Discrete Choice Modeling

USP 657

26 June, 2020

SIC

`A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models`

1 Introduction

2 Elements of the choice decision process

3 Utility-based choice theory

4 The Multinomial Logit Model

Extreme Value Type I Distribution The extreme value type I distribution has two forms. One is based on the smallest extreme and the other is based on the largest extreme. We call these the minimum and maximum cases, respectively. Formulas and plots for both cases are given. The extreme value type I distribution is also referred to as the Gumbel distribution.

	general (minimum)	standard(minimum)	general (maximum)	standard(maximum)
$f(x)$	$\frac{1}{\beta}e^{\frac{x-\mu}{\beta}}e^{-e^{\frac{x-\mu}{\beta}}}$	$e^x e^{-e^x}$	$\frac{1}{\beta}e^{-\frac{x-\mu}{\beta}}e^{-e^{-\frac{x-\mu}{\beta}}}$	$e^{-x}e^{-e^{-x}}$
$F(x)$		$1-e^{-e^x}$		$e^{-e^{-x}}$

where μ is the location parameter and β is the scale parameter.

Mean $\mu + 0.5772\beta$ (Euler's number)

Median $\mu - \beta\ln(\ln(2))$

Mode μ

Range $-\infty$ to ∞

Standard Deviation $\frac{\beta\pi}{\sqrt{6}}$

Skewness 1.13955

Kurtosis 5.4

Coefficient of Variation $\frac{\beta\pi}{\sqrt{6}(\mu+0.5772\beta)}$

Gumbel_pmf	Gumbel_cdf

5 : Data Assembly and Estimation of sample Multinomial Logit Model

5.2 Data requirements Overview Dependent, endogenous variable: mode choice

Independent, exogenous variables:(automobile ownership, time of day of travel, origin and destination of trip, and travel party size)

- Traveler related variables

Income, Number of automobiles in traveler's household, Number of workers in traveler's household, Sex, Age, Functions of these variables such as number of autos divided by number of workers

- Trip context variables

Trip purpose, Employment density at the traveler's workplace, Population density at the home location, and Dummy variable indicating whether the traveler;s workplace is the Central Business District (CBD)

- Mode (alternative) related variables:

Total travel time, IN-vehicle travel time, Out-of-vehicle travel time, Walk time, Wait time, Number of transfers, Transit headway(service frequency for carrier modes), and travel cost.

- Interaction of Mode and Traveler or Trip Related Variables

Travel cost divided by household income, Travel time or cost interacted with sex or age group for traveler, and Out-of-Vehicle time divided by total trip distance.

5.3 Sources and Methods for Traveler and Trip Related Data Collection

- Travel Survey Types

Household Travel Surveys, Workplace Surveys, Destination Surveys, Intercept Surveys,

- Sampling Design Considerations

Population of interest; Sampling units; The probability sampling; stratified random sampling; Sample size.

5.4 Methods for collecting mode related data Network analysis

5.5 Data structure for estimation The trip format: IDCase, each record contains all the information for mode choice over laternatives for a signle trip.

The trip-alternative format: IDCase-IDAlt, each record contains all the information for a single mode available to each trip maker so there is one record for each mode for each trip.

5.6 Application data for work mode choice in the San Francisco Bay Area

5.7 Estimation of MNL Model with Basic Specification

$$\begin{aligned} V_{DA} &= & \beta_1 TT_{DA} + \beta_2 TC_{DA} \\ V_{SR2} &= \beta_{SR2} + & \beta_1 TT_{SR2} + \beta_2 TC_{SR2} + \gamma_{SR2} Inc \\ V_{SR3+} &= \beta_{SR3+} + & \beta_1 TT_{SR3+} + \beta_2 TC_{SR3+} + \gamma_{SR3+} Inc \\ V_{TR} &= \beta_{TR} + & \beta_1 TT_{TR} + \beta_2 TC_{TR} + \gamma_{TR} Inc \\ V_{BK} &= \beta_{BK} + & \beta_1 TT_{BK} + \beta_2 TC_{BK} + \gamma_{BK} Inc \\ V_{WK} &= \beta_{WK} + & \beta_1 TT_{WK} + \beta_2 TC_{WK} + \gamma_{WK} Inc \end{aligned}$$

- Informal Tests

The sign of parameters (do the associated variables have a positive or negative effect on the alternatives with which they are associated?),

The difference (positive or negative) within sets of alternative specific variables (does the inclusion of this variable have a more or less positive effect on one alternative relative to another?) and

The ratio of pairs of parameters (is the ratio between the parameters of the correct sign and in a reasonable range?).

- Overall Goodness-of-Fit Measures

$\rho_0^2 = 1 - \frac{LL(\hat{\beta})}{LL(0)}$; $\quad \bar{\rho}_0^2 = 1 - \frac{LL(\hat{\beta}) - K}{LL(0)}$

$\rho_c^2 = 1 - \frac{LL(\hat{\beta})}{LL(c)}$; $\quad \bar{\rho}_c^2 = 1 - \frac{LL(\hat{\beta}) - K}{LL(c) - K_{MS}}$

- t test:

`Table 5-4 p.86`

Test of Individual Parameters $\frac{\hat{\beta}_k - \beta_k^*}{S_k}$

Test of Linear Relationship between Parameters $H_0 : \beta_k = \beta_l$, $\frac{\hat{\beta}_k - \hat{\beta}_l}{\sqrt{S_k^2 + S_l^2 - 2S_{k,l}}}$

$H_0 : \beta_{cost} = (VOT)\beta_{time}$, $\frac{\hat{\beta}_{cost} - (VOT)\hat{\beta}_{time}}{\sqrt{S_{cost}^2 + (VOT)^2 S_{time}^2 - 2S_{time,cost}}}$

Tests of Entire Models: LRT $-2[LL_R - LL_U]$

`5.7.3.4` The non-nested hypothesis test uses the adjusted likelihood ratio index $\bar{\rho}^2$ to test the hypothesis that the model with the lower $\bar{\rho}^2$ value is the true model

$\alpha = \Phi[-(-2(\bar{\rho}_H^2 - \bar{\rho}_L^2)LL(0) + (K_H - K_l))^{\frac{1}{2}}]$

5.8 Value of Time

- Value of Time for linear Utility Function

$V_i = ... + \beta_{TVT} TVT_i + \beta_{cost} Cost_i + ... = ... + \beta_{TVT} TVT_{it} + \beta_{costInc} \frac{Cost_{it}}{Income_t} + ...$

$VofT = \beta_{TVT} / \beta_{cost} = \frac{\partial V_i / \partial Time_i}{\partial V_i / \partial Cost_i} = \frac{\beta_{TVT}}{\beta_{CostInc} / Income_t}$

$VofT = \frac{0.6 \frac{\beta_{TVT}}{\beta_{Cost}}}{\ln[Income(\frac{\$1000}{year})]} \$/hour$

$VofT = \frac{\beta_{TVT}}{\beta_{CostInc}} Units_{VofT} = \frac{\beta_{TVT}}{\beta_{CostInc}} \frac{\frac{cents}{minute}}{\frac{\$1000}{year}} = \frac{\beta_{TVT}}{\beta_{CostInc}} 1.2$

- Value of Time for Time or Cost Transformation

$V_{it} = ... + \beta_{\ln(TVT)} \ln(TVT_{it}) + \beta_{cost}Cost_{it} + ...$

$VofT = \frac{\beta_{\ln(TVT)}}{\beta_{Cost}} \frac{1}{TVT_{it}} = \frac{\beta_{TVT}}{\beta_{\ln(Cost)}}Cost_{it}$

6

$H_0 : \underline{\beta}_1 = \underline{\beta}_2 = ... = \underline{\beta}_s ... = \underline{\beta}_S,$

$-2[l_R - l_U] = -2[l(\beta) - \sum_{s=1}^{S} l(\beta_s)] \geq \chi^2_{n,(p)}$

7 San Francisco Bay Area Shop/Other Mode Choice

8 Nested Logit Model

$$\begin{aligned}
U_{DA} =& V_{DA} & & +\varepsilon_{DA} \\
U_{SR} =& V_{GRP} + V_{SR} & +\varepsilon_{GRP} & +\varepsilon_{SR} \\
U_{Bus} =& V_{GRP} + V_{PT} + V_{Bus} & +\varepsilon_{GRP} + \varepsilon_{PT} & +\varepsilon_{Bus} \\
U_{LTR} =& V_{GRP} + V_{PT} + V_{LTR} & +\varepsilon_{GRP} + \varepsilon_{PT} & +\varepsilon_{LTR}
\end{aligned}$$

$Pr(Bus|PT) = \frac{\exp\left(\frac{V_{Bus}}{\theta_{PT}}\right)}{\exp\left(\frac{V_{Bus}}{\theta_{PT}}\right)+\exp\left(\frac{V_{LTR}}{\theta_{PT}}\right)}, Pr(LTR|PT) = \frac{\exp\left(\frac{V_{LTR}}{\theta_{PT}}\right)}{\exp\left(\frac{V_{Bus}}{\theta_{PT}}\right)+\exp\left(\frac{V_{LTR}}{\theta_{PT}}\right)}$

$Pr(SR|GRP) = \frac{\exp\left(\frac{V_{SR}}{\theta_{GRP}}\right)}{\exp\left(\frac{V_{SR}}{\theta_{GRP}}\right)+\exp\left(\frac{V_{PT}+\theta_{PT}\Gamma_{PT}}{\theta_{GRP}}\right)}, \; Pr(PT|GRP) = \frac{\exp\left(\frac{V_{PT}+\theta_{PT}\Gamma_{PT}}{\theta_{GRP}}\right)}{\exp\left(\frac{V_{SR}}{\theta_{GRP}}\right)+\exp\left(\frac{V_{PT}+\theta_{PT}\Gamma_{PT}}{\theta_{GRP}}\right)}, \; \Gamma_{PT} = \log[\exp\left(\frac{V_{Bus}}{\theta_{PT}}\right) + \exp\left(\frac{V_{LTR}}{\theta_{PT}}\right)]$

$Pr(DA) = \frac{\exp(V_{DA})}{\exp(V_{DA})+\exp(V_{GRP}+\theta_{GRPT}\Gamma_{GRP})}, \; Pr(GRP) = \frac{\exp(V_{GRP}+\theta_{GRPT}\Gamma_{GRP})}{\exp(V_{DA})+\exp(V_{GRP}+\theta_{GRPT}\Gamma_{GRP})}, \; \Gamma_{GRP} = \log[\exp\left(\frac{V_{SR}}{\theta_{GRP}}\right) + \exp\left(\frac{V_{PT}+\theta_{PT}\Gamma_{PT}}{\theta_{GRP}}\right)]$

$Pr(SR) = Pr(SR|GRP)Pr(GRP)$

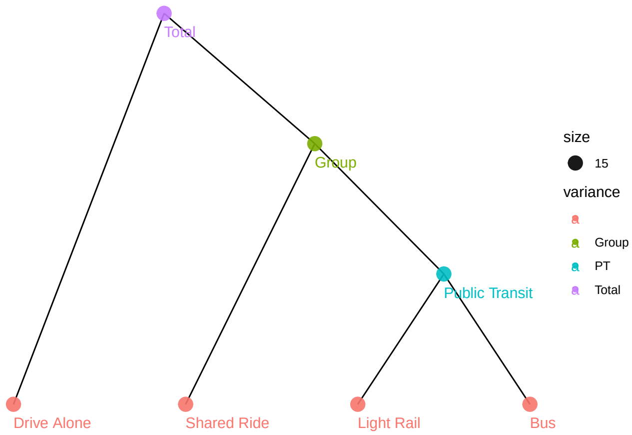
$Pr(Bus) = Pr(Bus|PT)Pr(PT|GRP)Pr(GRP)$

$Pr(LTR) = Pr(LTR|PT)Pr(PT|GRP)Pr(GRP)$

$Var(\varepsilon_{DA}) = Var(\varepsilon_{SR}) + Var(\varepsilon_{GRP}) = Var(\varepsilon_{GRP}) + Var(\varepsilon_{PT}) + Var(\varepsilon_{Bus}) = Var(\varepsilon_{GRP}) + Var(\varepsilon_{PT}) + Var(\varepsilon_{LTR}) = \frac{\pi^2}{6},$

$Var(\varepsilon_{Bus}) = Var(\varepsilon_{LTR}) = \frac{\pi^2}{6\mu_{PT}^2} = \frac{\pi^2\theta_{PT}^2}{6},$

$Var(\varepsilon_{GRP}) = \frac{\pi^2\theta_{Grp}^2}{6}, Var(\varepsilon_{PT}) = \frac{\pi^2\theta_{PT}^2}{6}, \; Var(\varepsilon_{Total} - \varepsilon_{GRP}) = \frac{\pi^2}{6}(1-\theta_{Grp}^2), Var(\varepsilon_{Grp} - \varepsilon_{PT}) = \frac{\pi^2}{6}(\theta_{Grp}^2 - \theta_{PT}^2)$



- Elasticity Comparison of Nested Logit vs. MNL Models

Direct Elasticity	Change in Non-Nested Alternative	Change in Nested Alternative
Multinomial Logit	$(1-P_j)\beta_{Los}Los_j$	Not Applicaable
Nested Logit	$(1-P_j)\beta_{Los}Los_j$	$(1-P_k+\frac{1-\theta_n}{\theta_n}(1-P_{k\|N}))\beta_{Los}Los_k$
Cross Elasticity	Effect on Non-Nested Alts	
Multinomial Logit	$-P_j\beta_{Los}Los_j$	Not Applicaable
Nested Logit	$-P_j\beta_{Los}Los_j$	$-P_k\beta_{Los}Los_k$
Cross Elasticity	Effect on Nested Alts	
Multinomial Logit	Not Applicaable	Not Applicaable
Nested Logit	$-P_j\beta_{Los}Los_j$	$-(P_k+\frac{1-\theta_n}{\theta_n}P_{k\|N})\beta_{Los}Los_k$

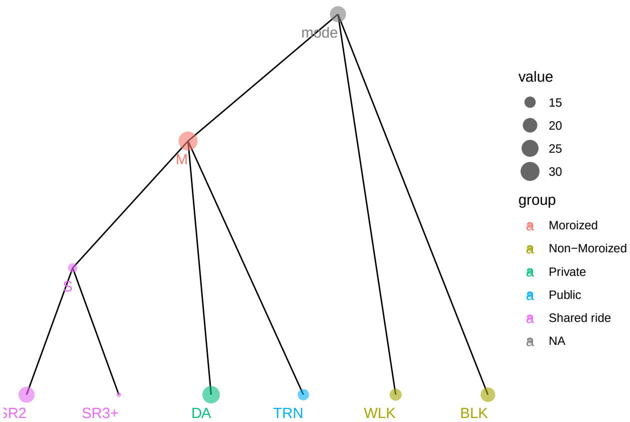
8.4 Statistical Testing of Nested Logit Structures Adopting a nested logit model implies rejection of the MNL. We reject the null hypothesis that the MNL model is the correct model if the calculated value is greater than the test or critical value for the distribution as:

$-2[l^{MNL}-l^{NL}] \geq \chi^2_n$,

for a top level nest, $\frac{\hat{\theta}_k-1}{S_k}$

for other nest, $\frac{\hat{\theta}_{k\subset j}-\hat{\theta}_j}{\sqrt{S^2_{k\subset j}+S^2_j-2S_{k\subset j,j}}}$

9 Selecting a Preferred Nesting Structure



10 Multiple Maxima in the Estimation of Nested Logit Models

11 Aggregate Forecasting, Assessment, and Application

11.2 Aggregate Forecasting

$\hat{N}_i = \sum_{n=1}^{N} P_i(x_{ni}, \hat{\theta}) = \sum_{n=1}^{N} \hat{P}_{ni}$

Where θ is the expected value of the vector of parameters obtained in the estimation phase and $\hat{P}_{ni}$ is the estimated probability of choosing mode i for individual n.

$\hat{S}_i = \frac{1}{N} \sum_{n=1}^{N} P_i(x_{ni}, \hat{\theta}) = \frac{1}{N} \sum_{n=1}^{N} \hat{P}_{ni}$

Two sources of variation:

the probabilistic form of the discrete choice model at the individual level. This sampling variance declines with the size N of the population for which an aggregate prediction is desired. N is likely to be large enough so that the sampling variance can be ignored.

An estimate vector $\hat{\theta}$ of the true value. An estimate of the asymptotic variance of the aggregate share prediction due to this estimation variance:

$Var(\hat{S}_i) = d_i' Var(\hat{\theta}) d_i$, where $d_i = \frac{1}{N} \sum_{n=1}^{N} [\hat{P}_{ni}(x_{ni} - \sum_j \hat{P}_{nj} x_{nj})]$

$Var(\hat{\theta})$ is the asymptotic variance-covariance matrix of the parameters obtained in estimation and d_i' is the transpose matrix of d_i. If the dimension of the vector x_{ni} is $K \times 1$, then the dimension of d_i is also $K \times 1$ and the dimension of $Var(\hat{\theta})$ is $K \times K$.

sample enumeration: uses a random sample of the population of interest and then applies equation to the sample.

synthetic population:construct a synthetic prediction sample for the entire population through micro-simulation.

$\tilde{S}_i = \frac{1}{M} \sum_{n=1}^{M} P_i(x_{ni}, \hat{\theta}) = \frac{1}{M} \sum_{n=1}^{M} \hat{P}_{ni}$

A two-step iterative proportional fitting (IPF) procedure is used to estimate simultaneously the multi-way distribution for each census tract within a PUMA (Miller, 1996), in such a way as to match the marginal distribution of each census tract and the multi-way correlation of the PUMA(Public Use Microdata Area) in the Public Use Microdata Sample (PUMS).

11.3 Aggregate Assessment of Travel Mode Choice Models

The testing of alternative model structures and specifications should be conducted at the disaggregate level to select a preferred model structure/specification. This testing and selection process should not be pursued at the aggregate level.

1. In some model structures such as the multinomial logit, the predicted aggregate share of each modal alternative in the estimation sample will be the same regardless of the model specification as long as a full set of alternative specific constants are included. Thus, aggregate testing of alternative models using the estimation sample is futile.

2. A model that performs poorly at the disaggregate level may perform as well or even better than a model that performs well at the disaggregate level when both models are applied to a hold-out sample to obtain aggregate predictions.

3. errors in individual-level predictions tend to average out in the aggregate, and so aggregate-level testing does not discriminate much among alternative models.

4. the same problems mentioned earlier are applicable (though to a lesser extent) even for semi-aggregate comparisons among alternative models.

Once a model has been selected, it becomes necessary to validate the models at the aggregate level. This validation is used to ensure that the prediction of aggregate behavior, which drives the modeling process, is consistent with the historically observed aggregate data and by extension future aggregate behavior. Effective validation requires matching prediction to observations at easily observable and measurable locations.

The predicted shares can be compared to the actual shares and summary measures such as the root mean square error or average absolute error percentage may be computed to assess the degree to which the predictions correspond to the aggregate observed data.

12 Recent Advances in Discrete Choice Modeling

(Bhat, 2005) The first assumption in the MNL model is that the random components of the utilities of the different alternatives are independent and identically distributed (IID).

The assumption of *independence* implies that there are no common unobserved factors affecting the utilities of the various alternatives. The same underlying unobserved factor (opportunity to socialize or lack of privacy) impacts the utilities of multiple modes. The presence of such common underlying factors across modal utilities has implications for competitive structure.

The assumption of *identically* distributed (across alternatives) random utility terms implies that the variation in unobserved factors affecting modal utility is the same across all modes. If comfort is an unobserved variable whose values vary considerably for the train mode but little for the automobile mode, then the random components for the automobile and train modes will have different variances. Unequal error variances have significant implications for competitive structure, as discussed in detail by Bhat (1995).

A second assumption of the MNL model is that it maintains *homogeneity* in responsiveness to attributes of alternatives across individuals

The MNL model does not allow sensitivity variations to an attribute (for example, travel cost or travel time in a mode choice model) due to unobserved individual characteristics. However, unobserved individual characteristics can and generally will affect responsiveness. Ignoring the effect of unobserved individual attributes can lead to biased and inconsistent parameter and choice probability estimates (see Chamberlain, 1980).

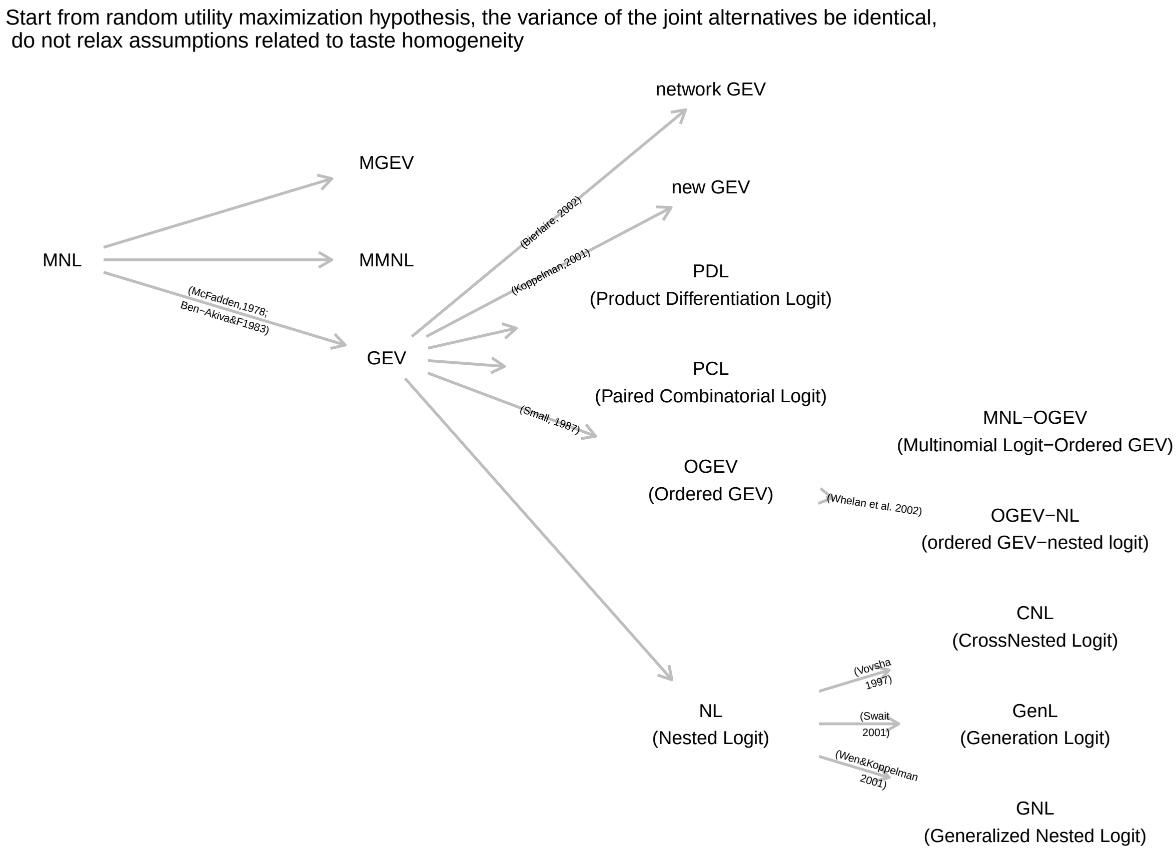
12.2 The GEV Class of Models The GEV-class of models relaxes the independence from irrelevant alternatives (IIA) property of the multinomial logit model by relaxing the independence assumption between the error terms of alternatives. In other words, a generalized extreme value error structure is used to characterize the unobserved components of utility as opposed to the univariate and independent extreme value error structure used in the multinomial logit model.

three important characteristics of all GEV models:

(1) the overall variances of the alternatives (i.e., the scale of the utilities of alternatives) are assumed to be identical across alternatives,

(2) the choice probability structure takes a closed-form expression, and

(3) all GEV models collapse to the MNL model when the parameters generating correlation take values that reduce the correlations between each pair of alternatives to zero.



12.2 The MMNL Class of Models $P_{qi}(\theta) = \int_{-\infty}^{+\infty} L_{qi}(\beta) f(\beta|\theta) d(\beta)$, where $L_{qi}(\beta) = \frac{e^{\beta' x_{qi}}}{\sum_j e^{\beta' x_{qi}}}$

P_{qi} is the probability that individual q chooses alternative i, x_{qi} is a vector of observed variables specific to individual q and alternative i, β represents parameters which are random realizations from a density function $f(.)$, and θ is a vector of underlying moment parameters characterizing $f(.)$.

Train (1986) and Ben-Akiva et al. (1993) applied the mixed logit to customer-level data, but considered only one or two random coefficients in their specifications. Thus, they were able to use *quadrature techniques* for estimation.

The first applications to realize the full potential of mixed logit by allowing several random coefficients simultaneously include Revelt and Train (1998) and Bhat (1998b), both of which were originally completed in the early 1996 and exploited the advances in simulation methods (Train,2003;Bhat, 2005).

The MMNL model structure can be motivated from two very different (but formally equivalent) perspectives (see Bhat, 2000a).

From an intrinsic motivation to allow flexible substitution patterns across alternatives (*error-components structure*) or from a need to accommodate unobserved heterogeneity across individuals in their sensitivity to observed exogenous variables (*random coefficients structure*) or a combination of the two.

Assumption for $f(.)$ are Normal, log-normal(β has to take the same sign for every individual), triangular and uniform(bounded on both sides;same sign for one or more coefficients), Rayleigh distribution(same sign of coefficients for all decision makers).

the MMNL model represents a computationally efficient structure when the number of error components (or factors) needed to generate the desired error covariance structure across alternatives is much smaller than the number of alternatives (see Bhat, 2003a,b

The MMNL model structure also serves as a comprehensive framework for relaxing both the IID error structure as well as the response homogeneity assumption.

	MMNL model	MNP model
capture random taste variations,substitution patterns	flexible	flexible
capture	temporal correlation over time	panel data
accommodate distributions	non-normal	only normal
understand the structure	easier	
estimate	multidimensional integrals	same
simulator	arising from	logit-smoothed Accept-Reject (AR)
code	simple, straightforward	
	appealing and broad	if number of normal random coefficients more than alternatives
dimensionality	order of the number of random coefficients	order of the number of alternatives

12.4 The Mixed GEV Class of Models There are instances when substantial computational efficiency gains may be achieved using a MGEV structure.

It is possible that the utility of spatial units that are close to each other will be correlated due to common unobserved spatial elements. A common specification in the spatial analysis literature for capturing such spatial correlation is to allow contiguous alternatives to be correlated.

In the MMNL structure, such a correlation structure may require multidimensional integration of the order of the number of spatial units.

A carefully specified GEV model can accommodate the spatial correlation structure within a closed-form formulation. However, the GEV model structure of Bhat and Guo cannot accommodate unobserved random heterogeneity across individuals.

The MGEV model involves multidimensional integration only of the order of the number of random coefficients, while the MMNL model would entail a multidimensional integration of the order of spatial units plus the number of random coefficients.

A closed-form analytic structures should be used whenever feasible, because they are always more accurate than the simulation evaluation of analytically intractable structures. Superimposing a mixing structure to accommodate random coefficients over a closed form analytic structure that accommodates a particular desired inter-alternative error correlation structure represents a powerful approach to capture random taste variations and complex substitution patterns.

DCM

`Discrete choice methods with simulation`

1 Introduction

5 Probit

The logit model is limited in three important ways. It cannot represent random taste variation.It exhibits restrictive substitution patterns due to the IIA property. And it can not be used with panel data when unobserved factors are correlated over time for each decision maker.

GEV models relax the second of these restrictions, but not the other two.

Probit models deal with all three. The only limitation of probit models is that they require normal distributions for all unobserved components of utility.

$U_{nj} = V_{nj} + \varepsilon_{nj},\ \forall j;\ \varepsilon'_n = \langle \varepsilon_{n1}, .., \varepsilon_{nJ} \rangle$

$\phi(\varepsilon_n) = \frac{1}{(2\pi)^{J/2}|\Omega|^{1/2}} e^{-\frac{1}{2}\varepsilon'_n \Omega^{-1} \varepsilon_n}$

$P_{ni} = Pr(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}, \forall j \neq i) = \int I(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}, \forall j \neq i)\phi(\varepsilon_n) d\varepsilon_n = \int_{\varepsilon_n \in B_{ni}} \phi(\varepsilon_n) d\varepsilon_n$

$\varepsilon_n \text{s.t.} V_{nj} + \varepsilon_{nj}, \forall j \neq i)$

Define $\tilde{U}_{nji} = U_{nj} - U_{ni},\ \tilde{V}_{nji} = V_{nj} - V_{ni},\ \tilde{\varepsilon}_{nji} = \varepsilon_{nj} - \varepsilon_{ni}$

$\phi(\tilde{\varepsilon}_n) = \frac{1}{(2\pi)^{\frac{1}{2}(J-1)}|\tilde{\Omega}_i|^{1/2}} e^{-\frac{1}{2}\tilde{\varepsilon}'_{ni} \tilde{\Omega}^{-1} \tilde{\varepsilon}_{ni}}$

$P_{ni} = \int I(\tilde{V}_{ni} + \tilde{\varepsilon}_{ni} > \tilde{V}_{nj} + \tilde{\varepsilon}_{nj}, \forall j \neq i)\phi(\tilde{\varepsilon}_{ni}) d\tilde{\varepsilon}_{ni} = \int_{\tilde{\varepsilon}_{ni} \in \tilde{B}_{ni}} \phi(\tilde{\varepsilon}_{ni}) d\tilde{\varepsilon}_{ni}$

5.2 Identification

5.5 Panel Data

5.6 Simulation of the Choice Probabilities

- Accept-Reject Simulator
- Smoothed AR Simulators
- GHK Simulator

$S^r = \frac{e^{U^r_{ni}/\lambda}}{\sum_j e^{U^r_{nj}/\lambda}}$

let L_1 be the Choleski factor of $\tilde{\Omega}_1$, a lower-triangular matrix $\begin{bmatrix} C_{aa} & 0 \\ C_{ab} & C_{bb} \end{bmatrix}$

$\tilde{\varepsilon}_{n21} = C_{aa}\eta_1$
$\tilde{\varepsilon}_{n31} = C_{ab}\eta_1 + C_{bb}\eta_2$

$\tilde{U}_{n21} = \tilde{V}_{n21} + C_{aa}\eta_1$
$\tilde{U}_{n31} = \tilde{V}_{n31} + C_{ab}\eta_1 + C_{bb}\eta_2$

$P_{n1} = Pr(\tilde{V}_{n21} + C_{aa}\eta_1 < 0, \tilde{V}_{n31} + C_{ab}\eta_1 + C_{bb}\eta_2 < 0)$
$= Pr(\tilde{V}_{n21} + C_{aa}\eta_1 < 0) Pr(\tilde{V}_{n31} + C_{ab}\eta_1 + C_{bb}\eta_2 < 0 | \tilde{V}_{n21} + C_{aa}\eta_1 < 0)$

$= Pr(\eta_1 < -\tilde{V}_{n21}/C_{aa}) Pr(\eta_2 < -(\tilde{V}_{n31} + C_{ab}\eta_1)/C_{bb} | \eta_1 < -\tilde{V}_{n21}/C_{aa})$
$= \Phi(\frac{-V_{n21}}{C_{aa}}) \int_{\eta_1=-\infty}^{-V_{n21}/C_{aa}} \Phi\left(\frac{-(V_{n31}+C_{ab}\eta_1)}{C_{bb}}\right) \bar{\phi}(\eta_1) d\eta_1$

$\bar{\phi}(\eta_1) = \phi(\eta_1)/\Phi(\frac{-V_{n21}}{C_{aa}})$ for $-\infty < \eta_1 < -\frac{V_{n21}}{C_{aa}}$, and 0 otherwise.

This probability is simulated as follows:

1. Calculate $k = \Phi(\frac{-V_{n21}}{C_{aa}})$.
2. Draw a value of η_1,labeled η_1^r,from a truncated standard normal truncated at $\frac{-V_{n21}}{C_{aa}}$. This is accomplished as follows:
   (a) Draw a standard uniform μ^r.
   (b) Calculate $\eta_1^r = \Phi^{-1}(\mu^r \Phi(\frac{-V_{n21}}{C_{aa}}))$.
3. Calculate $g^r = \Phi\left(\frac{-(V_{n31}+C_{ab}\eta_1)}{C_{bb}}\right)$.
4. The simulated probability for this draw is $\check{P}_{n1}^r = k \times g^r$.
5. Repeat steps 1-4 R times,and average the results.This average is the simulated probability: $\check{P}_{n1} = \frac{1}{R}\sum \check{P}_{n1}^r$

- General Model

$U_{in} = V_{in} + \varepsilon_{in},\ \forall i;\ \varepsilon_n' = \langle \varepsilon_{1n}, .., \varepsilon_{Jn} \rangle;\ \varepsilon_n \sim N(0, \Omega)$

Ω has $\frac{1}{2}J(J+1)$ elements. 5 alternatives with 15 elements.

Difference against first alternative

$\tilde{U}_{i1n} = \tilde{V}_{i1n} + \tilde{\varepsilon}_{i1n},\ \forall i;\ \tilde{\varepsilon}_{1n}' = \langle \tilde{\varepsilon}_{21n}, .., \tilde{\varepsilon}_{J1n} \rangle$ with $J-1$ length; $\tilde{\varepsilon}_{1n} \sim N(0, \Omega_1)$

$\tilde{\Omega}_1$ has $\frac{1}{2}(J-1)J$ elements. 5 alternatives with 10 elements.

- GHK Simulator with Maximum Likelihood Estimation `p.129`

1. Identification
2. Assure the covariance matrix is positive definite
3. Assure Ω_i is related for all i.

8 Numerical Maximization

$LL(\beta) = \sum_{n=1}^{N} \ln P_n(\beta)/N$

The gradient at β_t is the vector of first derivative of $LL(\beta)$ evaluated at β_t: $g_t = \left(\frac{\partial LL(\beta)}{\partial \beta}\right)_{\beta_t}$

The Hessian is the matrix of second derivatives: $H_t = \left(\frac{\partial g_t}{\partial \beta'}\right)_{\beta_t} = \left(\frac{\partial LL(\beta)}{\partial \beta}\right)_{\beta_t}$

- Newton Raphson

$LL(\beta_{t+1}) = LL(\beta_t) + (\beta_{t+1} - \beta_t)' g_t + \frac{1}{2}(\beta_{t+1} - \beta_t)' H_t (\beta_{t+1} - \beta_t)$

$\frac{\partial LL(\beta_{t+1})}{\partial \beta_{t+1}} = g_t + H_t(\beta_{t+1} - \beta_t) = 0$

$\beta_{t+1} = \beta_t + (-H_t^{-1}) g_t$

Quadratics: If $LL(\beta)$ were exactly quadratic in β, then the NR procedure would reach the maximum in one step from any starting value.

Concavity:

$LL(\beta_{t+1}) = LL(\beta_t) + (\lambda(-H_t^{-1})g_t)' g_t = LL(\beta_t) + \lambda g_t'((-H_t^{-1}))g_t$

- BHHH

Berndt, Hall, Hall, and Hausman (1974)

8.6 Variance of the Estimates $\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{\mathcal{D}} N(0, (-\mathbf{H})^{-1})$

If the model is not correctly specified, $\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{\mathcal{D}} N(0, \mathbf{H^{-1}VH^{-1}})$

where $\mathbf{V}$ is the variance of the scores in the population.

10 Simulation-Assisted Estimation

1. Maximum Simulated Likelihood: MSL

This procedure is the same as maximum likelihood (ML) except that simulated probabilities are used in lieu of the exact probabilities. The properties of MSL have been derived by, for example, Gourieroux and Monfort,(1993), Lee(1995), and Hajivassiliou and Ruud (1994).

The log-likelihood function is $LL(\theta) = \sum_n \ln P_n(\theta)$

the ML estimator $\sum_n s_n(\theta) = \sum_n \frac{\partial \ln P_n(\theta)}{\partial \theta} = 0$

2. Method of Simulated Moments: MSM

This procedure, suggested by McFadden (1989), is a simulated analog to the traditional method of moments (MOM). Under traditional MOM for discrete choice, residuals are defined as the difference between the 0–1 dependent variable that identifies the chosen alternative and the probability of the alternative. Exogenous variables are identified that are uncorrelated with the model residuals in the population. The estimates are the parameter values that make the variables and residuals uncorrelated in the sample. The simulated version of this procedure calculates residuals with the simulated probabilities rather than the exact probabilities.

$\sum_n \sum_j [d_{nj} - P_{nj}(\theta)] z_{nj} = 0$

d_{nj} is the dependent variable that identifies the chosen alternative: $d_{nj} = 1$ if n chose j, and $= 0$ otherwise, and z_{nj} is a vector of exogenous variables called instruments.

The residuals are $d_{nj} - P_{nj}(\theta)$, and the MOM estimator is the parameter values at which the residuals are uncorrelated with the instruments in the sample.

A regression model takes the form $y_n = x_n'\beta + \varepsilon_n$, $\sum_n [y_n - x_n'\beta] z_n = 0$

When the explanatory variables in the model are exogenous, $\sum_n [y_n - x_n'\beta] x_n = 0$, $\hat{\beta} = (\sum_n x_n x_n')^{-1} (\sum_n x_n y_n)$

MOM estimator becomes the standard instrumental variables estimator: $\hat{\beta} = (\sum_n z_n x_n')^{-1} (\sum_n z_n y_n)$

3. Method of Simulated Scores: MSS

As discussed in Chapter 8, the gradient of the log likelihood of an observation is called the score of the observation. The method of scores finds the parameter values that set the average score to zero. When exact probabilities are used, the method of scores is the same as maximum likelihood, since the log-likelihood function is maximized when the average score is zero. Hajivassiliou and McFadden (1998) suggested using simulated scores instead of the exact ones. They showed that, depending on how the scores are simulated, MSS can differ from MSL and, importantly, can attain consistency and efficiency under more relaxed conditions.

$\sum_n s_n(\theta) = \sum_n \frac{\partial \ln P_{nj}(\theta)}{\partial \theta} = \sum_n \frac{1}{P_{nj}} \frac{\partial P_{nj}}{\partial \theta} = 0$

Common Distributions

D	$E; EX^2; V$	$f(x); F(x), P(X \leq x)$	$MLE; T; I$	$M_x(t); I$
$Bern(p)$	p;p;pq	$p^x q^{1-x}, x = 1, 0; 0 \leq p \leq 1$	$\bar{X}$;$\sum x_i \sim Bin(n, p)$;$\frac{1}{pq}$; $Ep_{mle} = p$,$Vp_{mle} = \frac{pq}{n}$	$pe^t + q$
$Bino(n, p)$	np;$np(np + q)$;npq	$\binom{n}{x} p^x q^{n-x}$,$x = 0, 1..n; 0 \leq p \leq 1$;	$\bar{X}$or$k \geq X_{(n)}$;$\sum x_i \sim Bino(n, p)$;$1/pq$	$(pe^t + q$
$Geom(p)$	$1/p$;$(p + 2q)/p^2$;q/p^2	pq^{x-1},$x = 1, 2, ..; 0 \leq p \leq 1$;$1 - q^x$	$1/\bar{X}$;$\sum x_i$;	$\frac{pe^t}{1-qe^t}, t$
$NBino(r, p)$	r/p; ;rq/p^2;$0 \leq p \leq 1$	$\binom{x-1}{r-1} p^r q^{x-r}$,$x = r, r + 1..$	$\bar{X}$;$\sum x_i$	$(\frac{pe^t}{1-qe^t})^r$
$HGeom(N, m, k)$	$\frac{km}{N}$;;$\mu\frac{(N-m)(N-k)}{N(N-1)}$;$N, m, k \geq 0$	$\binom{m}{x}\binom{N-m}{k-x}/\binom{N}{k}$;	$m - (N - k) \leq x \leq m$	
$Pois(\mu)$	μ;$\mu^2 + \mu$;μ;$\mu \geq 0$	$\frac{\mu^x}{x!} e^{-\mu}$,$x = 0, 1..$;$e^{-\mu} \sum_{i=0}^{x} \frac{\mu^i}{i!}$	$\bar{X}$;$\sum x_i$;$\frac{1}{\mu}$;$P(x + 1) = \frac{\lambda}{\lambda+1} P(x)$	$e^{\mu(e^t-1)}$
$Unif(n)$	$\frac{n+1}{2}$;$\frac{(n+1)(2n+1)}{6}$;$\frac{n^2-1}{12}$	$\frac{1}{n}$,$x = 1, 2..n$,$n = b - a + 1$;$\frac{x-a+1}{n}$	;X_nComp	$\frac{1}{n}\sum_{i=1}^{n}$
$Unif(a, b)$	$\frac{a+b}{2}$; ;$\frac{(b-a)^2}{12}$	$\frac{1}{b-a}$,$a \leq x \leq b$;$\frac{x-a}{b-a}$	$\min x_{(1)}, x_{(n)}$;R ancillary	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
$Norm(\mu, \sigma^2)$	μ;$\mu^2 + \sigma^2$;σ^2	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$,$\sigma > 0$	$\bar{X}, \frac{\sum(x_i-\bar{x})^2}{n}$;;$I_\mu = \frac{1}{\sigma^2}$,$I_{\sigma^2} = \frac{1}{2\sigma^4}$; $E\hat{\sigma}^2_{mle} = \frac{(n-1)\sigma^2}{n}$,$V\hat{\sigma}_{mle}$	$e^{\mu t + \frac{\sigma^2 t^2}{2}}$
$SNorm(0, 1)$	0;1;1	$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$		$e^{\frac{t^2}{2}}$
$LNorm(\mu, \sigma^2)$	$e^{\mu+\frac{\sigma^2}{2}}$;$e^{2\mu+2\sigma^2}$;$E^2X(e^{\sigma^2} - 1)$	$\frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{-(\ln x-\mu)^2}{2\sigma^2}}$,$x \geq 0$,$\sigma > 0$	$\hat{\mu} = \frac{1}{n}\sum \ln x_i$,$\hat{\sigma}^2 = \frac{1}{n}\sum \ln(x_i - \hat{\mu})^2$	$EX^n =$
$Cauchy(\theta, \sigma)$		$\frac{1}{\pi\sigma} \frac{1}{1+(\frac{x-\theta}{\sigma})^2}$;;$\sigma > 0$	=t_1;=$\frac{Z_1}{Z_2}$	
$DExpo(\mu, \sigma^2)$	μ;$\mu^2 + 2\sigma^2$;$2\sigma^2$	$\frac{1}{2\sigma} e^{-\lvert\frac{x-\mu}{\sigma}\rvert}$,$\sigma > 0$;	$\hat{\mu} = median$,$\hat{\sigma} = \frac{1}{n}\sum \lvert x_i\rvert$	$\frac{e^{\mu t}}{1-\sigma^2 t^2}$
$Expo(\beta)$	β; ;β^2,$\beta > 0$	$\frac{1}{\beta} e^{-\frac{x}{\beta}}, x \geq 0$;$1 - e^{-\lambda x}$	$\frac{1}{\bar{X}}$;$\sum x_i$;$\frac{1}{\lambda^2}$;$E\lambda_{mle} = \frac{\lambda}{n-1}$, $V\lambda_{mle} = \frac{n^2\lambda^2}{(n-1)^2(n-2)}$,$V\lambda$	$\frac{\lambda}{\lambda-t}, t <$ $\frac{\lambda^2}{n-2}$
$Gamm(\alpha, \beta)$	$\alpha\beta$; ;$\alpha\beta^2$;$\alpha, \beta > 0$	$\frac{1}{\Gamma(a)\beta^\alpha} x^{a-1} e^{-x/\beta}$,$x \geq 0$	$\prod x_i$,$\sum x_i$	$(\frac{1}{1-\beta t})^a$
$Beta(a, b)$	$\frac{a}{a+b}$;$\frac{a(a+1)}{(a+b)(a+b+1)}$;$\frac{ab}{(a+b)^2(a+b+1)}$	$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1 - x)^{b-1}$;$0 \leq x \leq 1$	$\alpha > 0, \beta > 0$	$B(\alpha, \beta)$
χ^2_p	p;$2p + p^2$;$2p$	$\frac{1}{\Gamma\frac{p}{2} 2^{\frac{p}{2}}} x^{\frac{p}{2}-1} e^{-\frac{x}{2}}$,$x \geq 0; p = 1, 2..$	;$\sum \ln x_i$;	$(1 - 2t)$
t_p	$0, p > 1$; ;$\frac{p}{p-2}, p > 2$	$\frac{\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p}{2})\sqrt{p\pi}} (1 + \frac{x^2}{p})^{-\frac{p+1}{2}}$		
$F_{p,q}$	$\frac{q}{q-2}, q > 2$;;$2(\frac{q}{q-2})^2 \frac{p+q-2}{p(q-4)}, q > 4$	$\frac{\Gamma(\frac{p+q}{2})}{\Gamma(\frac{p}{2})\Gamma(\frac{q}{2})} (\frac{p}{q})^{\frac{p}{2}} \frac{x^{\frac{p}{2}-1}}{(1+\frac{p}{q}x)^{\frac{p+q}{2}}}$,$x \geq 0$	$F_{p,q} = \frac{\chi^2_p/p}{\chi^2_q/q}$;$F_{1,q} = t^2_q$	
$Arcsine$	$\frac{1}{2}$; ;$\frac{1}{8}$	$\frac{1}{\pi\sqrt{x(1-x)}}$,$x \in [0, 1]$;$\frac{1}{\pi \arcsin \sqrt{x}}$		$Beta(\frac{1}{2},$
$Dirichlet$	$\frac{a_i}{\sum_k a_k} \sum_{i=1}^{k} x_i = 1$;;$\frac{a_i(a_0-a_i)}{a_0^2(a_0+1)}$	$\frac{1}{B(a)} \Pi_{i=1}^{k} x_i^{a_i-1}$,$x \in (0, 1)$	$B(a) = \frac{\Pi_{i=1}^{k}\Gamma(a_i)}{\Gamma(\sum_{i=1}^{k} a_i)}$	$Cov(X_i$
$Weibull(\gamma, \beta)$	$\beta^{1/\gamma}\Gamma(1 + 1/\gamma)$;; $\beta^{2/\gamma}[\Gamma(1 + 2/\gamma) - \Gamma^2(1 + 1/\gamma)]$	$\frac{\gamma}{\beta} x^{\gamma-1} e^{-\frac{x^\gamma}{\beta}}$,$x \geq 0$,$\gamma > 0, \beta > 0$; $1 - e^{-(\frac{x}{\beta})^\gamma}$	x^γ,$\sum x^\gamma$,$\sum \ln x$	;;$\beta^{\frac{n}{\gamma}}\Gamma(1$
$Pareto(\alpha, \beta)$	$\frac{\beta\alpha}{\beta-1}$,$\beta > 1$;;$\frac{\beta\alpha^2}{(\beta-1)^2(\beta-2)}$,$\beta > 2$	$\frac{\beta\alpha^\beta}{x^{\beta+1}}$;$1 - (\frac{\alpha}{x})^\beta$,$x > \alpha$,$\alpha, \beta > 0$	$\hat{\alpha} = \min x_i$,$\hat{\beta} = \frac{n}{\sum \ln(x_i/x_{(1)})}$; $\sum x_i$comp/suff;$1/\beta^2$	