To: Dr. Gregory Macfarlane

From: Dr. Gregory Macfarlane

Subject: Mode Choice Lab Solution

Date: March 21, 2014

There are two parts to this unit's lab. The first part involves estimating multinomial logit models from actual data, and the second demonstrates the application of mode choice models in the Wasatch Front Travel Demand model.

1 Mode Choice Model Estimation

For this week's assignment, you will use data from the 2000 Bay Area Travel Survey to estimate multinomial logit models that predict mode choice for work trips. The data is available on T-Square attached to this assignment. The data are listed as `WorkTrips.Rdata`. We will also need to load the texttmlogit library package, which contains the tools necessary to estimate multinomial logit models.[1]

```
load("WorkTrips.Rdata")
library(mlogit)
```

Because multinomial logit models are so different from other models, we need to use a special function to coerce the data into a form the model software can use. To do this, there needs to be a unique `HH.Person` identification variable. Also, we should change the alternatives from their number to the name of the mode.

```
# Make new id field
worktrips$IDFIELD <- paste(worktrips$HHID, worktrips$PERID, sep = "-")

# Name alternatives
alternatives <- c(" Drive Alone", " Share 2", " Share 3+", " Transit",
    " Bike", " Walk")
worktrips$ALTNUM <- factor(worktrips$ALTNUM, labels = alternatives)

# make mlogit data frame
logitdata <- mlogit.data(worktrips, choice= "CHOSEN", alt.var = "ALTNUM",
                chid.var = "IDFIELD", shape = "long")

head(logitdata[,1:8], n=12)

##                     HHID PERID CASE        ALTNUM NUMALTS CHOSEN  IVTT OVTT
## 2-1. Drive Alone       2     1    1  Drive Alone        5   TRUE 13.38  2.0
## 2-1. Share 2           2     1    1      Share 2        5  FALSE 18.38  2.0
## 2-1. Share 3+          2     1    1     Share 3+        5  FALSE 20.38  2.0
## 2-1. Transit           2     1    1      Transit        5  FALSE 25.90 15.2
## 2-1. Bike              2     1    1         Bike        5  FALSE 40.50  2.0
## 3-1. Drive Alone       3     1    2  Drive Alone        5  FALSE 29.92 10.0
## 3-1. Share 2           3     1    2      Share 2        5  FALSE 34.92 10.0
## 3-1. Share 3+          3     1    2     Share 3+        5  FALSE 21.92 10.0
## 3-1. Transit           3     1    2      Transit        5   TRUE 22.96 14.2
## 3-1. Bike              3     1    2         Bike        5  FALSE 58.95 10.0
## 5-1. Drive Alone       5     1    3  Drive Alone        4   TRUE  8.60  6.0
## 5-1. Share 2           5     1    3      Share 2        4  FALSE 13.60  6.0
```

[1] and other discrete choice structures.

Now that your data is cleaned and formatted, you can estimate multinomial logit models. To do this, use the `mlogit()` function, in a manner sort of like you would use the `lm()` command. One thing to look out for: the difference between generic and alternative-specific variables.[2]

Generic Variables These are coefficients with a single estimated parameter. That is, the $\hat{\beta}$ coefficient for these variables has the same value in the utility equation for every alternative. These estimates come from variables that vary naturally across the alternatives, like the cost of travel.

Alternative-Specific Variables This type of coefficient has a unique estimate for each alternative. That is, $\hat{\beta}_{DA}$ is different from $\hat{\beta}_{Walk}$. This type of estimate comes from variables that are constant across alternatives, like the distance of the trip.

To specify the model, we use the following construction.

```
fit.mnl <- mlogit ( CHOICE ~ Generic | Alt.Specific, data = logitdata )
```

To examine the model output, the standard `summary()` command will produce a coefficients table and a few test statistics. A more robust course in discrete choice analysis would explain these in greater detail, but for our limited purposes you just need to understand that McFadden's R^2 statistic (or ρ^2) is roughly equivalent to that in a linear regression model, and that the t-values associated with the coefficients have the same interpretations. You want to have a high R^2 value and avoid insigificant explanatory variables as much as possible.

1.1 Models to Estimate

For this lab, you will need to estimate and interpret four different models on the BATS dataset.

1. **Value of Time** Estimate a model with just the total travel time (`TVTT`) and the cost of the trip (`COST`). These two parameter estimates will allow you to calculate the value of time for the sample population as

$$VOT = \frac{60\hat{\beta}_{TVTT}}{100\hat{\beta}_{COST}}$$

   Report the value of time you calculate. Is this reasonable?

```
vot.mnl <- mlogit(CHOSEN ~ COST + TVTT, data = logitdata)
# Value of time
vot.mnl$coefficients["TVTT"]/vot.mnl$coefficients["COST"] * 0.6

##  TVTT
## 6.321
```

2. **Ratio of Time** Estimate a model with just the out-of-vehicle travel time (`OVTT`) and the in-vehicle travel time (`IVTT`). What is the ratio of these parameters? What does this tell you about how people feel waiting for the bus?

```
tt.mnl <- mlogit(CHOSEN ~ OVTT + TVTT, data = logitdata)
```

3. **Density** Estimate a model with the residential population density (`RSPOPDEN`) and the workplace employment density (`WKEMPDEN`), controlling for the affordability of the trip (`COSTINC`). Does land use at the origin or the destination of the trip affect the choice problem more?

[2] This can be confusing for many students; just remember that the difference between generic and alternative-specific is in the coefficients, not the variables.

```
dens.mnl <- mlogit(CHOSEN ~ COSTINC | RSPOPDEN + WKEMPDEN, data = logitdata)
```

4. **The Model** The Wasatch Front HBW mode choice model is quite complicated, but the coefficients are given on page 57 of the model documentation (you should read this section). Using variables that you select, try and build a predictive model using data that could be available at the mode choice step in a travel demand model. Try and get a ρ^2 value above 0.25.

```
all.mnl <-  mlogit(CHOSEN ~ IVTT + OVTT + COSTINC | DIST +
        RSEMPDEN + WKEMPDEN + NUMVEH + NUMEMPHH + WKCCBD,
                data = logitdata)
        fixed <- attr(all.mnl$coefficients, 'fixed')
        modelcoefs <- format(round(all.mnl$coefficients[!fixed],digits = 4), format='f')
```

When you have estimated your final model, present it as a system of equations, such as the auto ownership models on page 34 of the Wasatch Front model documentation. Present your other three models in a model parameters table.

$$V_{\text{Auto}} = 0 + -0.0213(\text{IVTT}) + -0.0390(\text{OVTT}) + -0.0358(\text{Cost/Inc})$$

$$\begin{aligned} V_{\text{SR2}} = & -2.8348 + -0.0213(\text{IVTT}) + -0.0390(\text{OVTT}) + -0.0358(\text{Cost/Inc}) \\ & + 0.0060(\text{Distance}) + -0.0001(\text{Res. Emp. Density}) + 0.0007(\text{Work Emp. Density}) \\ & + -0.2017(\text{Num. Vehicles}) + 0.5531(\text{Num. Workers}) + 0.9001(\text{Work in CBD}) \end{aligned}$$

$$\begin{aligned} V_{\text{SR3+}} = & -4.7690 + -0.0213(\text{IVTT}) + -0.0390(\text{OVTT}) + -0.0358(\text{Cost/Inc}) \\ & + 0.0278(\text{Distance}) + -0.0023(\text{Res. Emp. Density}) + 0.0015(\text{Work Emp. Density}) \\ & + -0.1127(\text{Num. Vehicles}) + 0.5950(\text{Num. Workers}) + 1.3885(\text{Work in CBD}) \end{aligned}$$

$$\begin{aligned} V_{\text{Transit}} = & -1.9896 + -0.0213(\text{IVTT}) + -0.0390(\text{OVTT}) + -0.0358(\text{Cost/Inc}) \\ & + 0.0679(\text{Distance}) + 0.0023(\text{Res. Emp. Density}) + 0.0020(\text{Work Emp. Density}) \\ & + -0.7472(\text{Num. Vehicles}) + 0.5076(\text{Num. Workers}) + 1.5920(\text{Work in CBD}) \end{aligned}$$

$$\begin{aligned} V_{\text{Bike}} = & -2.3802 + -0.0213(\text{IVTT}) + -0.0390(\text{OVTT}) + -0.0358(\text{Cost/Inc}) \\ & + -0.1863(\text{Distance}) + 0.0054(\text{Res. Emp. Density}) + 0.0025(\text{Work Emp. Density}) \\ & + -0.3984(\text{Num. Vehicles}) + 0.2876(\text{Num. Workers}) + -0.5797(\text{Work in CBD}) \end{aligned}$$

$$\begin{aligned} V_{\text{Walk}} = & 0.1603 + -0.0213(\text{IVTT}) + -0.0390(\text{OVTT}) + -0.0358(\text{Cost/Inc}) \\ & + -1.2929(\text{Distance}) + 0.0029(\text{Res. Emp. Density}) + 0.0017(\text{Work Emp. Density}) \\ & + -0.5961(\text{Num. Vehicles}) + 0.3632(\text{Num. Workers}) + 0.3232(\text{Work in CBD}) \end{aligned}$$

Submit your analysis as a memorandum in PDF format. Your discussion should consider why certain parameters take certain values and whether your results are reasonable.

Table 1: Mode Choice Models

```
apsrtable(vot.mnl, tt.mnl, dens.mnl, digits = 4,
       model.names = c("Value of Time", "Time Ratio", "Density"),
       Sweave = TRUE, coef.rows = 1)
```

	Value of Time		Time Ratio		Density	
Share 2:(intercept)	-2.3083^*	(0.0547)	-1.9330^*	(0.0518)	-2.3263^*	(0.0659)
Share 3+:(intercept)	-3.7024^*	(0.0929)	-3.0379^*	(0.0851)	-3.5813^*	(0.1332)
Transit:(intercept)	-0.9739^*	(0.0885)	-0.2030^*	(0.1013)	-2.7920^*	(0.0732)
Bike:(intercept)	-3.0705^*	(0.1539)	-2.7384^*	(0.1544)	-3.9779^*	(0.2039)
Walk:(intercept)	-0.7040^*	(0.1293)	-0.8085^*	(0.1647)	-2.5314^*	(0.1174)
COST	-0.0049^*	(0.0002)				
TVTT	-0.0514^*	(0.0031)	-0.0375^*	(0.0040)		
OVTT			-0.0403^*	(0.0070)		
COSTINC					-0.0621^*	(0.0102)
Share 2:RSPOPDEN					-0.0019	(0.0017)
Share 3+:RSPOPDEN					-0.0108^*	(0.0047)
Transit:RSPOPDEN					0.0005	(0.0003)
Bike:RSPOPDEN					0.0150^*	(0.0046)
Walk:RSPOPDEN					0.0047^*	(0.0024)
Share 2:WKEMPDEN					0.0021^*	(0.0004)
Share 3+:WKEMPDEN					0.0039^*	(0.0004)
Transit:WKEMPDEN					0.0056^*	(0.0003)
Bike:WKEMPDEN					0.0015	(0.0011)
Walk:WKEMPDEN					0.0026^*	(0.0006)
N	5029		5029		5029	
LL(C)	-4857.1824		-4857.1824		-4857.1824	
LL($\hat{\beta}$)	-3637.5785		-3907.5062		-3707.4454	
ρ^2_C	0.251093		0.195520		0.236709	

Standard errors in parentheses

* indicates significance at $p < 0.05$