Lecture 5: Maximum Likelihood Estimation (MLE) Method

Anthony Chen
Professor of Transportation Engineering
Department of Civil & Environmental Engineering
The Hong Kong Polytechnic University
Email: anthony.chen@polyu.edu.hk

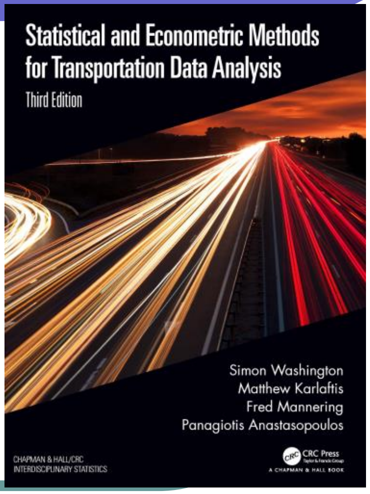
DEPARTMENT OF
CIVIL AND ENVIRONMENTAL ENGINEERING
土木及環境工程學系

Contents

- Maximum Likelihood Estimation
- Binary Logit Estimation
- BIOGEME
- Route Choice Estimation Results and Interpretation

Maximum Likelihood Estimation (MLE) Method

Washington, S., Karlaftis, M., Mannering, F., Anastasopoulos, P., 2020. Statistical and econometric methods for transportation data analysis. Third edition, CRC Press, Taylor and Francis Group, New York, NY.



What is Maximum Likelihood Estimation (MLE) Method?

- Random sample $X_1, X_2, \ldots, X_n$
- Probability function $p(X_k, \theta)$
- Probability density function $f(X_k, \theta)$
- Unknown parameter(s) θ
- MLE is a particular technique for estimating the value(s) of the θ

Maximum Likelihood Estimation

- A standard, classical problem in statistics and econometrics
- Criterion for conformity
  - Real world observations may be based on probability theory
  - Parameter values should maximize the probability or likelihood
- Can be used for both disaggregate and aggregate levels
  - **Disaggregate**: individual observations of traveler characteristics and choices
  - **Aggregate**: zonal characteristics and travel volumes

Likelihood & Log-likelihood Functions

Joint probability (density)

$$L(X,\theta)=\prod_{k=1}^{n} p(X_k,\theta) \qquad L(X,\theta)=\prod_{k=1}^{n} f(X_k,\theta)$$

Multiplication

Log function is monotone; hence, maximizing the log-likelihood function is **equivalent** to maximizing the likelihood function

$$\log L(X,\theta)=\log\prod_{k=1}^{n} f(X_k,\theta)=\sum_{k=1}^{n}\log f(X_k,\theta)$$

Multiplication Summation

The Maximum Likelihood Estimate is the value(s) of the parameter(s) θ which maximize the likelihood function

Parameters of Normal Distribution

When the random sample is Normally distributed

$$f(x_k,\theta)=\frac{1}{\sigma_k\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x_k-\mu_k(\theta)}{\sigma_k}\right)^2\right]$$

$$\log L(X,\theta)=\sum_{k=1}^{n}\log\left(\frac{1}{\sigma_k\sqrt{2\pi}}\right)-\frac{1}{2}\sum_{k=1}^{n}\left(\frac{X_k-\mu_k(\theta)}{\sigma_k}\right)^2$$

Score and Observed Information

The first derivative of the log-likelihood is the ***score***

$$s(X,\theta) = \frac{d}{d\theta} \log L(X,\theta)$$

The second derivative of the ***negative*** log-likelihood function is called the ***observed information***

$$I(X,\theta) = -\frac{d^2}{d\theta^2} \log L(X,\theta)$$

From elementary calculus, MLEs *maximize* the log-likelihood function so

- The ***score*** (derivative) must be *zero* => $s(X,\theta) = 0$ (or $\theta = g(X)$)
- The second derivative must be *negative*
- ***Observed information*** must be *positive*

Data are *random*

- Garbage in garbage out
- Random in random out
- Data X are random
- So *any* estimate $\theta = g(X)$ of θ must also be random
- In particular, MLEs, the score and observed information are all random

Statistical Properties of MLE

Expected Value $$\bar{\theta} = E[g(X)] = \int_{X_1} \dots \int_{X_n} g(X)L(X,\theta)dX_1 \dots dX_n$$

Variance $$\sigma^2 = \operatorname{var} \ \theta = E\left[\left(\theta - \bar{\theta}\right)^2\right] = E\left[\theta^2 - \bar{\theta}^2\right]$$

Unbiased Estimate $$\bar{\theta} = E[g(X)] = \theta$$

Minimum Variance Bound $$\sigma^2 = \operatorname{var}\theta \geq \frac{1}{E[I(X,\theta)]}$$

Observed information

Computing Multi-Parameter MLEs: Newton's Method

- Multi-parameter MLEs are computed by solving the *system* of nonlinear equations

$$s(X,\theta)=0$$

Setting the ***score*** (derivatives) to *zero*

- **Newton's method**: Idea is to approximate non-linear equations by linear system

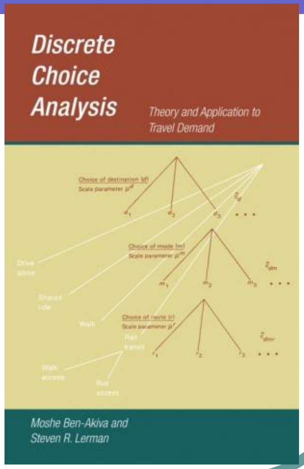
$$0 \approx s\left(X,\theta^{(0)}\right)+G\left(X,\theta^{(0)}\right)\left(\theta-\theta^{(0)}\right)$$

$$\theta \approx \theta^{(0)} - G\left(X,\theta^{(0)}\right)^{-1} s\left(X,\theta^{(0)}\right)$$

Binary Logit Estimation

Ben-Avika, M. and Lerman, S. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press.

See pages 79-93 (also uploaded to BB)



Maximum Likelihood Estimation

(Binary Logit Model) -- two modes

Choice indicator $y_{in} = \begin{cases} 1 & \text{if person n chose alternative i} \\ 0 & \text{if person n chose alternative j} \end{cases}$

Conservation $y_{in} + y_{jn} = 1$

Vector of Attributes x_{in} and x_{jn} (measurable quantities – travel time or cost)

Problem: Given a sample of N observations, the problem is to find the 'best' estimates $\beta_o, \beta_1, \cdots, \beta_k$ that corresponds to the attributes x

Likelihood Function

Likelihood function $$L(\beta_o, \beta_1, \cdots, \beta_k) = \prod_{n=1}^{N} P_n(i)^{y_{in}} P_n(j)^{y_{jn}}$$

Multiplication

where $$P_n(i) = \frac{e^{B^T x_{in}}}{e^{B^T x_{in}} + e^{B^T x_{jn}}} = \frac{1}{1 + e^{B^T (x_{jn} - x_{in})}} = \frac{1}{1 + e^{-B^T x_n}}$$

$$P_n(j) = \frac{e^{B^T x_{jn}}}{e^{B^T x_{in}} + e^{B^T x_{jn}}} = \frac{1}{1 + e^{B^T (x_{in} - x_{jn})}} = \frac{1}{1 + e^{B^T x_n}}$$

where $x_n = x_{in} - x_{jn}$

Maximize the Log Likelihood Function

Take the natural log of the likelihood function

$$\ln\{L(\beta_o, \beta_1, \cdots, \beta_k)\} = \ell n\{L(\boldsymbol{\beta})\} = \sum_{n=1}^{N} [y_{in} \ell n\{P_n(i)\} + y_{jn} \ell n\{P_n(j)\}]$$

To find the 'best' estimates $(\beta_o, \beta_1, \cdots, \beta_k)$, we maximize $\ln\{L\}$ by differentiating it with respect to each β's and setting the partial derivatives to zero.

$$\text{max.} \quad \ln\{L(\hat{\beta}_o,\ \hat{\beta}_1,\ \cdots\ ,\ \hat{\beta}_k)\}$$

$$\frac{\partial \ln\{L(\boldsymbol{\beta})\}}{\partial \hat{\beta}_k} = \sum_{n=1}^{N} \left\{ y_{in} \frac{\partial P_n(i)/\partial \hat{\beta}_k}{P_n(i)} + y_{jn} \frac{\partial P_n(j)/\partial \hat{\beta}_k}{P_n(j)} \right\} = 0 \quad \forall k = 0, 1, \cdots, k$$

First Derivatives

$$\frac{\partial P_n(i)}{\partial \hat{\beta}_k} = \frac{1}{(1+e^{-\beta^T x_n})^2} \cdot e^{-\beta^T x_n} x_{nk}$$ ← k^{th} attribute

$$\frac{\partial P_n(i)/\partial \hat{\beta}_k}{P_n(i)} = \frac{1}{(1+e^{-\beta^T x_n})} \cdot e^{-\beta^T x_n} x_{nk}$$

$$= P_n(j) x_{nk}$$

$$\frac{\partial P_n(j)/\partial \hat{\beta}_k}{P_n(j)} = \frac{1}{(1+e^{\beta^T x_n})} \cdot e^{\beta^T x_n} x_{nk}$$

$$= P_n(i) x_{nk}$$

Slide 14: $P_n(i) = \frac{e^{B^T x_{in}}}{e^{B^T x_{in}} + e^{B^T x_{jn}}} = \frac{1}{1+e^{B^T(x_{jn}-x_{in})}} = \frac{1}{1+e^{-B^T x_n}}$ $\quad P_n(j) = \frac{e^{B^T x_{jn}}}{e^{B^T x_{in}} + e^{B^T x_{jn}}} = \frac{1}{1+e^{B^T(x_{in}-x_{jn})}} = \frac{1}{1+e^{B^T x_n}}$

nonlinear equation from slide 15

$$\frac{\partial \ln\{L(\boldsymbol{\beta})\}}{\partial \beta_k} = \sum_{n=1}^{N} [y_{in} P_n(j) - y_{jn} P_n(i)] x_{nk}$$

$$= \sum_{n=1}^{N} [y_{in}(1 - P_n(i)) - (1 - y_{in}) P_n(i)] x_{nk}$$

$$y_{in} - y_{in} P_n(i) - P_n(i) + y_{in} P_n(i)$$

$$= \sum_{n=1}^{N} [y_{in} - P_n(i)] x_{nk}$$

Conservation

$$P_n(i) + P_n(j) = 1$$

$$y_{in} + y_{jn} = 1$$

Second Derivatives

$$\frac{\partial}{\partial \hat{\beta}_\ell}\left[\frac{\partial \ln\{L(\boldsymbol{\beta})\}}{\partial \hat{\beta}_k}\right] = \sum_{n=1}^{N}\left[0 - \frac{\partial P_n(i)}{\partial \hat{\beta}_\ell}\right] \cdot x_{nk}$$

$$\frac{\partial P_n(i)}{\partial \hat{\beta}_\ell} = \frac{1}{(1+e^{-B^T x_n})^2} \cdot e^{-B^T x_n} \cdot x_{n\ell}$$

$$= \frac{1}{1+e^{-B^T x_n}} \cdot \frac{e^{-B^T x_n}}{1+e^{-B^T x_n}} \cdot x_{n\ell}$$

$$= P_n(i) \cdot P_n(j) \cdot x_{n\ell}$$

$$\frac{\partial^2 \ln\{L(\boldsymbol{\beta})\}}{\partial \hat{\beta}_\ell \partial \hat{\beta}_k} = \sum_{n=1}^{N}[-P_n(i)P_n(j)]x_{n\ell}x_{nk}$$

$$= \sum_{n=1}^{N}[-P_n(i)(1-P_n(i))]x_{n\ell}x_{nk}$$

when $\ell = k$

$$\frac{\partial^2 \ln\{L(\boldsymbol{\beta})\}}{\partial \hat{\beta}_k^2} = \sum_{n=1}^{N}[-P_n(i)(1-P_n(i))]x_{nk}^2$$

Newton-Raphson Method

$m = 0$

Step 0 Choose $\hat{\beta}^m = [\beta_o^m,\ \beta_1^m,\ \cdots,\ \beta_k^m]$ as an initial guess

Step 1 Linearize $\nabla \ln\{L(\hat{\beta}^m)\}$ around $\hat{\beta}^m$

$$\nabla \ln\{L(\hat{\beta}^m)\} + \nabla^2 \ln\{L(\hat{\beta}^m)\}(\hat{\beta} - \hat{\beta}^m) = 0$$

Step 2 Solve the linearized form for

$$\hat{\beta}^{m+1} = \hat{\beta}^m - [\nabla^2 \ln\{L(\hat{\beta}^m)\}]^{-1} \nabla \ln\{L(\hat{\beta}^m)\}$$

Step 3 Check if $\left| \dfrac{\hat{\beta}_k^{m+1} - \hat{\beta}_k^m}{\hat{\beta}_k^m} \right| < \varepsilon \quad (e.g., 10^{-2})$ for all k = 0, 1, ... , k

then terminate with $\hat{\beta}^{m+1}$ as the solution

Otherwise, set m = m + 1 and go to step 1

Data for Binary Logit Example

Table 4.3
Simple binary example

	β_1	β_2
Auto utility, V_{An}	1	Auto travel time (min)
Transit utility, V_{Tn}	0	Transit travel time (min)

$V_A = b_1 + b_2 * X_2$

$V_T = \quad b_2 * X_2$

X_2 = travel time

Table 4.4
Data for simple binary example

Observation number	Auto time	Transit time	Chosen alternative
1	52.9	4.4	Transit
2	4.1	28.5	Transit
3	4.1	86.9	Auto
4	56.2	31.6	Transit
5	51.8	20.2	Transit
6	0.2	91.2	Auto
7	27.6	79.7	Auto
8	89.9	2.2	Transit
9	41.5	24.5	Transit
10	95.0	43.5	Transit
11	99.1	8.4	Transit
12	18.5	84.0	Auto
13	82.0	38.0	Auto
14	8.6	1.6	Transit
15	22.5	74.1	Auto
16	51.4	83.8	Auto
17	81.0	19.2	Transit
18	51.0	85.0	Auto
19	62.2	90.1	Auto
20	95.1	22.2	Transit
21	41.6	91.5	Auto

Observed data: n = 21

Log Likelihood and Estimation Results

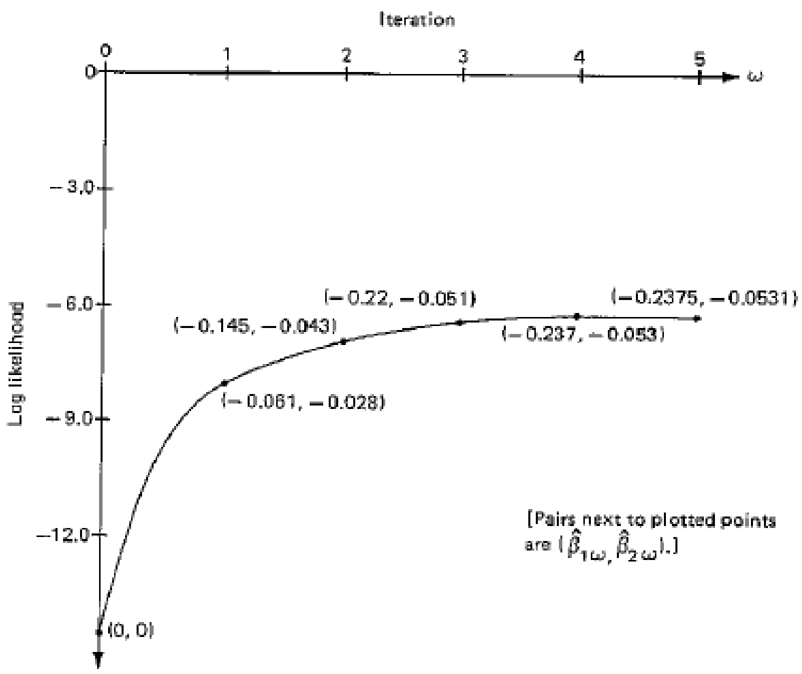


Figure 4.8
Log likelihood values in estimation of simple binary logit example

It takes 6 iterations in the Newton-Raphson method to converge to a stable solution.

Recall: Hypothesis testing in Lecture 2. We used *t* Test/ p-value for testing the significance of coefficients and F-test for the overall regression model.

Table 4.5
Estimation results for simple binary logit example

Variable number	Variable name	Coefficient estimate	Asymptotic standard error	*t* statistic
1	Auto constant	−0.2375	0.7505	−0.32
2	Travel time (min)	−0.0531	0.0206	−2.57

Summary statistics
Number of observations = 21
Number of cases = 21
$\mathscr{L}(0) = -14.556$
$\mathscr{L}(c) = -14.532$
$\mathscr{L}(\hat{\beta}) = -6.166$
$-2[\mathscr{L}(0) - \mathscr{L}(\hat{\beta})] = 16.780$
$-2[\mathscr{L}(c) - \mathscr{L}(\hat{\beta})] = 16.732$
$\rho^2 = 0.576$
$\bar{\rho}^2 = 0.439$

BIOGEME

https://biogeme.epfl.ch/

BIOGEME

- Biogeme is an open-source Python package, that relies on the version 3 of Python designed for the estimation of discrete choice models.
  - Logit
  - Binary probit
  - Nested logit
  - Cross-nested logit
  - Multivariate Extreme Value models
  - etc

BIOGEME

Installation

- Biogeme is an open-source Python package, that relies on the version 3 of Python.
- Install Python.
- Install Biogeme using pip.

```
1 ! pip install biogeme
```

```
Requirement already satisfied: biogeme in c:\users\umer mansoor\miniconda3\lib\site-packages (3.2.8)
Requirement already satisfied: unidecode in c:\users\umer mansoor\miniconda3\lib\site-packages (from biogeme) (1.3.1)
Requirement already satisfied: scipy in c:\users\umer mansoor\miniconda3\lib\site-packages (from biogeme) (1.5.4)
Requirement already satisfied: pandas in c:\users\umer mansoor\miniconda3\lib\site-packages (from biogeme) (1.1.2)
Requirement already satisfied: numpy in c:\users\umer mansoor\miniconda3\lib\site-packages (from biogeme) (1.19.2)
Requirement already satisfied: tqdm in c:\users\umer mansoor\miniconda3\lib\site-packages (from biogeme) (4.48.2)
Requirement already satisfied: cython in c:\users\umer mansoor\miniconda3\lib\site-packages (from biogeme) (0.29.24)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\umer mansoor\miniconda3\lib\site-packages (from pandas->bioge
me) (2.8.1)
Requirement already satisfied: pytz>=2017.2 in c:\users\umer mansoor\miniconda3\lib\site-packages (from pandas->biogeme) (2020.
1)
Requirement already satisfied: six>=1.5 in c:\users\umer mansoor\miniconda3\lib\site-packages (from python-dateutil>=2.7.3->pan
das->biogeme) (1.15.0)
```

Import Packages & Data

```python
import pandas as pd
import biogeme.database as db
import biogeme.biogeme as bio
from biogeme import models
from biogeme.expressions import Beta
```

Import Python Packages

```python
df = pd.read_csv('Ben_akiva_data.csv')
df.head(5)
```

Import Data

Choice selected

Choice Available

	Observ	Auto_time	Transit_time	Choice	Auto_AV	Transit_AV
0	1	52.9	4.4	2	1	1
1	2	4.1	28.5	2	1	1
2	3	4.1	86.9	1	1	1
3	4	56.2	31.6	2	1	1
4	5	51.8	20.2	2	1	1

Data frame

Parameters & Utility Functions

```
ASC_Auto = Beta('ASC_Auto', 0, None, None, 0)
B_TIME = Beta('B_TIME', 0, None, None, 0)
```

Parameters to be estimated

```
Auto_time_SCALED = Auto_time / 100
Transit_time_SCALED = Transit_time / 100
```

Parameters Scaling

```
V1 = ASC_Auto + B_TIME * Auto_time
V2 = B_TIME * Transit_time
```

Utility Functions

ASC_Auto = Alternative specific constant for Auto

B_Time = Parameter to be estimated for auto & Transit time

Estimating Logit Model

```
V = {1: V1, 2: V2}
av = {1: Auto_AV, 2: Transit_AV}
```

Availability conditions for alternatives

```
logprob = models.loglogit(V, av, Choice)
```

Log of Logit Model

```
biogeme = bio.BIOGEME(database, logprob)
biogeme.modelName = 'BEN_AKIVA_LOGIT'
```

Biogeme

```
results = biogeme.estimate()
pandasResults = results.getEstimatedParameters()
pandasResults
```

Estimate Biogeme

	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
ASC_Auto	-0.237573	0.750476	-0.316563	0.751575	0.805174	-0.295058	0.767950
B_TIME	-0.053110	0.020642	-2.572867	0.010086	0.021672	-2.450673	0.014259

Print Results

```
print(results)
```

Print results

```
Results for model BEN_AKIVA_LOGIT
Output file (HTML):                    BEN_AKIVA_LOGIT.html
Nbr of parameters:              2
Sample size:                    21
Excluded data:                  0
Null log likelihood:            -14.55609
Init log likelihood:            -14.55609
Final log likelihood:           -6.166042
Likelihood ratio test (null):           16.7801
Rho square (null):                      0.576
Rho bar square (null):                  0.439
Likelihood ratio test (init):           16.7801
Rho square (init):                      0.576
Rho bar square (init):                  0.439
Akaike Information Criterion:   16.33208
Bayesian Information Criterion: 18.42113
Final gradient norm:            4.608197e-06
ASC_Auto       : -0.238[0.75 -0.317 0.752][0.805 -0.295 0.768]
B_TIME         : -0.0531[0.0206 -2.57 0.0101][0.0217 -2.45 0.0143]
('B_TIME', 'ASC_Auto'): 0.00255 0.165   0.247   0.805   0.0108  0.618   0.233   0.816
```

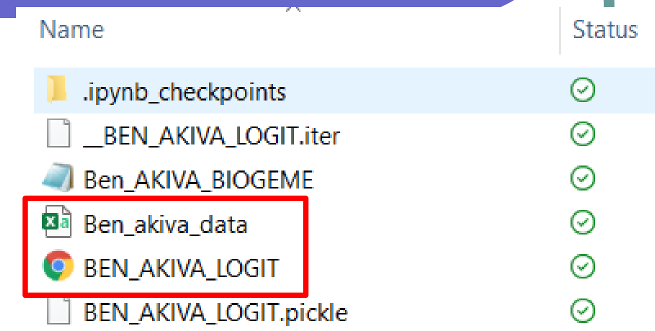
Output Report

Log Likelihood

Estimation Report (1)

Report file: BEN_AKIVA_LOGIT.html
Database name: Ben_akiva_data.csv

Database & Output



Estimation report

Number of estimated parameters:	2
Sample size:	21
Excluded observations:	0
Null log likelihood:	-14.55609
Init log likelihood:	-14.55609
Final log likelihood:	-6.166042
Likelihood ratio test for the null model:	16.7801
Rho-square for the null model:	0.576
Rho-square-bar for the null model:	0.439
Likelihood ratio test for the init. model:	16.7801
Rho-square for the init. model:	0.576
Rho-square-bar for the init. model:	0.439
Akaike Information Criterion:	16.33208
Bayesian Information Criterion:	18.42113
Final gradient norm:	4.6082E-06

Estimation Report (2)

Estimated parameters

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
ASC_Auto	-0.238	0.75	-0.317	0.752	0.805	-0.295	0.768
B_TIME	-0.0531	0.0206	-2.57	0.0101	0.0217	-2.45	0.0143

Correlation of coefficients

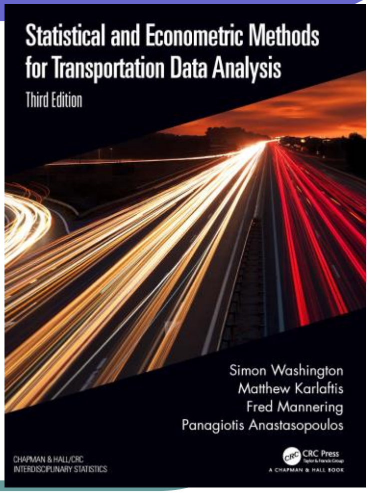
Coefficient1	Coefficient2	Covariance	Correlation	t-test	p-value	Rob. cov.	Rob. corr.	Rob. t-test	Rob. p-value
B_TIME	ASC_Auto	0.00255	0.165	0.247	0.805	0.0108	0.618	0.233	0.816

Biogeme

- More details: http://biogeme.epfl.ch/
  - Source code
  - Tutorial
  - Examples
  - Videos

Multinomial Logit Example: Estimation Results and Interpretation

Washington, S., Karlaftis, M., Mannering, F., Anastasopoulos, P., 2020. Statistical and econometric methods for transportation data analysis. Third edition, CRC Press, Taylor and Francis Group, New York, NY.



Route Choice Example: Problem Statement

A survey of **151 commuters** was conducted. Information was collected on their route selection on their morning trip from **home to work**. All commuters departed from the same origin (a large residential complex in suburban State College, Pennsylvania) and went to work in the downtown area of State College. Distance was measured precisely from parking lot of origin to parking lot of destination so there is a variance in distances among commuters even though they departed and arrived in the same general areas.

Commuters had a choice of **three alternate routes**: a **four-lane arterial** (speed limit=60 km/h, 2 lanes each direction), a **two-lane highway** (speed limit=60 km/h, 1 lane each direction) and a limited access **four-lane freeway** (speed limit=90 km/h, 2 lanes each direction). Each of these three routes shared some common portions for access and egress because, for example, the same road to the downtown area is used by both freeway and two-lane road alternatives since the freeway exits on to the same city street as the two-lane road.

Route Choice Example

Properties of the Multinomial Logit Model: Consider three alternatives: ***arterial***, ***two-lane highway***, and ***freeway***. The model probabilities for these are:

$$P(a) = \frac{e^{V_a}}{e^{V_a} + e^{V_t} + e^{V_f}} \qquad P(t) = \frac{e^{V_t}}{e^{V_a} + e^{V_t} + e^{V_f}} \qquad P(f) = \frac{e^{V_f}}{e^{V_a} + e^{V_t} + e^{V_f}}$$

In these equations the V's are direct utility functions, and reflect the (analyst assumed) linear relation between X's and utility.

$$\mathbf{V = b_0 + b_1 * X_1 + b_2 * X_2 + \ldots}$$

Variable Names & Explanation

Variable Number	Explanation
1	Route chosen: 1 if arterial, 2 if two-lane road, 3 if freeway
2	Traffic flow rate of arterial at time of departure (vehicles per hour)
3	Traffic flow rate of two-lane road at time of departure (vehicles per hour)
4	Traffic flow rate of freeway at time of departure (vehicles per hour)
5	Number of traffic signals on the arterial
6	Number of traffic signals on the two-lane road
7	Number of traffic signals on the freeway
8	Distance on the arterial in kilometers
9	Distance on the two-lane road in kilometers
10	Distance on the freeway in kilometers
11	Seat belts: 1 if wearing, 0 if not
12	Number of passengers in vehicle
13	Commuter age in years: 1 if less than 23, 2 if 24 to 29, 3 if 30 to 39, 4 if 40 to 49, 5 if 50 and over50
14	Gender: 1 if male, 0 if female
15	Marital status: 1 if single, 0 if married
16	Number of children in household (aged 16 or less)
17	Annual household income (US dollars per year): 1 if less than 20,000, 2 if 20,000 to 29,999, 3 if 30,000 to 39,999, 4 if 40,000 to 49,999, 5 if more than 50,000
18	Age of the vehicle used on the trip in years

Multinomial Logit Example

The variables defining utility functions are 1 or 2 groups 1) those that vary across outcome alternatives, and 2) those that do not.

Vary Across Outcomes	Constant Across Outcomes
Number of traffic signals	Income
Distance	Number children
	Number vehicles

Note:

(1) Utility functions are derived on the notion of differences in utility functions—the utility of different outcomes.

(2) If a variable does not vary across outcomes, it can only be entered in the model as (at most I-1) alternative-specific constants.

Estimation Results

Recall: Hypothesis testing in Lecture 2. We used *t* Test/ p-value for testing the significance of coefficients and F-test for the overall regression model.

Independent Variable	Variable Mnemonic	Estimated Parameter	t-statistic
Two-lane road constant		1.65	1.02
Freeway constant		-3.20	-1.16
Variables that vary across alternate outcomes			
Distance on the arterial in kilometers	*Dista*	-0.942	-3.99
Distance on the two-lane road in kilometers	*Distt*	-1.135	-5.75
Distance on the freeway in kilometers	*Distf*	-0.694	-2.24
Variables that do not vary across alternate outcomes			
Male indicator (1 if male commuter, 0 if not - defined for the freeway utility function)	*Male*	0.766	1.19
Vehicle age in years (defined for the two-lane road utility function)	*Vehage*	0.128	1.87
Vehicle age in years (defined for the two-lane freeway utility function)	*Vehage*	0.233	2.75
Number of observations		151	
Log likelihood at zero		-165.89	
Log likelihood at convergence		-92.51	

Interpretation: Alternative-Specific Constants

Based on this table, the estimated utility functions are

$$V_a = -\ 0.942(dista)$$
$$V_t = 1.65 - 1.135(distt) + 0.128(vehage)$$
$$V_f = -3.20 - 0.694(distf) + 0.233(vehage) + 0.764(male)$$

The lack of constant in the arterial function establishes it as a 0 baseline. Thus, all else being equal, the two-lane road is more likely to be selected (with its positive constant) relative to the arterial, and the freeway is less likely to be selected relative to the arterial (with its negative constant). Also, all else being equal, the freeway is less likely to be selected than the two-lane road.

Interpretation: Elasticities

Recall: Elasticity analysis in Lecture 4.

Elasticities are used to help understand and interpret the magnitude of effects. These are given respectively as:

$$E_{x_{ik}}^{P(i)} = \left[1 - P(i)\right] \beta_{ki} x_{ki}$$

Elasticity values are interpreted as the percent effect that a 1% change in x_{ki} has on the outcome probability P(i).

Elasticity with respect to Distance

To determine the elasticity's for distances on the three routes, elasticities are applied over all observations (N =151 commuters). It is found that the **average elasticity for distance** (averaged over N commuters) on the arterial is **-6.48**.

This means for the average commuter a 1% increase in distance on the arterial will decrease the probability of the arterial being selected by 6.48%.

Among the 151 commuters in the sample, **elasticity's range** from a high of -13.47 to a low of -1.03.

The computed average elasticity's for distances on the **two-lane road** and the **freeway** are -3.07 and -6.60, respectively. These findings show that distance on the two-lane road has the least effect on the selection probabilities.

Marginal Rates of Substitution

Marginal rates of substitution are also used to help understand and interpret the magnitude of effects. These are given respectively as:

$$MRS(i)_{ba} = \frac{\beta_{ia}}{\beta_{ib}}$$

Marginal rates of substitution are computed to determine the **relative magnitude** of any two parameters estimated in the model.

MRS between Distance and Vehicle Age

The marginal rate of substitution between distance and vehicle age on the two-lane road is estimated.

The estimated parameters are -0.942 for distance and 0.128 for vehicle age. The marginal rate of substitution between distance and vehicle age is -0.136 km/vehicle-year (0.128/-0.942).

Thus, each year a vehicle ages (which increases the probability of the two-lane route being selected) distance is increased 0.136 km on average while the same route choice probability is maintained.

Questions & Answers

https://www.polyu.edu.hk/cee/people/anthony.chen/