Decoding Commuter Choices

Ishi Jain (220464)
BS-SDS
Indian Institute of Technology, Kanpur

Contents

1 Discrete choice models

The utility that decision maker n obtains from alternative j is U_{nj}, $j = 1, \ldots, J$. The behavioral model is therefore: choose alternative i if and only if $U_{ni} > U_{nj} \; \forall \; j \neq i$. The researcher does not observe the decision maker's utility, instead observes some attributes of the alternatives as faced by the decision maker, labeled $x_{nj} \; \forall \; j$, and some attributes of the decision maker, labeled s_n. The representative utility is denoted as $V_{nj} = V(x_{nj}, s_n) \; \forall \; j$. Usually, V depends on parameters that are unknown to the researcher and therefore estimated statistically;

Utility is decomposed as $U_{nj} = V_{nj} + \epsilon_{nj}$, where ϵ_{nj} captures the factors that affect utility but are not included in V_{nj}. The joint density of the random vector $\boldsymbol{\epsilon}'_n = \langle \epsilon_{n1}, \ldots, \epsilon_{nJ} \rangle$ is denoted $f(\boldsymbol{\epsilon}_n)$. The probability that decision maker n chooses alternative i is

$$
\begin{aligned}
P_{ni} = \text{Prob}(U_{ni} > U_{nj} \; \forall \; j \neq i) &= \text{Prob}(V_{ni} + \epsilon_{ni} > V_{nj} + \epsilon_{nj} \; \forall \; j \neq i) \\
&= \text{Prob}(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj} \; \forall \; j \neq i). \\
&= \int I(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj} \; \forall \; j \neq i) f(\boldsymbol{\epsilon}_n) \, d\boldsymbol{\epsilon}_n.
\end{aligned}
$$

2 Logit Model

A decision maker, labeled n, faces J alternatives. The utility that the decision maker obtains from alternative j is decomposed into V_{nj} that is known by the researcher up to some parameters, and ϵ_{nj} that is treated by the researcher as random: $U_{nj} = V_{nj} + \epsilon_{nj} \; \forall \; j$. The logit model is obtained by assuming that each ϵ_{nj} is independently, identically distributed extreme value: Gumbel distribution

$$
f(\epsilon_{nj}) = e^{-\epsilon_{nj}} e^{-e^{-\epsilon_{nj}}},
$$

and the cumulative distribution is:

$$
F(\epsilon_{nj}) = e^{-e^{-\epsilon_{nj}}}
$$

The variance of this distribution is $\pi^2/6$. If ϵ_{nj} and ϵ_{ni} are i.i.d. extreme value, then $\epsilon^*_{nji} = \epsilon_{nj} - \epsilon_{ni}$ follows the logistic distribution.

$$
F(\epsilon^*_{nji}) = \frac{e^{\epsilon^*_{nji}}}{1 + e^{\epsilon^*_{nj}}}
$$

Using the extreme value distribution for the errors (and hence the logistic distribution for the error differences) is nearly the same as assuming that the errors are independently normal. The probability that decision maker n chooses alternative i is

$$P_{ni} = \text{Prob}(V_{ni} + \epsilon_{ni} > V_{nj} + \epsilon_{nj} \ \forall \ j \neq i) \tag{3.4}$$

$$= \text{Prob}(\epsilon_{nj} < \epsilon_{ni} + V_{ni} - V_{nj} \ \forall \ j \neq i).$$

If ϵ_{ni} is considered given, this expression is the cumulative distribution for each ϵ_{nj} evaluated at $\epsilon_{ni} + V_{ni} - V_{nj}$ is $e^{-e^{-(\epsilon_{ni}+V_{ni}-V_{nj})}}$. Since the ϵ's are independent, this cumulative distribution over all $j \neq i$ is the product of the individual cumulative distributions:

$$P_{ni}|\epsilon_{ni} = \prod_{j \neq i} e^{-e^{-(\epsilon_{ni}+V_{ni}-V_{nj})}}.$$

Since, ϵ_{ni} is not given, the choice probability is the integral of $P_{ni}|\epsilon_{ni}$ over all values of ϵ_{ni} weighted by its density:

$$P_{ni} = \int (\prod_{j \neq i} e^{-e^{-(\epsilon_{ni}+V_{ni}-V_{nj})}}) \cdot e^{-\epsilon_{ni}} \cdot e^{-e^{-\epsilon_{ni}}} \, d\epsilon_{ni}$$

The logit choice probability:

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} \tag{3.6}$$

$$\sum_j P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} = 1$$

The relation of the logit probability to representative utility is sigmoid, or S-shaped.

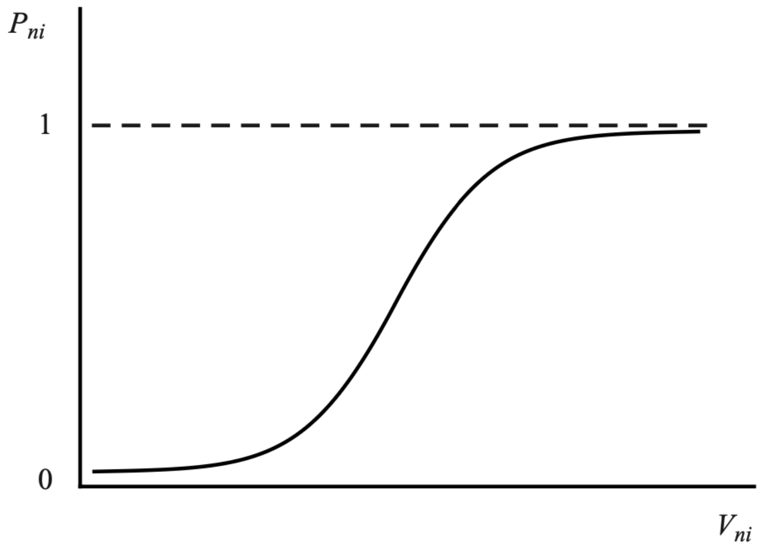

Figure 3.1. Graph of logit curve.

Derivation of Logit Probabilities:

$$P_{ni} = \int_{s=-\infty}^{\infty} \left(\prod_{j \neq i} e^{-e^{-(s+V_{ni}-V_{nj})}} \right) e^{-s} e^{-e^{-s}} \, ds,$$

where s is ϵ_{ni}. Our task is to evaluate this integral. Noting that $V_{ni} - V_{ni} = 0$ and then collecting terms in the exponent of e, we have

$$P_{ni} = \int_{s=-\infty}^{\infty} \left(\prod_{j} e^{-e^{-(s+V_{ni}-V_{nj})}} \right) e^{-s} \, ds$$

$$= \int_{s=-\infty}^{\infty} \exp\left(-\sum_{j} e^{-(s+V_{ni}-V_{nj})} \right) e^{-s} \, ds$$

$$= \int_{s=-\infty}^{\infty} \exp\left(-e^{-s} \sum_{j} e^{-(V_{ni}-V_{nj})} \right) e^{-s} \, ds.$$

Define $t = \exp(-s)$ such that $-\exp(-s)ds = dt$. Note that as s approaches infinity, t approaches zero, and as s approaches negative infinity, t becomes infinitely large.

$$
\begin{aligned}
P_{ni} &= \int_0^\infty \exp\left(-t\sum_j e^{-(V_{ni}-V_{nj})}\right)(-dt) \\
&= \int_0^\infty \exp\left(-t\sum_j e^{-(V_{ni}-V_{nj})}\right)dt \\
&= \left.\frac{\exp\left(-t\sum_j e^{-(V_{ni}-V_{nj})}\right)}{-\sum_j e^{-(V_{ni}-V_{nj})}}\right|_0^\infty \\
&= \frac{1}{\sum_j e^{-(V_{ni}-V_{nj})}} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}},
\end{aligned}
$$

3 Independence of irrelevant alternatives

If we consider the probabilities of choice for two alternatives l and m, we have

$$P_l = \frac{e^{V_l}}{\sum_j e^{V_j}} \quad \text{and} \quad P_m = \frac{e^{V_m}}{\sum_j e^{V_j}}.$$

The ratio of these two probabilities is:

$$\frac{P_l}{P_m} = \frac{e^{V_l}}{e^{V_m}} = e^{V_l - V_m}.$$

This probability ratio for the two alternatives depends only on the characteristics of these two alternatives and not on those of other alternatives. This is called the IIA property (for independence of irrelevant alternatives). IIA relies on the hypothesis that the errors are identical and independent. It is not a problem by itself and may even be considered as a useful feature for a well-specified model. However, this hypothesis may be in practice violated, especially if some important variables are omitted.

4 Consumer Surplus

A person's consumer surplus is the utility, in dollar terms, that the person receives in the choice situation. The decision maker chooses the alternative that provides the greatest utility. Consumer surplus is therefore

$$CS_n = \left(\frac{1}{\alpha_n}\right) \max_j (U_{nj}),$$

where α_n is the marginal utility of income:

$$\frac{dU_n}{dY_n} = \alpha_n,$$

with Y_n the income of person n. The division by α_n translates utility into dollars, since

$$\frac{1}{\alpha_n} = \frac{dY_n}{dU_n}.$$

The researcher does not observe U_{nj} and therefore cannot use this expression to calculate the decision maker's consumer surplus. Instead, the researcher observes V_{nj} and knows the distribution of the remaining portion of utility. With this information, the researcher is able to calculate the expected consumer surplus:

$$E(CS_n) = \frac{1}{\alpha_n} E\left[\max_j (V_{nj} + \epsilon_{nj})\right],$$

If each ϵ_{nj} is i.i.d. extreme value and utility is linear in income (so that α_n is constant with respect to income), then this expectation becomes:

$$E(CS_n) = \frac{1}{\alpha_n} \ln\left(\sum_{j=1}^{J} e^{V_{nj}}\right) + C,$$

where C is an unknown constant that represents the fact that the absolute level of utility cannot be measured.

$E(CS_n)$ is the average consumer surplus in the subpopulation of people who have the same representative utilities as person n.

The change in consumer surplus that results from a change in the alternatives and/or the choice set is calculated from:

$$\Delta E(CS_n) = \frac{1}{\alpha_n}\left[\ln\left(\sum_{j=1}^{J^1} e^{V_{nj}^1}\right) - \ln\left(\sum_{j=1}^{J^0} e^{V_{nj}^0}\right)\right],$$

5 McFadden's R^2

A statistic called the likelihood ratio index is often used with discrete choice models to measure how well the models fit the data. The likelihood ratio index is defined as

$$\rho = 1 - \frac{LL(\hat{\beta})}{LL(0)}$$

The likelihood ratio index is defined as

$$\rho = 1 - \frac{LL(\hat{\beta})}{LL(0)}$$

where $LL(\hat{\beta})$ is the value of the log-likelihood function at the estimated parameters and $LL(0)$ is its value when all the parameters are set equal to zero.

The likelihood ratio index ranges from zero, when the estimated parameters are no better than zero parameters, to one, when the estimated parameters perfectly predict the choices of the sampled decision makers.

Two models estimated on samples that are not identical or with a different set of alternatives for any sampled decision maker cannot be compared via their likelihood ratio index values.

The percentage of sampled decision makers for which the highest-probability alternative and the chosen alternative are the same is called the percent correctly predicted.

6 Multinomial logit model: "ModeCanada" dataset

ModeCanada, is an example of a data set in long format. It presents the choice of 3880 travellers for a transport mode for the Ontario-Quebec corridor.

There are four transport modes (air, train, bus and car) and most of the variables are alternative specific (cost for monetary cost, *ivt* for in-vehicle time, *ovt* for out-of-vehicle time, *freq* for frequency). The only choice situation specific variables are *dist* (the distance of the trip), *income* (household income), *urban* (a dummy for trips which have a large city at the origin or the destination), and *noalt* (the number of available alternatives). The advantage of this shape is that there are much fewer columns than in the wide format, the caveat being that values of *dist*, *income*, and *urban* are repeated four times.

For data in "long" format, the *shape* and the *choice* arguments are no more mandatory.

To replicate published results later in the text, we'll use only a subset of the choice situations, namely those for which the 4 alternatives are available. This can be done using the *subset* function with the *subset* argument set to *noalt == 4* while estimating the model. This can also be done within *dfidx*, using the *subset* argument.

The information about the structure of the data can be explicitly indicated using choice situations and alternative indexes (respectively *case* and *alt* in this data set) or, in part, guessed by the *dfidx* function. Here, after subsetting, we have 2779 choice situations with 4 alternatives, and the rows are ordered first by choice situation and then by alternative (train, air, bus and car in this order).

Random utility models are fitted using the mlogit function. Basically, only two arguments are mandatory, formula and data, if an dfidx object (and not an ordinary data.frame) is provided.

`mlogit` provides two further useful arguments:

- `reflevel` indicates which alternative is the "reference" alternative, i.e., the one for which the coefficients of choice situation specific covariates are set to 0.

- `alt.subset` indicates a subset of alternatives on which the estimation has to be performed; in this case, only the lines that correspond to the selected alternatives are used and all the choice situations where not selected alternatives have been chosen are removed.

We estimate the model on the subset of three alternatives (we exclude bus whose market share is negligible in our sample) and we set car as the reference alternative. Moreover, we use a total transport time variable computed as the sum of the in-vehicle and the out-of-vehicle time variables.

The summary of the multinomial logit model provides several key pieces of information. Firstly, it includes the estimated coefficients for each predictor variable for each alternative, with these coefficients relative to the reference alternative, typically "car". Positive coefficients signify that an increase in the predictor variable is associated with a higher probability of choosing the alternative compared to the reference. Conversely, negative coefficients indicate a lower probability. Secondly, it offers the standard errors of these coefficient estimates, where smaller standard errors denote more precise estimates. Additionally, it presents z-values and p-values, where z-values are the coefficients divided by their standard errors and p-values indicate the statistical significance of the coefficients, usually with a threshold of 0.05. The log-likelihood of the fitted model is also included, with higher values indicating a better fit to the data. Finally, other model fit statistics such as the AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) values are provided, serving as measures to compare models, with lower values indicating better-fitting models.

7 Mixed Logit Model

The mixed logit model, also known as the random parameters logit model, extends the standard logit model to allow for random variation in the coefficients (preferences) across individuals. This model can handle more complex choice behaviors and is more flexible.

Assumptions:

- Relaxation of IIA: The mixed logit model does not assume IIA. It allows for correlation in unobserved factors over alternatives and over time.
- Heterogeneity in Preferences: It accounts for unobserved heterogeneity by allowing the coefficients to vary randomly across individuals.

The probability P_i of choosing alternative i can be written as an integral of the standard logit probability over a distribution of parameters:

$$P_i = \int \left(\frac{\exp(\beta' x_i)}{\sum_{j=1}^{J} \exp(\beta' x_j)} \right) f(\beta \mid \theta) d\beta$$

where $f(\beta \mid \theta)$ is the density function of the random parameters β with hyperparameters θ.

Advantages:

- Flexibility: The mixed logit model can approximate any random utility model arbitrarily closely.
- Heterogeneity: It captures individual-specific taste variations, making it suitable for more realistic modeling of choice behavior.
- Complex Substitution Patterns: By allowing for correlated error terms, it can handle more complex substitution patterns between choices.

Mixed logit models are widely used in transportation research, marketing, environmental economics, and any field where understanding individual-level heterogeneity in choice behavior is crucial.

Case Study: The study aims to understand how different attributes of fishing sites affect anglers' choices. The sample includes 962 river trips by 258 anglers in Montana from July 1992 to August 1993. There are 59 possible river sites defined based on various factors.
The utility U_{njt} of angler n choosing site j on trip t is modeled as:

$$U_{njt} = \beta_n x_{njt} + \epsilon_{njt} \tag{1}$$

where x_{njt} represents the attributes of site j, and β_n are the coefficients that vary over anglers.
Site attributes:

- Fish stock: Measured in units of 100 fish per 1000 feet of river.
- Aesthetics rating: Scale from 0 to 3, with 3 being the highest.
- Trip cost: Cost of traveling to the site, including variable driving costs and value of time.
- Major fishing site indicator: Whether the site is listed as a major fishing site in the Angler's Guide to Montana.
- Campgrounds: Number of campgrounds per USGS block.
- Access areas: Number of state recreation access areas per USGS block.
- Restricted species: Number of restricted species at the site.
- Log of site size: Logarithm of the site size in USGS blocks.

The mixed logit model allows for variation in preferences (heterogeneity) across individuals by estimating both the mean and standard deviation of the coefficients. **The coefficients β_n vary among anglers but not over trips for each angler.**

The mixed logit provides more informa- tion than a standard logit, in that the mixed logit estimates the extent to which anglers differ in their preferences for site attributes. The standard deviations of the coefficients enter significantly, indicating that a mixed logit provides a significantly better representation of the choice situa- tion than standard logit, which assumes that coefficients are the same for all anglers. The mixed logit also allows for the fact that several trips are observed for each sampled angler and that each angler's preferences apply to each of the angler's trips.

8 Important Notes

Advanced level of heterogeneity: The existing few online applica- tions of discrete choice models in recommender systems were based on multinomial or nested logit/probit models, which do not account for preference heterogeneity. Such models can only be used in non- personalized recommendations. On the other hand, logit mixture models (which account for heterogeneity) cannot be estimated in real-time because estimation requires integration over multi- dimensional distributions (in Maximum Likelihood Estimation), or drawing from complex posteriors (in Hierarchical Bayes methods). Applications of logit mixture models were also limited to inter- consumer heterogeneity, and assumed that preferences are stable over time. The proposed methodology accounts for more complex patterns of heterogeneity (inter- and intra-consumer heterogeneity), which improves the quality of predictions and recommendations.

Advantages of using discrete choice models in personalized recommendations: First, these models represent utility as a function of the attributes of items (or alternatives), and the individual preferences towards each of these attributes. Therefore, utility is not inferred from measures of similarity obtained from item or user profiling. Second, since utility is modeled as a function of attributes, this method is able to handle cases where new items (with known attributes) could be recommended (e.g. items that have not been chosen or rated before), and cases where the attributes vary over time. The researcher decides on the specification of the utility functions, which may include the attributes, the individual preferences for attributes, contextual variables, and individual characteristics, thus making use of all the available data. Third, since the users' preferences are inferred from their previous choices, this reduces the burden on users because they are not required to rate or evaluate any items.

9 Swissmetro dataset description

This dataset consists of survey data collected on the trains between St. Gallen and Geneva, Switzer- land, during March 1998. The respondents provided information in order to analyze the impact of the modal innovation in transportation, represented by the Swissmetro, a revolutionary mag-lev un- derground system, against the usual transport modes represented by car and train.

The dataset includes responses from rail-based travelers and car users identified via license plate observations on motorways.

Observations: Each of the 1,191 respondents (441 rail users + 750 car users) provided responses for 9 hypothetical choice situations, resulting in a total of 10,729 records.
Filtered Data: After filtering out certain observations (e.g., non-commuters, unknown choices), the dataset used for analysis contains 6,768 records.
Variables:

- Demographic Information: Age, gender, income, travel purpose.
- Travel Characteristics: Travel time (TT), cost (CO), headway (HE), and availability (AV) for each mode.
- Modes of Transport: Train, Swissmetro (SM), and car.
- Unique Identifiers: IDs for each respondent and each choice situation.
- Group and Survey: Differentiates between current rail and road users.
- Purpose: Travel purpose, such as commuting, business, or leisure.
- Choice Indicator: Indicates the chosen mode of transport (Train, SM, Car)

Dataset Size: 6768 rows × 29 columns

10 Airline itinerary dataset description

The survey targeted customers using an Internet airline booking service for low-cost travel deals. While waiting for search results, randomly selected customers were asked to complete a survey based on their specific travel requests.
Each respondent was presented with three choices:

1. A non-stop flight.
2. A flight with one stop on the same airline.
3. A flight with one stop and a change of airline.

The respondents had to rank these choices and had the option to decline all of them.
Total Respondents: 3,609
Survey Responses: Each respondent provided one stated preference (SP) response.
The survey collected data, such as: Age, Gender, Income, Occupation, Education, Desired departure time, Trip purpose, Who is paying for the trip, Number in the travel party.
Dataset Size: 3609 rows x 54 columns

11 Results

Given that xlogit requires the dataset to be provided in the long format, we reshape the dataset using the wide-to-long utility provided by xlogit.

11.1 Swissmetro Dataset

Coefficient	**Description**
`asc_train`	Alternative Specific Constant for Train (1 if the alternative is Train, 0 otherwise)

Coefficient	Description
`asc_sm`	Alternative Specific Constant for Swissmetro (1 if the alternative is Swissmetro, 0 otherwise)
`cost_train`	Travel cost for Train
`cost_sm`	Travel cost for Swissmetro
`cost_car`	Travel cost for Car
`time_train_sm`	Travel time for Train and Swissmetro
`time_car`	Travel time for Car
`headway_train`	Headway for Train
`headway_sm`	Headway for Swissmetro
`seatconf_sm`	Seat configuration for Swissmetro
`survey_train_sm`	Train survey indicator for Train and Swissmetro
`regular_class_sm`	Regular class indicator for Swissmetro
`single_lug_car`	Single luggage indicator for Car
`multip_lug_car`	Multiple luggage indicator for Car

Multinomial Logit Results:

```
Estimation time= 0.1 seconds
---------------------------------------------------------------------------
Coefficient              Estimate      Std.Err.         z-val         P>|z|
---------------------------------------------------------------------------
asc_train              -1.2929512     0.1237556   -10.4476139      2.43e-24 ***
asc_sm                 -0.5026152     0.1032927    -4.8659312      5.87e-06 ***
time_train_sm          -0.6990098     0.0396510   -17.6290608      8.13e-67 ***
time_car               -0.7229887     0.0442625   -16.3340968      1.18e-57 ***
cost_train             -0.5618773     0.0807075    -6.9618968      2.59e-11 ***
cost_sm                -0.2816843     0.0417373    -6.7489902       1.1e-10 ***
cost_car               -0.5139009     0.0970406    -5.2957304      6.66e-07 ***
headway_train          -0.3143519     0.0505955    -6.2130370      3.49e-09 ***
headway_sm             -0.3773753     0.1652542    -2.2836046        0.0589 .
seatconf_sm            -0.7824379     0.0758912   -10.3100010      9.91e-24 ***
survey_train_sm         2.5424946     0.0921336    27.5957408      1.5e-157 ***
regular_class_sm        0.5650259     0.0652226     8.6630441      4.93e-17 ***
single_lug_car          0.4227658     0.0611684     6.9115077      3.66e-11 ***
multip_lug_car          1.4141058     0.2373032     5.9590672      1.62e-08 ***
---------------------------------------------------------------------------
Significance:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood= -5159.258
AIC= 10346.517
BIC= 10441.996
```

Mixed Logit Results: when 'time_train_sm' and 'time_car' are normally varied and other variables are kept constant

when 'time_train_sm' and 'time_car' are uniformly varied and other variables are kept constant

```
Estimation time= 81.9 seconds
----------------------------------------------------------------------------
Coefficient              Estimate      Std.Err.         z-val         P>|z|
----------------------------------------------------------------------------
asc_train              -2.6512858     0.5160897    -5.1372579      2.87e-07 ***
asc_sm                 -2.8361850     0.4168501    -6.8038487      1.11e-11 ***
time_train_sm          -3.5433134     0.2189289   -16.1847640      7.79e-58 ***
time_car               -4.1890838     0.1988197   -21.0697608      1.67e-95 ***
cost_train             -4.2478221     0.4053027   -10.4806171      1.66e-25 ***
cost_sm                -2.7432884     0.2486363   -11.0333397      4.57e-28 ***
cost_car               -2.9687530     0.2437387   -12.1800645      8.94e-34 ***
headway_train          -0.4602198     0.0906998    -5.0740991         4e-07 ***
headway_sm             -1.0814773     0.2742113    -3.9439560       8.1e-05 ***
seatconf_sm            -0.5918300     0.1364511    -4.3373044      1.46e-05 ***
survey_train_sm         5.2143260     0.4821355    10.8150625      4.86e-27 ***
regular_class_sm        0.6211679     0.1960884     3.1677959       0.00154 **
single_lug_car          0.8833470     0.3943789     2.2398433        0.0251 *
multip_lug_car          2.1603961     0.9496666     2.2748996        0.0229 *
sd.time_train_sm        2.7558560     0.1380437    19.9636522      3.33e-86 ***
sd.time_car             1.9328328     0.0893387    21.6348981        2e-100 ***
----------------------------------------------------------------------------
Significance:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood= -3761.443
AIC= 7554.885
BIC= 7664.004
```

```
----------------------------------------------------------------------------
Coefficient              Estimate      Std.Err.         z-val         P>|z|
----------------------------------------------------------------------------
asc_train              -1.7200344     0.5215918    -3.2976637       0.00098 ***
asc_sm                 -2.3565805     0.4384186    -5.3751843      7.91e-08 ***
time_train_sm          -4.1552127     0.2534804   -16.3926387      2.96e-59 ***
time_car               -4.2497950     0.2195137   -19.3600480      2.57e-81 ***
cost_train             -4.4188079     0.3691890   -11.9689586       1.1e-32 ***
cost_sm                -2.7771904     0.2398600   -11.5783792      1.03e-30 ***
cost_car               -3.2035364     0.2525415   -12.6851893      1.85e-36 ***
headway_train          -0.4192441     0.0875088    -4.7908773       1.7e-06 ***
headway_sm             -1.0902762     0.2733656    -3.9883443      6.72e-05 ***
seatconf_sm            -0.6128962     0.1338751    -4.5781180      4.78e-06 ***
survey_train_sm         4.2272153     0.5041457     8.3849081      6.12e-17 ***
regular_class_sm        0.3478338     0.1833754     1.8968398        0.0579 .
single_lug_car          0.5370716     0.4564415     1.1766492         0.239
multip_lug_car          2.2386582     1.0066568     2.2238545        0.0262 *
sd.time_train_sm        5.2284355     0.2620739    19.9502360      4.29e-86 ***
sd.time_car             4.2204566     0.2138090    19.7393748      2.26e-84 ***
----------------------------------------------------------------------------
Significance:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood= -3812.676
AIC= 7657.353
BIC= 7766.472
```

Coefficient	Estimate	Std.Err.	z-val	P>|z|	
asc_train	-5.4373560	0.3982463	-13.6532498	6.9e-42	***
asc_sm	-4.8600582	0.3692878	-13.1606232	4.45e-39	***
time_train_sm	-0.9552519	0.0657250	-14.5340651	3.76e-47	***
time_car	-3.2108039	0.1697851	-18.9109904	9.12e-78	***
cost_train	-2.3456985	0.2545032	-9.2167741	4.01e-20	***
cost_sm	-1.7353815	0.1899957	-9.1337933	8.59e-20	***
cost_car	-1.6256920	0.2075796	-7.8316560	5.55e-15	***
headway_train	-0.3297651	0.0634004	-5.2013130	2.04e-07	***
headway_sm	-0.6722847	0.2291111	-2.9343179	0.00335	**
seatconf_sm	-0.8043152	0.0915054	-8.7898138	1.88e-18	***
survey_train_sm	6.3765976	0.4033208	15.8102394	2.56e-55	***
regular_class_sm	0.5101349	0.0798138	6.3915597	1.75e-10	***
single_lug_car	1.2559236	0.2908567	4.3180152	1.6e-05	***
multip_lug_car	2.8247408	0.8442845	3.3457214	0.000825	***
sd.time_car	1.6916693	0.1024355	16.5144840	4.28e-60	***

Significance: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood= -4443.121
AIC= 8916.242
BIC= 9018.542

Figure 1: Enter Caption

when 'time_car' is normally varied and other variables are kept constant

11.2 Air Itinerary Dataset

Coefficient	Description
asc_one	Alternative Specific Constant for alternative one (1 if the alternative is one, 0 otherwise)
asc_two	Alternative Specific Constant for alternative two (1 if the alternative is two, 0 otherwise)
cost_one	Travel cost for alternative one
cost_two	Travel cost for alternative two
cost_three	Travel cost for alternative three
flytime_one_two	Flying time for alternatives one and two
flytime_three	Flying time for alternative three
triptime_one_two	Trip time for alternatives one and two
triptime_three	Trip time for alternative three
legroom_one	Legroom for alternative one
legroom_two	Legroom for alternative two
legroom_three	Legroom for alternative three
arrival_one	Arrival time for alternative one
arrival_two	Arrival time for alternative two
arrival_three	Arrival time for alternative three
dep_one	Departure time for alternative one
dep_two	Departure time for alternative two
dep_three	Departure time for alternative three

Multinomial Logit Results:

Replace 'example-image' with the filename of your image

Mixed Logit Results: Results when 'flytime_one_two','flytime_three', 'arrival_one', 'arrival_two', 'arrival_three' are varied normally and rest are kept constant.

Coefficient	Estimate	Std.Err.	z-val	P>|z|	
asc_one	1.3218455	84.0294376	0.0157307	0.987	
asc_two	1.1642401	0.4391041	2.6513989	0.00805	**
flytime_one_two	−0.4340280	168.0579238	−0.0025826	0.998	
flytime_three	−0.2268955	168.0579192	−0.0013501	0.999	
cost_one	−0.0194305	0.0007036	−27.6139512	1.92e-152	***
cost_two	−0.0208069	0.0008095	−25.7045882	5.96e-134	***
cost_three	−0.0210308	0.0008426	−24.9587569	5.66e-127	***
legroom_one	0.2425505	0.0366831	6.6120549	4.35e-11	***
legroom_two	0.2101237	0.0450403	4.6652329	3.19e-06	***
legroom_three	0.1678312	0.0478797	3.5052676	0.000462	***
arrival_one	−0.0975048	43.3975799	−0.0022468	0.998	
arrival_two	−0.1002657	43.3975747	−0.0023104	0.998	
arrival_three	−0.0494488	43.3975859	−0.0011394	0.999	
dep_one	0.0963431	43.3975820	0.0022200	0.998	
dep_two	0.0666618	43.3975862	0.0015361	0.999	
dep_three	0.0761969	43.3975735	0.0017558	0.999	
triptime_one_two	−0.1777532	43.3976807	−0.0040959	0.997	
triptime_three	−0.3086682	43.3976304	−0.0071126	0.994	

Significance: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood= -2366.380
AIC= 4768.760
BIC= 4880.202

Based on the log likelihood over the Swissmetro dataset, it is very clear that the mixed logit model very clearly outperforms multinomial logit model with a clear margin of around 27%. While for the Boeing airlines dataset, mixed logit performs nearly similar to multinomial logit with a minor difference 0.16%.

12 Conclusions

From the above results we can conclude that depending on the complexity of the dataset both mixed and multinomial logit models have their own advantages. Mixed models are more efficient in incorporating heterogeneity for large datasets. However,while implementing mixed models, it turns out that varying probability distributions for coefficients of selected features produces better results.

References

[1] Discrete Choice Methods with Simulation by Kenneth Train
https://eml.berkeley.edu/books/choice2.html

[2] CRAN package- mlogit: Multinomial Logit Models
https://cran.r-project.org/web/packages/mlogit/vignettes/c3.rum.html

[3] Larch Documentation
https://larch.newman.me/v5.7.0/intro.html

[4] Online discrete choice models: Applications in personalized recommendations
https://www.sciencedirect.com/science/article/pii/S0167923619300181

[5] xlogit: Python Package
https://xlogit.readthedocs.io/en/latest/

[6] Airline itinerary dataset description
https://transp-or.epfl.ch/documents/technicalReports/CS_AirlineDescription.pdf

[7] Swissmetro dataset description
https://transp-or.epfl.ch/documents/technicalReports/CS_SwissmetroDescription.pdf