

# USP 657 HW2

qushen@pdx.edu

Shen Qu

## Objectives

The topic of homework 2 is Multinomial Logit Model (MNL) estimation, specification testing and application:

Specify and estimate an MNL model;

Conduct reasonableness check and specification testing;

Derive Value of Time, user benefit from model estimation results.

## Datasets

The Swissmetro dataset (Bierlaire, M., Axhausen, K., and G, Abay (2001)) consists entirely of survey data collected on the train between St. Gallen and Geneva, Switzerland, 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 maglev underground system, against the usual transportation modes represented by car and train.

1. Formulate a hypothesis: selection of explanatory variables and how they affect the utilities (generic and/or alternative-specific), inclusion of socioeconomic variables, etc.;

According the data description and Bierlaire et al. (2001), we select several factors as candidate variables and divide them into three groups: attributes of generic alternatives and specific alternatives, characteristics of decision-makers.

Suppose the model specification as below:

Alt	Generic			Specific		Characteristics						
	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$	$\beta_9$	$\beta_{10}$	$\beta_{11}$
Tra	0	$Av_{tra}$	$Fr_{tra}$	$TT_{pub}$	$Co_{tra}$	Ga	Age	0	male	inc	0	0
Sm	$1_{sm}$	$Av_{sm}$	$Fr_{sm}$	$TT_{pub}$	$Co_{sm}$	Ga	0	Seat	0	0	1st	0
Car	$1_{car}$	0	0	$TT_{car}$	$Co_{car}$	0	0	0	0	0	0	Lug

Let "Alt" = "alternative"; "sm" = "Swissmetro"; "tra" = "train"; "pub" = "transit";

"Av" = "availability"; "Fr" = "Headway"; "TT" = "Travel Time"; "CO" = "Travel Cost"

"Ga" = "General abonnement", annual season ticket for transit

"Seats" is the parameter associated with the seat configuration in the Swissmetro

"Lug" = "luggage" the parameter associated with presence of relevant amounts of luggage during the trip. "inc" = "income"; "1st" = "First Class"

$$\begin{aligned}
 V_{i,Train} &= 0 + \beta_{av,tra} + \beta_{fr,tra} Fr_{tra} + \beta_{tt,pub} TT_{tra} + \beta_{tc,tra} Co_{tra} + 0 + \beta_{age} Age + 0 \\
 V_{i,sm} &= \beta_{0,sm} + \beta_{av,sm} + \beta_{fr,sm} Fr_{sm} + \beta_{tt,pub} TT_{sm} + \beta_{tc,sm} Co_{sm} + \beta_{seat} Seat_{sm} + 0 + 0 \\
 V_{i,Car} &= \beta_{0,car} + 0 + 0 + \beta_{tt,car} TT_{car} + \beta_{tc,car} Co_{car} + 0 + 0 + \beta_{lug} Lug
 \end{aligned}$$

<https://stats.stackexchange.com/questions/100007/alternative-specific-variables-in-r> <https://notebook.community/timothyb0912/pylogit/examples/notebooks/Main%20PyLogit%20Example>

2. Estimate the related model. Give an interpretation of the obtained results and check if they are plausible. Note that car is not available to traveler without car (indicated by CAR\_AV).

Removed the cases of traveler without car, the LRT test shows that model 1.1 is significant better than the ASC model.

```

m0 <- mlogit(choice~1|1|0, sm, reflevel="train")
sm.av <- sm %>% filter(av == 1)
m1 <- mlogit(choice~1|1|0, sm.av, reflevel="train")
m1.1 <- mlogit(choice~1+fr|1|0, sm.av, reflevel="train")
lrtest(m0,m1,m1.1)
## Likelihood ratio test
##
## Model 1: choice ~ 1 | 1 | 0
## Model 2: choice ~ 1 | 1 | 0
## Model 3: choice ~ 1 + fr | 1 | 0
##      #Df LogLik Df Chisq Pr(>Chisq)
## 1      2 -10101
## 2      2 -10101 0      0          1
## 3      3 -9439 1 1326 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

3. Try to adjust the model specification, improving the final log-likelihood value while keeping coefficient estimates coherent with the behavioral hypothesis. It is suggested that you start with a simple model specification, and introduce new variables into your model specification using a stepwise modeling strategy, which increases the complexity by adding different variables at each step.

Chosen the ASC model as the benchmark, the comparison of models show that the more predictors give the larger log-likelihood value and each adding has a significant improvement.

```

m1.2 <- mlogit(choice~1+fr|1|tt, sm.av, reflevel="train")
m1.3 <- mlogit(choice~1+fr|ga|tt, sm.av, reflevel="train")
m1.4 <- mlogit(choice~1+fr|ga|tt+co, sm.av, reflevel="train")
m1.5 <- mlogit(choice~1+fr|ga+male|tt+co, sm.av, reflevel="train")
m1.6 <- mlogit(choice~1+fr|ga+male+age|tt+co, sm.av, reflevel="train")
m1.7 <- mlogit(choice~1+fr|ga+male+age+luggage|tt+co, sm.av, reflevel="train")
m1.8 <- mlogit(choice~1+fr|ga+male+age+luggage+first|tt+co, sm.av, reflevel="train")
m1.9 <- mlogit(choice~1+fr+seats|ga+male+age+luggage+first|tt+co, sm.av, reflevel="train")
m1.10 <- mlogit(choice~1+fr+seats|ga+male+age+luggage+first+income|tt+co, sm.av, reflevel="train")

```

```

lrtest(m1.1, m1.2, m1.3, m1.4, m1.5, m1.6, m1.7, m1.8, m1.9, m1.10)
## Likelihood ratio test
##
## Model 1: choice ~ 1 + fr | 1 | 0
## Model 2: choice ~ 1 + fr | 1 | tt
## Model 3: choice ~ 1 + fr | ga | tt
## Model 4: choice ~ 1 + fr | ga | tt + co
## Model 5: choice ~ 1 + fr | ga + male | tt + co

```

```
## Model 6: choice ~ 1 + fr | ga + male + age | tt + co
## Model 7: choice ~ 1 + fr | ga + male + age + luggage | tt + co
## Model 8: choice ~ 1 + fr | ga + male + age + luggage + first | tt + co
## Model 9: choice ~ 1 + fr + seats | ga + male + age + luggage + first |
##      tt + co
## Model 10: choice ~ 1 + fr + seats | ga + male + age + luggage + first +
##      income | tt + co
##      #Df LogLik Df      Chisq Pr(>Chisq)
## 1      3  -9439
## 2      6  -8839 3 1199.62 < 2e-16 ***
## 3      8  -8458 2  761.08 < 2e-16 ***
## 4     11  -8352 3  213.16 < 2e-16 ***
## 5     13  -8278 2  147.85 < 2e-16 ***
## 6     15  -8218 2  119.45 < 2e-16 ***
## 7     17  -8204 2   28.77 0.00000056 ***
## 8     19  -8196 2   16.14 0.00031 ***
## 9     20  -8193 1    5.70 0.01693 *
## 10    22  -8188 2    8.61 0.01352 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Among the predictors, the coefficient of 'income' is not significant. We can choose model 1.9.

```
summary(m1.10)
##
## Call:
## mlogit(formula = choice ~ 1 + fr + seats | ga + male + age +
##      luggage + first + income | tt + co, data = sm.av, reflevel = "train",
##      method = "nr")
##
## Frequencies of alternatives:choice
## train  car  sm
## 0.133 0.287 0.580
##
## nr method
## 6 iterations, 0h:0m:1s
## g'(-H)^-1g = 0.000314
## successive function values within tolerance limits
##
## Coefficients :
##              Estimate Std. Error z-value Pr(>|z|)
## (Intercept):car  0.2695459  0.1806794    1.49  0.13574
## (Intercept):sm   0.9792242  0.1525414    6.42  1.4e-10 ***
## fr              -0.0071950  0.0008395   -8.57 < 2e-16 ***
## seats           0.1669159  0.0712290    2.34  0.01911 *
## ga:car          -7.7730728  0.4396754  -17.68 < 2e-16 ***
## ga:sm           -1.9104250  0.2349420   -8.13  4.4e-16 ***
## male:car        0.8484585  0.0844646   10.05 < 2e-16 ***
## male:sm         0.7046964  0.0687979   10.24 < 2e-16 ***
## age:car         -0.1323742  0.0361860   -3.66  0.00025 ***
## age:sm          -0.3134580  0.0304270  -10.30 < 2e-16 ***
## luggage:car     -0.2832419  0.0621523   -4.56  5.2e-06 ***
## luggage:sm     -0.2712188  0.0529633   -5.12  3.0e-07 ***
## first:car       0.3399792  0.0863344    3.94  8.2e-05 ***
## first:sm       0.2975256  0.0799681    3.72  0.00020 ***
## income:car     -0.0563654  0.0402716   -1.40  0.16162
## income:sm      0.0268958  0.0336402    0.80  0.42399
## tt:train       -0.0136548  0.0006288  -21.71 < 2e-16 ***
## tt:car         -0.0091745  0.0005646  -16.25 < 2e-16 ***
## tt:sm          -0.0137823  0.0006199  -22.23 < 2e-16 ***
## co:train       -0.0013479  0.0001119  -12.04 < 2e-16 ***
## co:car         -0.0039868  0.0007719   -5.16  2.4e-07 ***
## co:sm         -0.0008728  0.0000765  -11.41 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Log-Likelihood: -8190
## McFadden R^2: 0.189
## Likelihood ratio test : chisq = 3830 (p.value = <2e-16)
```

4. Perform a test on whether the coefficients for travel time attributes should be generic or alternative-specific, and a test on whether the coefficients for drive cost is equal to those for train cost or Swiss Metro cost.

```
m1.21 <- mlogit(choice~1+fr+tt|1, sm.av, reflevel="train")
lrtest(m1.2,m1.21)
## Likelihood ratio test
##
## Model 1: choice ~ 1 + fr | 1 | tt
## Model 2: choice ~ 1 + fr + tt | 1
## #Df LogLik Df Chisq Pr(>Chisq)
## 1 6 -8839
## 2 4 -8920 -2 162 <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
m1.41 <- mlogit(choice~1+fr+co|ga|tt, sm.av, reflevel="train")
lrtest(m1.4,m1.41)
## Likelihood ratio test
##
## Model 1: choice ~ 1 + fr | ga | tt + co
## Model 2: choice ~ 1 + fr + co | ga | tt
## #Df LogLik Df Chisq Pr(>Chisq)
## 1 11 -8352
## 2 9 -8390 -2 76.9 <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

If we change travel time from specific alternative to generic, the new model is worse at .05 significance level.

By the same way, the model treating travel cost as generic alternative is worse than before at .05 significance level.

Therefore, travel time and travel cost should be both alternative-specific variables.

5. Derive Value of Time and Value of Frequency; depending on results of your tests above, you may have up to 3 sets of values.

```
coe<- coef(m1.9)
(vot.car<- coe["tt:car"]/coe["co:car"])
## tt:car
## 2.24
(vot.sm<- coe["tt:sm"]/coe["co:sm"])
## tt:sm
## 15.8
(vot.train<- coe["tt:train"]/coe["co:train"])
## tt:train
## 10.1
# (vof.sm <- coe["fr"]/coe.fr["co:sm"])
# (vof.train <- coe["fr"]/coe.fr["co:train"])
```

The Value of Time are 2.237, 15.805, 10.126 for Car, Sm, and Train respectively.

```
sm.av.fr <- sm.av %>%
  mutate(fr_train = ifelse(idx(sm.av, 2)=="train", fr, 0),
         fr_sm = ifelse(idx(sm.av, 2)=="sm", fr, 0))
m1.91 <- mlogit(choice~1+fr_train+fr_sm+seats|ga+male+age+luggage+first|tt+co, sm.av.fr) #, reflevel="car"
coe.fr<- coef(m1.91)
(vof.sm <- coe.fr["fr_sm"]/coe.fr["co:sm"])
## fr_sm
## 8.13
(vof.train <- coe.fr["fr_train"]/coe.fr["co:train"])
## fr_train
## 5.35
```

The Value of Frequency are 8.135, 5.35 for Sm and Train respectively.

```
m1.92 <- update(m1.91, alt.subset=c("train", "sm"))
hmfetest(m1.91, m1.92)
##
## Hausman-McFadden test
##
## data: sm.av.fr
## chisq = 117, df = 13, p-value <2e-16
## alternative hypothesis: IIA is rejected
```

- Design one market segmentation scheme based on decision maker's socio-economic characteristics (e.g. age, income) and perform a market segmentation test to determine if the segments are statistically different from one another.

```
levels(factor(sm$income)) # levels(factor(sm$age))
## [1] "0" "1" "2" "3" "4"
m1.71 <- mlogit(choice~1+fr|ga+male+age+luggage|tt+co, sm.av%>%filter(income==1), reflevel="train") #
m1.72 <- mlogit(choice~1+fr|ga+male+age+luggage|tt+co, sm.av%>%filter(income==2), reflevel="train")
m1.73 <- mlogit(choice~1+fr|ga+male+age+luggage|tt+co, sm.av%>%filter(income==3), reflevel="train")
m1.74 <- mlogit(choice~1+fr|ga+male+age+luggage|tt+co, sm.av%>%filter(income==4), reflevel="train")

lrt.seg <- 2*(logLik(m1.71)+logLik(m1.72)+logLik(m1.73)+logLik(m1.74)-logLik(m1.7))
df.seg <- 17*4-17
pchisq(lrt.seg,df.seg,lower.tail = F)
## 'log Lik.' 1.03e-175 (df=17)

# DCA p.195
qchisq(.05,12,lower.tail = F)
## [1] 21
qchisq(.05,24,lower.tail = F)
## [1] 36.4
```

The market segmentation of 'income' contains four levels. The degree of freedom of four fitted models and pooled model are all 17.

The market segmentation test shows that the p-value is smaller than 0.05. The evidence can reject equality of coefficients across the income market segments.

- Revert to your best model without market segmentation, and evaluate the user benefit of reducing the train headway (TRAIN\_FR) by half.

```
levels(factor(sm$fr))
## [1] "0" "10" "20" "30" "60" "120"
m1.93 <- mlogit(choice~1+fr+seats|ga+male+age+luggage+first|tt+co, sm.av, alt.subset=c("train", "sm"))
m1.94 <- mlogit(choice~1+I(fr*0.5)+seats|ga+male+age+luggage+first|tt+co, sm.av, alt.subset=c("train", "sm"))
coef(m1.94)
## (Intercept):train      I(fr * 0.5)      seats      ga:train
##      -1.32226      -0.01523      -0.43916      2.11047
```

```
##      male:train      age:train      luggage:train      first:train
##      -0.71311      0.32875      0.20722      -0.31764
##      tt:sm      tt:train      co:sm      co:train
##      -0.00776      -0.00902      -0.00101      -0.00161

beta93 <- Vectorize(coef(m1.93)[c(-9,-11)])
beta94 <- as.vector(coef(m1.94)[c(-9,-11)])
train<- sm.av%>%filter(idx(sm.av, 2)=="train")%>% select(fr,seats,ga,male,age,luggage,first,tt,co) %>% data.frame()
X <- cbind(1,train[, -10])%>%as.matrix()
unique(round(X%*%beta94-X%*%beta93,3))
##      [,1]
## 1 -0.914
## 2 -0.228
## 3 -0.457
```

## References

Bierlaire, M., Axhausen, K., and G, Abay (2001). The acceptance of modal innovation: The case of Swissmetro. Proceedings of the 1st Swiss Transportation Research Conference, Ascona, Switzerland. Available at <http://www.strc.ch/conferences/2001/bierlaire1.pdf>