Modeling Heterogeneity in Discrete Choice: Recent Developments and Contrasts with Bayesian Estimation

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

Abstract

This study examines some aspects of mixed (random parameters) logit modeling. We present some familiar results in specification and classical estimation of the random parameters model. We then describe several extensions of the mixed logit model developed in recent papers. The relationship of the mixed logit model to Bayesian treatments of the simple multinomial logit model is noted, and comparisons and contrasts of the two methods are described. The techniques described here are applied to two data sets, a stated/revealed choice survey of commuters and one simulated data set on brand choice.

Random Parameters Models of Discrete Choice

- Econometric Methodology for Discrete Choice Models
  - Classical Estimation and Inference
  - Bayesian Methodology
- Model Building Developments
  - The Mixed Logit Model
  - Extensions of the Standard Model
  - Modeling Individual Heterogeneity
  - ‘Estimation’ of Individual Taste Parameters

Useful References

- **Classical**
  - Train, K., *Discrete Choice Methods with Simulation*, Cambridge, 2003. (Train)
  - Hensher, D., Rose, J., Greene, W., *Applied Choice Analysis*, Cambridge, 2005.
  - Hensher, D., Greene, misc. papers, 2003-2005, http://www.stern.nyu.edu/~wgreene
- **Bayesian**
  - Allenby, G., Lenk, P., "Modeling Household Purchase Behavior with Logistic Normal Regression," *JASA*, 1997.
  - Allenby, G., Rossi, P., "Marketing Models of Consumer Heterogeneity," *Journal of Econometrics*, 1999. (A&R)
  - Yang, S., Allenby, G., "A Model for Observation, Structural, and Household Heterogeneity in Panel Data," Marketing Letters, 2000.

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

OHIO STATE

The Multinomial Logit Model

Independent type 1 extreme value (Gumbel):

- **F(εitj) = 1 – Exp(-Exp(εitj))**
- **Independence across utility functions**
- **Identical variances, σ^2 = p²/6**
- **Same taste parameters for all individuals**

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

What's Wrong with this MNL Model?

- I.I.D. → IIA (Independence from irrelevant alternatives)
  - Peculiar behavioral assumption
  - Leads to skewed, implausible empirical results
  - Functional forms, e.g., nested logit, avoid IIA
  - IIA will be a nonissue in what follows.
- Insufficiently heterogeneous:

  "… economists are often more interested in aggregate effects and regard heterogeneity as a statistical nuisance parameter problem which must be addressed but not emphasized. Econometricians frequently employ methods which do not allow for the estimation of individual level parameters." (A&R, 1999)

Heterogeneity

- **Observational: Observable differences across choice makers**
- **Choice strategy: How consumers make decisions**
- **Structure: Model frameworks**
- **Preferences: Model ‘parameters’**

Accommodating Heterogeneity

- Observed? Enter in the model in familiar (and unfamiliar) ways.
- Unobserved? The purpose of this study.

Observable (Quantifiable) Heterogeneity in Utility Levels

$$U_{ijt} = \alpha_j + \beta' x_{itj} + \gamma_j' z_{it} + \varepsilon_{ijt}$$

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

Choice, e.g., among brands of cars

x_{itj} = attributes: price, features

Z_{it} = observable characteristics: age, sex, income

Observable Heterogeneity in Preference Weights

$$U_{ijt} = \alpha_j + \beta_i' x_{itj} + \gamma_j' z_{it} + \varepsilon_{ijt}$$

$$\beta_i = \beta + \Phi h_i$$

$$\beta_{i,k} = \beta_k + \varphi_k' h_i$$

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

‘Quantifiable’ Heterogeneity in Scaling

$$U_{ijt} = \alpha_j + \beta' x_{itj} + \gamma_j' z_{it} + \varepsilon_{ijt}$$

$$\text{Var}[\varepsilon_{ijt}] = \sigma_j^2 \exp(\delta_j' w_{it}), \ \sigma_1^2 = \pi^2/6$$

w_{it} = observable characteristics: age, sex, income, etc.

Attention to Heterogeneity

- Modeling heterogeneity is important
- Scaling is extremely important
- Attention to heterogeneity – an informal survey of four literatures

	Levels	Scaling
Economics	□	None
Education	□	None
Marketing	□	Jordan Louviere
Transport	□	□

Heterogeneity in Choice Strategy

- **Consumers avoid 'complexity'**
  - **Lexicographic preferences eliminate certain choices → choice set may be endogenously determined**
  - **Simplification strategies may eliminate certain attributes**
- **Information processing strategy is a source of heterogeneity in the model.**

Modeling Attribute Choice

- **Conventional: $U_{ijt} = \beta' x_{ijt}$. For ignored attributes, set $x_{k,ijt} = 0$. Eliminates $x_{k,ijt}$ from utility function**
  - **Price = 0 is not a reasonable datum. Distorts choice probabilities**
- **Appropriate: Formally set $\beta_k = 0$**
  - **Requires a 'person specific' model**
  - **Accommodate as part of model estimation**
  - **(Work in progress) Stochastic determination of attribution choices**

Choice Strategy Heterogeneity

- **Methodologically, a rather minor point – construct appropriate likelihood given known information**

$$\log L = \sum_{m=1}^{M} \sum_{i \in M} \log L_i(\boldsymbol{\theta} \mid \text{data}, m)$$

- **Not a latent class model. Classes are not latent.**
- **Not the ‘variable selection’ issue (the worst form of “stepwise” modeling)**
- **Familiar strategy gives the wrong answer.**

Application of Information Strategy

- **Stated/Revealed preference study, Sydney car commuters. 500+ surveyed, about 10 choice situations for each.**
- **Existing route vs. 3 proposed alternatives.**
- **Attribute design**
  - **Original: respondents presented with 3, 4, 5, or 6 attributes**
  - **Attributes – four level design.**
    - **Free flow time**
    - **Slowed down time**
    - **Stop/start time**
    - **Trip time variability**
    - **Toll cost**
    - **Running cost**
  - **Final: respondents use only some attributes <u>and indicate when surveyed which ones they ignored</u>**

OHIO STATE

Estimation Results

Mean values of travel time savings inclusive and exclusive of individuals who ignored specific attributes

time = random parameter, cost = fixed parameter

Attribute	'Zero Fill'	'Model Adjusted'
Free flow time	10.71	9.09
Slowed time	8.25	6.58
Stop/start time	10.61	8.69
Slowed/stop/start time	11.96	8.12

Ratio of non-ignored to ignored mean Value of Travel Time Saved

Attribute	Ratio NI/I
Free flow time	1.18
Slowed time	1.25
Stop/start time	1.22
Slowed/stop/start time	1.62

Note: for all times, running cost is the cost parameter.

OHIO STATE

“Structural Heterogeneity”

- Marketing literature
- Latent class structures
  - Yang/Allenby - latent class random parameters models
  - Kamkura et al – latent class nested logit models with fixed parameters

Latent Classes and Random Parameters

Heterogeneity with respect to 'latent' consumer classes

$$\Pr(\text{Choice}_i) = \sum_{q=1}^{Q} \Pr(\text{choice}_i \mid \text{class} = q)\Pr(\text{class} = q)$$

$$\Pr(\text{choice}_i \mid \text{class} = q) = \frac{\exp(x'_{i,\text{choice}}\beta_{\text{class}})}{\Sigma_{j=\text{choice}}\exp(x'_{i,j}\beta_{\text{class}})}$$

$$\Pr(\text{class} = q \mid i) = F_{i,q},\ \text{e.g., } F_{i,q} = \frac{\exp(z'_i\delta_q)}{\Sigma_{q=\text{classes}}\exp(z'_i\delta_q)}$$

Simple discrete random parameter variation

$$\Pr(\text{choice}_i \mid \beta_i) = \frac{\exp(x'_{i,\text{choice}}\beta_i)}{\Sigma_{j=\text{choice}}\exp(x'_{i,j}\beta_i)}$$

$$\Pr(\beta_i = \beta_q) = F_{i,q} = \frac{\exp(z'_i\delta_q)}{\Sigma_{q=\text{classes}}\exp(z'_i\delta_q)},\ q = 1,\ldots,Q$$

$$\Pr(\text{Choice}_i) = \sum_{q=1}^{Q} \Pr(\text{choice} \mid \beta_i = \beta_q)\Pr(\beta_q)$$

Latent Class Probabilities

- Ambiguous at face value – Classical Bayesian model?
- Equivalent to random parameters models with discrete parameter variation
  - Using nested logits, etc. does not change this
  - Precisely analogous to continuous 'random parameter' models
- Not always equivalent – zero inflation models

Unobserved Preference Heterogeneity

- What is it?
- How does it enter the model?

$$U_{ijt} = \alpha_j + \beta' x_{itj} + \gamma_j' z_{it} + \varepsilon_{ijt} + w_i$$

Random Parameters? **Random 'Effects'?**

Random Parameters?

Stochastic Frontier Models with *Random Coefficients*. M. Tsionas, *Journal of Applied Econometrics*, 17, 2, 2002.

Bayesian analysis of a production model

What do we (does he) mean by "random?"

What Do We Mean by Random Parameters?

- **Classical**
  - Distribution across individuals
  - Model of heterogeneity across individuals
  - Characterization of the population
  - "Superpopulation?" (A&R)

- **Bayesian**
  - Parameter Uncertainty? (A&R) Whose?
  - Distribution defined by a 'prior?' Whose prior? Is it unique? Is one 'right?'
  - Definitely NOT heterogeneity. That is handled by individual specific 'random' parameters in a hierarchical model.

Continuous Random Variation in Preference Weights

$$U_{ijt} = \alpha_j + \beta_i' x_{itj} + \gamma_j' z_{it} + \varepsilon_{ijt}$$

$$\beta_i = \beta + \Phi h_i + w_i$$

$$\beta_{i,k} = \beta_k + \varphi_k' h_i + w_{i,k}$$

Most treatments set $\Phi = 0$

$$\beta_i = \beta + w_i$$

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

Heterogeneity arises from continuous variation in β_i across individuals. (Classical and Bayesian)

What Do We ‘Estimate?’

Classical

$f(\beta_i|\Omega,Z_i)$=population

‘Estimate’ Ω, then

$E[\beta_i|\Omega,Z_i)$=cond’l mean

$V[\beta_i|\Omega,Z_i)$=cond’l var.

Estimation Paradigm

“Asymptotic (normal)”

“Approximate”

“Imaginary samples”

Bayesian

$f(\beta|\Omega^0)$=prior

$L(\text{data}|\ \beta)$=Likelihood

$f(\beta|\text{data}, \Omega^0)$=Posterior

$E(\beta|\text{data}, \Omega^0)$=Posterior Mean

$V(\beta|\text{data}, \Omega^0)$=Posterior Var.

Estimation Paradigm

“Exact”

“More accurate”

(“Not general beyond this prior and this sample…”)

OHIO STATE

How Do We ‘Estimate It?’

- **Objective**
  - **Bayesian: Posterior means**
  - **Classical: Conditional Means**
- **Mechanics: Simulation based estimator**
  - **Bayesian: Random sampling from the posterior distribution. Estimate the mean of a distribution. Always easy.**
  - **Classical: Maximum simulated likelihood. Find the maximum of a function. Sometimes very difficult.**
- **These will look suspiciously similar.**

A Practical Model Selection Strategy

What self contained device is available to suggest that the analyst is fitting the wrong model to the data?

- **Classical: The iterations fail to converge. The optimization otherwise breaks down. The model doesn't 'work.'**
- **Bayesian? E.g., Yang/Allenby Structural/Preference Heterogeneity has both discrete and continuous variation in the same model. Is this identified? How would you know? The MCMC approach is too easy. It always works.**

Bayesian Estimation Platform: The Posterior (to the data) Density

Prior : $f(\beta \mid \Omega^0)$

Likelihood : $L(\beta \mid \text{data}) \; ^{\circ} \; f(\text{data} \mid \beta)$

Joint density : $f(\beta, \text{data} \mid \Omega^0) = L(\beta \mid \text{data}) f(\beta \mid \Omega^0)$

Posterior : $$f(\beta \mid \text{data}, \Omega^0) = \frac{f(\beta, \text{data} \mid \Omega^0)}{f(\text{data})}$$

$$= \frac{L(\beta \mid \text{data}) f(\beta \mid \Omega^0)}{\int_\beta L(\beta \mid \text{data}) f(\beta \mid \Omega^0) d\beta}$$

Posterior density of β given data and prior Ω^0

The Estimator is the Posterior Mean

$$E[\beta \mid data, \Omega^0] = \int_\beta \beta \, f(\beta \mid data, \Omega^0)\, d\beta$$

$$= \int_\beta \beta \left(\frac{L(\beta \mid data) f(\beta \mid \Omega^0)}{\int_\beta L(\beta \mid data) f(\beta \mid \Omega^0) d\beta} \right) d\beta$$

Simulation based (MCMC) estimation: Empirically,

$$\hat{E}[\beta] = \frac{1}{R} \sum_{r=1}^{R} \beta_r \mid \text{known posterior population}$$

This is not 'exact.' It is the mean of a random sample.

Classical Estimation Platform: The Likelihood

$$\text{Marginal}: f(\beta_i \mid \text{data}, \Omega)$$

$$\text{Population Mean} = E[\beta_i \mid \text{data}, \Omega]$$

$$= \int_{\beta_i} \beta_i f(\beta_i \mid \Omega) d\beta_i$$

$$= \overline{\beta} = \text{a subvector of } \Omega$$

$$\hat{\Omega} = \text{Argmax } L(\beta_i, i = 1,\ldots,N \mid \text{data}, \Omega)$$

$$\text{Estimator} = \hat{\overline{\beta}}$$

Expected value over all possible realizations of β_i (according to the estimated asymptotic distribution). I.e., over all possible samples.

Maximum Simulated Likelihood

True log likelihood

$$L_i(\beta_i \mid data_i) = \prod_{t=1}^{T_i} f(data_i \mid \beta_i)$$

$$L_i(\Omega \mid data_i) = \int_{\beta_i} \prod_{t=1}^{T_i} f(data_i \mid \beta_i) f(\beta_i \mid \Omega) d\beta_i$$

$$\log L = \sum_{i=1}^{N} \log \int_{\beta_i} L_i(\beta_i \mid data_i) f(\beta_i \mid \Omega) d\beta_i$$

Simulated log likelihood

$$\log L_S = \sum_{i=1}^{N} \log \frac{1}{R} \sum_{r=1}^{R} L_i(\beta_{iR} \mid data_i, \Omega)$$

$$\hat{\Omega} = \text{argmax}(\log L_S)$$

Individual Parameters

$\beta_i \sim N(\bar{\beta}, \bar{\Omega})$, i.e., $\beta_i = \bar{\beta} + w_i$

$\bar{\beta} \sim N(\beta_0, \Omega_0)$, i.e., $\bar{\beta} = \beta_0 + w_0$

$\bar{\Omega} \sim \text{Inverse Wishart}(G_0, g_0)$

$\hat{\beta}_i = \text{Posterior Mean} = E[\beta_i \mid \text{data}, \beta_0, \Omega_0, G_0, g_0]$

Computed using a Gibbs sample (MCMC)

Estimating β_i

"… In contrast, classical approaches to modeling heterogeneity yield only aggregate summaries of heterogeneity and do not provide actionable information about specific groups. The classical approach is therefore of limited value to marketers." (A&R p. 72)

A Bayesian View: Not All Possible Samples, Just This Sample

Based on any 'classical' random parameters model,

$$E[\beta_i \mid \text{This sample}] = \int_{\beta_i} \beta_i f(\beta_i \mid \text{data}_i, \Omega) d\beta_i$$

$$\boxed{\hat{\bar{\beta}}_i}$$

$$= \text{conditional mean in } f(\beta_i \mid \text{data}_i, \Omega)$$

$$= \int_{\beta_i} \beta_i \left(\frac{L(\beta_i \mid \text{data}_i) f(\beta_i \mid \Omega)}{\int_{\beta} L(\beta_i \mid \text{data}_i) f(\beta_i \mid \Omega) d\beta_i} \right) d\beta_i$$

= conditional mean conditioned on the data observed for individual i.

Looks like the posterior mean

OHIO STATE

THE Random Parameters Logit Model

Random Utility :

$$U_{ijt} = \alpha_{i,j} + \beta_i' x_{itj} + \gamma_{i,j}' z_{it} + \varepsilon_{ijt}$$

Random parameters :

$$\theta_{i,k} = \bar{\theta}_k + \delta_k' w_i + \sigma_k u_{i,k}$$

$$\Theta_i = \bar{\Theta} + \Delta w_i + \Sigma u_i, \ \Sigma \text{ a diagonal matrix}$$

Extensions :

Correlation : Σ = a lower triangular matrix

Autocorrelation : $u_{i,k,t} = \rho u_{i,k,t-1} + v_{i,k,t}$

Variance Heterogeneity : $\sigma_{i,k} = \sigma_k \exp(\gamma_k' f_i)$

Structural parameters : $\Omega = [\bar{\theta}, \Delta, \Sigma, \rho, \Gamma]$

Heterosecedastic Logit Kernels

$U_{ijt} = V_{ijt} + \Sigma_{m=1}^{M} c_{jm} \lambda_{im} K_{im}$

Kernel K_{im} = an individual choice specific effect

$c_{jm} = 1$ or 0

λ_m = scale (standard deviation) for K_{im}

Example

$$U_{i,1,t} = V_{i,1,t} + \boxed{\lambda_{i1} K_{i1}} + \boxed{\lambda_{i2} K_{i2}}$$

$$U_{i,2,t} = V_{i,2,t} + \boxed{\lambda_{i1} K_{i1}} + \lambda_{i3} K_{i3}$$

$$U_{i,3,t} = V_{i,3,t} + \boxed{\lambda_{i2} K_{i2}}$$

$$\boxed{\begin{array}{l} K_{im} \sim N[0, \lambda_{im}^2], \\ \lambda_{im} = \lambda_m \exp(\varphi_m' h_i) \end{array}}$$

Produces a stochastic underpinning for nested logit added to the full random parameters model. More general. Allows nests to overlap.

Much easier to estimate than nested logit

Conditional Estimators

$$\hat{\Omega} = \text{argmax} \sum_{i=1}^{N} \log \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_i} P_{ijt}(\beta_{ir} \mid \Omega, \text{data}_{it})$$

$$\hat{L}_i = \prod_{t=1}^{T_i} P_{ijt}(\hat{\beta}_i \mid \hat{\Omega}, \text{data}_{it})$$

$$\hat{E}[\beta_{i,k} \mid \text{data}_i] = \frac{(1/R)\Sigma_{r=1}^{R} \beta_{i,k,r} \Pi_{t=1}^{T} P_{ijt}(\hat{\beta}_i \mid \hat{\Omega}, \text{data}_{it})}{(1/R)\Sigma_{r=1}^{R} \Pi_{t=1}^{T} P_{ijt}(\hat{\beta}_i \mid \hat{\Omega}, \text{data}_{it})} = \frac{1}{R} \sum_{r=1}^{R} \hat{w}_{i,r} \beta_{i,k,r}$$

$$\hat{E}[\beta_{i,k}^2 \mid \text{data}_i] = \frac{(1/R)\Sigma_{r=1}^{R} \beta_{i,k,r}^2 \Pi_{t=1}^{T} P_{ijt}(\hat{\beta}_i \mid \hat{\Omega}, \text{data}_{it})}{(1/R)\Sigma_{r=1}^{R} \Pi_{t=1}^{T} P_{ijt}(\hat{\beta}_i \mid \hat{\Omega}, \text{data}_{it})} = \frac{1}{R} \sum_{r=1}^{R} \hat{w}_{i,r} \beta_{i,k,r}^2$$

$$\hat{V}ar[\beta_{i,k} \mid \text{data}_i] = \hat{E}[\beta_{i,k}^2 \mid \text{data}_i] - \left\{ \hat{E}[\beta_{i,k} \mid \text{data}_i] \right\}^2$$

$\hat{E}[\beta_{i,k} \mid \text{data}_i] \pm 2\sqrt{\hat{V}ar[\beta_{i,k} \mid \text{data}_i]}$ will encompass 95% of any reasonable distribution

Simulation Based Estimation

- **Bayesian: Limited to convenient priors (normal, inverse gamma and Wishart) that produce mathematically tractable posteriors. Largely simple RPM's without heterogeneity.**
- **Classical: Use any distributions for any parts of the heterogeneity that can be simulated. Rich layered model specifications.**
  - **Comparable to Bayesian (Normal)**
  - **Constrain parameters to be positive. (Triangular, Lognormal)**
  - **Limit ranges of parameters. (Uniform, Triangular)**
  - **Produce particular shapes of distributions such as small tails. (Beta, Weibull, Johnson S_B)**
  - **Heteroscedasticity and scaling heterogeneity**
  - **Nesting and multilayered correlation structures**

Computational Difficulty?

"Outside of normal linear models with normal random coefficient distributions, performing the integral can be computationally challenging." (A&R, p. 62)

(No longer even remotely true)

- MSL with dozens of parameters is simple
- Multivariate normal (multinomial probit) is no longer the benchmark alternative. (See McFadden and Train)
- Intelligent methods of integration (Halton sequences) speed up integration by factors of as much as 10. (These could be used by Bayesians.)

Individual Estimates

- **Bayesian… "exact…" what do we mean by "the exact posterior"**
- **Classical… "asymptotic…"**
- **These will be very similar. Counterpoint is not a crippled LCM or MNP. Same model, similar values.**
- **A theorem of Bernstein-von Mises: Bayesian ------ > Classical as N $\rightarrow\infty$ (The likelihood function dominates; posterior mean $\rightarrow$ mode of the likelihood; the more so as we are able to specify flat priors.)**

OHIO STATE

Application 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 generated by a 3 class latent class process (100, 200, 100 in classes)
- Thanks to www.statisticalinnovations.com (Latent Gold and Jordan Louviere)

A Discrete (4 Brand) Choice Model

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

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

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

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

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

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

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

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

Random Parameters Model

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

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

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

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

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

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

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

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

Heterogeneous (in the Means) Random Parameters Model

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Multinomial Logit Model Estimates

```
+---------------------------------------------+
| Dependent variable                   Choice |
| Log likelihood function         -4158.503   |
| Akaike IC= 8325.006  Bayes IC= 8349.289     |
| R2=1-LogL/LogL*  Log-L fncn  R-sqrd  RsqAdj |
| Constants only   -4391.1804  .05299  .05160 |
+---------------------------------------------+
```

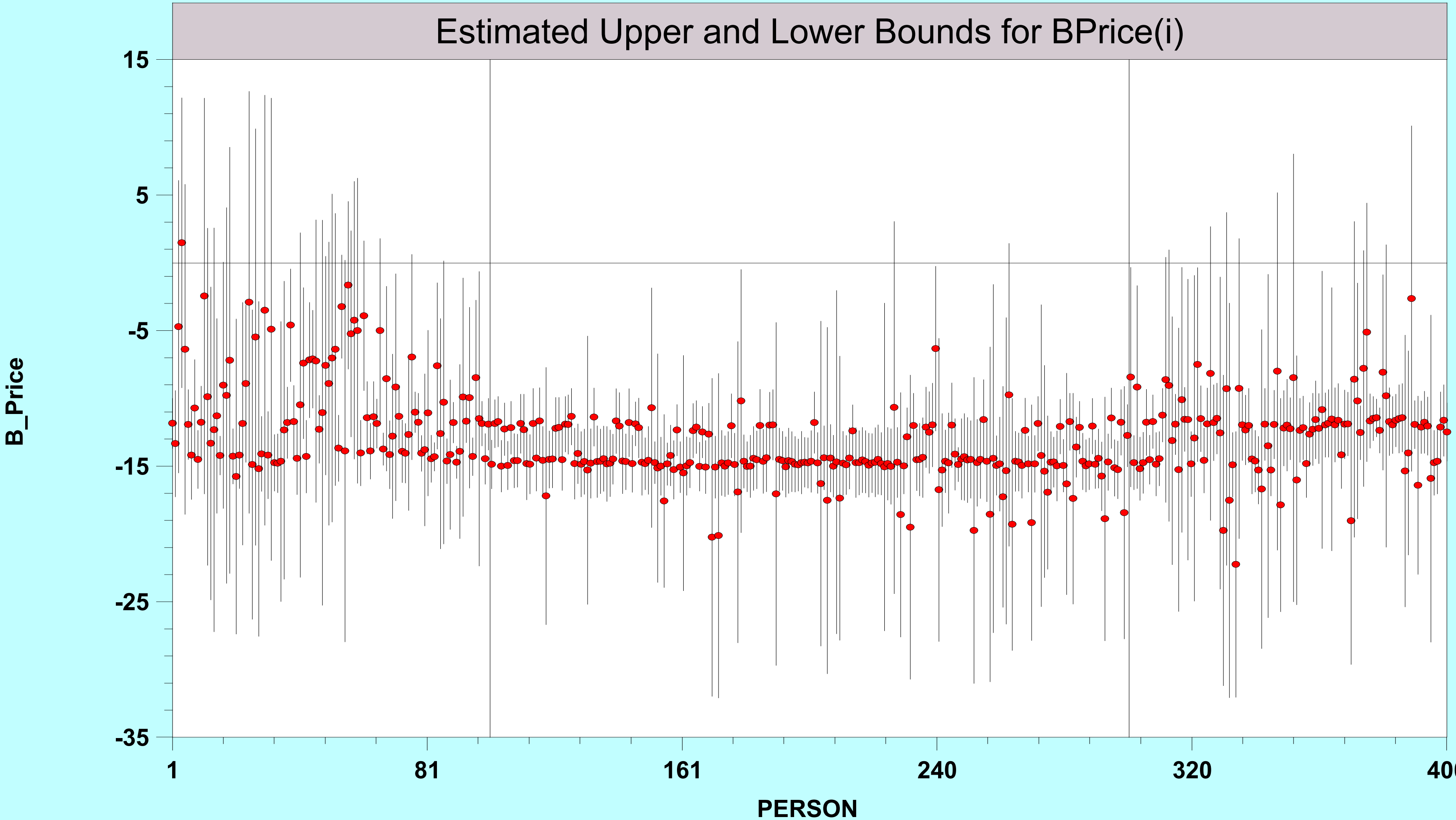
Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]
FASHION	1.47890473	.06776814	21.823	.0000
PRICE	-11.8023376	.80406103	-14.678	.0000
QUALITY	1.01372755	.06444532	15.730	.0000
NONEASC	.03679254	.07176387	.513	.6082

OHIO STATE

Random Parameters/Kernel Logit Model			
Log likelihood function	-4053.216		
Akaike IC= 8134.433 Bayes IC= 8219.426			
Restricted log likelihood	-4436.142		
Chi squared	765.8509		
Degrees of freedom	14		
Prob[ChiSqd > value] =	.0000000		
R2=1-LogL/LogL*	Log-L fncn	R-sqrd	RsqAdj
No coefficients	-4436.1420	.08632	.08499
Constants only	-4391.1804	.07696	.07562
At start values	-4158.5029	.02532	.02389

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]
	Random parameters in utility functions			
FASHION	1.39889187	.13014724	10.749	.0000
PRICE	-11.9782479	1.08721068	-11.017	.0000
	Nonrandom parameters in utility functions			
QUALITY	1.07407105	.06821565	15.745	.0000
NONEASC	-.03395153	.08660849	-.392	.6950
	Heterogeneity in mean, Parameter:Variable			
FASH:MAL	.52976749	.13061519	4.056	.0000
PRIC:MAL	-2.74812084	.92940764	-2.957	.0031
	Derived standard deviations of parameter distributions			
NsFASHIO	1.47073413	.17287203	8.508	.0000
NsPRICE	7.51259238	1.15472257	6.506	.0000
	Heteroscedasticity in random parameters			
sFASH|A1	-.57092895	.17036022	-3.351	.0008
sFASH|A2	-.70991688	.26247499	-2.705	.0068
sPRIC|A1	-1.74346569	1.44125385	-1.210	.2264
sPRIC|A2	-1.79724134	2.33495668	-.770	.4415
	Standard deviations of latent kernel effects			
SigmaK01	.50852209	.08064434	6.306	.0000
SigmaK02	.20560141	.14363266	1.431	.1523

Individual E[β_i|data$_i$] Estimates*



The intervals could be made wider to account for the sampling variability of the underlying (classical) parameter estimators.

Extending the RP Model to WTP

- Use the model to estimate conditional distributions for any function of parameters
- Willingness to pay = $\beta_{i,time} / \beta_{i,cost}$
- Use same method

$$\hat{E}[WTP_i \mid data_i] = \frac{(1/R)\Sigma_{r=1}^{R} WTP_{ir} \Pi_{t=1}^{T} P_{ijt}(\hat{\beta}_{ir} \mid \hat{\Omega}, data_{it})}{(1/R)\Sigma_{r=1}^{R} \Pi_{t=1}^{T} P_{ijt}(\hat{\beta}_{ir} \mid \hat{\Omega}, data_{it})} = \frac{1}{R}\sum_{r=1}^{R} \hat{w}_{i,r} WTP_{ir}$$

OHIO STATE

What is the 'Individual Estimate?'

- Point estimate of mean, variance and range of random variable β_i | data$_i$.
- Value is NOT an estimate of β_i ; it is an estimate of E[β_i | data$_i$]
- What would be the best estimate of the actual realization β_i|data$_i$?
- An interval estimate would account for the sampling 'variation' in the estimator of Ω that enters the computation.
- Bayesian counterpart to the preceding? Posterior mean and variance? Same kind of plot could be done.

Methodological Differences

Focal point of the discussion in the literature is the simplest possible MNL with random coefficients,

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

$$\begin{pmatrix} \alpha_{i,j} \\ \beta_i \end{pmatrix} = \begin{pmatrix} \overline{\alpha}_j \\ \overline{\beta} \end{pmatrix} + \begin{pmatrix} w_{i,\alpha j} \\ w_i \end{pmatrix}$$

This is far from adequate to capture the forms of heterogeneity discussed here. Many of the models discussed here are inconvenient or impossible with received Bayesian methods.

Standard Criticisms

- **Of the Classical Approach**
  - Computationally difficult (ML vs. MCMC)
  - No attention is paid to household level parameters.
  - There is no natural estimator of individual or household level parameters
  - Responses: See the preceding and Train (2003, ch. 10)

- **Of Classical Inference in this Setting**
  - Asymptotics are “only approximate” and rely on “imaginary samples.” Bayesian procedures are “exact.”
  - Response: The inexactness results from acknowledging that we try to extend these results outside the sample. The Bayesian results are “exact” but have no generality and are useless except for this sample, these data and this prior. (Or are they? Trying to extend them outside the sample is a distinctly classical exercise.)

Standard Criticisms

- **Of the Bayesian Approach**
  - **Computationally difficult.**
  - **Response: Not really, with MCMC and Metropolis-Hastings**
  - **The prior (conjugate or not) is a canard. It has nothing to do with "prior knowledge" or the uncertainty of the investigator.**
  - **Response: In fact, the prior usually has little influence on the results. (Bernstein and von Mises Theorem)**
- **Of Bayesian 'Inference'**
  - **It is not statistical inference**
  - **How do we discern any uncertainty in the results? This is precisely the underpinning of the Bayesian method. There is no uncertainty. It is 'exact.'**

A Preconclusion

"The advantage of hierarchical Bayes models of heterogeneity is that they yield disaggregate estimates of model parameters. These estimates are of particular interest to marketers pursuing product differentiation strategies in which products are designed and offered to specific groups of individuals with specific needs. In contrast, classical approaches to modeling heterogeneity yield only aggregate summaries of heterogeneity and do not provide actionable information about specific groups. The classical approach is therefore of limited value to marketers." (A&R p. 72)

OHIO STATE

Disaggregated Parameters

- The description of classical methods as only producing aggregate results is obviously untrue.
- As regards "targeting specific groups…" both of these sets of methods produce estimates for the specific data in hand. Unless we want to trot out the specific individuals in this sample to do the analysis and marketing, any extension is problematic. This should be understood in both paradigms.
- NEITHER METHOD PRODUCES ESTIMATES OF INDIVIDUAL PARAMETERS, CLAIMS TO THE CONTRARY NOTWITHSTANDING. BOTH PRODUCE ESTIMATES OF THE MEAN OF THE CONDITIONAL (POSTERIOR) DISTRIBUTION OF POSSIBLE PARAMETER DRAWS CONDITIONED ON THE PRECISE SPECIFIC DATA FOR INDIVIDUAL I.

Conclusions

- When estimates of the same model are compared, they rarely differ by enough to matter. See Train, Chapter 12 for a nice illustration
- Classical methods shown here provide rich model specifications and do admit ‘individual’ estimates. Have yet to be emulated by Bayesian methods
- Just two different algorithms. The philosophical differences in interpretation is a red herring. Appears that each has some advantages and disadvantages