INTRODUCTION TO TRANSPORTATION SYSTEMS

Lectures 5/6:

Modeling/Equilibrium/Demand

OUTLINE

1. **Conceptual view of TSA**
2. **Models: different roles and different types**
3. **Equilibrium**
4. **Demand Modeling**

References:

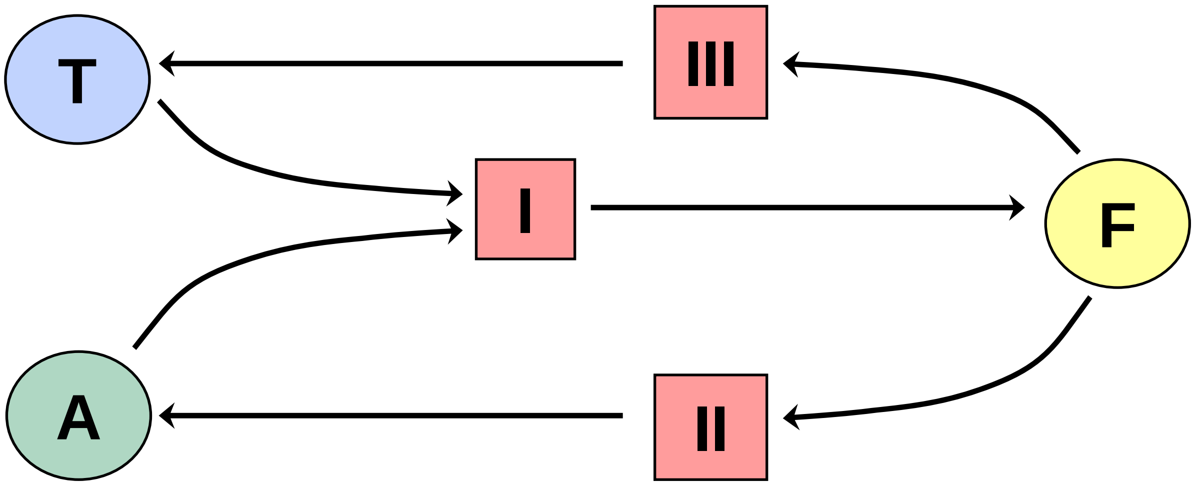
Manheim, *Fundamentals of Transportation Systems Analysis*, Chapter 1

Gomez-Ibañez et al., *Essays in Transportation Economics and Policy*, Chapter 2

CONCEPTUAL VIEW OF TSA

3 elements in transport system problems:

- **Transport system, *T***
- **Activity system, *A***
- **Flow pattern, *F***



CONCEPTUAL VIEW OF TSA

3 types of inter-relationships:

- **Type I:** **Direct interaction between *T* and *A* to produce *F***
  - **The short-run "equlibrium" or outcome**
  - **Many problems are dynamic rather than static**

- **Type II:** **Feedback from *F* to *A***
  - **A is continually in flux with some changes resulting from F**

- **Type III:** **Transport system changes as a result of *F***
  - **Transport operator adds service on a heavily-used route**
  - **New highway link constructed**

MODELS: DIFFERENT ROLES AND DIFFERENT TYPES

- **Models represent real system to predict impacts if specific actions are taken**

Key elements of a model:

- **Control variables: the decision variables**
- **Indirect control variables: these are indirectly affected by decisions**
- **Exogenous variables: known *a priori*, not affected by interactions**
- **Relationships between variables**
- **Parameters or coefficients**

MODELS: DIFFERENT ROLES AND DIFFERENT TYPES

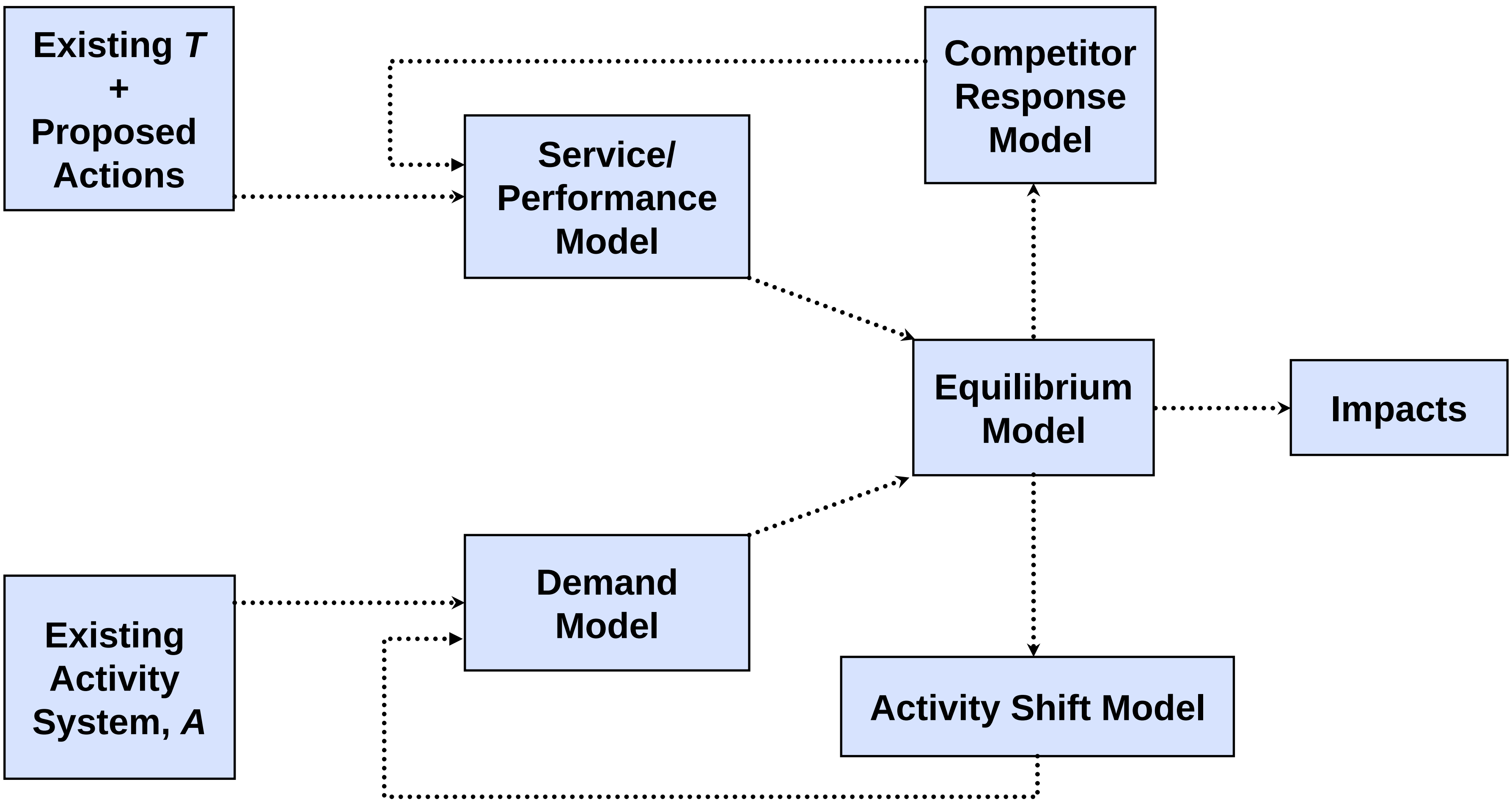
Attributes of a model:

- Complexity
- Accuracy
- Data Requirements
- Computational Requirements
- Estimation Requirements

ROLES FOR MODELS IN TSA

- Performance models: predicts performance or service equality at different flow levels
- Demand models: predicts the flows that result at different levels of service quality and price
- Equilibrium models: predicts *F*, given *T* and *A*, or finds flow which simultaneously satisfies performance and demand relationships
- Activity shift models: predicts changes in *A* over time
- Competitor response models: predicts response by other operators to *F* and changes in *T*

PREDICTION REVISITED



TYPES OF MODELS

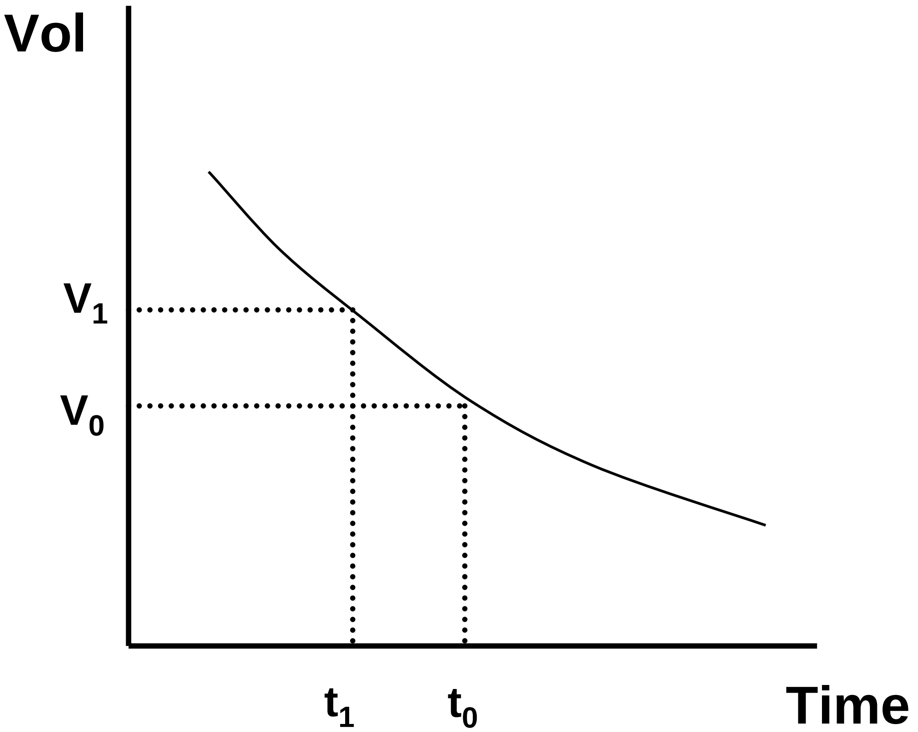
- **Descriptive: typical models for performance and demand**
  - **simulation models**
  - **systems of linear or non-linear equations**
  - **cross-sectional vs time-series**

- **Optimization: used in designing some aspects of the transportation system**
  - **continuous or discrete variables**
  - **linear or non-linear functions**

TRANSPORT DEMAND

Basic premise: transport is a derived demand

Classic simple demand function for a single O-D pair with fixed activity system



DEMAND AND SUPPLY:
Classical Microeconomic View

Market demand function

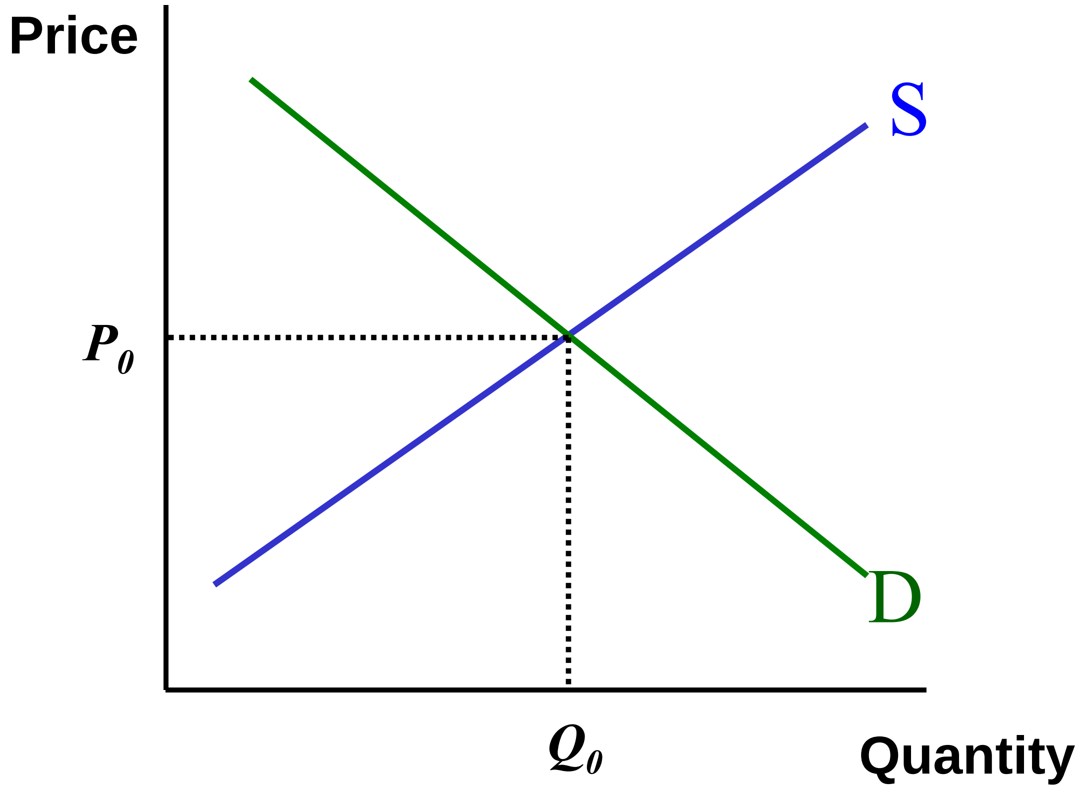
- Represents behavior of users

Market supply function

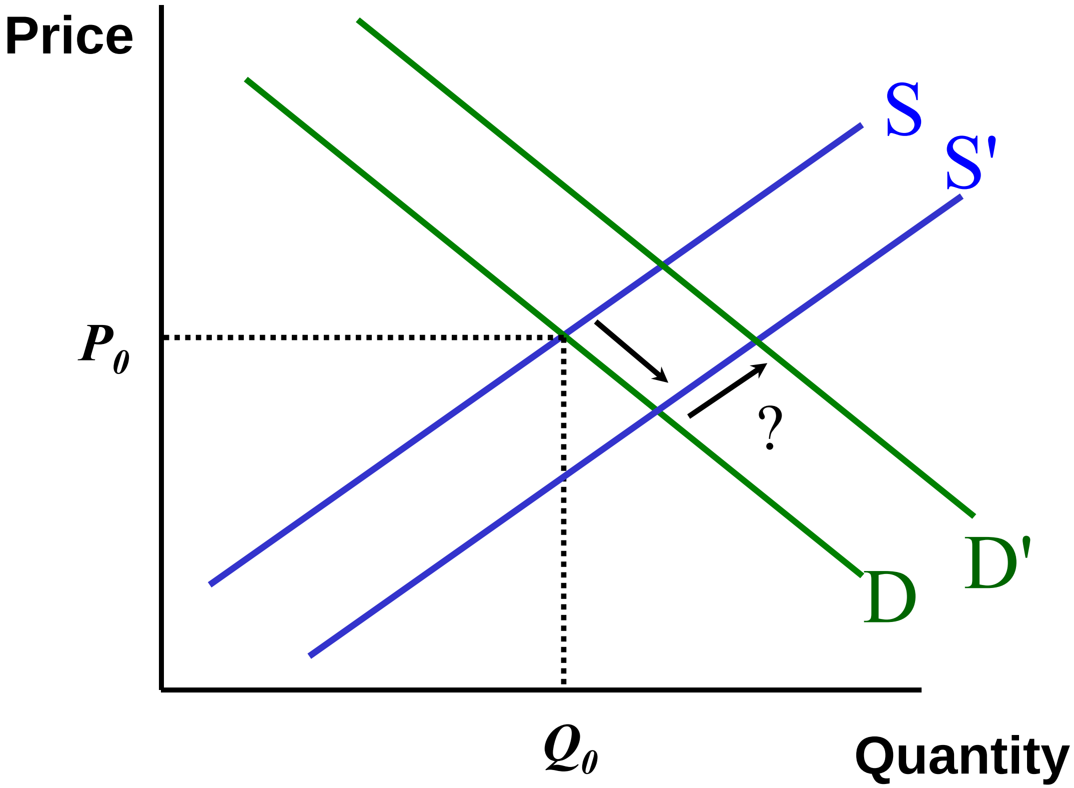
- Represents congestion and behavior of service providers

Supply/Demand Interaction: Equilibrium

EQUILIBRIUM



SHIFTING CURVES



COMPARATIVE STATICS

- **Create a *model* of market behavior:**
  - **Explain consumer and firm choices as a function of exogenous variables, such as income and government policy**
- **Develop scenarios:**
  - **Changes in exogenous variables**
- **Derive changes in the endogenous variables**

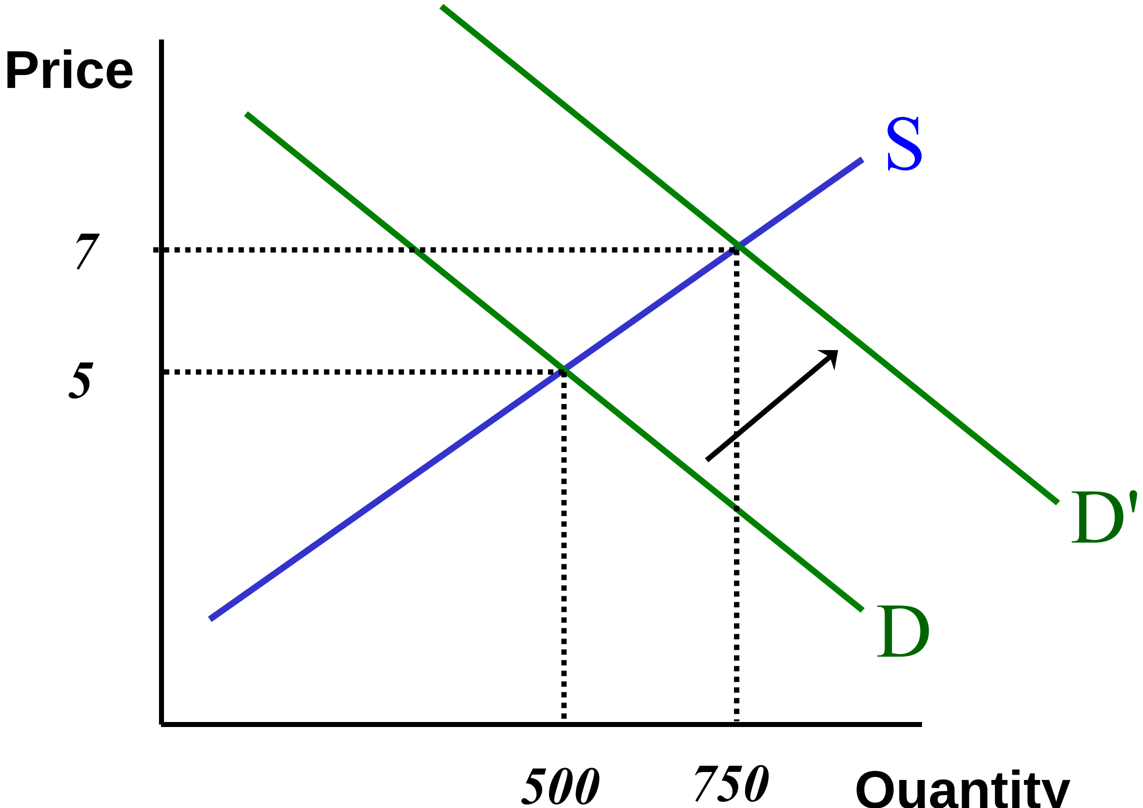
COMPARATIVE STATICS EXAMPLE

The market for taxi service:

- **Supply model: $Q_S = -125 + 125P$**
- **Demand model: $Q_D = 1000 - 100P$**
- **Where does the market clear?**

- **What happens if demand shifts such that now $Q_D = 1450 - 100P$?**

THE SOLUTION



TRANSPORTATION DEMAND ANALYSIS

- **Use models to understand complex processes**
  - **Transit ridership**
  - **Sprawl**
  - **Congestion pricing**
  - **Traveler information systems**
  - **Jobs-housing balance**

- **Assist decision making**

COMPLEXITY OF TRANSPORT DEMAND

- **Valued as input to other activities (derived demand)**
- **Encompasses many interrelated decisions**
  - **Very long-term to very short-term**
- **Large number of distinct services differentiated by location and time**
- **Demographics & socioeconomic matter**
- **Sensitivity to service quality**
- **Supply and demand interact via congestion**

***Complexity* and *Variety* → wide assortment of models to analyze transportation users' behavior.**

CHOICES IMPACTING TRANSPORT DEMAND

- **Decisions made by Organizations**
  - **Firm locates in Boston – Firm locates in Waltham**
  - **Firm invests in home offices, high speed connections**
  - **Developer builds in suburbs – Developer fills in in downtown**
- **Decisions made by Individual/Households**
  - **Live in mixed use area in Boston – Live in residential suburb**
  - **Don't work – Work (and where to work)**
  - **Own a car but not a bike – Own a bike but not a car**
  - **Own an in-vehicle navigation system**
  - **Work Monday-Friday 9-5 – Work evenings and weekends**
  - **Daily activity and travel choices:**
    what, where, when, for how long, in what order, by which mode and route, using what telecommunications

ROLE OF DEMAND MODELS

- Forecasts, parameter estimates, elasticities, values of time, and consumer surplus measures obtained from demand models are used to improve understanding of the ramifications of alternative investment and policy decisions
- Many uncertainties affect transport demand and the models are about to do the impossible

UTILITY FUNCTION

- **A function that represents the consumer's preferences ordering**

- **Utility functions give only an *ordinal* ranking:**
  - **Utility values have no inherent meaning**
  - **Utility function is not unique**
  - **Utility function is unaffected by monotonic transformation**

UTILITY OF A TRANSPORTATION MODE

- **Consumption bundles: auto, bus, train, etc.**
- **Utility function**

$$U_{bus} = \beta_0 + \beta_1 WT_{bus} + \beta_2 TT_{bus} + \beta_3 C_{bus}$$

  - WT_{bus} **-- waiting time (minutes)**
  - TT_{bus} **-- total travel time (minutes)**
  - C_{bus} **-- total cost of trip (dollars)**

- **Parameters β represent tastes, and vary by education, gender, trip purpose, etc.**

TIME BUDGETS AND VALUE OF TIME

- **Along with *income constraint*, there is also a *time constraint* (e.g., 24 hours in a day)**

  → **Gives time *value*.**

- **Value of time is the marginal rate of substitution between time and cost**

$$U_{bus} = \beta_0 + \beta_1 WT_{bus} + \beta_2 TT_{bus} + \beta_3 C_{bus}$$

$$VOT = \frac{MU_{TT}}{MU_C} = \frac{\beta_2}{\beta_3} \ \$/\text{min}$$

VALUE OF TIME

The monetary value of a unit of time for a user.

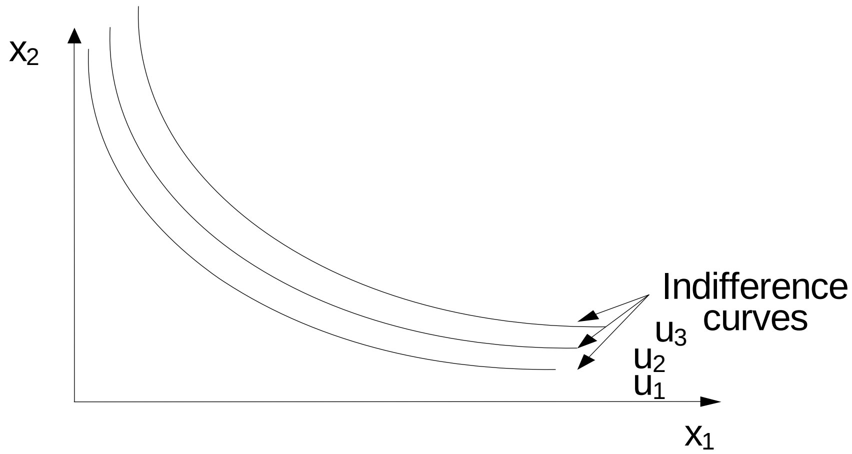
Work Trips (San Francisco)	Auto	Bus	
In-vehicle time	140	76	*Percentage of after tax wage*
Walk access time		273	
Transfer wait time		195	

Vacation Trips (U.S.)	Auto	Bus	Rail	Air	
Total travel time	6	79-87	54-69	149	*Percentage of pretax wage*

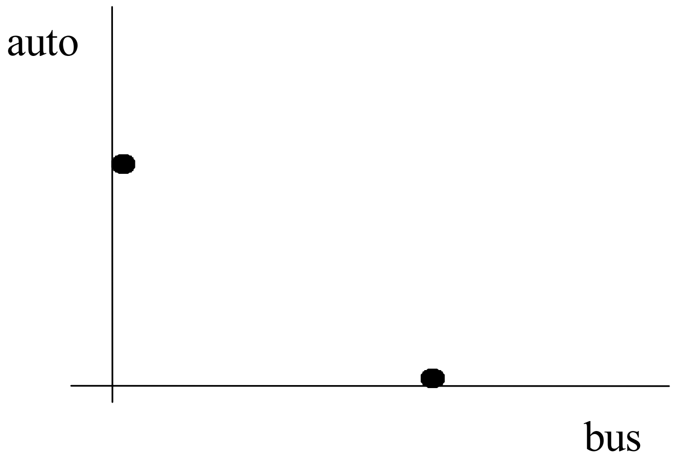
Freight	Rail	Truck	
Total transit time	6-21	8-18	*Percentage of daily shipment value*

CONTINUOUS VS. DISCRETE OPTIONS

Continuous options



Discrete options



DISCRETE CHOICE ANALYSIS

- **Method for modeling choices from among discrete alternatives**
- **Components**
  - **Decision-makers and their socio-economic characteristics**
  - **Alternatives and their attributes**
- **Example: Mode Choice to Work**
  - **Decision maker: Worker**
  - **Characteristics: Income, Age**
  - **Alternatives: Auto and Bus**
  - **Attributes: Travel Cost, Travel Time**

DISCRETE CHOICE FRAMEWORK

- **Decision-Maker**
  - **Individual (person/household)**
  - **Socio-economic characteristics (e.g. Age, gender,income, vehicle ownership)**
- **Alternatives**
  - **Decision-maker *n* selects one and only one alternative from a choice set $C_n=\{1,2,\ldots,i,\ldots,J_n\}$ with J_n alternatives**
  - **Attributes of alternatives (e.g.Travel time, cost)**
- **Decision Rule**
  - **Dominance, satisfaction, utility etc.**

CHOICE: TRAVEL MODE TO WORK

• **Decision maker:**	**an individual worker**
• **Choice:**	**whether to drive to work or take the bus to work**
• **Goods:**	**bus, auto**
• **Utility function:**	$U(X) = U(\text{bus, auto})$
• **Consumption: bundles**	**{1,0} (person takes bus)** **{0,1} (person drives)**

CONSUMER CHOICE

- **Consumers maximize utility**
  - **Choose the alternative that has the maximum utility (and falls within the income constraint)**

    If *U*(bus) > *U*(auto) → choose bus

    If *U*(bus) < *U*(auto) → choose auto

***U*(bus)=?**

***U*(auto)=?**

CONSTRUCTING THE UTILITY FUNCTION

- **Use attribute approach**
- ***U*(bus) = *U*(walk time, in-vehicle time, fare, …)**

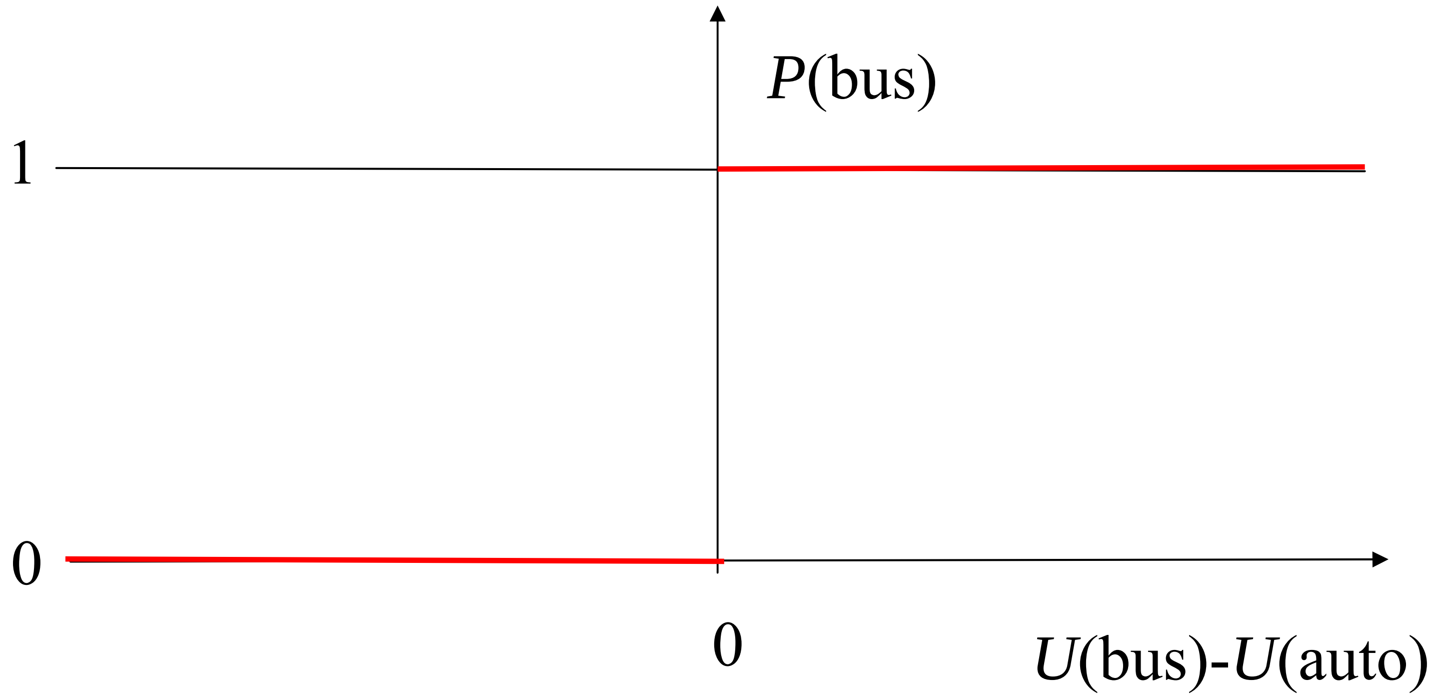
  ***U*(auto) = *U*(travel time, parking cost, …)**
- **Assume linear (in the parameters)**

  $U(\text{bus}) = \beta_1 \times (\text{walk time}) + \beta_2 \times (\text{in-vehicle time}) + \ldots$
- **Parameters represent tastes, which may vary over people.**
  - **-- Include socio-economic characteristics (e.g., age, gender, income)**
  - $U(\text{bus}) = \beta_1 \times (\text{walk time}) + \beta_2 \times (\text{in-vehicle time}) + \beta_3 \times (\text{cost/income}) + \ldots$

DETERMINISTIC BINARY CHOICE

If U(bus) - U(auto) > 0 , Probability(bus) = 1

If U(bus) - U(auto) < 0 , Probability(bus) = 0



PROBABILISTIC CHOICE

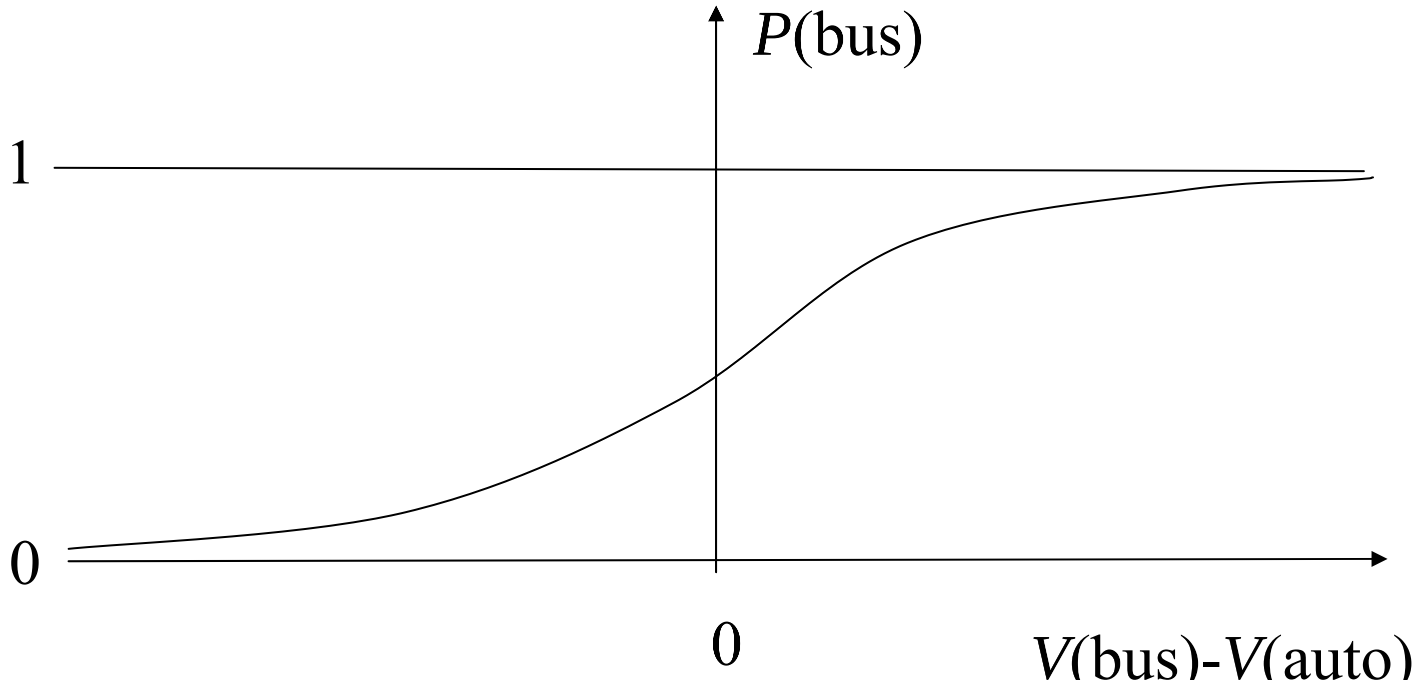
- **'Random' utility**
- **Random utility model**

  U_i = V(attributes of i; parameters) + $epsilon_i$

- **What is in the epsilon?**

  Analysts' imperfect knowledge:
  - **Unobserved attributes**
  - **Unobserved taste variations**
  - **Measurement errors**
  - **Use of proxy variables**
- U(bus) = $\beta_1 \times$(walk time) + $\beta_2 \times$(in-vehicle time + $\beta_3 \times$(cost/income) + … + *epsilon_bus*

PROBABILISTIC BINARY CHOICE



A SIMPLE EXAMPLE: ROUTE CHOICE

- **Sample size: *N* = 600**
- **Alternatives: Tolled, Free**
- **Income: Low, Medium, High**

Route choice	Income			
	Low (k=1)	Medium (k=2)	High (k=3)	
Tolled (i=1)	**10**	**100**	**90**	**200**
Free (i=2)	**140**	**200**	**60**	**400**
	150	**300**	**150**	**600**

ROUTE CHOICE EXAMPLE (cont'd)

Probabilities

- (Marginal) probability of choosing toll road *P*(*i* = 1)

$$\hat{P}(i = 1) = 200\ /\ 600 = 1/3$$

- (Joint) probability of choosing toll road and having medium income: *P*(*i*=1, *k*=2)

$$\hat{P}(i = 1, k = 2) = 100\ /\ 600 = 1/6$$

$$\sum_{i=1}^{2} \sum_{k=1}^{3} P(i,k) = 1$$

CONDITIONAL PROBABILITY *P*(*I*|*K*)

$$P(i,k) = P(i) \cdot P(k \mid i)$$
$$= P(k) \cdot P(i \mid k)$$

Independence

$$P(i \mid k) = P(i)$$

$$P(k \mid i) = P(k)$$

$$P(i) = \sum_k P(i,k)$$

$$P(k) = \sum_i P(i,k)$$

$$P(k \mid i) = \frac{P(i,k)}{P(i)}, \qquad P(i) \neq 0$$

$$P(i \mid k) = \frac{P(i,k)}{P(k)}, \qquad P(k) \neq 0$$

MODEL : *P*(*i*|*k*)

- **Behavioral Model~**

  Probability (Route Choice|Income) = *P*(*i*|*k*)

- **Unknown parameters**

$$P(i=1 \mid k=1)=\pi_1$$

$$P(i=1 \mid k=2)=\pi_2$$

$$P(i=1 \mid k=3)=\pi_3$$

Estimated Values:

π_1 = 1/15 = 0.067

π_2 = 1/3 = 0.333

π_3 = 3/5 = 0.6

EXAMPLE: FORECASTING

- **Toll Road share under existing income distribution: 33%**
- **New income distribution**

Route choice	**Income** Low (*k*=1)	Medium (*k*=2)	High (*k*=3)		
Tolled (*i*=1)	**1/15*45=3**	**1/3*300=100**	**3/5*255=153**	**256**	**43%**
Free (*i*=2)	**42**	**200**	**102**	**344**	**57%**
New income distribution	45	300	255	600	
Existing income distribution	***150***	***300***	***150***	***600***	

- **Toll road share: 33%→43%**

THE RANDOM UTILITY MODEL

- **Decision rule: Utility maximization**
  - **Decision maker *n* selects the alternative *i* with the highest utility U_{in} among J_n alternatives in the choice set C_n.**

    $U_{in} = V_{in} + \varepsilon_{in}$

    V_{in} **=Systematic utility : function of observable variables**

    ε_{in} **=Random utility**

THE RANDOM UTILITY MODEL (cont'd)

- **Choice probability:**

$$P(i|C_n) = P(U_{in} \geq U_{jn}, \ \forall j \in C_n)$$

$$= P(U_{in} - U_{jn} \geq 0, \ \forall j \in C_n)$$

$$= P(U_{in} = \max_j U_{jn}, \forall j \in C_n)$$

- **For binary choice:**

$$P_n(1) = P(U_{1n} \geq U_{2n})$$

$$= P(U_{1n} - U_{2n} \geq 0)$$

THE RANDOM UTILITY MODEL (contd.)

Routes	Attributes		Utility (utils)
	Travel time (t)	Travel cost (c)	
Tolled (i=1)	t_1	c_1	U_1
Free (i=2)	t_2	c_2	U_2

$$U_1 = -\beta_1 t_1 - \beta_2 c_1 + \varepsilon_1$$

$$U_2 = -\beta_1 t_2 - \beta_2 c_2 + \varepsilon_2$$

$$\beta_1, \beta_2 > 0$$

THE RANDOM UTILITY MODEL (cont'd)

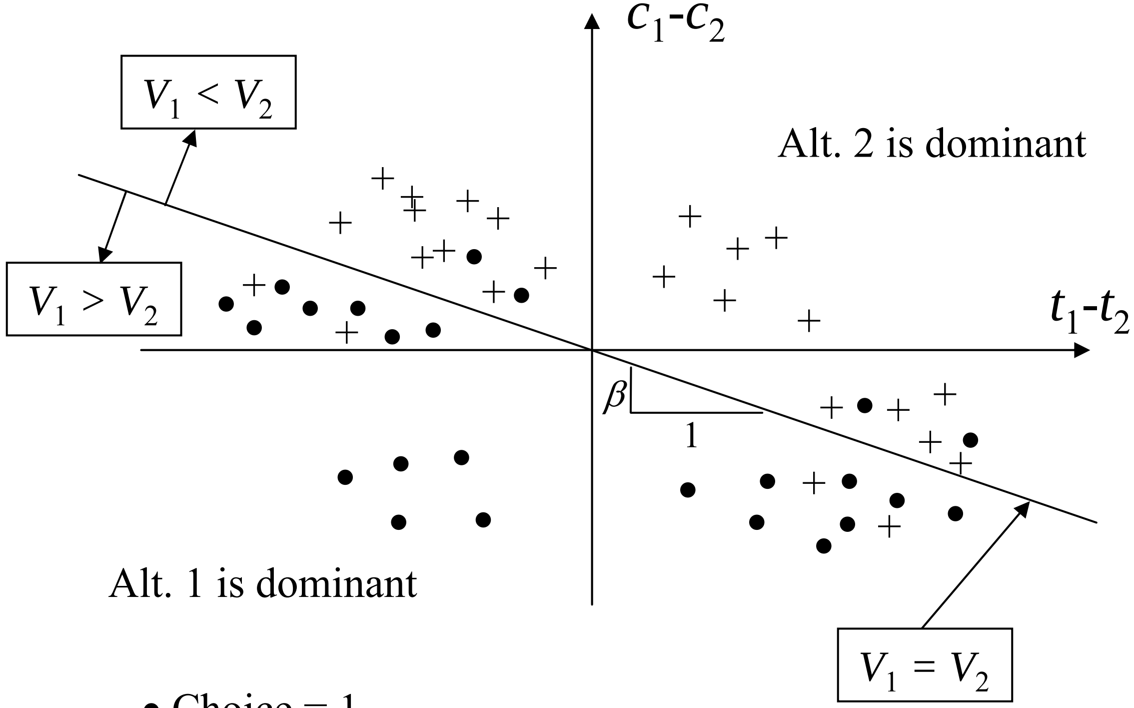
- **Ordinal utility**
  - **Decisions are based on utility differences**
  - **Unique up to order preserving transformation**

$$U_1 = (-\beta_1 t_1 - \beta_2 c_1 + \varepsilon_1 + K)\lambda$$

$$U_2 = (-\beta_1 t_2 - \beta_2 c_2 + \varepsilon_2 + K)\lambda$$

$$\beta_1, \beta_2, \lambda > 0$$

THE RANDOM UTILITY MODEL (contd.)



$$U_1 = -\frac{\beta_1}{\beta_2} \cdot t_1 - c_1 + \varepsilon_1$$

$$U_2 = -\frac{\beta_1}{\beta_2} \cdot t_2 - c_2 + \varepsilon_2$$

$$\beta = \frac{\beta_1}{\beta_2} = \text{"value of time"}$$

• Choice = 1

+ Choice = 2

$$U_1 - U_2 = -\frac{\beta_1}{\beta_2} \cdot (t_1 - t_2) - (c_1 - c_2) + (\varepsilon_1 - \varepsilon_2)$$

THE SYSTEMATIC UTILITY

- **Attributes: describing the alternative**
  - **Generic vs. Specific**
    - **Examples: travel time, travel cost, frequency**
  - **Quantitative vs. Qualitative**
    - **Examples: comfort, reliability, level of service**
  - **Perception**
  - **Data availability**
- **Characteristics: describing the decision-maker**
  - **Socio-economic variables**
    - **Examples: income,gender,education**

RANDOM TERMS

- **Capture imperfectness of information**
- **Distribution of *epsilons***
- **Typical models**
  - **Logit model (i.i.d. "Extreme Value" error terms, a.k.a. Gumbel)**
  - **Probit model (Normal error terms)**

BINARY CHOICE

- **Choice set $C_n = \{1,2\}\ \forall n$**

$$P_n(1) = P(1|C_n) = P(U_{1n} \geq U_{2n})$$
$$= P(V_{1n} + \varepsilon_{1n} \geq V_{2n} + \varepsilon_{2n})$$
$$= P(V_{1n} - V_{2n} \geq \varepsilon_{2n} - \varepsilon_{1n})$$
$$= P(V_{1n} - V_{2n} \geq \varepsilon_n) = P(V_n \geq \varepsilon_n) = F_\varepsilon(V_n)$$

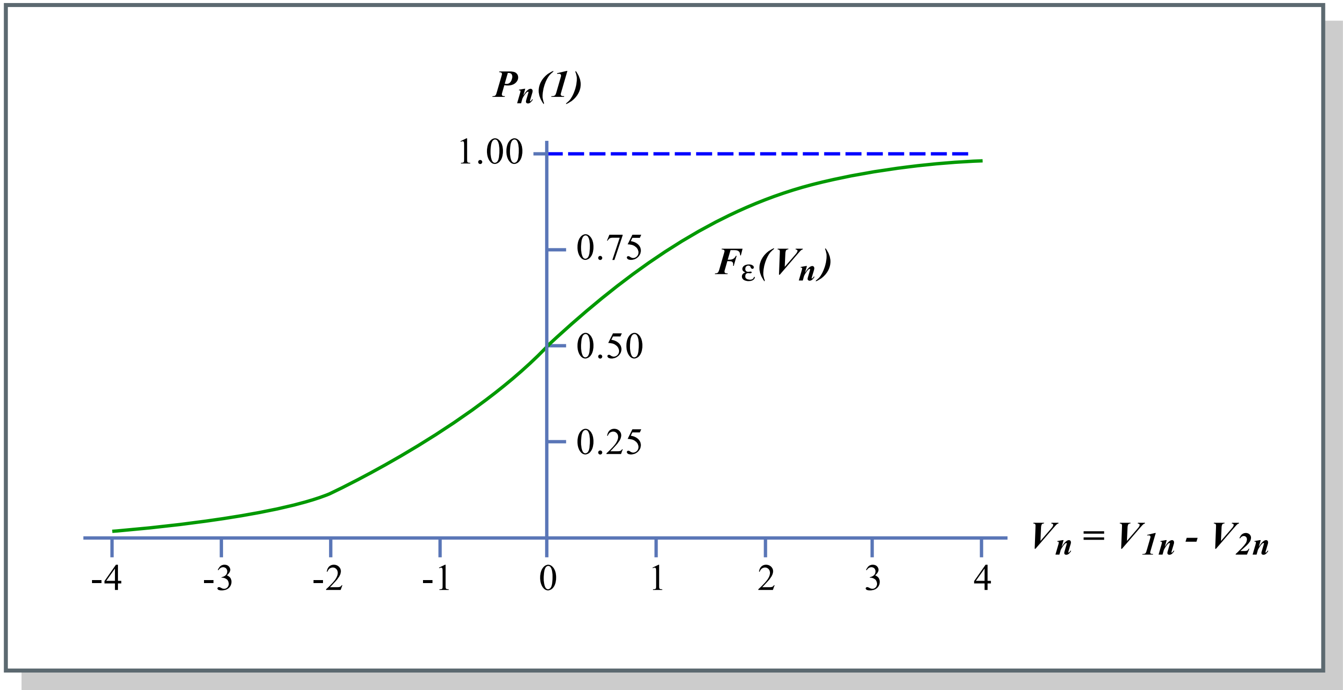


Figure by MIT OCW.

BINARY LOGIT MODEL

- **"Logit" name comes from *Log*istic Probability Un*it***

ε_{1n} ~ ExtremeValue (0,μ)

$$F_\varepsilon(\varepsilon_{1n}) = \exp\left[-e^{-\mu\varepsilon_{1n}}\right]$$

ε_{2n} ~ ExtremeValue (0,μ)

$$F_\varepsilon(\varepsilon_{2n}) = \exp\left[-e^{-\mu\varepsilon_{2n}}\right]$$

ε_n ~ Logistic (0,μ)

$$F_\varepsilon(\varepsilon_n) = \frac{1}{1+e^{-\mu\varepsilon_n}}$$

$$P_n(1) = F_\varepsilon(V_n) = \frac{1}{1+e^{-\mu V_n}}$$

WHY LOGIT?

- Probit does not have a closed form – the choice probability is an integral.
- The logistic distribution is used because:
  - It approximates a normal distribution quite well.
  - It is analytically convenient

LIMITING CASES

- **Recall:** $P_n(1) = P(V_n \geqslant \varepsilon_n)$

  $= F_\varepsilon(V_{1n} - V_{2n})$

- **With logit,** $$F_\varepsilon(V_n) = \frac{1}{1+e^{-\mu V_n}} = \frac{e^{\mu V_{1n}}}{e^{\mu V_{1n}} + e^{\mu V_{2n}}}$$

- **What happens as $\mu \rightarrow \infty$?**

- **What happens as $\mu \rightarrow 0$?**

RE-FORMULATION

- $P_n(i) \quad = P(U_{in} \geqslant U_{jn})$

$$= \frac{1}{1 + e^{-\mu(V_{in} - V_{jn})}}$$

$$= \frac{e^{\mu V_{in}}}{e^{\mu V_{in}} + e^{\mu V_{jn}}}$$

- **If V_{in} *and* V_{jn} are linear in their parameters:**

$$P_n(i) = \frac{e^{\mu\beta' x_{in}}}{e^{\mu\beta' x_{in}} + e^{\mu\beta' x_{jn}}}$$

MULTIPLE CHOICE

- **Choice set C_n: J_n alternatives, $J_n \geq 2$**

$$P(i \mid C_n) = P[V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}, \forall j \in C_n]$$

$$= P\left[\left(V_{in} + \varepsilon_{in}\right) = \max_{j \in C_n} \left(V_{jn} + \varepsilon_{jn}\right)\right]$$

$$= P\left[\varepsilon_{jn} - \varepsilon_{in} \leq V_{in} - V_{jn}, \forall j \in C_n\right]$$

MULTIPLE CHOICE (cont.)

- **Multinomial Logit Model**
  - ε_{jn} **independent and identically distributed (i.i.d.)**
  - ε_{jn} **~ *ExtremeValue(0,μ)*** $\forall j$

$$f(\varepsilon) = \mu e^{-\mu\varepsilon} \exp\left[-e^{-\mu\varepsilon}\right]$$

$$f(\varepsilon) = \mu e^{-\mu\varepsilon} \exp\left[-e^{-\mu\varepsilon}\right]$$

  - **Variance:** $\pi^2/6\mu^2$

$$\mathrm{P}\left(i \mid C_n\right) = \frac{e^{\mu V_{in}}}{\sum_{j \in C_n} e^{\mu V_{jn}}}$$

MULTIPLE CHOICE – AN EXAMPLE

- **Choice Set $C_n = \{1,2,3\} \;\forall n$**

$$P(1 \mid C_n) = \frac{e^{\mu V_{1n}}}{e^{\mu V_{1n}} + e^{\mu V_{2n}} + e^{\mu V_{3n}}}$$

SPECIFICATION OF SYSTEMATIC COMPONENTS

- **Types of Variables**
  - **Attributes of alternatives: Z_{in}, e.g., travel time, travel cost**
  - **Characteristics of decision-makers: S_n, e.g., age, gender, income, occupation**
  - **Therefore: $X_{in} = h(Z_{in}, S_n)$**
- **Examples:**
  - $X_{in1} = Z_{in1}$ **= travel cost**
  - $X_{in2} = \log(Z_{in2})$ **= log (travel time)**
  - $X_{in3} = Z_{in1}/S_{n1}$ **= travel cost / income**
- **Functional Form: Linear in the Parameters**

$$V_{in} = \beta_1 X_{in1} + \beta_2 X_{in2} + \ldots + \beta_k X_{inK}$$

$$V_{jn} = \beta_1 X_{jn1} + \beta_2 X_{jn2} + \ldots + \beta_k X_{jnK}$$

DATA COLLECTION

- **Data collection for each individual in the sample:**
  - ***Choice set: available alternatives***
  - ***Socio-economic characteristics***
  - ***Attributes of available alternatives***
  - ***Actual choice***

n	**Age**	**Auto Time**	**Transit Time**	**Choice**
1	35	15.4	58.2	Auto
2	45	14.2	31.0	Transit
3	37	19.6	43.6	Auto
4	42	50.8	59.9	Auto
5	32	55.5	33.8	Transit
6	15	N/A	48.4	Transit

MODEL SPECIFICATION EXAMPLE

$$V_{auto} = \beta_0 + \beta_1 \, TT_{auto} + \beta_2 \, age_1 + \beta_3 \, age_2$$

$$V_{transit} = \beta_1 \, TT_{transit}$$

where $age_1 = 1$ *if* $age \leq 20$, *0 otherwise*

$age_2 = 1$ *if* $age > 40$, *0 otherwise*

	β_0	β_1	β_2	β_3
Auto	***1***	TT_{auto}	age_1	age_2
Transit	***0***	$TT_{transit}$	***0***	***0***

PROBABILITIES OF OBSERVED CHOICES

- **Individual 1:**

$V_{auto} = \beta_0 + \beta_1\ 15.4 + \beta_2\ 0 + \beta_3\ 0$

$V_{transit} = \beta_1\ 58.2$

$$P(Auto) = \frac{e^{\beta_0 + 15.4\beta_1}}{e^{\beta_0 + 15.4\beta_1} + e^{58.2\beta_1}}$$

- **Individual 2:**

$V_{auto} = \beta_0 + \beta_1\ 14.2 + \beta_2\ 0 + \beta_3\ 1$

$V_{transit} = \beta_1\ 31.0$

$$P(Transit) = \frac{e^{31.0\beta_1}}{e^{\beta_0 + 14.2\beta_1 + \beta_3} + e^{31.0\beta_1}}$$

MAXIMUM LIKELIHOOD ESTIMATION

- **Find the values of β that are most likely to result in the choices observed in the sample:**
  - **max $L^*(\beta) = P_1$(Auto)P_2(Transit)…P_6(Transit)**

- **If**

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

- **Then we maximize, over choices of $\{\beta_1, \beta_2 \ldots, \beta_k\}$, the following expression:**

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

- β^* $= \arg\max_{\beta} L^*(\beta_1, \beta_2 \ldots, \beta_k)$

  $= \arg\max_{\beta} \log L^*(\beta_1, \beta_2 \ldots, \beta_k)$