



# Evidence-Based Transportation Demand Analysis

## *Aggregate Demand Modelling*

Khandker Nurul Habib, PhD. PEng

---

# Outline

- Approaches of Travel Demand Modelling
- Aggregate Demand Models:
  - Total Demand Models
  - Issues with Total Demand Modelling
  - Modelling Shares of Alternative Demands
  - Modelling Total Demand along with Shares
- Ad-hoc Modelling approaches:
  - Elasticity-based Model
  - Pivot-point Updating

# Travel Demands

- Urban travel demands
  - Number of trips per day by individuals/households/traffic zones
  - Proportions of trips by different modes: Transit ridership
  - Proportion of trips on different routes/links/corridors: Transit lines
- Inter-city travel demands:
  - Total trips by different modes (bus, rail, air, car) between cities
- International travel demands:
  - Total number of passenger travel within region, inter-region, between countries, between continents
  - Passenger arrival rates by ground transportation modes (bus, rail, car)
- Tourism travel demands / Special travel generators:
  - Trips generated by hotels
  - Trips attracted by hospitals
  - Trips attracted to historical sites, recreational locations, etc.

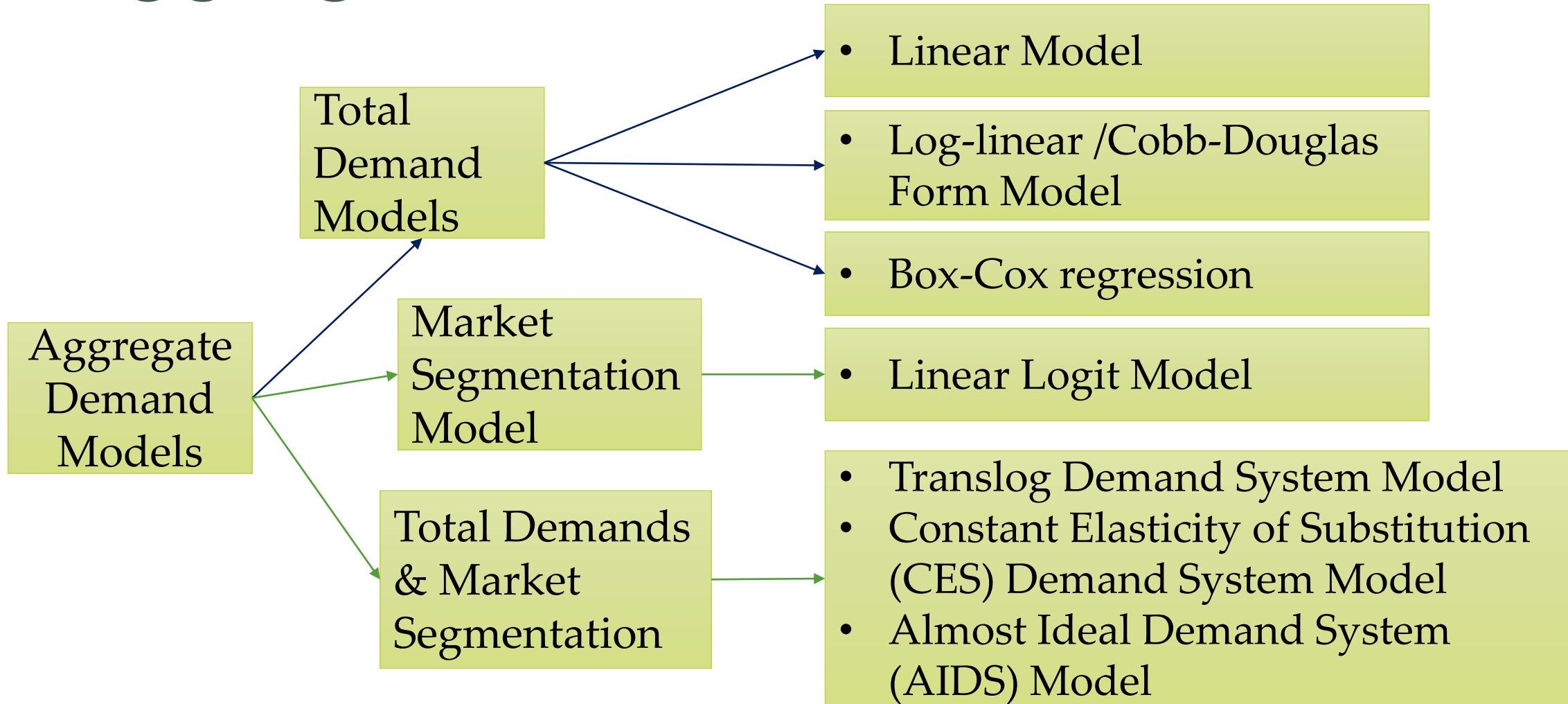
# Approaches of Travel Demand Models

- Based on the concept of utility:
  - Direct utility: when utility function is a function of quantity demands
  - Indirect utility: when utility function is not a function of quantity
- Based on uncertainty in predictions:
  - Deterministic method: trip rate tables, diversion curves etc.
  - Stochastic method: application of stochastic econometric models
- Based on mathematical optimization approach:
  - Random utility maximization
  - Fitting linear/non-linear demand curves
- Based on level of aggregate of travel demands:
  - Aggregate demand model: modelling aggregation of choices
  - Disaggregate demand model: modelling individual choices

# Aggregate versus Disaggregate Model

- Choice of aggregate versus disaggregate demand models:
  - Purpose and context of study
  - Data availability and resource available for model development
  - Disaggregate models are always better than aggregate models if and only if data are available and computational burdens are allowed
- In practice, aggregate models complements disaggregate models
- Aggregate demand models:
  - Suppresses heterogeneity in travel demand
  - Suppresses variability in travel demands
  - Future predictions are uncertain future when extrapolation is problematic

# Aggregate Demand Models



# Linear Model

- Linear regression model:

$$\text{Total Demand, } D \text{ (e.g. number of trips)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_N x_n + \varepsilon = \beta_0 + \sum_i \beta_i x_i + \varepsilon$$

$$\Pr(D) = f(\varepsilon) = \text{the pdf of } \varepsilon ; \text{ considering that the mean value of } \varepsilon \text{ is } \left( \beta_0 + \sum_i \beta_i x_i \right)$$

- Elasticity of demand (function of  $x_j$ )

$$E = \frac{\partial D / D}{\partial x_j / x} = \frac{\partial D}{\partial x_j} \times \frac{x_j}{D} = \frac{\beta_j}{D} x_j$$

- Model estimations: Least-Square or Maximum Likelihood

- Limitations:

- Assumption of linear functional form
- No restriction on 'zero value' prediction
- Difficult to handle data with large portion of 'zero values' of dependent variables

# Log-Linear / Cobb-Douglas Form Model

- Multiplicative functional form:

$$\text{Total Demand, } D \text{ (e.g. number of trips)} = e^{\beta_0} x_1^{\beta_1} x_2^{\beta_2} \dots e^{\beta_k x_k} \dots x_n^{\beta_n} e^{\varepsilon}$$

$$\ln(D) = \beta_0 + \beta_1 \ln(x_1) + \beta_2 \ln(x_2) + \dots + \beta_k x_k + \dots \beta_n \ln(x_n) + \varepsilon$$

$$\Pr(D) = f(\varepsilon) = \text{the pdf of } \varepsilon ; \text{ considering that the mean value of } \varepsilon \text{ is } \left( \beta_0 + \sum_i \beta_i \log(x_i) \right)$$

- Elasticity of demand

$$E = \frac{\partial D}{\partial x_j} \frac{x_j}{D} = \beta_j$$

- Model estimations: Least-Square or Maximum Likelihood
- Limitations:
  - Fixed elasticity
  - Handling 'zero value' in the observed data needs extra care



# Log-Linear / Cobb-Douglas Form Model

- Multiplicative functional form gives flexibility of various possible forms (easy to handle categorical/dummy variable)

$$\begin{aligned} D &= x_1^{\beta_1} x_2^{\beta_2} x_3^{\beta_3} \dots x_n^{\beta_n} \dots \exp(\beta_0 + \gamma_1 y_1 + \dots + \gamma_n y_n + \varepsilon) \\ &= \prod_{i=1}^n x_i^{\beta_i} \prod_{j=1}^m \exp(\beta_0 + \gamma_j y_j) \end{aligned}$$

$$\ln(D) = \beta_0 + \sum_{i=1}^n \beta_i \ln(x_i) + \sum_{j=1}^m \gamma_j y_j + \varepsilon$$

# Box-Cox Regression Model

- Non-linear transformation of dependent and/or independent variables

$$\text{Total Demand, } D(\lambda) = \sum_j \beta x_j(\lambda) + \sum_j \beta z_j + \varepsilon$$

$$D(\lambda) = D^\lambda = \begin{cases} \frac{D^\lambda - 1}{\lambda} & \text{for } \lambda \neq 0 \\ \ln(D) & \text{for } \lambda = 0 \end{cases} \quad x(\lambda) = x^\lambda = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{for } \lambda \neq 0 \\ \ln(x) & \text{for } \lambda = 0 \end{cases}$$

- Elasticity of demand

$$E = \frac{\partial D}{\partial x_j} \frac{x_j}{D} = \beta_j \left( \frac{D}{x_j} \right)^\lambda$$

- Model estimations: Least-Square or Maximum Likelihood

# Handling Zero Demand

- Presence of 'zero value' of observed demand in the dataset causes problem in model estimation as well as prediction
- Ad-hoc way of handling 'zero value':
  - For Cob-Douglas formulation: offset all observation by 1 so that for the 'zero value' it becomes  $\log(D+1)=\log(0+1)=0$
  - For linear regression cannot handle 'zero demand'. So, one has to remove data of 'zero demand'
- In case, the 'zero demands' are legitimate and has a reasonable share in the observed data, a sample selection approach needs to be taken:

$$\text{Total Demand, } D = (\Pr(D = 0) + \Pr(D \neq 0) \times \Pr(D))$$

$$\Pr(D = 0) = \Phi(\gamma_o + \gamma_1 z_1 + \dots)$$

$$\Pr(D \neq 0) = 1 - \Phi(\gamma_o + \gamma_1 z_1 + \dots)$$

$$\text{Total Demand, } D = (\Phi(\gamma_o + \gamma_1 z_1 + \dots) + (1 - \Phi(\gamma_o + \gamma_1 z_1 + \dots)) \times \Pr(D))$$

# Lag effect in Aggregate Demand Modelling

- Lag effects:
  - Lag effect of changes in supply (cost/price), context (socio-economic)
  - Lag effect of changes in total demand
- Aggregate demand model can handle lag effect:
  - Total observed demand,  $D$  is the equilibrium demand
  - Actual demand or Desired Demand,  $D^*$  is unobserved
  - Assumption is that demand, the observed or equilibrium demand ( $D$ ) adjusts partially towards desired level over the time steps of analysis ( $t$ ):

$$\ln(D_t) - \ln(D_{t-1}) = \mu(\ln(D_t^*) - \ln(D_{t-1}))$$

$\mu$  is the rate of adjustment and  $0 < \mu < 1$

- ✓ Higher the value of  $\mu$  quicker is the adjustment.
- ✓ Instant adjustment if the value of  $\mu = 1$
- ✓ No adjustment, if  $\mu = 0$

# Lag effect in Aggregate Demand Modelling

- Since actual demand or desired demand is unobserved, we can consider it as a random variable

$$\text{Latent unobserved desired demand, } D_t^* = \prod_{i=1}^n x_i^{\beta_i} \times \prod_{j=1}^m \exp(\beta_0 + \gamma_j y_j) \times \exp(\varepsilon)$$

$$\text{Taking logarithm: } \ln(D_t^*) = \beta_0 + \sum_{i=1}^n \ln(x_i^{\beta_i}) + \sum_{j=1}^m \gamma_j y_j + \varepsilon \dots\dots\dots(1)$$

$$\text{Considering lagged demand effects, } \ln(D_t) - \ln(D_{t-1}) = \mu(\ln(D_t^*) - \ln(D_{t-1})) \quad \text{where } 0 < \mu < 1$$

$$\Rightarrow \mu \ln(D_t^*) = (\mu - 1) \ln(D_{t-1}) + \ln(D_t) \quad \Rightarrow \ln(D_t^*) = (1 - 1/\mu) \ln(D_{t-1}) + \ln(D_t) / \mu$$

$$\text{Replacing } D_t^* \text{ in original demand function (1): } (1 - 1/\mu) \ln(D_{t-1}) + \ln(D_t) / \mu = \beta_0 + \sum_{i=1}^n \beta_i \ln(x_i) + \sum_{j=1}^m \gamma_j y_j + \varepsilon$$

$$\ln(D_t) = (1 - \mu) \ln(D_{t-1}) + \mu(\beta_0 + \sum_{i=1}^n \beta_i \ln(x_i) + \sum_{j=1}^m \gamma_j y_j + \varepsilon)$$

$$\text{Final Demand Model: } \ln(D_t) = \mu' \ln(D_{t-1}) + \beta_0' + \sum_{i=1}^n \beta_i' \ln(x_i) + \sum_{j=1}^m \gamma_j' y_j + \varepsilon'$$

# Lag effect in Aggregate Demand Modelling

Aggregate Model with lagged demand

$$\ln(D_t) = \mu' \ln(D_{t-1}) + \beta'_0 + \sum_{i=1}^n \beta'_i \ln(x_i) + \sum_{j=1}^m \gamma'_j y_j + \varepsilon'$$

Aggregate Model without lagged demand

$$\ln(D_t) = \beta_0 + \sum_{i=1}^n \beta_i \ln(x_i) + \sum_{j=1}^m \gamma_j y_j + \varepsilon$$

- Lagged effect model can be estimated a linear regression model
- Model estimated without considering lagged demand will give static elasticity
- Model estimated with lagged demand effect will give long-term elasticity

# Example: Modelling Demand of Inter-City Train

Cobb-Douglas Specification:

$$J_t = \beta_0 J_{t-1}^{\beta_1} G_t^{\beta_2} F_t^{\beta_3} T_t^{\beta_4} \exp\left(\beta_5 t + \sum_{k=1}^K \gamma_k x_{kt}\right)$$

- $J_t$  is the total demand (passenger-miles) at time  $t$
- $J_{t-1}$  is the total demand (passenger-miles) at time  $t-1$
- $G_t$  is the income variable at time  $t$
- $F_t$  is the rail fare price variable
- $T_t$  is a rail performance measure
- $t$  is the linear time trend
- $x_k$  is a vector of other variables, including dummies and variables relating to competing modes
- $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \gamma_k$  etc. are the parameters of the model (to be estimated)

# Example: Modelling Demand of Inter-City Train

- Observed dataset does not have any 'zero value': Used log-normal

$$\begin{aligned} \ln(J_t) = & \ln(\beta_0) + \beta_1 \ln(J_{t-1}) + \beta_2 \ln(G_t) + \beta_3 \ln(F_t) + \beta_4 \ln(T_t) \\ & + (\beta_5 t + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3 + \gamma_4 x_4 + \dots) \end{aligned}$$

$$J_t = \exp \left( \begin{aligned} & \ln(\beta_0) + \beta_1 \ln(J_{t-1}) + \beta_2 \ln(G_t) + \beta_3 \ln(F_t) + \beta_4 \ln(T_t) \\ & + \beta_5 t + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3 + \gamma_4 x_4 + \dots \end{aligned} \right)$$



# Example: Modelling Demand of Inter-City Train

Variable selection: Based on data availability

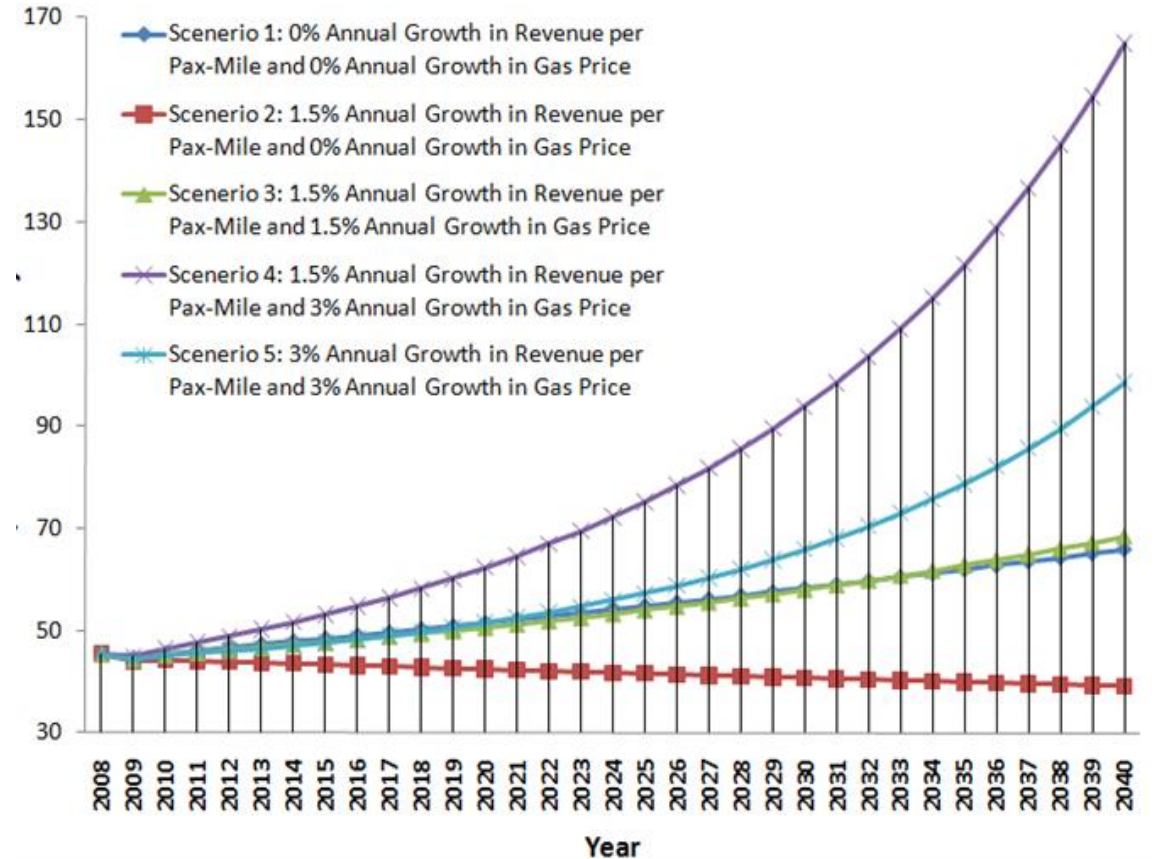
- It was considered essential to include variables covering a number of market and service characteristics:
  - Demand lag, and time trend
  - Own-price elasticity
  - Cross-elasticity to auto and air, especially fares
  - Service quality in terms of on-time performance
  - Macroeconomic factors (GDP, income)
  - Seasonality

# Example: Modelling Demand of Inter-City Train

Variable	Period	Form
Lag demand: Ln of Pax-mile	Quarter	Lag (years) = $1 / (1 - \text{parameter})$
Rail Fare: Rev. per pax-mi	Quarter	Coefficient $> 1$ : <i>elastic</i>
On-Time Performance (OTP)	Quarter	Ln (%)
Average Employment	Quarter	Ln (Millions)
Income Proxy (Avg. Market's Emplmt. x GDP / Employed)	Quarter	Ln (\$ Millions)
Time Trend	Quarter	units
Quarterly Dummies	Quarter	0 or 1
Gasoline Price	Quarter	Deflated (real) price
Airfare	Annual	Deflated (real) price
GDP of European Union	Quarter	Real GDP

# Example: Modelling Demand of Inter-City Train

R-squared = 0.93			
Variables	Coef.	Robust Std. Err.	t-Stat
Total Pax-Miles in Previous Quarter	0.518	0.154	3.360
Avg Revenue per Pax-Mile	-0.533	0.288	-1.850
OTP	0.132	0.100	1.320
Avg Employment in Millions	2.045	0.841	2.430
Time Step	-0.029	0.019	-1.490
Quarter 1 Dummy	0.054	0.037	1.450
Quarter 2 Dummy	0.101	0.066	1.520
Quarter 3 Dummy	0.252	0.060	4.200
Gas Price	0.418	0.135	3.080
Constant	3.387	1.385	2.450



# Market Segmentation: Aggregate Logit Model

- Consider three alternatives: Air(A), Bus(B), Car(C) for an intercity case

$D_A$  = Total demand (number of trips) for A

$D_B$  = Total demand (number of trips) for B

$D_C$  = Total demand (number of trips) for C

- Market shares of the alternatives

$S_A$  = Ratio of Total demand for A =  $D_A / (D_A + D_B + D_C)$

$S_B$  = Total demand for alternative B =  $D_B / (D_A + D_B + D_C)$

$S_C$  = Total demand for alternative C =  $D_C / (D_A + D_B + D_C)$

- Consider observed shares as observed (pseudo) probabilities & use Logit function with fixing one alternative as the base alternative (A in this example):
  - Logarithm of the ratio (log of odd-ratio) of the shares become a linear regression model

## 2-Alternative Aggregate Logit Model

- Consider two alternative urban commuting modes: Car ( $C_1$ ), Transit ( $C_2$ )

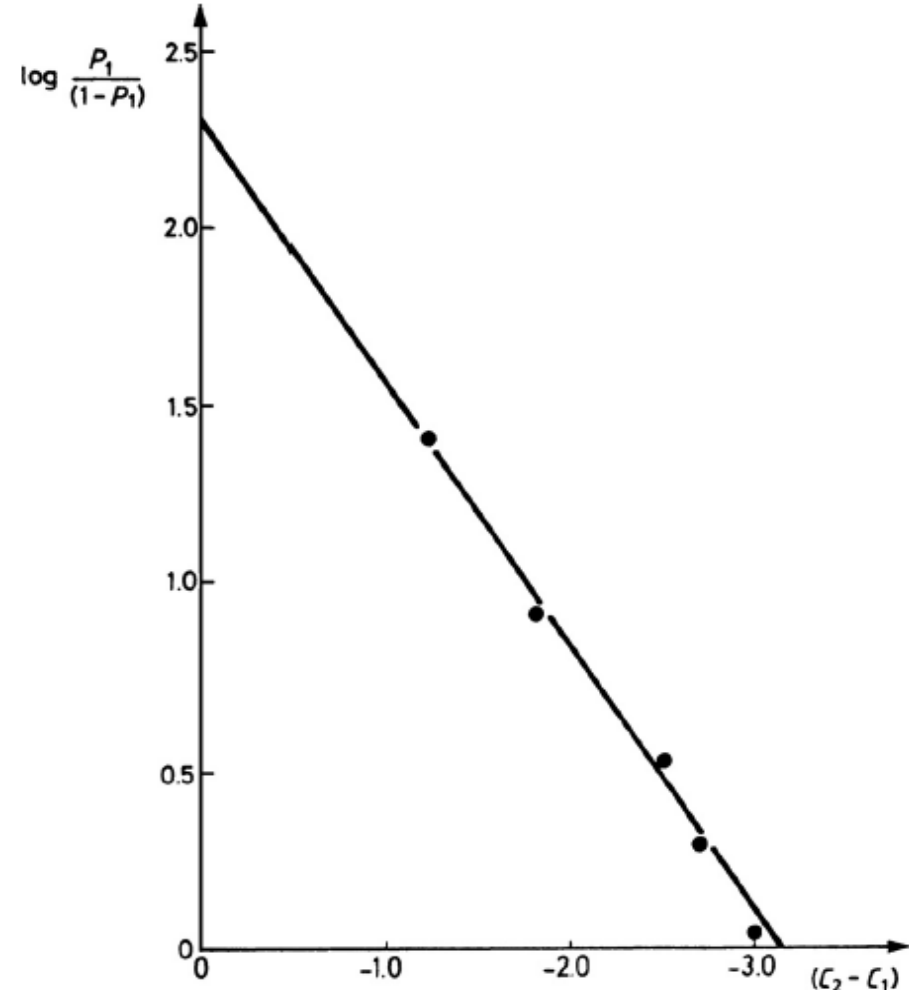
$$\text{Log - Odd ratio, } \log\left(\frac{\Pr(\text{transit})}{1 - \Pr(\text{transit})}\right) = \beta_0 + \beta_c (C_{\text{transit}} - C_{\text{car}})$$

- Data: Observation

Zone pair	$P_1$ (%)	$P_2$ (%)	$C_1$	$C_2$	$\log [P_1/(1 - P_1)]$
1	51.0	49.0	21.0	18.0	0.04
2	57.0	43.0	15.8	13.1	0.29
3	80.0	20.0	15.9	14.7	1.39
4	71.0	29.0	18.2	16.4	0.90
5	63.0	37.0	11.0	8.5	0.53

# 2-Alternative Aggregate Logit Model

- Linear-Regression of log-odd ratio against the cost difference will give a model linear logit model



# Multi-Alternative Aggregate Logit Model

- Based on indirect utility function and logit formulation ( A as the reference alternative)

$$\Pr(C) = S_C = \frac{\exp\left(\sum_c \beta x\right)}{\exp\left(\sum_c \beta x\right) + \exp\left(\sum_B \beta x\right) + 1} \quad \Pr(B) = S_B = \frac{\exp\left(\sum_B \beta x\right)}{\exp\left(\sum_c \beta x\right) + \exp\left(\sum_B \beta x\right) + 1} \quad \Pr(A) = S_A = 1 - \Pr(B) - \Pr(C)$$

- Log-odd ratios (with respect to the reference alternative): Can be modelled as Log-linear regression model

$$\begin{aligned} \log(S_C / S_A) &= \sum_c \beta x - 0 \\ \log(S_B / S_A) &= \sum_B \beta x - 0 \\ \log(S_A / S_A) &= 0 \end{aligned}$$

- Estimation:
  - Maximum likelihood estimation (for more than 2 alternatives)
  - Least square regression as regression of log-odd ratio (convenient for 2 alternatives)

# Aggregate Logit Model with Capturing Competition

- Key point is to set the reference alternative. With respect to the reference alternative two types of formulations are possible
  - Ratio of common-attributes (e.g. cost, time) format
$$\ln(S_C / S_A) = \beta_{0-CA} + \sum_{k=1}^K \beta_{CA-k} (x_{C-k} / x_{A-k}) + \sum_{j=1}^J \gamma_{C-j} x_j$$
$$\ln(S_B / S_A) = \beta_{0-BA} + \sum_{k=1}^K \beta_{BA-k} (x_{B-k} / x_{A-k}) + \sum_{j=1}^J \gamma_{B-j} x_j$$
  - Difference of common-attributes (e.g. cost, time) format
$$\ln(S_C / S_A) = \beta_{0-CA} + \sum_{k=1}^K \beta_{CA-k} (x_{C-k} - x_{A-k}) + \sum_{j=1}^J \gamma_{C-j} x_j$$
$$\ln(S_B / S_A) = \beta_{0-BA} + \sum_{k=1}^K \beta_{BA-k} (x_{B-k} - x_{A-k}) + \sum_{j=1}^J \gamma_{B-j} x_j$$
- $X_{CA-k}$  = the common attributes of C and A
- $X_{C-j}$  = the attributes of only C
- $X_{BA-k}$  = the common attributes of B and A
- $X_{B-j}$  = the attributes of only B
- Estimation: either by maximum likelihood or least-square for log-odd ratio



# Aggregate Logit Model with Capturing Competition

- Elasticity of substitution of a variable between an alternative (B,C) against the reference alternative (A)

$$E_{B-A} = -\beta_{BA}(x_B / x_A) \quad E_{C-A} = -\beta_{CA}(x_C / x_A)$$

- Elasticity of substitution between two non-reference alternatives (B and C)

$$E_{B-C} = -\beta_{BA}(x_B / x_A) \text{ if } d(x_C) = 0 \text{ \& } d(x_B) \neq 0$$

$$E_{B-C} = -\beta_{CA}(x_C / x_A) \text{ if } d(x_C) \neq 0 \text{ \& } d(x_B) = 0$$

- Such elasticity measure is problematic as there no consistent measurement of elasticity when attributes of both non-reference alternative change
- Elasticity of substitution depends on reference alternative

# Aggregate Logit Model for Capturing Competition

- Elasticity of substitution between two non-reference alternatives (B and C)

$$E_{B-C} = -\beta_{BA}(x_B) \text{ if } d(x_c) = 0 \text{ \& } d(x_B) \neq 0$$

$$E_{B-C} = -\beta_{CA}(x_C) \text{ if } d(x_c) \neq 0 \text{ \& } d(x_B) = 0$$

- Such elasticity measure is also problematic as there no consistent measurement of elasticity when attributes of both non-reference alternative change

# Total Demand & Market Segmentation: Translog Model

- Translog demand system model specifies indirect utility (generalized cost),  $V$  function of a demand generation process

$$\ln(V_i) = \alpha_0 + \sum_j \alpha_i \ln(\text{Pe}_{ij} / Y_i) + \frac{1}{2} \sum_j \sum_{k \neq j} \beta_{jk} \ln(\text{Pe}_{ij} / Y_i) \ln(\text{Pe}_{ik} / Y_i) + \dots$$

- Once specified
  - Total quantity demand for alternative  $j$ ,  $x_j^*$  is estimated by applying Roy's Identity,

$$x_{ij}^* = -(\partial(V_i) / \partial(\text{Pe}_{ij})) / (\partial(V_i) / \partial(Y_i))$$

- Finally the demand share of alternative  $j$ ,  $P_i(j)$

$$P_i(j) = \frac{x_{ij}^*}{\sum_j x_{ij}^*}$$

- $\text{Pe}$  = Effective price or cost
- $i$  is the individual
- $j, k = 1, 2, 3, 4$ , are the alternatives
- $Y$  = Income or Budget
- $X_j^*$  = Total demands of  $j$

# Empirical Example: Translog Model

- Modelling demands for making trips of j destinations

$$\ln(V_i) = \alpha_0 + \sum_j \alpha_j \ln(Pe_{ij} / Y_i) + \frac{1}{2} \sum_j \sum_{k \neq j} \beta_{jk} \ln(Pe_{ij} / Y_i) \ln(Pe_{ik} / Y_i) + \dots$$

$$x_{ij}^* = \frac{\alpha_j + \sum_k \beta_{jk} \ln(Pe_k / Y)}{(Pe_j / Y) \left( \sum_j \alpha_j + \sum_j \sum_k \beta_{jk} \ln(Pe_k / Y) \right)}$$

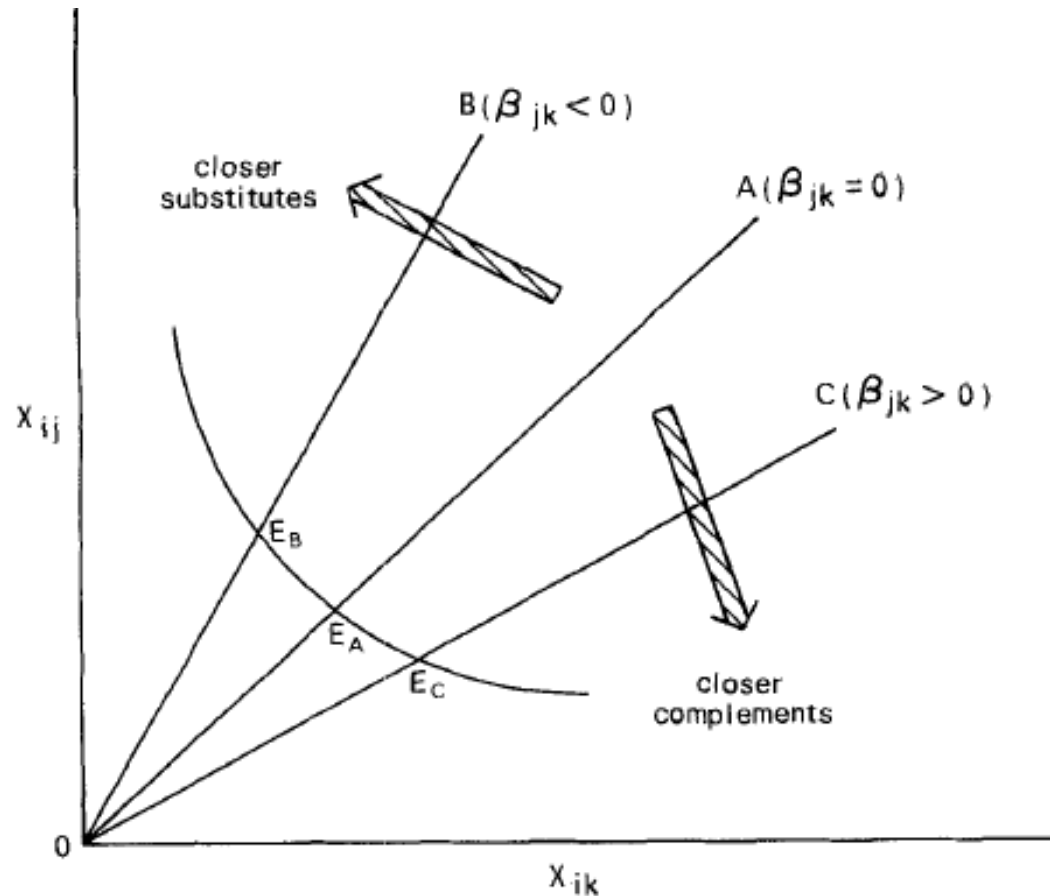
- Assumed that observed demand for alternative j by individual i has,  $x_{ij}^*$  has randomness and so follows a distribution

$x_{ij}^*$  has lognormally or Poisson distribution

- Use Poisson or Lognormal regression model

- Pe = Generalized cost of making a trip
- i is the individual
- j, k = 1, 2, 3, 4, are the alternatives
- Y = Composite Income / Budget
- $X_j^*$  = Total trip to alternative j

# Translog Model Provides Superior Specification



- Trip generation to alternative destinations (j and k) by an individual (i) at varying substitution scenarios that can be captured by a TransLog model:
  - A zero interaction coefficient ( $\beta_{jk}=0$ ) indicates no substitution (independent demands)
  - A negative interaction coefficient ( $\beta_{jk}<0$ ) indicates an increase in Trips demand to j at the expense of those to k
  - A positive interaction coefficient ( $\beta_{jk}>0$ ) indicates complementary relationship

# Market Segmentation: Translog Model

- Translog demand system model is based on microeconomic principle of utility maximization
- Substitution patterns between choice alternatives can be accommodate flexibly
- Model can be estimated by using least-square estimation method or maximum likelihood estimation technique

# Predicting Incremental Changes: Elasticity-based Model

- Predicting changes in total demands:

$$T_t = T_{t-1} + E_x \times T_{t-1} \times \frac{x_t - x_{t-1}}{x_{t-1}} \quad E = \text{Elasticity} = \frac{(T - T_0) / T_0}{(x - x_0) / x_0}$$

- $T_t$  is the total demand after change of  $x$  at time  $t$  (e.g. transit ridership at time  $t$ )
  - $T_{t-1}$  is the total demand before change of  $x$  (e.g. transit ridership time  $t-1$ )
  - $x_t$  is the attribute after change at time  $t$  (e.g. transit fare at time  $t$ )
  - $x_{t-1}$  is the attribute before change at time  $t$  (e.g. previous transit fare)
- Elasticity of demand ( $E$ ) needs to be known and a fixed value
  - Useful for short term analysis when the expectation of 'no big change in behaviour' is valid

# Predicting Incremental Changes: Elasticity-based Model

- Example of typical functional forms of total demand and corresponding elasticity:

Type	Functional Form	Elasticity
Linear	$T = \alpha + \beta S$	$E = \frac{\beta S}{T} = \frac{1}{1 + \alpha/\beta S}$
Product	$T = \alpha S^\beta$	$E = \beta$
Exponential	$T = \alpha \exp(\beta S)$	$E = \beta S$
Share	$p_i = \frac{T_i}{\sum_j T_j}$	$E_{S_i}(p_i) = 1 - p_i$ $E_{S_j}(p_i) = -p_j$



# Predicting Incremental Changes: Pivot-Point Model

$$P'(j) = \frac{P^0(j) \exp(v_j - v_j^0)}{\sum_k P^0(k) \exp(v_k - v_k^0)}$$

- $P'(j)$  is probability/proportion of choosing  $j$  after change in systematic utility  $v_j^0$
  - $P^0(j)$  is probability/proportion of choosing  $j$  before change
  - $v_j$  is the systematic utility function after change
  - $v_j^0$  is the systematic utility before change
- 
- Systematic utility function needs to be known and pre-defined
  - Considers that preference structure and competition do not change

# Aggregate Demand models

- Aggregate models are often very useful:
  - When quick estimation of changes is necessary
  - Lack of detailed micro data for disaggregate modelling
  - Forecasting scenario analysis without precise specification of scenario contexts
- Aggregate and Disaggregate models are complementary:
  - Should not be considered either or



# Thank You