MODE CHOICE

Outline

- Introduction to discrete choice models
- General formulation
- Binary choice models
- Specification
- Model estimation
- Application Case Study

Discrete Choice Introduction (1)

- Discrete or nominal scale data often play a dominant role in transportation
  - because many interesting analyses deal with such data.
- Examples of discrete data in transportation include
  - the mode of travel (automobile, bus, rail transit),
  - place to relocate (urban, sub-urban, local)
  - lane changing (lane to left, right or stay on the same lane)
  - the type or class of vehicle owned, and
  - the type of a vehicular crash (run-off-road, rear-end, head-on, etc.).

Discrete Choice Introduction (2)

- From a conceptual perspective,
  - such data are classified as those involving a behavioral choice (choice of mode or type of vehicle to own) or
  - those simply describing discrete outcomes of a physical event (type of vehicle accident).

Models for Discrete Data

- The concept of discrete choice model is
  - the individual decision maker who, faced with a set of feasible discrete alternatives, selects the one that yields greatest utility
  - A set of discrete alternatives form a choice set
- For a variety of reasons the utility of any alternative is, from the perspective of the analyst, best viewed as a random variable.

Random Utility

- In a random utility model the probability of any alternative *i* being selected by person *n* from choice set C*n* is given by

$$P(i|\mathcal{C}_n) = \Pr(U_{in} \geq U_{jn}, \forall j \in \mathcal{C}_n).$$

- Where
  - *i*, and *j* are two alternatives
  - *Uin*->utility of alternative *i* as perceived by decision maker *n*
  - C*n*-> choice set

Random Utility

- We ignore situations where $U_{in} = U_{jn}$ for any i and j in the choice set because
  - if U_{in} and U_{jn} are continuous random variables then the probability $Pr(U_{in} = U_{jn})$ that they are equal is zero.

- Let us pursue the basic idea further by considering the special case where the choice set C_n contains exactly two alternatives.
  - Such situations lead to what are termed binary choice models.

Random Utility

- For convenience we denote the choice set *Cn* as {*i*, *j*}, where, for example,
  - alternative *i* might be the option of driving to work and
  - alternative *j* would be taking the train.
- The probability of person n choosing *i* is

$$P_n(i) = Pr(U_{in} \geq U_{jn}),$$

- the probability of choosina alternative j is

$$P_n(j) = 1 - P_n(i).$$

Binary Choice

- Let us develop the basic theory of random utility models into a class of operational binary choice models
- A detailed discussion of binary models serves a number of purposes.
  - First, the simplicity of binary choice situations makes it possible to develop a range of practical models, which is more tedious in more complicated choice situations.
  - Second, there are many basic conceptual problems that are easiest to illustrate in the context of binary choice.
  - Many of the solutions can be directly applied to situations with more than two alternatives.

Systematic component and disturbances

- *Uin* and *Ujn* are random variables, we begin by dividing each of the utilities into two additive parts as follows

$$
\begin{aligned}
U_{in} &= V_{in} + \varepsilon_{in}, \\
U_{jn} &= V_{jn} + \varepsilon_{jn}.
\end{aligned}
$$

- Where
  - Vin and Vjn are called the systematic (or representative) components of the utility of i and j;
  - εin and εjn are the random parts and are called the disturbances (or random components).

Systematic component and disturbances

- It is important to stress that Vin and Vjn are functions and are assumed here to be deterministic (i.e., nonrandom).
- The terms εin and εjn may also be functions, but they are random from the observational perspective of the analyst.

Systematic component and disturbances

- Probability that alternative I is selected by decision maker n is

$$\begin{aligned} P_n(i) = \Pr(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}) &= \Pr(\varepsilon_{in} - \varepsilon_{jn} \geq V_{jn} - V_{in}) \\ &= \Pr(\varepsilon_{jn} - \varepsilon_{in} \leq V_{in} - V_{jn}). \end{aligned}$$

- We can see that the absolute levels of V's and ε's do not matter; all that matters is the relative values of the differences

Specification of the Systematic Component

- The first issue in specifying *Vin* and *Vjn* is to ask, what types of variables can enter these functions?
- For any individual *n* any alternative *i* can be represented by a vector of attributes z_{in}.
  - In a choice of travel mode, z_{in} might include travel time, cost, comfort, convenience, and safety.
- It is also useful to characterize the decision maker *n* by another vector of characteristics, which we shall denote by *Sn*.
  - These are often variables such as income, auto ownership, household size, age, occupation, and gender.

Specification of the Systematic Component

- The problem of specifying the functions V_{in} and V_{jn} consists of defining combinations of z_{in}, z_{jn}, and S_n that reflect reasonable hypotheses about the effects of such variables
- It is generally convenient to define a new vector of variables, which includes both z_{in} and S_n.
- We write the vectors $x_{in} = h(z_{in}, S_n)$ and $x_{jn} = h(z_{jn}, S_n)$, where h is a function

Specification of the Systematic Component

- The function h can be as simple as a pure attribute model, with $x_{in} = z_{in}$,
- but can also involve non-trivial interactions of z_{in} with elements of S_n such as price, or travel cost, divided by income, or the log of income minus price.
- Now we can write the systematic components of the utilities of i and j

$$V_{in} = V(x_{in}) \quad \text{and} \quad V_{jn} = V(x_{jn}).$$

Specification of the Systematic Component

- If we denote $\beta^T = (\beta 1, \beta 2, \ldots, \beta K)$ as the (row) vector of K unknown

$$
\begin{aligned}
V_{in}(x_{in}, \beta) &= \beta^T x_{in} &= \beta_1 x_{in1} + \beta_2 x_{in2} + \cdots + \beta_K x_{inK}, \\
V_{jn}(x_{jn}, \beta) &= \beta^T x_{jn} &= \beta_1 x_{jn1} + \beta_2 x_{jn2} + \cdots + \beta_K x_{jnK}.
\end{aligned}
$$

- When such a linear formulation is adopted, parameters $\beta 1, \ldots, \beta K$ are called coefficients.

Specification of the Systematic Component

- A coefficient appearing in all utility functions is generic,
- And a coefficient appearing in only one utility function is alternative specific.
- Consider a binary mode choice example, where one alternative is auto (A) and the other is transit (T), and where the utility functions are defined as

$$\begin{aligned} V_{An} &= 0.37 - 2.13 t_{An} \\ V_{Tn} &= - 2.13 t_{Tn}. \end{aligned}$$

Specification of the Systematic Component

- In this case it appears as though the auto utility has an additional term equal to 0.37. We can "convert" this model into the form of equation by defining our x's as follows

$$
\begin{aligned}
x_{An1} &= 1, \\
x_{Tn1} &= 0, \\
x_{An2} &= t_{An}, \\
x_{Tn2} &= t_{Tn},
\end{aligned}
$$

- with K = 2, β1 = 0.37 is alternative specific, and β2 = −2.13 is generic. Thus

$$
\begin{aligned}
V_{An} &= \beta^T x_{An} = \beta_1 x_{An1} + \beta_2 x_{An2} = 0.37 - 2.13 t_{An}, \\
V_{Tn} &= \beta^T x_{Tn} = \beta_1 x_{Tn1} + \beta_2 x_{Tn2} = - 2.13 t_{Tn}.
\end{aligned}
$$

Specification of the Systematic Component

- In this example, the variable xAn1 is an alternative specific (i.e., auto) dummy variable and $\beta 1$ is called an alternative specific constant.

Linearity in Parameters

- A model with a linear-in-parameter formulation can be described in a specification table.
- A specification table has
  - as many columns as alternatives in the model (two in the specific context of binary choice), and
  - as many rows as coefficients (K).
  - Entry (k, i) of the table contains xik, the variable k for alternative i.

		Auto	Train
β_1	0.37	1	0
β_2	-2.13	t_{An}	t_{Tn}

Linearity in Parameters

- Linearity in the parameters is not as restrictive an assumption as one might first think. Linearity in the parameters is not equivalent to linearity in the variables z and S.
- We allow for any function h of the variables so that polynomial, piecewise linear, logarithmic, exponential, and other transformations of the attributes are valid for inclusion as elements of x.

Linearity in Parameters

- We note that we have implicitly assumed that the parameters β1, β2,. . . , βK are the same for all members of the population.
- Again this is not as restrictive as it may seem at first glance.
- If different socioeconomic groups are believed to have entirely different parameters β, then it is possible to develop a distinct model for each subgroup.
- This is termed market segmentation.

Linearity in Parameters

- In the extreme case a market segment corresponds to a single individual, and a vector of parameters is specific to an individual.
- In addition, if the preferences or tastes of different members of the population vary systematically with some known socioeconomic characteristics, we can define some of the elements in x to reflect this.
- For example, it is not unusual to define as a variable
  - cost divided by income, reflecting the a priori belief that the importance of cost declines as the inverse of income.

Specification of the Disturbances

- Our last remaining component of an operational binary choice model is the disturbance terms.
- As with the systematic components Vin and Vjn, we can discuss the specification of binary choice models by considering only the difference εjn −εin rather than each element εin and εjn separately.

Specification of the Disturbances

- This implies that as long as one can add a constant to the systematic component, the means of disturbances can be defined as equal to any constant without loss of generality.
- We can define new random variables

$$\epsilon'_{in}=\varepsilon_{in}-E[\varepsilon_{in}]$$

$$\epsilon'_{jn}=\varepsilon_{jn}-E[\varepsilon_{jn}]$$

Specification of the Disturbances

- Alternatively,

$$\epsilon'_{in} = \varepsilon_{in} - a_{in}$$

$$\epsilon'_{jn} = \varepsilon_{jn} - a_{jn}$$

- So that $E[\epsilon'_{in}] = E[\epsilon'_{jn}] = 0$

Specification of the Disturbances

- The revised utility equation becomes

$$\begin{aligned} U_{in} &= V_{in} + a_{in} + \varepsilon'_{in}, \\ U_{jn} &= V_{jn} + a_{jn} + \varepsilon'_{jn}, \end{aligned}$$

- Where a_{in}, and a_{jn} are unknown constants

- Typically, we assume that the error components εin are identically distributed across n, so that ain = ai and ajn = aj, for all decision makers n, and ai and aj are unknown parameters to be estimated.
- They are called alternative specific constants, and play the same role as intercepts in linear regression.

Specification of the Disturbances

- As only the difference $\varepsilon_{jn} - \varepsilon_{in}$ matters in this context, only the difference between the two constants can be estimated.
- In practice, ***one of the two constants is constrained to 0 and the other one is estimated***:

$$
\begin{aligned}
U_{in} &= V_{in} + a_i && + \varepsilon'_{in}, \\
U_{jn} &= V_{jn} && + \varepsilon'_{jn},
\end{aligned}
$$

$$
\begin{aligned}
U_{in} &= V_{in} && + \varepsilon'_{in}, \\
U_{jn} &= V_{jn} + a_j && + \varepsilon'_{jn}.
\end{aligned}
$$

Illustrative Example

- Let us consider the same example of choosing between auto and transit

$$
\begin{aligned}
U_{An} &= \beta_0 + \beta_1 t_{An} + \beta_2 c_{An}, \\
U_{Tn} &= \beta_1 t_{Tn} + \beta_2 c_{Tn},
\end{aligned}
$$

- Let us consider the traveler has only information about time and not the cost.
- So the cost is added to the error term.
- Depending on what unobserved variables we have the distribution of the error term will change.
- Let us explore more on the functional forms later.

Common Binary Choice Models

- Let us derive operational models by introducing
- the most common binary choice models:
  - the binary probit and
  - the binary logit models.

- In each subsection we begin by making some assumption about the distribution of the two disturbances, εin and εjn, or about the difference between them.
- Given one of these assumptions, we then solve for the probability that alternative i is chosen.

Common Binary Choice Models

- Let us re-specify the random utility model

$$\begin{aligned} P_n(i) &= \Pr(\varepsilon_{jn} - \varepsilon_{in} \leq V_{in} - V_{jn}) \\ &= \Pr(\varepsilon_n \leq V_{in} - V_{jn}), \end{aligned}$$

  - Where $\varepsilon_n = \varepsilon_{in} - \varepsilon_{jn}$

- It means that the probability for individual n to choose alternative i is equal to the probability that the difference Vin – Vjn exceeds the value of εn.
- We need to know how εn is distributed

Common Binary Choice Models

- A function providing the probability that the value of a random variable Ɛn is below a given threshold is called a Cumulative Distribution Function (CDF), and is denoted by FƐn

$$\Pr(\varepsilon_n \leq c) = F_{\varepsilon_n}(c).$$

- The probability expression on the right hand side of utility equation is equal to the cumulative distribution function (CDF) of Ɛn evaluated at Vin − Vjn as follows:

$$P_n(i) = F_{\varepsilon_n}(V_{in} - V_{jn}).$$

- The choice model is obtained by deriving the CDF of Ɛn.

Binary Probit

- One possible assumption is to view the disturbances as the sum of a large number of unobserved but independent components.
  - By the central limit theorem the distribution of the disturbances would tend to be normal.
- To be more specific, suppose that ε_{in} and ε_{jn} are both normal with zero means and variances σ_{2i} and σ_{2j} respectively, and further that they have covariance σij

Binary Probit

- Under these assumptions the term $\varepsilon n = \varepsilon jn - \varepsilon in$ is also normally distributed with mean zero but with variance $\sigma 2i + \sigma 2j - 2\sigma ij = \sigma 2$.
- Note that we implicitly assume here that the random variables $\varepsilon jn - \varepsilon in$ are independent and identically distributed (i.i.d.) across individuals, and independent of the attributes xn.

Binary Probit

- The choice probabilities can be solved as follows:

$$
\begin{aligned}
P_n(i) &= \Pr(\varepsilon_{jn} - \varepsilon_{in} \leq V_{in} - V_{jn}) \\
&= \int_{\varepsilon=-\infty}^{V_{in}-V_{jn}} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon}{\sigma}\right)^2\right] d\varepsilon,\ \sigma > 0, \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(V_{in}-V_{jn})/\sigma} \exp\left[-\frac{1}{2}u^2\right] du \\
&= \Phi\left(\frac{V_{in}-V_{jn}}{\sigma}\right),
\end{aligned}
$$

- Where, $u = \varepsilon/\sigma$, and $\Phi(\cdot)$ denotes the standardized cumulative normal distribution. This model is called binary probability unit or binary probit.

Binary Probit

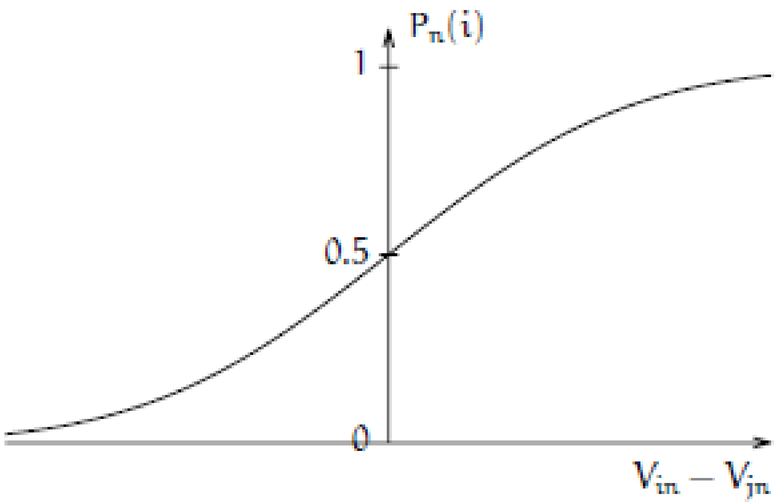
- In the case where $V_{in} = \beta^T x_{in}$ and $V_{jn} = \beta^T x_{jn}$,

$$P_n(i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta^T(x_{in}-x_{jn})/\sigma} \exp\left[-\frac{1}{2}u^2\right] du = \Phi\left(\frac{\beta^T(x_{in}-x_{jn})}{\sigma}\right)$$

- $1/\sigma$ is the scale of the utility function that can be set to an arbitrary positive value, usually $\sigma = 1$

Binary Probit Shape

- Note that the choice function has a characteristic sigmoidal shape and that the choice probabilities are never zero or one.
- They approach zero and one as the systematic components of the utilities become more and more different



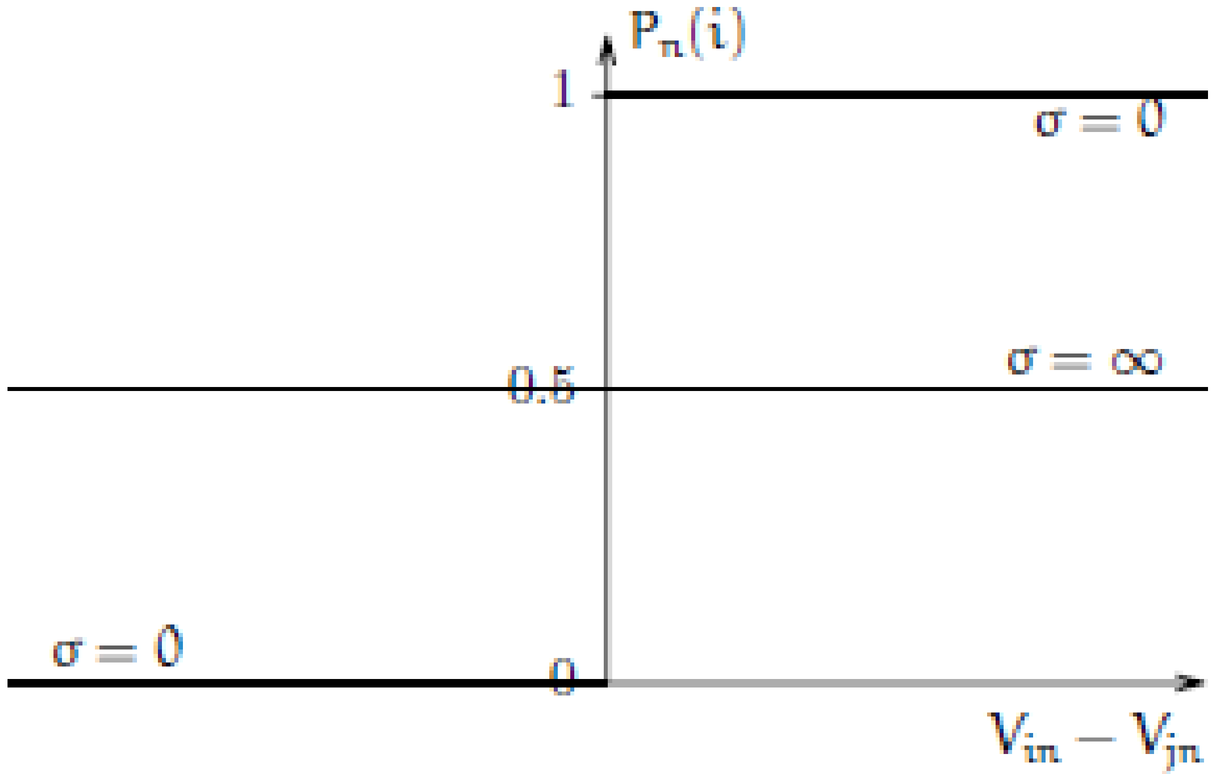
Probit Model: Limiting Case

- There are two limiting cases of a probit model of special interest, both involving extreme values of the scale parameter. The first case is for $\sigma \to 0$:

$$\lim_{\sigma \to 0} P_n(i) = \begin{cases} 1 & \text{if } V_{in} - V_{jn} > 0, \\ 0 & \text{if } V_{in} - V_{jn} < 0; \end{cases}$$

- As $\sigma \to 0$, the choice model is deterministic. On the other hand, when $\sigma \to \infty$, the choice probability of i becomes 1/2. Intuitively the model predicts equal probability of choice for each alternative, irrespectively of Vin and Vjn

Limiting Cases of Binary Probit



Limiting Cases of Binary Probit

- Although binary probit is intuitively reasonable and there are at least some theoretical grounds for its assumptions about the distribution of ε_{in} and ε_{jn}, it has the unfortunate property of not having a closed form.
- Instead, we must express the choice probability as an integral.
- Although it is not really an issue in the binary case, it becomes problematic when we consider more alternatives.

Limiting Cases of Binary Probit

- This aspect of binary probit provides the motivation for searching for a choice model that is more convenient analytically.
- One such model is binary logit.
- Its derivation from the random utility model is justified by viewing the disturbances as the *maximum* of a large number of unobserved but independent utility components.

Extreme Value

- The extreme value distribution, also called Gumbel distribution (Gumbel, 1958) has two forms.
- One is based on the smallest extreme and the other is based on the largest extreme.
- For utility maximization we consider the largest extreme value
- Such distributions are called as Extreme Value Distribution.

Extreme Value

- Similarly to the Central Limit Theorem which justifies the normal distribution as the limit distribution of the sum of many random variables,
- The extreme value distribution is obtained as the limiting distribution of the maximum of many random variables
- The random variable ε is said to be extreme value distributed with location parameter η and scale parameter μ > 0 if its cumulative distribution function (CDF) is given by

$$F(\varepsilon) = e^{\left(-e^{-\mu(\varepsilon-\eta)}\right)}$$

- Pdf

$$f(\varepsilon) = \mu e^{-\mu(\varepsilon-\eta)} e^{\left(-e^{-\mu(\varepsilon-\eta)}\right)}.$$

Extreme Value

- $\varepsilon n = \varepsilon jn - \varepsilon in$ is logistically distributed.

If $\varepsilon_a \sim EV(\eta_a, \mu)$ and $\varepsilon_b \sim EV(\eta_b, \mu)$ are independent with scale parameter μ, then

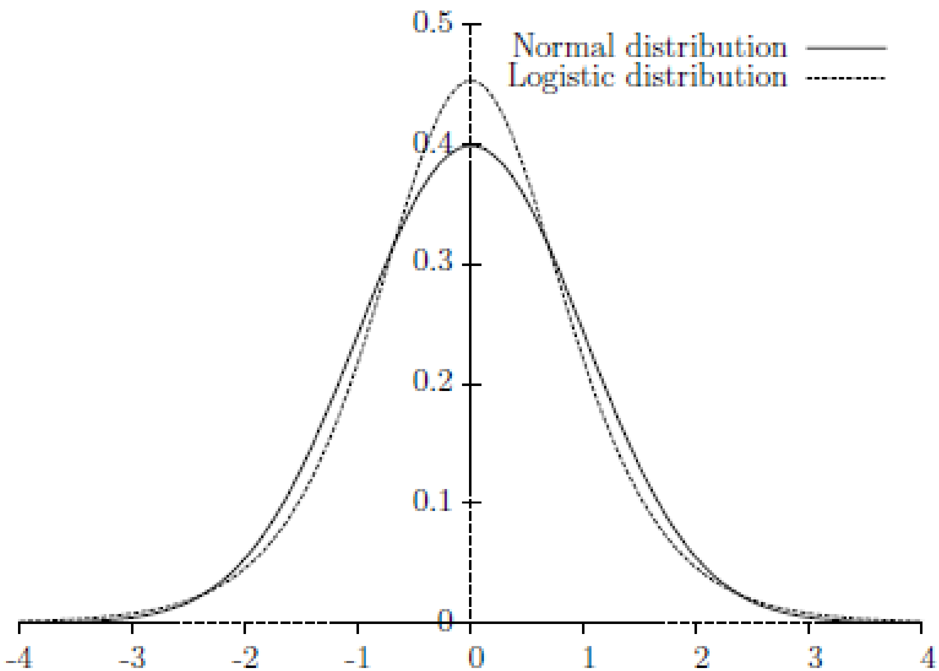
$$\varepsilon = \varepsilon_a - \varepsilon_b \sim \text{Logistic}(\eta_a - \eta_b, \mu),$$

namely

$$\begin{aligned} f(\varepsilon) &= \frac{\mu e^{-\mu(\varepsilon - \eta_a + \eta_b)}}{(1 + e^{-\mu(\varepsilon - \eta_a + \eta_b)})^2}, \\ F(\varepsilon) &= \frac{1}{1 + e^{-\mu(\varepsilon - \eta_a + \eta_b)}}, \ \mu > 0, -\infty < \varepsilon < \infty. \end{aligned}$$

Normal versus Logistic Distribution

- The logistic distribution has heavier tails than the normal
- Normal and logistic distribution with mean 0 and variance 1

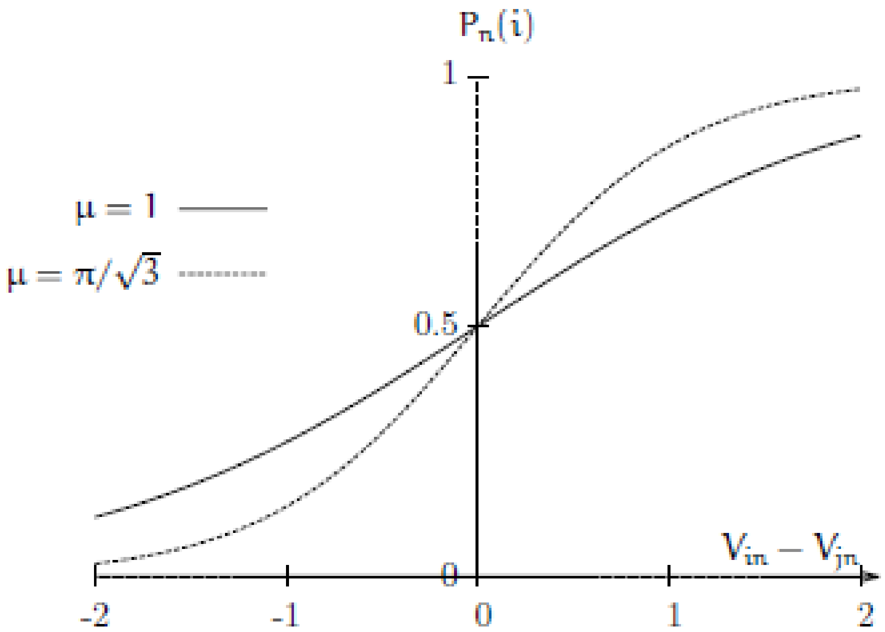


Binary Logit

- For binary logit the choice probability for alternative *i* is given by

$$
\begin{aligned}
P_n(i) &= \Pr(\varepsilon_n \le V_{in} - V_{jn}) \\
&= F(V_{in} - V_{jn}) \\
&= \frac{1}{1 + e^{-\mu(V_{in} - V_{jn})}} \\
&= \frac{e^{\mu V_{in}}}{e^{\mu V_{in}} + e^{\mu V_{jn}}}.
\end{aligned}
$$

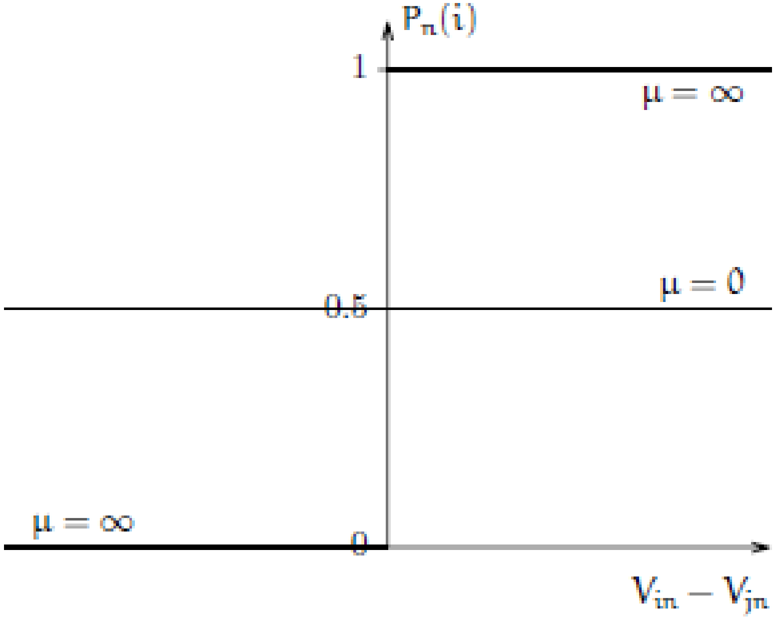
Binary Logit Shape



Limiting Case of Binary Logit

- If Vin and Vjn are linear in their parameters
- μ is the scale parameter

$$P_n(i) = \frac{e^{\mu\beta^T x_{in}}}{e^{\mu\beta^T x_{in}} + e^{\mu\beta^T x_{jn}}} = \frac{1}{1 + e^{-\mu\beta^T (x_{in} - x_{jn})}}.$$



Limiting Case of Binary Logit

- In the case of linear-in-parameters utilities, the parameter μ cannot be distinguished from the overall scale of the β's.
- For convenience we generally make an arbitrary assumption that $\mu = 1$.
- This corresponds to assuming the variances of εin and εjn are both $\pi^2/6$, implying that $Var(\varepsilon jn - \varepsilon in) = \pi^2/3$.

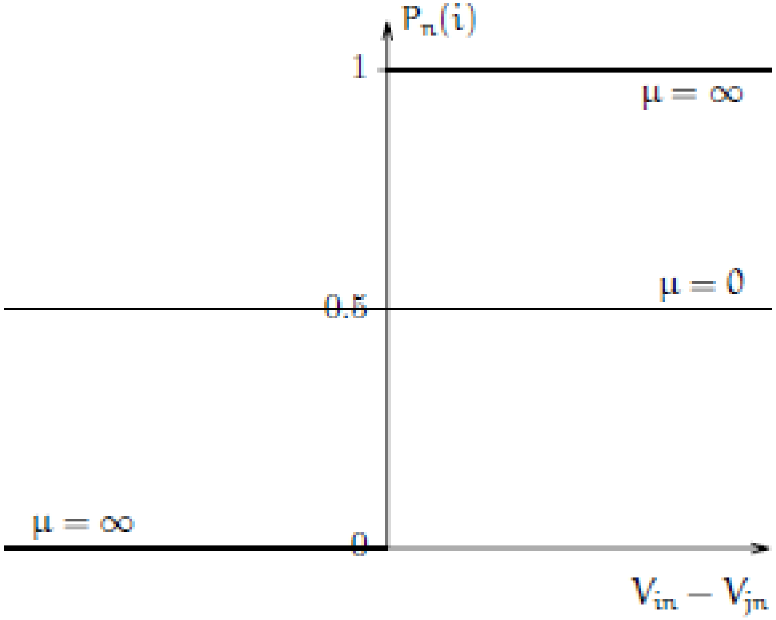
Limiting Case of Binary Logit

- Note that this differs from the standard scaling of binary probit models, where we set $Var(\varepsilon_{jn} - \varepsilon_{in}) = 1$, and it implies that the scaled logit coefficients are $\pi/\sqrt{3}$ times larger than the scaled probit coefficients.
- A rescaling of either the logit or probit utilities is therefore required when comparing coefficients from the two models.

Limiting Case of Binary Logit

- that is, as μ ⟶ ∞, the choice model is deterministic. On the other hand,
- when μ ⟶ 0, the choice probability of i becomes

$$\lim_{\mu \to \infty} P_n(i) = \begin{cases} 1 & \text{if } V_{in} - V_{jn} > 0, \\ 0 & \text{if } V_{in} - V_{jn} < 0; \end{cases}$$



Estimation Approach

- The model coefficients reflect the sensitivity of the behavior to the variables.
- To identify them, we use data on behavioral choices describing individuals, what they faced, and what they chose.
- Therefore, we turn now to the problem of estimating the values of the unknown parameters $\beta 1, \ldots, \beta K$ from a sample of observations.

Estimation Approach

- Each observation consists of the following

An indicator variable defined as

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

- Two vectors of attributes xin = h(zin, Sn) and xjn = h(zjn, Sn), each containing K values of the relevant variables.

Estimation Approach

- Given a sample of N observations, our problem then becomes one of finding estimates $\hat{\beta}_1, \ldots, \hat{\beta}_K$ that have some or all of the desirable properties of statistical estimators.
- We consider in detail the most widely used estimation procedure — maximum likelihood.

Maximum Likelihood

- The maximum likelihood estimation (MLE) procedure is conceptually quite straightforward.
- It consists in identifying the value of the unknown parameters such that the joint probability of the observed choices as predicted by the model is the highest possible.
- This joint probability is called the likelihood of the sample.

Maximum Likelihood

- Consider the likelihood of a sample of N observations assumed to be independently drawn from the population.
- The likelihood of the sample is the product of the likelihoods (or probabilities) of the individual observations
- Let us define the likelihood function as

$$\mathcal{L}^*(\beta_1, \beta_2, \ldots, \beta_K) = \prod_{n=1}^{N} P_n(i)^{y_{in}} P_n(j)^{y_{jn}},$$

- Where, Pn(i) and Pn(j) are functions of β1,. . . ,βK.

Maximum Likelihood

- Note

$$P_n(i)^{y_{in}} P_n(j)^{y_{jn}} = \begin{cases} P_n(i) & \text{if } y_{in} = 1, y_{jn} = 0 \\ P_n(j) & \text{if } y_{in} = 0, y_{jn} = 1. \end{cases}$$

- The log likelihood is written as follows

$$\mathcal{L}(\beta_1, \ldots, \beta_K) = \sum_{n=1}^{N} (y_{in} \ln P_n(i) + y_{jn} \ln P_n(j)),$$

- Noting that

noting that $y_{jn} = 1 - y_{in}$ and $P_n(j) = 1 - P_n(i)$,

Maximum Likelihood

- The log-likelihood function is given by

$$\mathcal{L}(\beta) = \mathcal{L}(\beta_1, \ldots, \beta_K) = \sum_{n=1}^{N} (y_{in} \ln P_n(i) + (1 - y_{in}) \ln(1 - P_n(i))),$$

- Maximize the log-likelihood

$$\max \mathcal{L}(\hat{\beta}) = \mathcal{L}(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_K),$$

- First order conditions

$$\frac{\partial \mathcal{L}}{\partial \beta_k}(\hat{\beta}) = \sum_{n=1}^{N} \left(y_{in} \frac{\partial P_n(i)/\partial \beta_k}{P_n(i)} + y_{jn} \frac{\partial P_n(j)/\partial \beta_k}{P_n(j)} \right) = 0, \ k = 1, \ldots, K,$$

- Or

$$\frac{\partial \mathcal{L}}{\partial \beta}(\hat{\beta}) = 0.$$

Maximum Likelihood

- Each entry k of the vector $\partial L(b\beta)/\partial\beta$ represents the slope of the multi-dimensional log likelihood function along the corresponding kth axis.
- If $b\beta$ corresponds to a maximum of the function, all these slopes must be zero
- Essentially an optimization problem requires efficient techniques to solve for estimates

Example-1: Netherland Mode Choice

- The example deals with mode choice behavior for intercity travelers in the city of Nijmegen (the Netherlands) using revealed preference data.
- The survey was conducted during 1987 for the Netherlands Railways to assess factors that influence the choice between rail and car for intercity travel

Example-1: Netherland Mode Choice

	Car	Train
β_1	1	0
β_2	cost of trip by car (in Guilders)	cost of trip by train (in Guilders)
β_3	travel time by car (hours) if trip purpose is work, 0 otherwise	0
β_4	travel time by car (hours) if trip purpose is not work, 0 otherwise	0
β_5	0	travel time by train (hours)
β_6	0	1 if first class is preferred, 0 otherwise
β_7	1 if commuter is male, 0 otherwise	0
β_8	1 if commuter is the main earner in the family, 0 otherwise	0
β_9	1 if commuter had a fixed arrival time, 0 otherwise	0

Example-1: Netherland Mode Choice

- Coefficient β1 is the alternative specific constant
- β2 is the coefficient of travel cost
- β3 and β4 are coefficients of car travel time.
- β5 is the coefficient of train travel time
- Coefficient β6 measures the impact on the utility of the train if the class preference for rail travel is first class.
- β7, β8 and β9 are coefficients of alternative-specific socioeconomic variables

Example-1: Netherland Mode Choice

- Input data format

	Individual 1	Individual 2	Individual 3
Train cost	40.00	7.80	40.00
Car cost	5.00	8.33	3.20
Train travel time	2.50	1.75	2.67
Car travel time	1.17	2.00	2.55
Gender	M	F	F
Trip purpose	Not work	Work	Not work
Class	Second	First	Second
Main earner	No	Yes	Yes
Arrival time	Variable	Fixed	Variable

Binary Probit

Variables	Coef.	Value	Individual 1		Individual 2		Individual 3	
			Car	Train	Car	Train	Car	Train
Car dummy	β_1	1.77	1	0	1	0	1	0
Cost	β_2	-0.0296	5.00	40.00	8.33	7.80	3.20	40.00
Travel time by car (work)	β_3	-1.51	0	0	2.00	0	0	0
Travel time by car (not work)	β_4	-1.26	1.17	0	0	0	2.55	0
Travel time by train	β_5	-0.308	0	2.50	0	1.75	0	2.67
First class dummy	β_6	0.545	0	0	0	1	0	0
Male dummy	β_7	-0.471	1	0	0	0	0	0
Main earner dummy	β_8	0.213	0	0	1	0	1	0
Fixed arrival time dummy	β_9	-0.355	0	0	1	0	0	0
V_{in}			-0.3120	-1.9551	-1.6354	-0.2252	-1.3126	-2.0065
$P_n(i)$			0.950	0.0502	0.0792	0.921	0.756	0.244

$$
\begin{aligned}
P_1(\text{car}) &= \Pr(-0.3120 + \varepsilon_{\text{car}1} \geq -1.9551 + \varepsilon_{\text{train}1}) \\
&= \Pr(1.6431 \geq \varepsilon_1),
\end{aligned}
$$

Binary Probit

- P1(car) = $\Phi(1.6431) = 0.950$.
- We compute similarly that P2(car) = 0.0792 and P3(car) = 0.756

Binary Logit

			Individual 1		Individual 2		Individual 3	
Variables	Coef.	Value	Car	Train	Car	Train	Car	Train
Car dummy	β_1	3.04	1	0	1	0	1	0
Cost	β_2	-0.0527	5.00	40.00	8.33	7.80	3.20	40.00
Travel time by car (work)	β_3	-2.66	0	0	2	0	0	0
Travel time by car (not work)	β_4	-2.22	1.17	0	0	0	2.55	0
Travel time by train	β_5	-0.576	0	2.50	0	1.75	0	2.67
First class dummy	β_6	0.961	0	0	0	1	0	0
Male dummy	β_7	-0.850	1	0	0	0	0	0
Main earner dummy	β_8	0.383	0	0	1	0	1	0
Fixed arrival time dummy	β_9	-0.624	0	0	1	0	0	0
V_{in}			-0.6642	-3.5504	-2.9596	-0.4589	-2.4072	-3.6464
$P_n(i)$			0.947	0.0528	0.0758	0.924	0.775	0.225

$$P_1(\text{car}) = \frac{e^{-0.6642}}{e^{-0.6642} + e^{-3.5504}} = 0.947,$$

$$P_1(\text{train}) = 1 - P_1(\text{car}) = 0.0528$$

Comparison

- the coefficients of the binary logit must be divided by $\pi/\sqrt{3}$ in order to be compared to the coefficients of the binary probit model

	Logit	Scaled logit	Probit
β_1	3.04	1.68	1.77
β_2	-0.0527	-0.0291	-0.0296
β_3	-2.66	-1.47	-1.51
β_4	-2.22	-1.22	-1.26
β_5	-0.576	-0.318	-0.308
β_6	0.961	0.53	0.545
β_7	-0.85	-0.469	-0.471
β_8	0.383	0.211	0.213
β_9	-0.624	-0.344	-0.355

Review: Log-likelihood function

$$L = \log\left(L^*\right) = \log\left(\prod_{i=1}^{n} \Pr_n(i)^{y_{in}} \Pr_n(j)^{y_{jn}}\right)$$

$$= \sum_{i=1}^{n} \log[\Pr_n(i)^{y_{in}} \Pr_n(j)^{y_{jn}}]$$

$$= \sum_{i=1}^{n} (y_{in} \log[\Pr_n(i)] + y_{jn} \log[\Pr_n(j)])$$

$$= \sum_{i=1}^{n} (y_{in} \log[\Pr_n(i)] + (1 - y_{in}) \log[1 - \Pr_n(i)])$$

Spreadsheet Example-1

Modal Split

Method: Logit Model

Inputs: Travel Time between zones, cost, etc.

Outputs: Trips for each mode of travel

Logit Model:

$$P_i = \frac{\exp(U_i)}{\sum_j \exp(U_j)}$$

P_i = probability of using mode i

U_i = Utility of using mode i

j represents different modes (Auto, HOV, Transit, etc.)

Examples of Logit Model Equations:

$P_{auto} + P_{bus} = 1$

$$P_{auto} = \frac{\exp(U_{auto})}{\exp(U_{auto}) + \exp(U_{bus})}$$

$$P_{bus} = \frac{\exp(U_{bus})}{\exp(U_{auto}) + \exp(U_{bus})}$$

Simple Utility Function:

$$Utility = Beta(TT)$$

We can use survey data to calibrate our utility function, as seen below.

Calibration Process: Without modal constant

δ_{jt} : Which modes they chose

We need to find the Beta coefficient which best predicts traveler choice.

Utility of each mode:

Our Prediction of their Choice:

Survey Data:

Traveler	Auto TT (min)	Bus TT (min)	Chosen Mode_Auto	Chosen Mode_Bus	U_Auto	U_Bus	SUM_Exp(U)	Prob_Auto	Prob_Bus	LL_Auto	LL_Bus
1	30	50	1	0	-1.13	-1.89	0.47	0.68	0.32	-0.38485	0.00000
2	20	10	1	0	-0.76	0.00	1.47	0.32	0.68	-1.14116	0.00000
3	40	30	0	1	-1.51	0.00	1.22	0.18	0.82	0.00000	-0.19912

Log-Likelihood

To find the best Beta coefficient, we use the Solver to maximize the log likelihood function.

Function Variables:

Beta -0.0378154

Optimization Objective:

Obj _LL -1.72513

$$\sum_t \sum_j \delta_{jt} \ln(P_{jt})$$

j = mode

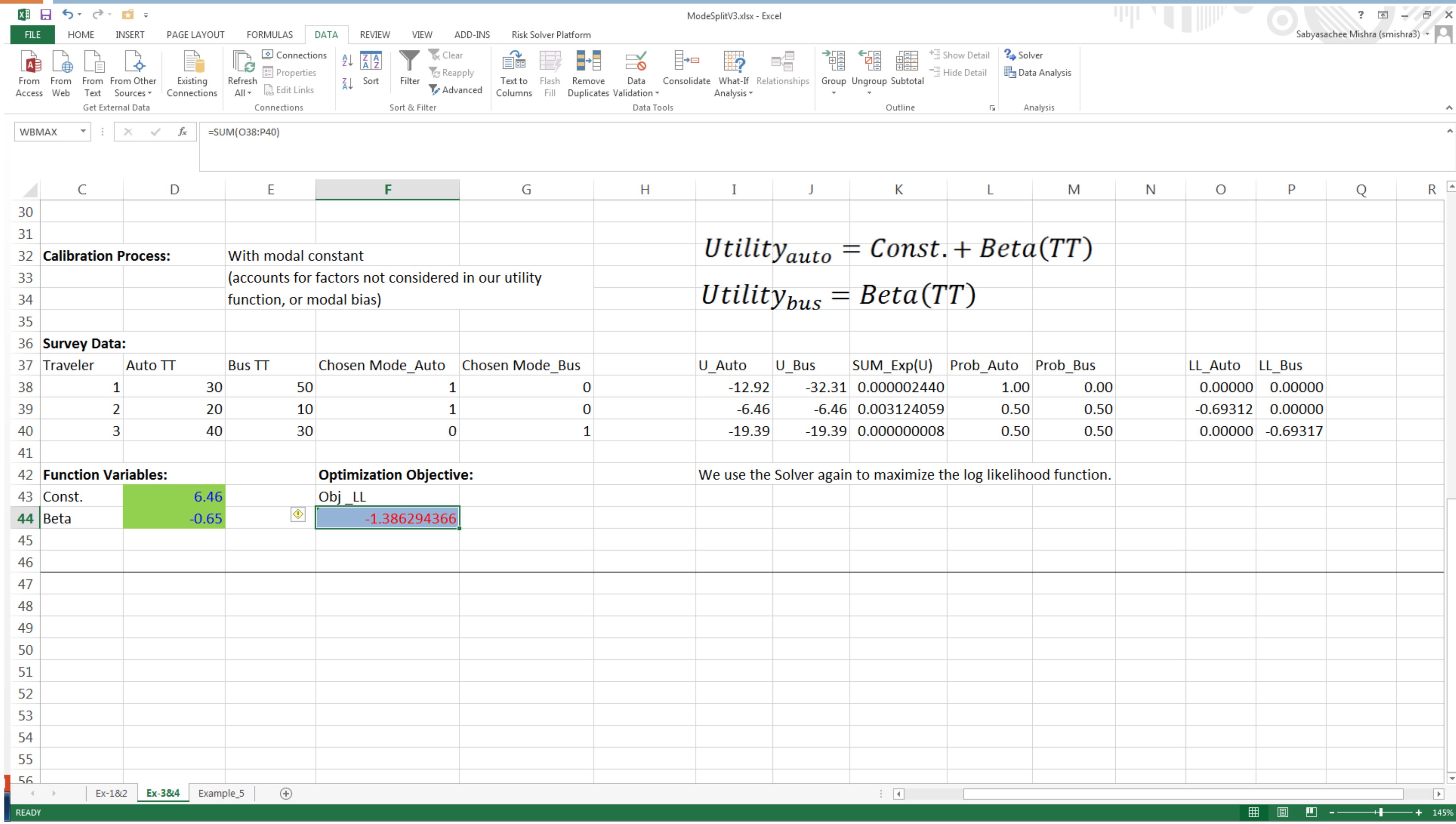
t = traveler

P_{jt} = Probabity of traveler t using mode j

δ_{jt} = 1 if traveler t chose mode j, 0 if they did not

Ex-1&2 | Ex-3&4 | Example_5

Spreadsheet Example-2



Spreadsheet Example-3

ModeSplitV3.xlsx - Excel

E30 =SUM(N4:O28)

Obs.	transit time (tTN)	auto time (tAN)	Choice (Car=1; Rail=0)	Choice (Rail=1; Car=0)	U_Auto	U_Transit	SUM_Exp(U)	Prob_Auto	Prob_Transit	LL_Auto	LL_Bus
1	1.916	1.283	1	0	1.92605	-3.94078	-2.0147342	0.997176	0.00282383	-0.00283	0
2	1.833	1.416	1	0	1.6525	0	1.65249827	0.839228	0.16077159	-0.17527	0
3	2.084	1.417	1	0	1.65044	0	1.6504415	0.838951	0.16104929	-0.1756	0
4	2.25	1.533	1	0	1.41186	0	1.41185542	0.804058	0.19594157	-0.21808	0
5	2.083	1.517	1	0	1.44476	0	1.44476385	0.809191	0.19080872	-0.21172	0
6	1.583	1.583	0	1	1.30902	0	1.3090166	0.787349	0.21265145	0	-1.5481
7	1.25	1.517	1	0	1.44476	0	1.44476385	0.809191	0.19080872	-0.21172	0
8	1.917	2.084	1	0	0.27857	0	0.27857158	0.569196	0.43080401	-0.56353	0
9	2.416	1.283	0	1	1.92605	0	1.92604955	0.872812	0.12718848	0	-2.06209
10	1.75	1.583	1	0	1.30902	0	1.3090166	0.787349	0.21265145	-0.23908	0
11	1.883	1.667	1	0	1.13625	0	1.13624737	0.75699	0.24301001	-0.27841	0
12	2.416	1.583	1	0	1.30902	0	1.3090166	0.787349	0.21265145	-0.23908	0
13	1.717	1.583	0	1	1.30902	0	1.3090166	0.787349	0.21265145	0	-1.5481
14	2.75	2.083	1	0	0.28063	0	0.28062835	0.5697	0.43029973	-0.56264	0
15	1.866	1.583	1	0	1.30902	0	1.3090166	0.787349	0.21265145	-0.23908	0
16	1.834	2.583	0	1	-0.74776	0	-0.7477599	0.32131	0.67869039	0	-0.38759
17	2.5	1.517	1	0	1.44476	0	1.44476385	0.809191	0.19080872	-0.21172	0
18	2.167	1.583	1	0	1.30902	0	1.3090166	0.787349	0.21265145	-0.23908	0
19	1.5	2.333	0	1	-0.23357	0	-0.2335658	0.441873	0.55812743	0	-0.58317
20	2.25	1.434	1	0	1.61548	0	1.6154763	0.83417	0.16582969	-0.18132	0
21	2.25	1.533	1	0	1.41186	0	1.41185542	0.804058	0.19594157	-0.21808	0
22	2.333	1.55	1	0	1.37689	0	1.37689022	0.798491	0.20150891	-0.22503	0
23	2.75	2.583	1	0	-0.74776	0	-0.7477599	0.32131	0.67869039	-1.13535	0
24	2.117	2.25	0	1	-0.06285	0	-0.0628533	0.484292	0.51570816	0	-0.66221
25	2.5	2.033	0	1	0.38347	0	0.38346718	0.594709	0.40529093	0	-0.90315

Sum LL	-13.02205638
Constant	4.564893784
Beta	-2.05677649

Ex-1&2 | Ex-3&4 | Example_5

Estimation Results Goodness-of-fit

- Number of parameters: The number K of estimated parameters
- Number of observations: The number N of observations actually used for the estimation.
- Null log likelihood: the value L(0) of the log likelihood function when all the parameters are zero.
- Constant log likelihood: the value L(c) of the log likelihood function when only an alternative-specific constant is included

Estimation Results Goodness-of-fit

- Final log likelihood: the value of the log likelihood function at its maximum, L(β_hat).
- Likelihood ratio: test statistic used to test the null hypothesis that all the parameters are zero, and
  - is defined as $-2(L(0) - L(\hat{\beta}))$.
  - asymptotically distributed as χ_2 with K degrees of freedom
- Rho-square: Denoted by ρ2, it is an informal goodness-of-fit index that measures the fraction of an initial log likelihood value explained by the model.
  - It is defined as $1 - (L(\hat{\beta})/L(0))$.

Estimation Results Goodness-of-fit

- Adjusted rho-square: Denoted $\bar{\rho}^2$, it is another informal goodness-of-fit measure that is similar to ρ^2 but corrected for the number of parameters estimated.
  - this measure is defined as $\bar{\rho}^2 = 1 - (L(\hat{\beta}) - K)/L(0)$.
- Value: Estimated value $\hat{\beta}_k$.
- Std. Err.: Estimated standard error.
- t-test: Ratio between the estimated value of the parameter and the estimated standard error.
- p-value: Probability of obtaining a t-test at least as large at the one reported, given that the true value of the parameter is 0.

Estimation Results Goodness-of-fit

- Let us take the same example of choice of mode between auto and transit

$$\begin{aligned} V_{An} &= 0.37 - 2.13 t_{An} \\ V_{Tn} &= - 2.13 t_{Tn}. \end{aligned}$$

Number of estimated parameters	:	2
Number of observations	:	25
$\mathcal{L}(0)$	:	-17.329
$\mathcal{L}(c)$	:	-14.824
$\mathcal{L}(\widehat{\beta})$	:	-12.377
$-2(\mathcal{L}(0) - \mathcal{L}(\widehat{\beta}))$	:	9.904
ρ^2	:	0.286
$\bar{\rho}^2$	:	0.170

Standard Representation of Results

- Binary logit

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	Auto constant	0.372	0.492	0.75	0.45
2	Travel time	-2.13	1.22	-1.75	0.08

Summary statistics

Number of observations = 25

$$
\begin{aligned}
\mathcal{L}(0) &= -17.329 \\
\mathcal{L}(c) &= -14.824 \\
\mathcal{L}(\hat{\beta}) &= -12.377 \\
-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] &= 9.904 \\
\rho^2 &= 0.286 \\
\bar{\rho}^2 &= 0.170
\end{aligned}
$$

Mode choice in the Netherlands, Revisited

- Binary probit

Param. number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	Car dummy	1.77	0.632	2.81	0.00
2	Cost	-0.0296	0.00706	-4.20	0.00
3	Travel time by car (work)	-1.51	0.347	-4.35	0.00
4	Travel time by car (not work)	-1.26	0.312	-4.03	0.00
5	Travel time by train	-0.308	0.258	-1.20	0.23
6	First class dummy	0.545	0.414	1.32	0.19
7	Male dummy	-0.471	0.206	-2.29	0.02
8	Main earner dummy	0.213	0.208	1.02	0.31
9	Fixed arrival time dummy	-0.355	0.211	-1.68	0.09

Summary statistics

Number of observations = 228

$$\begin{aligned} \mathcal{L}(0) &= -158.038 \\ \mathcal{L}(c) &= -148.347 \\ \mathcal{L}(\hat{\beta}) &= -109.544 \\ -2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] &= 96.987 \\ \rho^2 &= 0.307 \\ \bar{\rho}^2 &= 0.250 \end{aligned}$$

Mode choice in the Netherlands, Revisited

- Binary logit

Param. number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	Car dummy	3.04	1.09	2.78	0.01
2	Cost	-0.0527	0.0127	-4.17	0.00
3	Travel time by car (work)	-2.66	0.578	-4.60	0.00
4	Travel time by car (not work)	-2.22	0.499	-4.46	0.00
5	Travel time by train	-0.576	0.460	-1.25	0.21
6	First class dummy	0.961	0.768	1.25	0.21
7	Male dummy	-0.850	0.358	-2.37	0.02
8	Main earner dummy	0.383	0.353	1.09	0.28
9	Fixed arrival time dummy	-0.624	0.370	-1.69	0.09

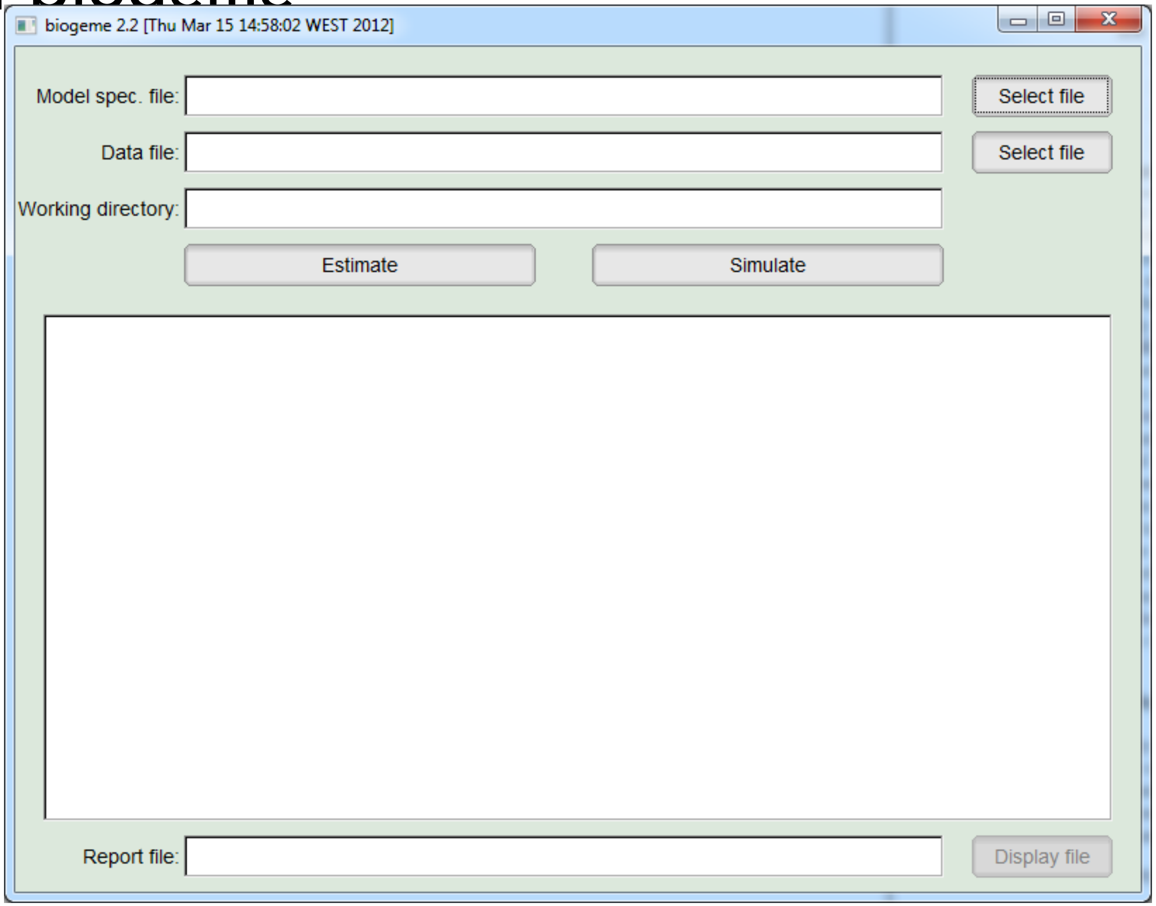
Summary statistics

Number of observations = 228

$$\begin{aligned}
\mathcal{L}(0) &= -158.038 \\
\mathcal{L}(c) &= -148.347 \\
\mathcal{L}(\hat{\beta}) &= -108.836 \\
-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] &= 98.404 \\
\rho^2 &= 0.311 \\
\bar{\rho}^2 &= 0.254
\end{aligned}$$

Software Demonstration: Biogeme

- You can use any software you like
- GUI biogeme



Example Data File

id	choice	rail_cost	rail_time	car_cost	car_time
1	0	40	2.5	5	1.167
2	0	35	2.016	9	1.517
3	0	24	2.017	11.5	1.966
4	0	7.8	1.75	8.333	2
5	0	28	2.034	5	1.267
219	1	35	2.416	6.4	1.283
220	1	30	2.334	2.083	1.667
221	1	35.7	1.834	16.667	2.017
222	1	47	1.833	72	1.533
223	1	30	1.967	30	1.267

Example Model File


```
[Choice]
choice

[Beta]
// Name       DefaultValue LowerBound UpperBound  status
ASC_CAR       0.0          -100.0     100.0         0
ASC_RAIL      0.0          -100.0     100.0         1
BETA_COST     0.0          -100.0     100.0         0
BETA_TIME     0.0          -100.0     100.0         0

[Utilities]
//Id Name Avail linear-in-parameter expression
0    Car  one   ASC_CAR * one + BETA_COST * car_cost +
                BETA_TIME * car_time
1    Rail one   ASC_RAIL * one + BETA_COST * rail_cost +
                BETA_TIME * rail_time

[Expressions]
// Define here arithmetic expressions for name that are not directly
// available from the data
one = 1

[Model]
// Currently, only $MNL (multinomial logit), $NL (nested logit), $CNL
// (cross-nested logit) and $NGEV (Network GEV model) are valid keywords
//
$MNL
```

Example Results

Estimation results				
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	ASC_{car}	2.85	1.02	2.80
2	β_{cost}	-0.130	0.0265	-4.89
3	$\beta_{gender1}$	-0.338	5.80e+06	0.00
4	$\beta_{gender2}$	0.338	5.80e+06	0.00
5	β_{time_car}	-2.34	0.495	-4.73
6	β_{time_rail}	-0.529	0.414	-1.28

Summary statistics

Number of observations $= 228$

$\mathcal{L}(0) = -158.038$

$\mathcal{L}(\hat{\beta}) = -115.880$

$\bar{\rho}^2 = 0.229$

Specification Testing

- Model-1: Generic Attributes
- Model-2: Alternative Specific Attributes
- Model-3: Attributes and Characteristics

Model-1: Generic Attributes

- Model form

$$V_{car} = ASC_{car} + \beta_{time} car_{time} + \beta_{cost} car_{cost}$$
$$V_{rail} = \beta_{time} rail_{time} + \beta_{cost} rail_{cost}$$

- Results:

Estimation results

Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	ASC_{car}	-0.798	0.275	-2.90
2	β_{cost}	-0.113	0.0241	-4.67
3	β_{time}	-1.33	0.354	-3.75

Summary statistics

Number of observations = 228

$\mathcal{L}(0) = -158.038$

$\mathcal{L}(\hat{\beta}) = -123.133$

$\bar{\rho}^2 = 0.202$

Model-2: Alternate Specific Attributes

- Model form
$$V_{car} = ASC_{car} + \beta_{time\_car} car_{time} + \beta_{cost} car_{cost}$$
$$V_{rail} = \beta_{time\_rail} rail_{time} + \beta_{cost} rail_{cost}$$

- Results:

Estimation results

Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	ASC_{car}	2.43	0.973	2.50
2	β_{cost}	-0.123	0.0256	-4.79
3	β_{time_car}	-2.26	0.485	-4.66
4	β_{time_rail}	-0.543	0.396	-1.37^{*}

Summary statistics

Number of observations = 228

$\mathcal{L}(0) = -158.038$

$\mathcal{L}(\hat{\beta}) = -118.023$

$\bar{\rho}^2 = 0.228$

Model-2: Attributes and Characteristics

- Model form

$$V_{car} = ASC_{car} + \beta_{time\_car} car_{time} + \beta_{cost} car_{cost}$$
$$V_{rail} = \beta_{time\_rail} rail_{time} + \beta_{cost} rail_{cost} + \beta_{gender} gender$$

- Results:

Estimation results				
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	ASC_{car}	2.85	1.02	2.80
2	β_{gender}	0.675	0.329	2.05
3	β_{cost}	-0.130	0.0265	-4.89
4	β_{time_car}	-2.34	0.495	-4.73
5	β_{time_rail}	-0.529	0.414	-1.28^{*}

Summary statistics

Number of observations = 228

$\mathcal{L}(0) = -158.038$

$\mathcal{L}(\hat{\beta}) = -115.880$

$\bar{\rho}^2 = 0.235$

Comparison between Generic and Alternate Specific

$$-2(L(\hat{\beta}_G) - L(\hat{\beta}_{AS}))$$

- where G and AS denote the generic and alternative-specific models, respectively.
- It is chi-square distributed with the number of degrees of freedom equal to the number of restrictions (KAS − KG). In this case, −2(−123.133 + 118.023) = 10.220. Since 2 0.95,1 = 3.841 at a 95% level of confidence,
- we can conclude that the model with the alternative-specific coefficients has a significant improvement in fit.

Choice with Multiple Alternatives

- Corresponding logit model is known as multinomial logit model

- Probability:

$$P_n(i) = \frac{e^{\mu V_{in}}}{\sum_{j \in C_n} e^{\mu V_{jn}}},$$

- Bound

$$0 \leq P_n(i) \leq 1, \text{ for all } i \in C_n,$$

- Sum

$$\sum_{i \in C_n} P_n(i) = 1.$$

Netherland Mode Choice Case

- Mode

$$V_{car} = ASC_{car} + \beta_{time} CAR\_TT + \beta_{cost} CAR\_CO$$
$$V_{train} = \beta_{time} TRAIN\_TT + \beta_{cost} TRAIN\_COST + \beta_{he} TRAIN\_HE$$
$$V_{SM} = ASC_{SM} + \beta_{time} SM\_TT + \beta_{cost} SM\_COST + \beta_{he} SM\_HE$$

- Results

Logit model with generic attributes

Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	ASC_{car}	0.189	0.0798	2.37
2	ASC_{SM}	0.451	0.0932	4.84
3	β_{cost}	-0.0108	0.000682	-15.90
4	β_{he}	-0.00535	0.000983	-5.45
5	β_{time}	-0.0128	0.00104	-12.23

Summary statistics

Number of observations = 6768

$\mathcal{L}(0) = -6964.663$

$\mathcal{L}(\hat{\beta}) = -5315.386$

$\bar{\rho}^2 = 0.236$

Interpretation

- Alternate specific constants:
  - The estimated values for the alternative specific constants ASC_{car} and ASC_{SM} show that, all the rest remaining constant, there is a preference in the choice of car and Swissmetro with respect to train.
  - Moreover, the higher value of ASC_{SM} shows a greater preference for Swissmetro compared to car.

Interpretation

- Generic Coefficients:
  - The higher the travel time or the cost of an alternative, the lower the related utility.
- The negative estimate of the headway coefficient beta-$_{he}$ indicates that the higher the headway, the lower the frequency of service, and thus the lower the utility.

MNL Model-Alternate Specific Attributes

- Mode

$$
\begin{aligned}
V_{car} &= ASC_{car} + \beta_{time} CAR\_TT + \beta_{car\_cost} CAR\_CO \\
V_{train} &= \beta_{time} TRAIN\_TT + \beta_{train\_cost} TRAIN\_COST + \beta_{he} TRAIN\_HE \\
V_{SM} &= ASC_{SM} + \beta_{time} SM\_TT + \beta_{SM\_cost} SM\_COST + \beta_{he} SM\_HE.
\end{aligned}
$$

- Result

Logit model with alternative specific travel cost

Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	ASC_{car}	-0.971	0.134	-7.22
2	ASC_{SM}	-0.444	0.102	-4.34
3	β_{car_cost}	-0.00949	0.00116	-8.21
4	β_{he}	-0.00542	0.00101	-5.36
5	β_{SM_cost}	-0.0109	0.000703	-15.49
6	β_{time}	-0.0111	0.00120	-9.26
7	β_{train_cost}	-0.0293	0.00169	-17.32

Summary statistics

Number of observations = 6768

$\mathcal{L}(0) = -6964.663$

$\mathcal{L}(\hat{\beta}) = -5068.559$

$\bar{\rho}^2 = 0.271$

Interpretation

- Alternate Specific Constants:
  - In this model, the ASC's are negative implying a preference, with all the rest constant, for the train alternative.
  - These results are different from those of the previous model where ASCcar and ASCSM were positive and significant.
  - The larger negative value of ASCcar implies that this alternative is more negatively perceived with respect to train than the Swissmetro alternative.

Interpretation

- Alternate Specific Attributes: The influence of the cost is different, showing a larger negative impact on the train alternative with respect to car and Swissmetro.

Generic vs. Alternate Specific Model Comparison

- To test whether a coefficient should be generic or alternative-specific, we use the likelihood ratio test
- The restricted model includes generic travel cost coefficients over the three alternatives, and the unrestricted model includes alternative-specific travel cost coefficients.
- Null hypothesis $H_0 : \beta_{\text{car\_cost}} = \beta_{\text{train\_cost}} = \beta_{\text{SM\_cost}}$

$$-2(\mathcal{L}_R - \mathcal{L}_U)$$

- Test statistic:

Generic vs. Alternate Specific Model Comparison

- Reject the null hypothesis if

$$-2(\mathcal{L}_R - \mathcal{L}_U) > X^2_{((1-\alpha),df)}$$

- For the example case

$$-2(-5315.386 + 5068.559) = 493.654 > 5.991$$

- Reject the null hypothesis and conclude the travel cost coefficient should be alternative-specific

Model Specification with Characteristics

- Mode

$$
\begin{aligned}
V_{\text{car}} &= \text{ASC}_{\text{car}} + \beta_{\text{time}}\text{CAR\_TT} + \beta_{\text{car\_cost}}\text{CAR\_CO} + \beta_{\text{senior}}\text{SENIOR} \\
V_{\text{train}} &= \beta_{\text{time}}\text{TRAIN\_TT} + \beta_{\text{train\_cost}}\text{TRAIN\_COST} + \beta_{\text{he}}\text{TRAIN\_HE} + \\
&\quad \beta_{\text{ga}}\text{GA} \\
V_{\text{SM}} &= \text{ASC}_{\text{SM}} + \beta_{\text{time}}\text{SM\_TT} + \beta_{\text{SM\_cost}}\text{SM\_COST} + \beta_{\text{he}}\text{SM\_HE} + \\
&\quad \beta_{\text{senior}}\text{SENIOR} + \beta_{\text{ga}}\text{GA}
\end{aligned}
$$

- Results

Logit model with socio-economic variables

Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	ASC_{car}	-0.608	0.143	-4.24
2	ASC_{SM}	-0.135	0.106	-1.26
3	$\beta_{\text{car\_cost}}$	-0.00936	0.00117	-8.02
4	β_{he}	-0.00586	0.00106	-5.55
5	$\beta_{\text{SM\_cost}}$	-0.0104	0.000744	-14.02
6	β_{time}	-0.0111	0.00121	-9.20
7	$\beta_{\text{train\_cost}}$	-0.0268	0.00176	-15.24
8	β_{senior}	-1.88	0.109	-17.31
9	β_{ga}	0.557	0.191	2.91

Summary statistics
Number of observations = 6759
$\mathcal{L}(0) = -6958.425$
$\mathcal{L}(\hat{\beta}) = -4927.167$
$\bar{\rho}^2 = 0.291$

Interpretation

- Interpretation (SENIOR)
  - The negative sign of the age coefficient (referring to SENIOR dummy variable) reflects a preference of older individuals for the train alternative
  - It seems a reasonable conclusion, dictated probably by safety reasons with respect to the car choice and a kind of "inertia" with respect to the modal innovation represented by the Swissmetro alternative.

Interpretation

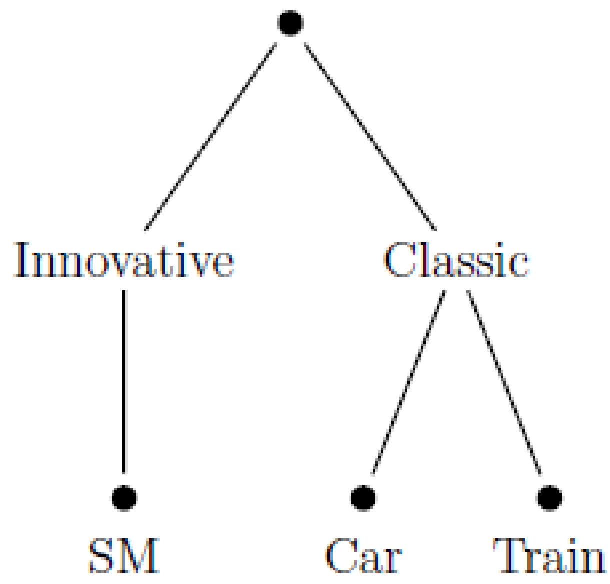
- Interpretation (GA)
  - The coefficient related to the ownership of the Swiss annual season ticket (GA) is positive, as expected. It reflects a preference for the SM and train alternative with respect to car, given that the traveler possesses a season ticket.
- ASCs: The interpretation of the alternative specific constants is similar to that of the previous model specification.

Generalized Extreme Value (GEV) Models

- MNL has IIA properties
  - Remember the blue bus and red bus paradox
- Alternatives
  - Nested Logit
  - Cross Nested Logit

NL Model Example

- Structure



- Model

$$
\begin{aligned}
V_{car} &= ASC_{car} + \beta_{CAR\_time} CAR\_TT + \beta_{cost} CAR\_CO \\
V_{train} &= \beta_{TRAIN\_time} TRAIN\_TT + \beta_{cost} TRAIN\_CO + \beta_{he} TRAIN\_HE + \\
&\quad \beta_{GA} GA \\
V_{sm} &= ASC_{SM} + \beta_{SM\_time} SM\_TT + \beta_{cost} SM\_CO + \beta_{he} SM\_HE \\
&\quad \beta_{GA} GA,
\end{aligned}
$$

NL Model Results

NL model					
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t-stat. 0*	Robust *t-stat. 1*
1	ASC_{car}	0.0272	0.119	0.23	
2	ASC_{SM}	0.243	0.119	2.05	
3	β_{cost}	-0.000986	0.000105	-9.36	
4	β_{car_time}	-0.00874	0.00101	-8.64	
5	β_{train_time}	-0.0113	0.000958	-11.77	
6	β_{SM_time}	-0.00995	0.00163	-6.09	
7	β_{he}	-0.00472	0.000862	-5.48	
8	β_{ga}	5.39	0.582	9.26	
9	$\mu_{classic}$	1.64	0.132	12.42	4.86

Summary statistics

Number of observations = 6759

$\mathcal{L}(0) = -6958.425$

$\mathcal{L}(\hat{\beta}) = -5207.794$

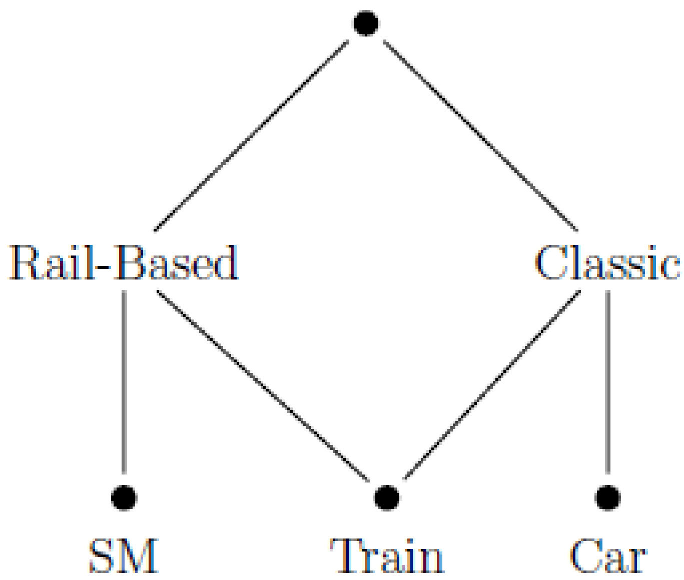
$\bar{\rho}^2 = 0.250$

Interpretation

- ASC:
  - The alternative specific constants show a preference for the Swissmetro alternative compared to the other modes, all the rest remaining constant.

Cross Nested Logit Model

- Structure:



Results

CNL model with fixed α's					
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t-stat. 0*	Robust *t-stat. 1*
1	$\mathrm{ASC}_{\mathrm{car}}$	-0.838	0.0787	-10.65	
2	$\mathrm{ASC}_{\mathrm{SM}}$	-0.457	0.0744	-6.15	
3	β_{cost}	-0.00705	0.000526	-13.39	
4	$\beta_{\mathrm{car\_time}}$	-0.00628	0.00122	-5.17	
5	$\beta_{\mathrm{train\_time}}$	-0.00863	0.00105	-8.18	
6	$\beta_{\mathrm{SM\_time}}$	-0.00715	0.00151	-4.74	
7	β_{he}	-0.00298	0.000533	-5.58	
8	β_{ga}	0.618	0.0940	6.57	
9	μ_{classic}	2.85	0.260	10.93	7.09
10	$\mu_{\mathrm{rail\_based}}$	4.73	0.483	9.78	7.71

Summary statistics

Number of observations = 6759

$\mathcal{L}(0) = -6958.425$

$\mathcal{L}(\hat{\beta}) = -5120.738$

$\bar{\rho}^2 = 0.263$

Typical Steps for Choice Models