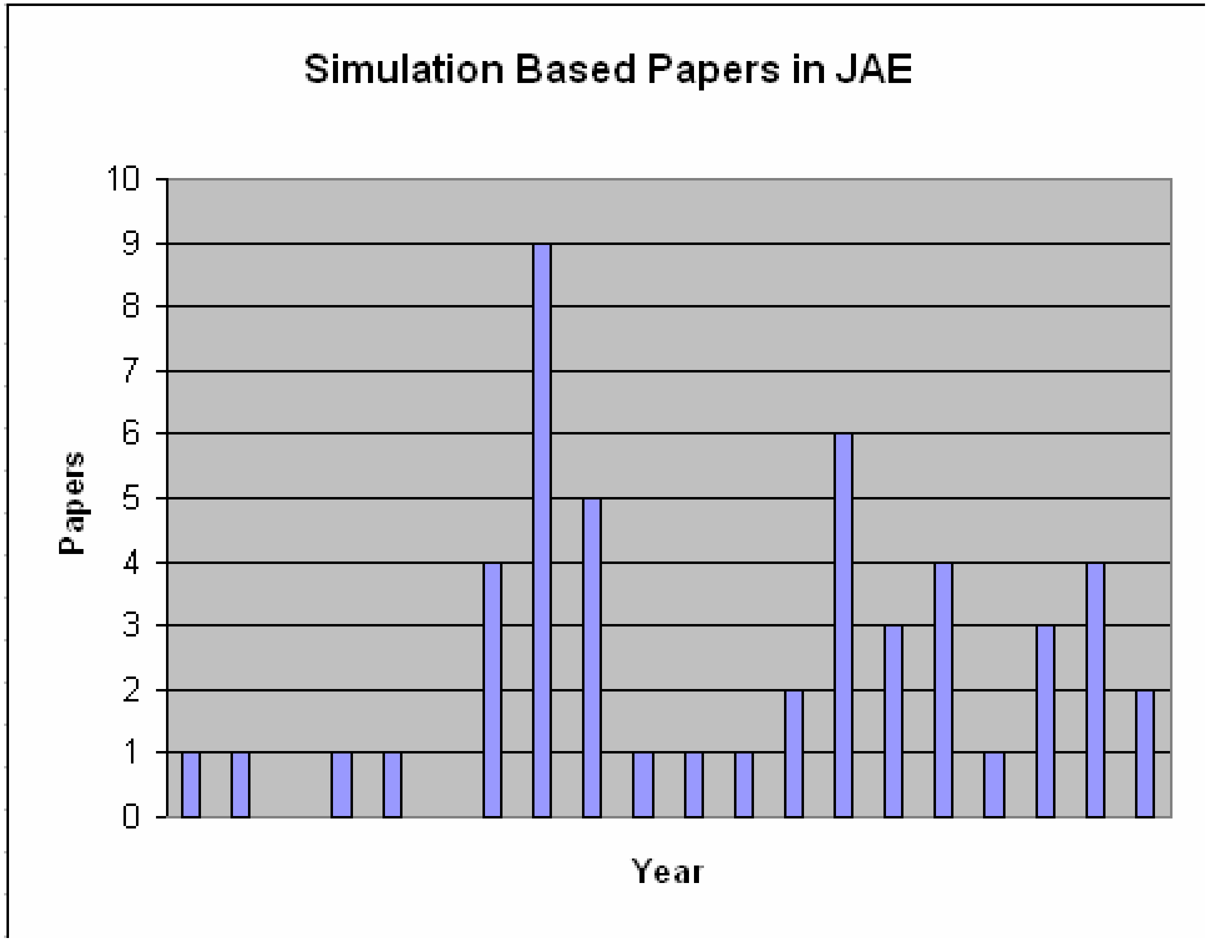
William Greene
Department of Economics
Stern School of Business
New York University

Camp Econometrics. Saratoga Springs, 2006

Simulation Based Estimation

- History:
  - 1981: Lerman and Manski, Multivariate normal probabilities
  - 1984: GHK simulator and the multinomial probit model (vastly oversold)
  - 1996 Gourieroux and Monfort: Simulation Based Econometrics
  - 1996 – present: Trivedi, Train, Greene, et. al. applications
  - 2003: Train, Discrete Choice Methods with Simulation
- Wave of the future?
  - A census of the JAE
  - Outside Economics?

Journal of Applied Econometrics



Agenda

- Methodology
- Maximum simulated likelihood
- Contrast to Bayesian MCMC estimation
- Mechanics of MSL, MCMC and ML
- Applications - several
  - Panel model for counts: GEE (whiz)
  - Health care – binary choice, counts
  - Random parameters multinomial logit
  - Logit kernels and bank distress

Simulation Based Estimation

- Range of applications
  - Microeconometric applications
  - Typically discrete choice
  - No limitation as such
- Typical implementation – modeling unobserved heterogeneity

The Conditional Density

- Data Structure
  - Panel: $\mathbf{y}_i = (y_{i1},\ldots,y_{i,Ti})'$
  - Multivariate: $\mathbf{y}_i = (s_i, r_i, h_i)'$
  - Both…
- Unobserved heterogeneity: $\mathbf{w}_i$.
- Conditional density
  - $f_i(\mathbf{y}) = f(\mathbf{y}_i \mid \mathbf{X}_i, \mathbf{w}_i, \boldsymbol{\beta})$ conditioned on $\mathbf{w}_i$

Log Likelihood

Conditional

$$\log L \mid \{(\mathbf{w}_i), i=1,\ldots,N\} = \sum_{i=1}^{N} \log f(\mathbf{y}_i \mid \mathbf{X}_i, \mathbf{w}_i, \boldsymbol{\beta})$$

Unconditional

$$\log L(\boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{i=1}^{N} \log \int_{\mathbf{w}_i} f(\mathbf{y}_i \mid \mathbf{X}_i, \mathbf{w}_i, \boldsymbol{\beta}) p(\mathbf{w}_i \mid \boldsymbol{\theta}) d\mathbf{w}_i$$

To be maximized over $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ (parameters of marginal distribution of $\mathbf{w}_i$).

Simulated Log Likelihood

The unconditional log likelihood is an expected value

$$\log L(\boldsymbol{\beta},\boldsymbol{\theta}) = \sum_{i=1}^{N} \log \int_{\mathbf{w}_i} f(\mathbf{y}_i \mid \mathbf{X}_i, \mathbf{w}_i, \boldsymbol{\beta}) p(\mathbf{w}_i \mid \boldsymbol{\theta}) d\mathbf{w}_i$$

$$= \sum_{i=1}^{N} \log E_{\mathbf{w}_i}\left[f(\mathbf{y}_i \mid \mathbf{X}_i, \mathbf{w}_i, \boldsymbol{\beta})\right]$$

The expected value can be estimated by random sampling

$$\log L^S(\boldsymbol{\beta},\boldsymbol{\theta}) = \sum_{i=1}^{N} \log \frac{1}{R} \sum_{r=1}^{R} \left[f(\mathbf{y}_i \mid \mathbf{X}_i, \mathbf{w}_{ir}, \boldsymbol{\beta})\right]$$

By the law of large numbers,

$$\log L^S(\boldsymbol{\beta},\boldsymbol{\theta}) \xrightarrow{p} \log L(\boldsymbol{\beta},\boldsymbol{\theta})$$

Sample of draws from the population $\mathbf{w}_i$.

Properties of the MSLE

- Assume
  - Likelihood is “regular”
  - MLE achieves the usual desirable properties
  - Smoothness with respect to the heterogeneity
- Simulation (For useful details, see M&T)
  - Rate of increase in draws is faster than $N^{1/2}$ (to be revisited below)
- Same asymptotic properties as MLE (See Munkin and Trivedi (1999) – cited below.)

A Random Parameters Model

Random parameters model

$f(\mathbf{y}_i \mid \mathbf{X}_i, \mathbf{w}_i, \boldsymbol{\beta}) = p(\mathbf{y}_i \mid \mathbf{X}_i, \boldsymbol{\beta}_\mathbf{i})$

$\boldsymbol{\beta}_\mathbf{i} = \boldsymbol{\beta}^\mathbf{0} + \mathbf{w}_i$

Partially observed heterogeneity

$E[\mathbf{w}_i] = \boldsymbol{\Delta}\mathbf{z}_i, \quad Var[\mathbf{w}_i] = \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}'$

It is assumed that $\mathbf{w}_i \perp \mathbf{X}_i$. Perhaps a Mundlak correction in a panel, $\mathbf{z}_i = \bar{\mathbf{x}}_i$

Structural parameters

$\mathbf{w}_i = \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i \quad \text{where} \quad E[\mathbf{v}_i] = \mathbf{0}, \; Var[\mathbf{v}_i] = \mathbf{I}$

Simulated Log Likelihood for a Panel

Simulated Log Likelihood

$$\log L^S(\boldsymbol{\beta^0}, \boldsymbol{\Delta}, \boldsymbol{\Gamma}) = \sum_{i=1}^{N} \log \frac{1}{R} \sum_{r=1}^{R} f(\mathbf{y}_i \mid \mathbf{X}_i, \mathbf{w}_{ir}, \boldsymbol{\beta})$$

$$= \sum_{i=1}^{N} \log \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_i} f[y_{it} \mid \mathbf{x}_{it}, (\boldsymbol{\beta^0} + \boldsymbol{\Delta}\mathbf{z_i} + \boldsymbol{\Gamma}\mathbf{v_{ir}})]$$

Simulation is over draws of $\mathbf{v}_{ir}$. Standard Normal, or other...

$\log L^S(\boldsymbol{\beta^0}, \boldsymbol{\Delta}, \boldsymbol{\Gamma})$ is maximized using conventional gradient methods

RP Binomial Logit Model

$$\log L^S(\boldsymbol{\beta}^0, \boldsymbol{\Delta}, \boldsymbol{\Gamma}) = \sum_{i=1}^{N} \log \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_i} f[y_{it} \mid \mathbf{x}_{it}, (\boldsymbol{\beta}^0 + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_{ir})]$$

$$f[y_{it} \mid \mathbf{x}_{it}, (\boldsymbol{\beta}^0 + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_{ir})] = \frac{\exp[(2y_{it} - 1)\mathbf{x}_{it}'(\boldsymbol{\beta}^0 + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_{ir})]}{1 + \exp[(2y_{it} - 1)\mathbf{x}_{it}'(\boldsymbol{\beta}^0 + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_{ir})]}$$

An Application

Data Source: Regina T. Riphahn, Achim Wambach, and Andreas Million, "Incentive Effects in the Demand for Health Care: A Bivariate Panel Count Data Estimation", Journal of Applied Econometrics, Vol. 18, No. 4, 2003, pp. 387-405.

German Health Care Usage Data (JAE Data Archive)

N = 7,293 Individuals,

Unbalanced Panel

T_i : 1=1525, 2=2158, 3=825, 4=926, 5=1051, 6=1000, 7=987

There are altogether 27,326 observations

Variables in the file are

HHNINC = household nominal monthly net income in German marks / 10000.
(4 observations with income=0 were dropped)
HHKIDS = children under age 16 in the household = 1; otherwise = 0
EDUC = years of schooling
MARRIED = marital Status
DOCTOR = Number of doctor visits > 0.

Estimated Random Parameters Logit Model for DOCTOR

	Fixed	Random Means	Random Std.Devs.
Constant	1.239 (.068)	1.185 (.057)	1.504 (.020)
HHNINC	-.191 (.075)	.045 (.063)	.441 (.038)
HHKIDS	-.429 (.027)	-.369 (.022)	.298 (.024)
EDUC	-.058 (.006)	-.057 (.005)	.013 (.001)
MARRIED	.251 (.031)	.191 (.026)	.045 (.017)

Log Likelihood -17804.1 -16408.7

Note: In this model, Γ = I and Δ = 0.

Bayesian Estimation

Random Parameters? Not "randomly distributed across observations."

What is "randomness" in this context?

The posterior mean is a conditional mean

$$\text{Posterior density for } \boldsymbol{\beta} = \frac{L(\text{Data}|\boldsymbol{\beta})\text{Prior}(\boldsymbol{\beta})}{p(\text{Data})}$$

$$= \frac{L(\text{Data}|\boldsymbol{\beta})\text{Prior}(\boldsymbol{\beta})}{\int_{\boldsymbol{\beta}} L(\text{Data}|\boldsymbol{\beta})\text{Prior}(\boldsymbol{\beta})d\boldsymbol{\beta}}$$

Bayesian estimator is the posterior mean:

$$= \frac{\int_{\boldsymbol{\beta}} \boldsymbol{\beta}\, L(\text{Data}|\boldsymbol{\beta})\text{Prior}(\boldsymbol{\beta})d\boldsymbol{\beta}}{\int_{\boldsymbol{\beta}} L(\text{Data}|\boldsymbol{\beta})\text{Prior}(\boldsymbol{\beta})d\boldsymbol{\beta}}$$

Bayesian Estimation

Posterior density via Bayes Theorem

$$p(\boldsymbol{\beta}\,|\,\textbf{Data}) = \frac{p(\text{Data}|\boldsymbol{\beta})\text{Prior}(\boldsymbol{\beta})}{p(\text{Data})} = \frac{p(\text{Data},\boldsymbol{\beta})}{p(\text{Data})} = \frac{p(\text{Data},\boldsymbol{\beta})}{\int_{\boldsymbol{\beta}} p(\text{Data},\boldsymbol{\beta})d\boldsymbol{\beta}}$$

$$= \frac{L(\text{Data}|\boldsymbol{\beta})\text{Prior}(\boldsymbol{\beta})}{\int_{\boldsymbol{\beta}} L(\text{Data}|\boldsymbol{\beta})\text{Prior}(\boldsymbol{\beta})d\boldsymbol{\beta}} \quad \text{(Note: Fn of L and Prior)}$$

Posterior mean

$$\hat{\boldsymbol{\beta}} = \frac{\int_{\boldsymbol{\beta}} \boldsymbol{\beta}\, L(\text{Data}|\boldsymbol{\beta})\text{Prior}(\boldsymbol{\beta})d\boldsymbol{\beta}}{\int_{\boldsymbol{\beta}} L(\text{Data}|\boldsymbol{\beta})\text{Prior}(\boldsymbol{\beta})d\boldsymbol{\beta}}$$

MCMC (Gibbs sampling) estimator

$$\hat{\boldsymbol{\beta}}_{\mathbf{MCMC}} = \frac{1}{R}\sum_{r=1}^{R} \boldsymbol{\beta}_{\mathbf{r}} \,|\, \text{posterior density}$$

By the law of large numbers, $\hat{\boldsymbol{\beta}}_{\mathbf{MCMC}} \xrightarrow{P} E[\boldsymbol{\beta}\,|\,\textbf{Data}]$

MCMC vs. MSL

Neglecting the interpretative difference - in purely numerical terms,

With noninformative (flat) priors, the posterior mean is

$$\hat{\boldsymbol{\beta}} = \frac{\int_{\boldsymbol{\beta}} \boldsymbol{\beta}\, L(\text{Data}|\boldsymbol{\beta})\text{Prior}(\boldsymbol{\beta})d\boldsymbol{\beta}}{\int_{\boldsymbol{\beta}} L(\text{Data}|\boldsymbol{\beta})\text{Prior}(\boldsymbol{\beta})d\boldsymbol{\beta}} = \frac{\int_{\boldsymbol{\beta}} \boldsymbol{\beta}\, L(\text{Data}|\boldsymbol{\beta})d\boldsymbol{\beta}}{\int_{\boldsymbol{\beta}} L(\text{Data}|\boldsymbol{\beta})d\boldsymbol{\beta}} = \int_{\boldsymbol{\beta}} \boldsymbol{\beta}\, p(\boldsymbol{\beta}\,|\,\text{Data})d\boldsymbol{\beta}$$

- The Bayesian estimator is the mean of the population of vectors whose distribution is defined by the conditional (on the data) likelihood.

- The MLE is the vector which sits at the mode of the same likelihood.

Why do MCMC and MSL differ numerically?

- Pure sampling variability – they are different statistics
- Asymmetry of the log likelihood function; mean ≠ mode
- Strength of the prior: Posterior ≠ Likelihood
- … only…   by and large, they don't differ.
- A Theorem of Von-Mises: Under weak priors, the "estimators converge" to the same thing, and the posterior variance converges to the inverse of the information matrix.

Application: Panel (RE) Probit

Harris, M., Rogers, M., Siouclis, A., "Modelling Firm Innovation Using Panel Probit Estimators," Melbourne Institute of Applied Economic and Social Research, WP 20/01, 2002.

Sample = 3,757 Australian firms, 3 years:

Random Effects Probit Model: Innovation reported in the survey year

	ML	MCMC
Constant	-1.646 (.052)	-1.646 (.052)
Labor	0.075 (.017)	0.075 (.017)
Lagged Profit Margin	-0.063 (.056)	-0.063 (.056)
Business Plan	0.485 (.042)	0.485 (.042)
Network	0.410 (.044)	0.409 (.043)
Export	0.072 (.050)	0.071 (.051)
Startup Firm	0.005 (.081)	0.001 (.082)
R&D	1.613 (.061)	1.613 (.063)
Correlation	0.380 (.020)	0.381 (.053)

(ML obtained by Hermite quadrature - Butler and Moffitt method.)

The Simulations

- An arms race in the MCMC literature: How many draws for the Gibbs sampler? 5,000? 50,000? There is no theory to define R.
- MSL – Considerations for good properties
  - $R/N^{1/2} \rightarrow \infty$ Many draws needed to invoke LLN
  - High quality random number generators with long periods. (Unclear why these are important.)
  - It's a moot point
    - Randomness is not the issue. Even coverage of the range of integration is what is needed.
    - Halton draws and "intelligent" draws reduce the number of draws needed by a large percent. 90 if K=1 and N is large.

Quadrature and Simulation

General Problem: How to compute

$$\int_{\mathbf{w}_i} f(\mathbf{y}_i \,|\, \mathbf{X}_i, \mathbf{w}_i, \boldsymbol{\beta}) p(\mathbf{w}_i \,|\, \boldsymbol{\theta}) d\mathbf{w}_i$$

Thus far, we have considered simulation and Gibbs sampling.

If $p(\mathbf{w}_i \,|\, \boldsymbol{\theta}) = \text{Exp}(-\Sigma_{k=1}^{K} w_{i,k}^2)$ (Multivariate spherical normal)

then Cartesian Hermite quadrature is a third possibility.

$$\int_{\mathbf{w}_i} f(\mathbf{y}_i \,|\, \mathbf{X}_i, \mathbf{w}_i, \boldsymbol{\beta}) p(\mathbf{w}_i \,|\, \boldsymbol{\theta}) d\mathbf{w}_i$$

$$\approx \sum_{h_K=1}^{H} w_{h_K} \sum_{h_{K-1}=1}^{H} w_{h_{K-1}} \ldots \sum_{h_1=1}^{H} w_{h_1} f(\mathbf{y}_i \,|\, \mathbf{X}_i, z_{h_1}, \ldots z_{h_{K-1}}, z_{h_K}, \boldsymbol{\beta})$$

Note: H^K function evaluations per observation per iteration,...

Discussion

Rabe-Heketh, S., Skrondal, A., Pickles, A., “Maximum Likelihood Estimation of Limited and Discrete Dependent Variable Models with Nested Random Effects,” Journal of Econometrics, 128, 2, October, 2005, pp. 301-323.

Develops an “Adaptive Quadrature” method for these problems.

On Adaptive Quadrature

- “Performs well for large cluster sizes”
- “Gives empirical Bayes predictions for cluster or individual specific random effects and their standard errors” (We will return to this claim.) (A claim usually made by Bayesians, typically followed by an untrue assertion that classical methods cannot do this.)

(Page 320)

On Alternatives

"Computer intensive alternatives to adaptive quadrature include simulation based approaches such as MCMC and MSL."

Rabe-Hesketh on MCMC

- “The hierarchical structure of multilevel models lends itself naturally to MCMC using Gibbs sampling. If vague priors are specified, the method essentially yields maximum likelihood estimates. (We knew that.)
- Quibbles with MCMC estimation
  - How to ensure that a truly stationary distribution has been obtained? (MCMC always works, even when it shouldn’t)
  - There is no diagnostic for assessing empirical identification. (Same issue. How do you know if a problem is not identified?)

Rabe-Hesketh on MSL

- “Regarding simulated maximum likelihood, a merit is that conditional independence specifications implicit in standard multilevel models may be relaxed….
- “Futhermore, unlike methods based on quadrature, simulation methods allow analysis of the approximation error” (See Munkin and Trivedi)
- (No shortcomings are cited)

Quadrature

- Virtues
  - Quite accurate for relatively few nodes
- Vices
  - Excruciatingly slow if K > 2 (Compare hours vs. seconds for simulation or Gibbs sampling)
  - Torturously complicated
  - Limited to normal densities for integrands
  - Limited to independent effects

Why Consider Quadrature?

- The last paragraph of the 21 page paper states "Maximum likelihood estimation and empirical Bayes predictions for all of these models using adaptive quadrature is implemented in **`gllamm`** which runs in Stata"
- Stata does not contain a simulation estimator nor any Gibbs sampling routines.

Parlor Tricks: MSL Estimation of a Panel Poisson Model

How to construct a bivariate Poisson model?

(1) Traditional: Observed Y1 = Z1 + U

Observed Y2 = Z2 + U

Z1, Z2, U all Poisson with means λ_1, λ_2, μ

A bit clumsy and only allows positive covariance μ

(Not good for Y1 = doctor visits, Y2 = hospital visits)

(2) Munkin, M. and Trivedi, P., "Simulated Maximum Likelihood Estimation of Multivariate Mixed Poisson Regression Models, with Application," The Econometrics Journal, 2, 1, 1999, pp. 29-48.

Bivariate Poisson

$$P(y_1 \mid x_1) = \frac{\exp(-\lambda_1)\lambda_1^{y_1}}{\Gamma(y_1 + 1)}, \quad P(y_2 \mid x_2) = \frac{\exp(-\lambda_2)\lambda_2^{y_2}}{\Gamma(y_2 + 1)}$$

$$\lambda_1 = \exp(\beta_1' x_1 + \varepsilon_1), \qquad \lambda_2 = \exp(\beta_2' x_2 + \varepsilon_2)$$

$(\varepsilon_1, \varepsilon_2)$ have a bivariate distribution with nonzero covariance that can be simulated. M&T assume joint normality.

The likelihood does not have a closed form, but the conditional likelihood can be easily simulated. The model is estimated by MSL

Application: Health care utilization:

y1 = emergency room visits

y2 = hospitalizations

A Panel Poisson Model

German Health Care Data: Unbalanced panel, up to 7 periods

$$P(y_{it} \mid x_{it}, \varepsilon_{it}) = \frac{\exp[-(\lambda_{it} \mid \varepsilon_{it})](\lambda_{it} \mid \varepsilon_{it})^{y_{it}}}{\Gamma(y_{it}+1)},$$

$$(\lambda_{it} \mid \varepsilon_{it}) = \exp(\boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}),\ t = 1,...,T_i$$

$(\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{i7}) \sim$ 7 variate normal with free correlation matrix

There are 7 specific dates in the data set. $d_{its} =$ a set of time dummies

$$(\lambda_{it} \mid \varepsilon_{it}) = \exp(\boldsymbol{\beta}'\mathbf{x}_{it} + \delta_{i1}d_{it1} + \delta_{i2}d_{it2} + \delta_{i3}d_{it3} + \delta_{i4}d_{it4} + \delta_{i5}d_{it5} + \delta_{i6}d_{it6} + \delta_{i7}d_{it7})$$

$$\boldsymbol{\delta}_i = \boldsymbol{\delta}^0 + \boldsymbol{\Gamma}\mathbf{v}_i$$

Estimated Models

	Fixed Parameters	Random Constants	Uncorrelated Effects	Random Constants	Correlated Effects
	Means	Means	Standard Deviations	Means	Standard Deviations
HHNINC	-0.683 (.024)	-0.633 (.008)		-0.322 (.011)	
HHKIDS	-0.363 (.007)	-0.369 (.002)		-0.117 (.003)	
EDUC	-0.056 (.002)	-0.065 (.001)		-0.098 (.001)	
MARRIED	0.142 (.009)	0.117 (.003)		0.074 (.004)	
Period 1=1984	2.006 (.022)	1.772 (.010)	1.376 (.006)	2.347 (.012)	1.696
Period 2=1985	1.984 (.022)	1.892 (.009)	1.052 (.005)	2.395 (.012)	1.379
Period 3=1986	2.121 (.022)	1.920 (.0090)	1.046 (.005)	2.541 (.012)	1.590
Period 4=1987	2.064 (.022)	2.208 (.008)	0.909 (.005)	2.426 (.012)	1.404
Period 5=1988	1.936 (.022)	1.889 (.008)	0.910 (.004)	2.482 (.012)	1.191
Period 6=1991	1.933 (.022)	1.897 (.009)	0.761 (.004)	2.350 (.012)	1/045
Period 7=1994	2.285 (.022)	2.293 (.008)	0.707 (.004)	2.449 (.012)	1.316
Log Likelihood	-105,520.20	-82,897.02		-72,467.9	

Cross Period Correlation Matrix

Matrix - COR_BETA

[7, 7] Cell: 1

	1	2	3	4	5	6	7
1	1	0.993387	0.841339	0.870112	0.894636	0.70838	0.594901
2	0.993387	1	0.832583	0.875792	0.924224	0.701134	0.673626
3	0.841339	0.832583	1	0.983707	0.808092	0.91454	0.567072
4	0.870112	0.875792	0.983707	1	0.835362	0.882269	0.677045
5	0.894636	0.924224	0.808092	0.835362	1	0.738865	0.763522
6	0.70838	0.701134	0.91454	0.882269	0.738865	1	0.583798
7	0.594901	0.673626	0.567072	0.677045	0.763522	0.583798	1

Multinomial Logit Model

- Workhorse for empirical analysis of discrete choice among unordered alternatives
- Numerous extensions have been developed for different situations and microeconomic assumptions

A Random Utility Model

Random Utility Model for Discrete Choice Among J alternatives at time t by person i.

$U_{itj} \;=\; \alpha_j \;+\; \beta' x_{itj} \;+\; \varepsilon_{ijt}$

α_j = **Choice specific constant**

x_{itj} = ***Attributes*** **of choice presented to person (<u>Information processing strategy</u>. Not all attributes will be evaluated. E.g., lexicographic utility functions over certain attributes.)**

β = **'Taste weights,' 'Part worths,' marginal utilities**

ε_{ijt} = **Unobserved <u>*random*</u> component of utility**

Mean=E[ε_{ijt}] = 0; Variance=Var[ε_{ijt}] = σ^2

The Multinomial Logit Model

Independent type 1 extreme value (Gumbel):

- $F(\varepsilon_{itj}) = 1 - \text{Exp}(-\text{Exp}(\varepsilon_{itj}))$
- **Independence** across utility functions
- **Identical variances**, $\sigma^2 = \pi^2/6$
- **Same taste parameters** for all individuals

$$\text{Prob}[\text{choice } j \mid i,t] = \frac{\exp(\alpha_j + \beta' x_{itj})}{\sum_{j=1}^{J_t(i)} \exp(\alpha_j + \beta' x_{itj})}$$

Model extensions build around IIA assumptions.

Not our interest here.

Building Individual Effects into Utilities for the MNL Model

Repeated Choice Situations

A counterpart to individual specific, random effects?

$$U_{i,t,j} = \boldsymbol{\beta}'\mathbf{x}_{itj} + \varepsilon_{itj} \quad + \quad \sigma u_i$$

Same effect in all utility functions is undesirable

Choice specific effect is not the objective.

Define a set of $M \le J$ "kernels" $u_{i,1},\ldots,u_{i,M}$

$$U_{i,t,j} = \boldsymbol{\beta}'\mathbf{x}_{itj} + \varepsilon_{itj} \quad + \quad C_{1,1}\sigma_1 u_{i,1} + C_{1,2}\sigma_2 u_{i,2} + \ldots + C_{1,M}\sigma_M u_{i,M}$$

$$U_{i,t,j} = \boldsymbol{\beta}'\mathbf{x}_{itj} + \varepsilon_{itj} \quad + \quad C_{2,1}\sigma_1 u_{i,1} + C_{2,2}\sigma_2 u_{i,2} + \ldots + C_{2,M}\sigma_M u_{i,M}$$

...

$$U_{i,t,j} = \boldsymbol{\beta}'\mathbf{x}_{itj} + \varepsilon_{itj} \quad + \quad C_{J,1}\sigma_1 u_{i,1} + C_{J,2}\sigma_2 u_{i,2} + \ldots + C_{J,M}\sigma_M u_{i,M}$$

$C_{j,m} =$ dummy variable for placing kernel m in utility j

The Kernel Logit Model

- Individual effects in utility functions, not tied to the choice structure
- Provides a stochastic specification that produces the nested logit model – simpler to estimate and always satisfies utility maximization without parameter constraints
- Estimable as a random parameters multinomial logit model with MSL.

Application: Australian Banks

- Hensher, D., Jones, S., Greene, W., “Corporate Bankruptcy and Insolvency Risk in Australia: A Kernel Logit Analysis,” Department of Accounting, University of Sydney, WP, 2006.
- Application
  - Nonfailed; Insolvent; Distressed; Bankrupt and in receivership
  - Unordered choice – MNL model.
  - Quibles: Are these choices ordered? How should this outcome be modeled?

A Complicated Application of MSL: MNL for Shoe Brand Choice

- Simulated Data: Stated Choice, 400 respondents, 8 choice situations
- 3 choice/attributes + NONE
  - Fashion = High / Low
  - Quality = High / Low
  - Price = 25/50/75,100 coded 1,2,3,4
- Heterogeneity: Sex, Age (<25, 25-39, 40+)
- Underlying data **actually** generated by a 3 class latent class process (100, 200, 100 in classes) (True process is known.)
- Thanks to www.statisticalinnovations.com (Latent Gold and Jordan Louviere)

A Discrete (4 Brand) Choice Model

$$U_{i,1,t} = \beta_{F,i}\,Fashion_{i,1,t} + \beta_Q Quality_{i,1,t} + \beta_{P,i}\,Price_{i,1,t} + \lambda_{Brand}K_{i,Brand} + \varepsilon_{i,1,t}$$

$$U_{i,2,t} = \beta_{F,i}\,Fashion_{i,2,t} + \beta_Q Quality_{i,2,t} + \beta_{P,i}\,Price_{i,2,t} + \lambda_{Brand}K_{i,Brand} + \varepsilon_{i,2,t}$$

$$U_{i,3,t} = \beta_{F,i}\,Fashion_{i,3,t} + \beta_Q Quality_{i,3,t} + \beta_{P,i}\,Price_{i,3,t} + \lambda_{Brand}K_{i,Brand} + \varepsilon_{i,3,t}$$

$$U_{i,NONE,t} = \alpha_{NONE} + \lambda_{NONE}K_{i,NONE} + \varepsilon_{i,NONE,t}$$

$$\beta_{F,i} = \overline{\beta}_F + \delta_F Sex_i + [\sigma_F \exp(\gamma_{F1} AgeL25_i + \gamma_{F2} Age2539_i)]\,w_{F,i};\ w_{F,i} \sim N[0,1]$$

$$\beta_{P,i} = \overline{\beta}_P + \delta_P Sex_i + [\sigma_P \exp(\gamma_{P1} AgeL25_i + \gamma_{P2} Age2539_i)]\,w_{P,i};\ w_{P,i} \sim N[0,1]$$

$$K_{Brand,i} \sim N[0,1]$$

$$K_{NONE,i} \sim N[0,1]$$

Random Parameters Model

$$U_{i,1,t} = \beta_{F,i}\,Fashion_{i,1,t} + \beta_Q Quality_{i,1,t} + \beta_{P,i}\,Price_{i,1,t} + \lambda_{Brand} K_{i,Brand} + \varepsilon_{i,1,t}$$

$$U_{i,2,t} = \beta_{F,i}\,Fashion_{i,2,t} + \beta_Q Quality_{i,2,t} + \beta_{P,i}\,Price_{i,2,t} + \lambda_{Brand} K_{i,Brand} + \varepsilon_{i,2,t}$$

$$U_{i,3,t} = \beta_{F,i}\,Fashion_{i,3,t} + \beta_Q Quality_{i,3,t} + \beta_{P,i}\,Price_{i,3,t} + \lambda_{Brand} K_{i,Brand} + \varepsilon_{i,3,t}$$

$$U_{i,NONE,t} = \alpha_{NONE} + \lambda_{NONE} K_{i,NONE} + \varepsilon_{i,NONE,t}$$

$$\beta_{F,i} = \bar{\beta}_F + \delta_F Sex_i + [\sigma_F \exp(\gamma_{F1} AgeL25_i + \gamma_{F2} Age2539_i)]\, w_{F,i};\ w_{F,i} \sim N[0,1]$$

$$\beta_{P,i} = \bar{\beta}_P + \delta_P Sex_i + [\sigma_P \exp(\gamma_{P1} AgeL25_i + \gamma_{P2} Age2539_i)]\, w_{P,i};\ w_{P,i} \sim N[0,1]$$

$$K_{Brand,i} \sim N[0,1]$$

$$K_{NONE,i} \sim N[0,1]$$

Heterogeneous (in the Means) Random Parameters Model

$$U_{i,1,t} = \beta_{F,i}\,Fashion_{i,1,t} + \beta_Q Quality_{i,1,t} + \beta_{P,i}\,Price_{i,1,t} + \lambda_{Brand}K_{i,Brand} + \varepsilon_{i,1,t}$$

$$U_{i,2,t} = \beta_{F,i}\,Fashion_{i,2,t} + \beta_Q Quality_{i,2,t} + \beta_{P,i}\,Price_{i,2,t} + \lambda_{Brand}K_{i,Brand} + \varepsilon_{i,2,t}$$

$$U_{i,3,t} = \beta_{F,i}\,Fashion_{i,3,t} + \beta_Q Quality_{i,3,t} + \beta_{P,i}\,Price_{i,3,t} + \lambda_{Brand}K_{i,Brand} + \varepsilon_{i,3,t}$$

$$U_{i,NONE,t} = \alpha_{NONE} + \lambda_{NONE}K_{i,NONE} + \varepsilon_{i,NONE,t}$$

$$\beta_{F,i} = \bar{\beta}_F + \delta_F Sex_i + [\sigma_F \exp(\gamma_{F1} AgeL25_i + \gamma_{F2} Age2539_i)]\, w_{F,i};\ w_{F,i} \sim N[0,1]$$

$$\beta_{P,i} = \bar{\beta}_P + \delta_P Sex_i + [\sigma_P \exp(\gamma_{P1} AgeL25_i + \gamma_{P2} Age2539_i)]\, w_{P,i};\ w_{P,i} \sim N[0,1]$$

$$K_{Brand,i} \sim N[0,1]$$

$$K_{NONE,i} \sim N[0,1]$$

Heterogeneity in Both Means and Variances

$$U_{i,1,t} = \beta_{F,i}\,\text{Fashion}_{i,1,t} + \beta_Q \text{Quality}_{i,1,t} + \beta_{P,i}\,\text{Price}_{i,1,t} + \lambda_{Brand} K_{i,Brand} + \varepsilon_{i,1,t}$$

$$U_{i,2,t} = \beta_{F,i}\,\text{Fashion}_{i,2,t} + \beta_Q \text{Quality}_{i,2,t} + \beta_{P,i}\,\text{Price}_{i,2,t} + \lambda_{Brand} K_{i,Brand} + \varepsilon_{i,2,t}$$

$$U_{i,3,t} = \beta_{F,i}\,\text{Fashion}_{i,3,t} + \beta_Q \text{Quality}_{i,3,t} + \beta_{P,i}\,\text{Price}_{i,3,t} + \lambda_{Brand} K_{i,Brand} + \varepsilon_{i,3,t}$$

$$U_{i,NONE,t} = \alpha_{NONE} + \lambda_{NONE} K_{i,NONE} + \varepsilon_{i,NONE,t}$$

$$\boxed{\beta_{F,i}} = \bar{\beta}_F + \boxed{\delta_F \text{Sex}_i} + \boxed{[\sigma_F \exp(\gamma_{F1}\text{AgeL25}_i + \gamma_{F2}\text{Age2539}_i)]}\, w_{F,i};\ w_{F,i} \sim N[0,1]$$

$$\boxed{\beta_{P,i}} = \bar{\beta}_P + \boxed{\delta_P \text{Sex}_i} + \boxed{[\sigma_P \exp(\gamma_{P1}\text{AgeL25}_i + \gamma_{P2}\text{Age2539}_i)]}\, w_{P,i};\ w_{P,i} \sim N[0,1]$$

$$K_{Brand,i} \sim N[0,1]$$

$$K_{NONE,i} \sim N[0,1]$$

Individual (Kernel) Effects Model

$$U_{i,1,t} = \beta_{F,i}\,\text{Fashion}_{i,1,t} + \beta_Q\,\text{Quality}_{i,1,t} + \beta_{P,i}\,\text{Price}_{i,1,t} + \lambda_{Brand}K_{i,Brand} + \varepsilon_{i,1,t}$$

$$U_{i,2,t} = \beta_{F,i}\,\text{Fashion}_{i,2,t} + \beta_Q\,\text{Quality}_{i,2,t} + \beta_{P,i}\,\text{Price}_{i,2,t} + \lambda_{Brand}K_{i,Brand} + \varepsilon_{i,2,t}$$

$$U_{i,3,t} = \beta_{F,i}\,\text{Fashion}_{i,3,t} + \beta_Q\,\text{Quality}_{i,3,t} + \beta_{P,i}\,\text{Price}_{i,3,t} + \lambda_{Brand}K_{i,Brand} + \varepsilon_{i,3,t}$$

$$U_{i,NONE,t} = \alpha_{NONE} + \lambda_{NONE}K_{i,NONE} + \varepsilon_{i,NONE,t}$$

$$\beta_{F,i} = \bar{\beta}_F + \delta_F\,\text{Sex}_i + [\sigma_F \exp(\gamma_{F1}\,\text{AgeL25}_i + \gamma_{F2}\,\text{Age2539}_i)]\,w_{F,i};\; w_{F,i} \sim N[0,1]$$

$$\beta_{P,i} = \bar{\beta}_P + \delta_P\,\text{Sex}_i + [\sigma_P \exp(\gamma_{P1}\,\text{AgeL25}_i + \gamma_{P2}\,\text{Age2539}_i)]\,w_{P,i};\; w_{P,i} \sim N[0,1]$$

$$K_{Brand,i} \sim N[0,1]$$

$$K_{NONE,i} \sim N[0,1]$$

Estimated Brand Choice Models

	Fashion	Price	Quality	No Choice	Log L
Multinomial Logit Model					
Main Effects	1.479 (.068)	-11.802 (.804)	1.014 (.064)	0.037 (.072)	-4158.503
Random Parameters / Kernel Logit Model					
Main Effects	1.399 (.130)	-11.978 (1.087)	1.074 (.068)	-0.034 (.086)	-4053.216
Heterogeneity in Means of Random Parameters					
Male	0.529 (.130)	-2.784 (.929)			
Standard Devs.	1.471 (.173)	7.512 (1.155)			
Heteroscedasticity in Variances of Random Parameters					
Age < 25	-0.571 (.170)	-0.710 (.262)			
Age 25-39	-1.743 (1.441)	-1.797 (2.335)			
Kernels – s.d.	Kernel 1: Brand 0.508 (.081)		Kernel 2: No Brand 0.206 (.144)		

“Estimating Individual Effects”

Hierarchical Bayes

$f(y_i|\text{parameters}) = g(\text{data},\boldsymbol{\beta}_i)$

$p(\boldsymbol{\beta}_i) = h(\boldsymbol{\beta},\boldsymbol{\Sigma})$ (Typically normal)

$p(\boldsymbol{\beta}) = \text{Normal}(\mu^0,\boldsymbol{\Sigma}^{\mathbf{0}})$

$p(\boldsymbol{\Sigma}) = \text{Inverse Wishart}(a^0,\mathbf{A}^{\mathbf{0}})$

Individual Estimate? Posterior mean is an estimate of $E[\boldsymbol{\beta}_i|\text{Data}_i]$

Classical Random Parameters

$f(y_i|\text{parameters}) = g(\text{data},\boldsymbol{\beta}_i)$

$\boldsymbol{\beta}_i = \boldsymbol{\beta}^{\mathbf{0}} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i$

density of is $\mathbf{v}_i$ specified...

Individual Estimate? We can estimate $E[\boldsymbol{\beta}_i|\text{Data}_i]$

Neither is an "estimate" of $\boldsymbol{\beta}_i$ any more than $\overline{x}$ is an estimate of $x_{\mathbf{i}}$ in a sample $(x_1,\ldots,x_N)$. It is an estimate of the conditional mean of the population from which $\boldsymbol{\beta}_i$ is a draw.

Simulation Estimation of $E[\boldsymbol{\beta}_i|\text{Data}_i]$

$p(\text{Data}_i \mid \boldsymbol{\beta}_i) =$ Individual contribution to the likelihood, $L(\text{Data}_i \mid \boldsymbol{\beta}_i)$

Marginal $p(\boldsymbol{\beta}_i) =$ Specified in the RP model $= g(\boldsymbol{\beta}_i \mid \boldsymbol{\theta}, \mathbf{z_i})$

$p(\text{Data}_i, \boldsymbol{\beta}_i) = L(\text{Data}_i \mid \boldsymbol{\beta}_i) g(\boldsymbol{\beta}_i \mid \boldsymbol{\theta}, \mathbf{z_i})$

Marginal $p(\text{Data}_i) = \int_{\boldsymbol{\beta}_i} p(\text{Data}_i, \boldsymbol{\beta}_i) d\boldsymbol{\beta}_i = \int_{\boldsymbol{\beta}_i} L(\text{Data}_i \mid \boldsymbol{\beta}_i) g(\boldsymbol{\beta}_i \mid \boldsymbol{\theta}, \mathbf{z_i})\, d\boldsymbol{\beta}_i$

Now, use Bayes Theorem

$$p(\boldsymbol{\beta}_i \mid \text{Data}_i) = \frac{L(\text{Data}_i \mid \boldsymbol{\beta}_i) g(\boldsymbol{\beta}_i \mid \boldsymbol{\theta}, \mathbf{z_i})}{\int_{\boldsymbol{\beta}_i} L(\text{Data}_i \mid \boldsymbol{\beta}_i) g(\boldsymbol{\beta}_i \mid \boldsymbol{\theta}, \mathbf{z_i})\, d\boldsymbol{\beta}_i}$$

The Conditional Mean Estimator

$$E(\boldsymbol{\beta}_i \mid \text{Data}_i) = \frac{\int_{\boldsymbol{\beta}_i} \boldsymbol{\beta}_i\, L(\text{Data}_i \mid \boldsymbol{\beta}_i) g(\boldsymbol{\beta}_i \mid \boldsymbol{\theta}, \mathbf{z_i}) d\boldsymbol{\beta}_i}{\int_{\boldsymbol{\beta}_i} L(\text{Data}_i \mid \boldsymbol{\beta}_i) g(\boldsymbol{\beta}_i \mid \boldsymbol{\theta}, \mathbf{z_i})\, d\boldsymbol{\beta}_i}$$

We can also estimate the conditional variance this way.

This is precisely the counterpart to the hierarchical Bayes estimator

RP Estimator for MNL

Likelihood is formed for the Multinomial Logit Model | $\boldsymbol{\beta}_i$

Marginal density for $\boldsymbol{\beta}_i$ is formed from that of $\mathbf{v}_i$

$$\boldsymbol{\beta}_i = \boldsymbol{\beta}^0 + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i \quad \text{where } \mathbf{v}_i \sim N[\mathbf{0},\mathbf{I}] = f(\mathbf{v}_i)$$

The estimator is

$$E[\boldsymbol{\beta}_i|\text{Data}_i] = \frac{\int_{\mathbf{v}_i} (\boldsymbol{\beta}^0 + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i)\ \text{MNL}[\text{Data}_i \,|\, (\boldsymbol{\beta}^0 + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i)] f(\mathbf{v}_i) d\mathbf{v}_i}{\int_{\mathbf{v}_i} \text{MNL}[\text{Data}_i \,|\, (\boldsymbol{\beta}^0 + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i)] f(\mathbf{v}_i) d\mathbf{v}_i}$$

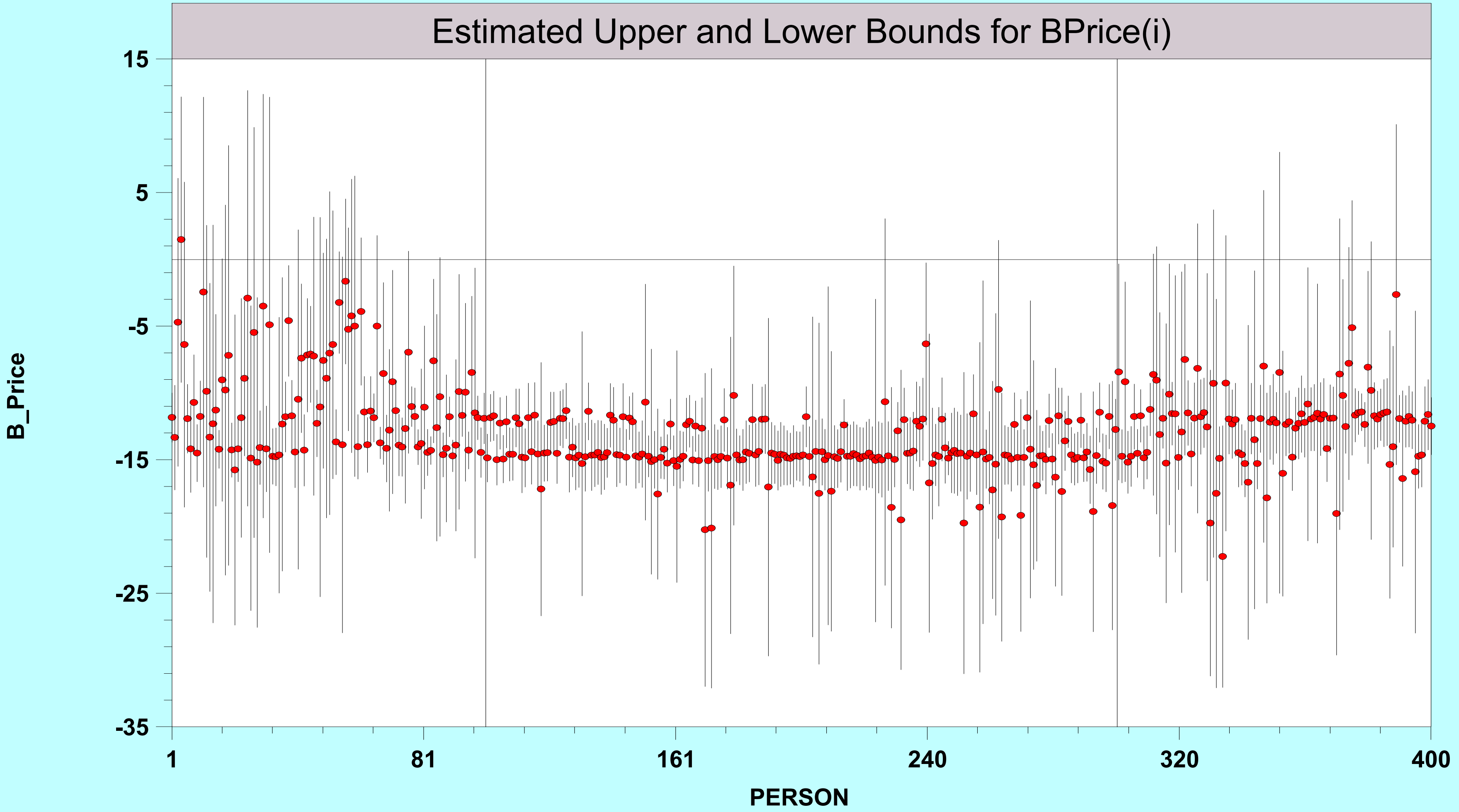
Simulation Estimator

$$E[\boldsymbol{\beta}_i|\text{Data}_i] = \frac{\frac{1}{R}\sum_{r=1}^{R} (\boldsymbol{\beta}^0 + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_{ir})\text{MNL}[\text{Data}_i \,|\, (\boldsymbol{\beta}^0 + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i)]}{\frac{1}{R}\sum_{r=1}^{R} \text{MNL}[\text{Data}_i \,|\, (\boldsymbol{\beta}^0 + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_{ir})]}$$

We do likewise for the expected square and then compute an estimate of $\text{Var}[\boldsymbol{\beta}_i|\text{Data}_i]$

A Caveat: The "plug-in" estimator uses the MLEs of $\boldsymbol{\beta}^0, \boldsymbol{\Delta}\mathbf{z}_i, \boldsymbol{\Gamma}$. Statistical properties may be questionable. The same caveat applies to the hierarchical Bayes estimators.

Individual $E[\beta_i|data_i]$ Estimates



Analyzing Willingness to Pay

Extending the idea of estimating $E[\boldsymbol{\beta}_{\mathbf{i}}|\text{Data}_i]$

Simulation estimator for the mean is

$$\hat{E}[\boldsymbol{\beta}_i|\text{Data}_i] = \frac{\frac{1}{R}\sum_{r=1}^{R}(\hat{\boldsymbol{\beta}}^{\mathbf{0}} + \hat{\boldsymbol{\Delta}}\mathbf{z}_i + \hat{\boldsymbol{\Gamma}}\mathbf{v}_{ir})\text{MNL}[\text{Data}_i \,|\, (\hat{\boldsymbol{\beta}}^{\mathbf{0}} + \hat{\boldsymbol{\Delta}}\mathbf{z}_i + \hat{\boldsymbol{\Gamma}}\mathbf{v}_i)]}{\frac{1}{R}\sum_{r=1}^{R}\text{MNL}[\text{Data}_i \,|\, (\hat{\boldsymbol{\beta}}^{\mathbf{0}} + \hat{\boldsymbol{\Delta}}\mathbf{z}_i + \hat{\boldsymbol{\Gamma}}\mathbf{v}_{ir})]}$$

Extend this to a function of the parameters

$$\hat{E}[g(\boldsymbol{\beta}_i)|\text{Data}_i] = \frac{\frac{1}{R}\sum_{r=1}^{R}g(\hat{\boldsymbol{\beta}}^{\mathbf{0}} + \hat{\boldsymbol{\Delta}}\mathbf{z}_i + \hat{\boldsymbol{\Gamma}}\mathbf{v}_{ir})\text{MNL}[\text{Data}_i \,|\, (\hat{\boldsymbol{\beta}}^{\mathbf{0}} + \hat{\boldsymbol{\Delta}}\mathbf{z}_i + \hat{\boldsymbol{\Gamma}}\mathbf{v}_i)]}{\frac{1}{R}\sum_{r=1}^{R}\text{MNL}[\text{Data}_i \,|\, (\hat{\boldsymbol{\beta}}^{\mathbf{0}} + \hat{\boldsymbol{\Delta}}\mathbf{z}_i + \hat{\boldsymbol{\Gamma}}\mathbf{v}_{ir})]}$$

WTP

Willingness to pay measures in a MNL model:

Usually: $-\frac{\hat{\beta}_{Attribute}}{\hat{\beta}_{Price}}$. Where no price is available, $\frac{\hat{\beta}_{Attribute}}{\hat{\beta}_{Income}}$

$$\text{Estimation}: \frac{\frac{1}{R}\sum_{r=1}^{R} \frac{\hat{\boldsymbol{\beta}}_{Attribute,i}}{\hat{\boldsymbol{\beta}}_{Price,i}} \text{MNL}[\text{Data}_i \,|\, (\hat{\boldsymbol{\beta}}_{ir})]}{\frac{1}{R}\sum_{r=1}^{R} \text{MNL}[\text{Data}_i \,|\, (\hat{\boldsymbol{\beta}}_{ir})]}$$

WTP Application

Greene, W., Hensher, D., Rose, J., “Using Classical Simulation Based Estimators to Estimate Individual WTP Values,” in R. Scarpa and A. Alberini, eds., Applications of Simulation Methods in Environmental and Resource Economics, Springer, 2005

Stated Choice Experiment

- Survey Data
- 223 Commuting Trips in Northwest Sydney
- T = 10 repetitions (scenarios)
- Choices:
  - Existing: Car, bus, train, busway
  - Proposed: light rail, heavy rail, new busway
- Demographics: Age, Hours worked per week, Income, Household size, Number of kids, Gender
- Attributes: Fare, In vehicle time, Waiting time, Access Mode/Time, Egress Time (various)
- Experimental design in several levels for each.

Mixed Logit Model

Simulation is tailored to the application:

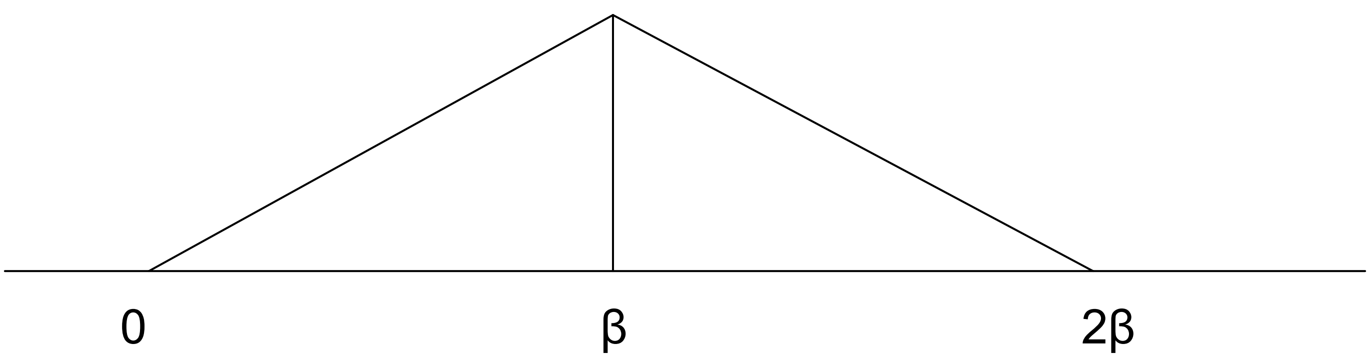
$\beta_{k,i} = \beta_k^0(1 + v_{k,i})$ where $v_{k,i}$ has a "tent" distribution from -1 to +1.

β_k^0 is free.

$v_{k,i} =$ a transformation of a standard uniform U(0,1). Simulation is trivial.

$v_{k,i} = \sqrt{2u_i}$ - 1 if u $\leq$ .5

$v_{k,i} =$ 1 - $\sqrt{2(1-u_i)}$ if u > .5



Value of Travel Time Saved

Willingness to Pay Attribute	VTTS Mean	VTTS Range
Main mode in vehicle time – car	28.80	6.37 - 39.80
Egress time – car	31.61	15.80 - 75.60
Main mode in vehicle time – public transport	15.71	2.60 - 31.50
Waiting time – all bus	19.10	9.90 - 40.30
Access time – all bus	17.10	8.70 - 31.70
Access plus wait time – all rail	10.50	9.60 - 22.90
Egress time – all public transport	3.40	2.00 - 7.40

Observations

- MSL is a fast, flexible method of handling different forms of heterogeneity.
- MCMC (current practice) and MSL are alternative approaches. Statistical "aspects" are a red herring
- Individual effects are equally straightforward with either approach.