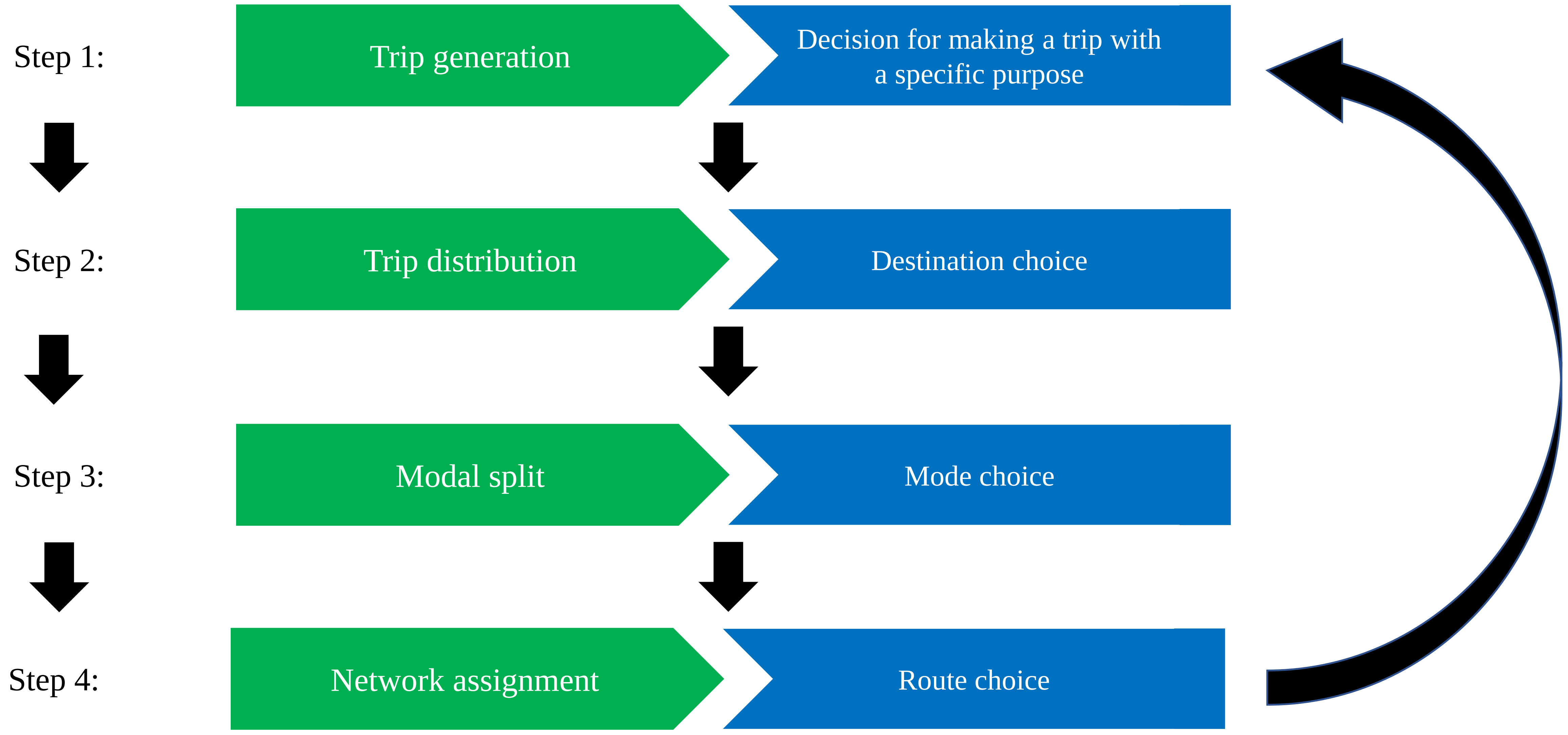
Decision-aid methodologies in transportation

CIVIL-557

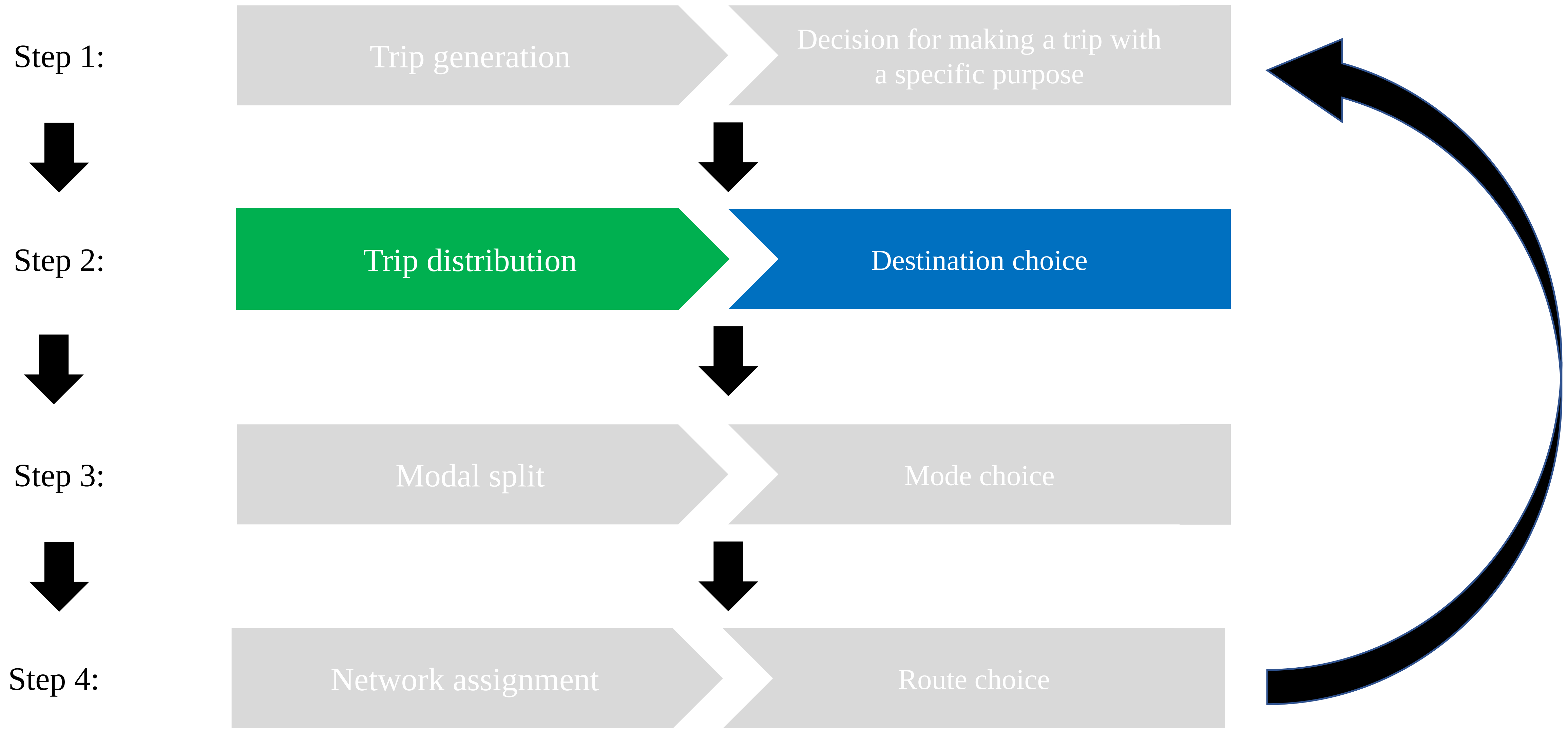
Modelling transportation systems
3. Trip distribution models

Evangelos Paschalidis

The 4-step model



The 4-step model



Previously – Trip generation models

Trip generation modelling

- Aim – motivation
- Terminology
- Models
  - Cross classification – category analysis
  - Growth factor models
  - Linear regression

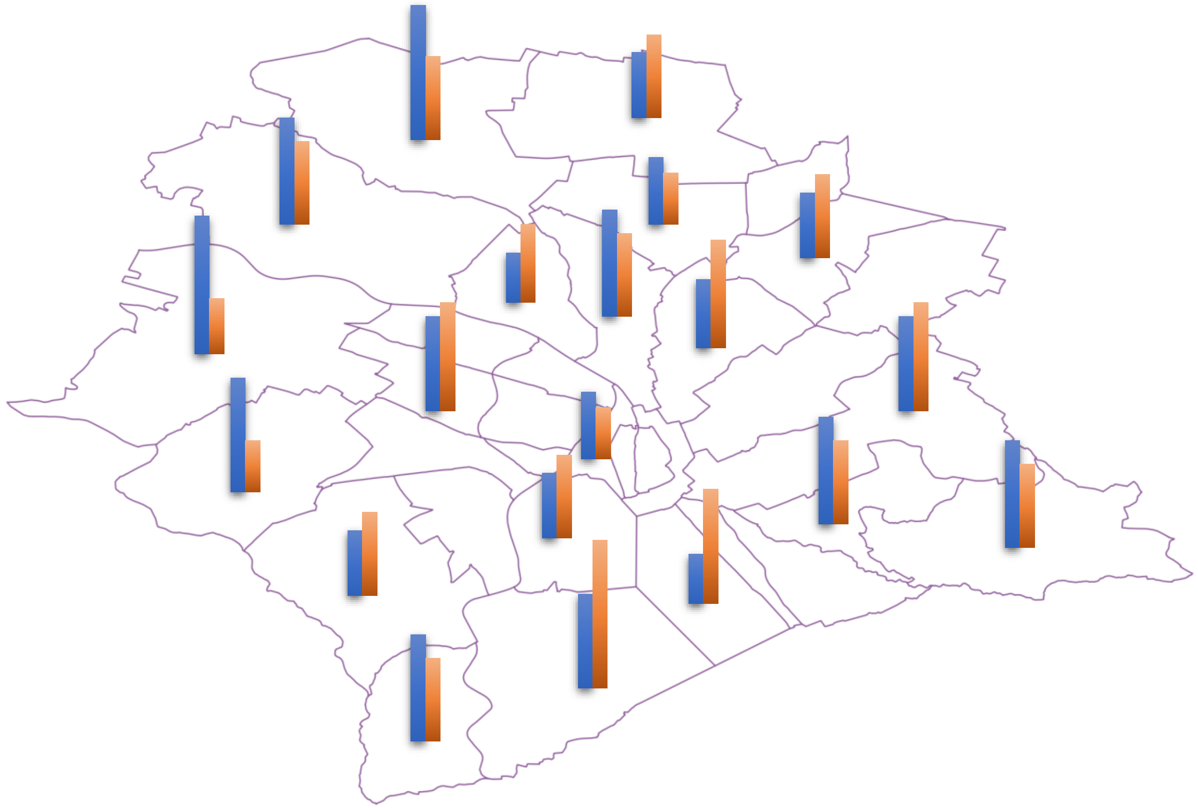
Trip distribution models

Trip distribution modelling

- Aim – background
- Models
  - Growth factor methods – Furness method
  - Synthetic methods – gravity model

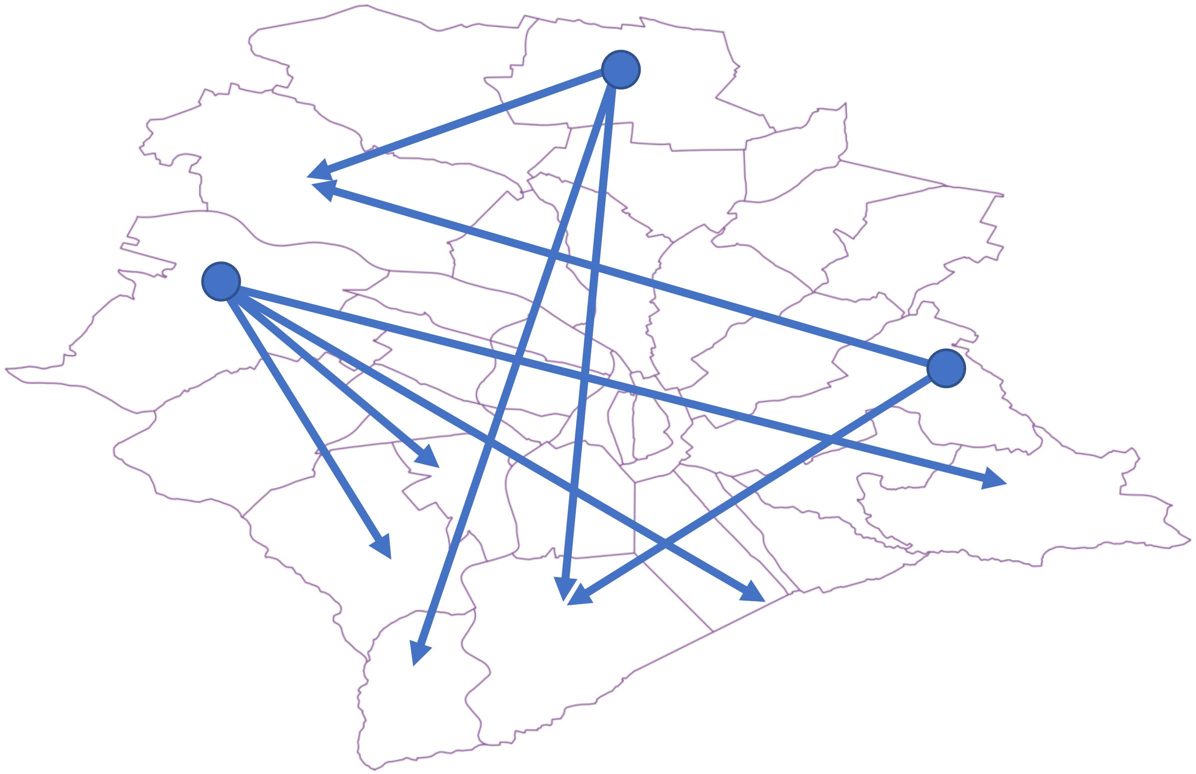
Step 1: Trip generation

- Trip generation step:
  - The number of trips generated in each zone
  - The number of trips attracted in each zone
- Generated trips are typically a function of socioeconomic characteristics and land use
- Attracted trips are typically a function of land use characteristics
- Output: The number of trips generated in and attracted to each zone



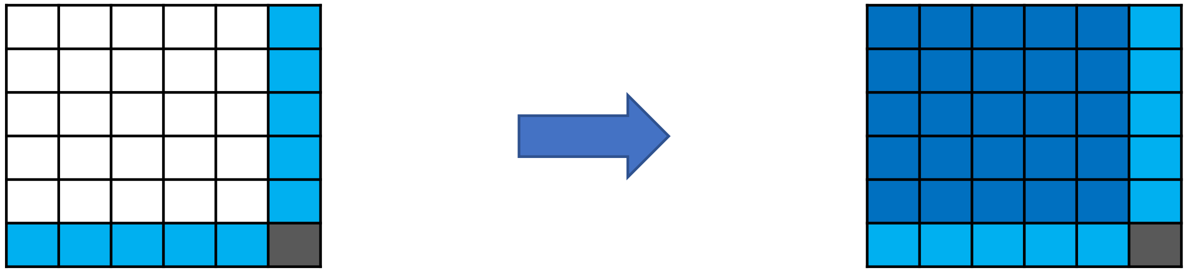
Step 2: Trip distribution

- Trip distribution step:
  - The number of trips between each origin-destination pair
  - The number of trips typically depends on the productivity of the origin zone and the attractiveness of the destination zone
  - Some typical factors that affect trip distribution are the size of a zone, the land use, and the trip cost between the origin and destination zones
  - Input: Trip production/attraction (from step 1), travel cost matrix
  - Output: Origin-destination matrix (typically by trip purpose)



Trip distribution – Aim

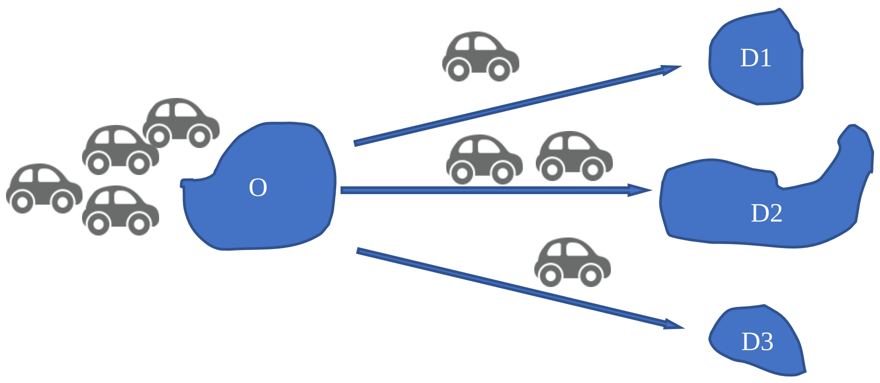
- Where the generated trips go & the attracted trips come from
- We match the O zones with D zones
- Sum of each row and column are known (From our trip generation models)
- Challenge: Fill in the blanks-cells
  - We begin from our estimates of trip ends in each zone



Trip distribution – Aim

Inputs:

- Trip productions and attractions
- Travel (generalised) cost matrix (if we have one)



Output:

- Full trip OD matrix T_{ij} (T_{ij} is the number of trips from origin zone *i* to destination zone *j*)

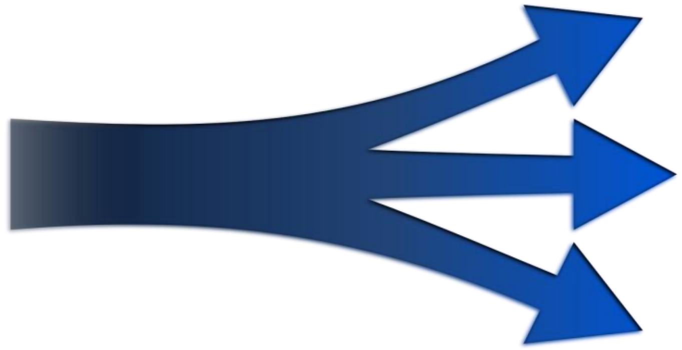
We should generate separate OD matrixes for each trip purpose, travel mode, scenario, and year in our analysis (we use the same demand segmentation from the trip generation step)

We do not examine the exact path from *i* to *j* yet. This is done by the assignment model in step 4

Trip characteristics

Different OD matrices per trip purpose:

- When trip purpose home to work – confident both for production and attraction trip ends
  - Travellers have specific home and work locations
- Numbers computed in the trip generation stage (residential characteristics and employment figures)
- We must convert production-attraction to origin-destination
  - The assignment step requires the origin and destination zones



Trip purpose

- Factors affecting distribution: Congestion, distance, cost, frequency of public transport etc.
- Factors take the form of generalised cost:
  - Comes from the assignment model and/or
  - Discrete choice models
  - For today we will assume its value known – more in the next lecture
- For commute trips we take input from the trip generation model
  - Generalised cost has secondary role
  - People most likely will not change work location due to the generalised cost

Trip purpose

“Home to other” trips (we exclude work and education) productions from the generation step:

- We care less about trip attractions in this case, lower importance
- People are more flexible to travel to a different zone – generalised cost is important here
- E.g. If purpose is shopping or leisure, one can easier change the destination zone, if generalised cost changes considerably
- Trip production and generalised cost the most important factors
- Fixed totals for the rows but not the columns (destination choice modelling)
- The opposite for home to work trips (totals of all rows and columns are fixed)

Trip purpose - Summary

- Home to work (or education), we fixed columns and rows totals, we just need to fill in the matrix
  - Less choice flexibility
- For other trips: Distribution process selects which columns have more trips (destination choice modelling)
- Generalised cost: must include as many as possible variables that make a destination attractive; otherwise may overrepresent irrelevant destinations
- May have too many calculations – comes at the expense of behavioural aspects. We primarily care about executing calculations fast

Defining the OD matrix

- Gravity model: assign trips based on the smaller generalised cost
  - How do we get the cost? At the assignment step and/or using choice models.

- In the first model iteration there is no assignment.
  - We need to fill the trips without any knowledge of the generalised cost.
  - Maybe we are lucky and we can use the cost from an older model

- Maybe I do not trust the generalised cost, using it gives me unexpected results:
  - Methods that do not depend on the generalised cost e.g., Furness or Fratar methods
  - Numbers add up correctly to the known totals

- Rebalancing: balance the matrix based on new totals and a previous OD matrix

Matrix conversion: from production–attraction (PA) to origin–destination (OD)

The distribution matrix is in PA format; we need to convert to OD to feed the assignment model. Why?

- PA matrix is not directional; we know the production and attraction zone but not the direction of trip e.g. home location is always considered as the trip production location
- Route choice depends on the direction of the travel
- Traffic conditions depend on the direction of the travel.
- Assignment model requires exact origin and destination

Solution:

- Reverse trips going towards home so we get the correct travel direction
- For every zone use shares from observed data (e.g. what % of productions are leaving or entering the zone) – we will do a simplified approach if this technique in the lab session

Matrix conversion: from production–attraction (PA) to origin–destination (OD)

Reasons for preferring the PA format:

- Consistency with Home-Based Trip Behaviour: people consider both the *from* and *to* costs
- More Meaningful Trip Generation Modelling: productions are based on residential characteristics and attractions are based on activity locations
- Easier Calibration of Trip Rates: PA allows calibration of trip rates by trip purpose and socioeconomic group, using household survey data (the production end always at home).
- Better input for Trip Distribution models (clear which zone attracts and which zone produces)

Matrix conversion: from production–attraction (PA) to origin–destination (OD)

- We can directly use OD if data is poor quality at first place (rule of thumb)
- We need different matrices for time of the day:
  - Information about trip generation, distribution, and mode choice typically available for 24h or average weekday
  - 24h window may be too long and generic for our project
  - Assignment (traffic levels, passenger numbers) at smaller time periods e.g. morning peak
  - Detailed approach: time of day choice model on how to split the number of trips in morning-peak, afternoon-peak, rest of the time
  - Usually we simplify: we use the data to see what proportion of the trips is made at each time of the day, based on some fixed factors
  - Caution! We must ensure that the factors will not lead to behaviour change in the scenario that we investigate (e.g. a new tram line that changes proportion between work/other reason trips starting from a zone)

Trips within the same zone

- Some trips never leave the zone (internal trips)
- Internal trips are usually ignored during network assignment (Step 4)
  - All trips are assumed to start and end at a centroid – internal trips have the same centroid
  - This is one more reason that it is important to correctly specify the zones!
- Important to include internal trips in the distribution step
  - Destinations may change due to changes in the transport system and some of the trips will leave the zone
- What is the generalised cost of an internal trip?
  - Assumption e.g. 70% of the generalised cost travelling to the nearest zone

Sanity checks on demand matrix

For large projects it is impossible to check cell by cell a trip distribution matrix produced by a model.

We check some specific aspects of the matrix.

- Trip length (either distance or travel time) distribution
  - We can compare our data to old data or simply to our expectation (not very scientific)
  - It is applied at the network level (not per unique OD) but gives an overall picture
- Maps are useful but can show only one type of information at the time
  - Demand by origin zones
  - Demand by destination zones
  - Demand from one origin to all destinations
  - Demand from all origins to a destination

Limitations of trip distribution models

- Filling the matrix has black box elements – always some OD pairs will not make sense
- Easy to miss obvious problems due to matrix size (check sanity checks)
- Missing behavioural consistency when combining data from multiple sources
- Generalised cost not only time-cost but also other factors – perception that drives choices (we must approximate well all these aspects)
- Lower credibility of trips identified as internal, due to simplified generalised cost. Number of trips to adjacent zones may be affected by this because there is redistribution of internal trips to these zones.

Reminder – trip distribution

	1	2	3	4	5	O_i
1	?	?	?	?	?	O_1
2	?	?	?	?	?	O_2
3	?	?	?	?	?	O_3
4	?	?	?	?	?	O_4
5	?	?	?	?	?	O_5
D_j	D_1	D_2	D_3	D_4	D_5	

Trip generation: Production O_i and attraction D_j

	1	2	3	4	5	O_i
1	T_{11}	T_{12}	T_{13}	T_{14}	T_{15}	O_1
2	T_{21}	T_{22}	T_{23}	T_{24}	T_{25}	O_2
3	T_{31}	T_{32}	T_{33}	T_{34}	T_{35}	O_3
4	T_{41}	T_{42}	T_{43}	T_{44}	T_{45}	O_4
5	T_{51}	T_{52}	T_{53}	T_{54}	T_{55}	O_5
D_j	D_1	D_2	D_3	D_4	D_5	

Trip distribution: Origin-destination matrix T_{ij}

Model types

1. We update an existing OD matrix with known T_{ij} values

- Growth factor methods
- Furness (or Fratar in the USA) method

2. We know the O_i and D_j at the aggregate level and we have to find T_{ij} (unknown) – Synthetic methods

- Gravity models

3. Logit model (next lecture)

Growth factor methods

- Uniform growth factor method
  - Unrealistic except perhaps for very short time spans (e.g. 2 years)
- Singly constrained growth factor methods
- Doubly constrained growth factor methods

Uniform growth factor method

Same concept with trip generation

$$T_{ij} = T_{ij}^0 \times \tau$$

where

T_{ij} = trips from origin zone *i* to destination zone *j* in the future year

T_{ij}^0 = trips from origin zone *i* to destination zone *j* in the base year

τ = a general growth factor from base year to future year applied to all *ij* pairs

Based on past trends or forecasts

Base year	1	2
1	T_{11}^0	T_{12}^0
2	T_{21}^0	T_{22}^0

$\times \tau$

Future year	1	2
1	T_{11}	T_{12}
2	T_{21}	T_{22}

$$T_{11} = T_{11}^0 \times \tau$$

Uniform growth factor method: Example

Base year matrix

	A	B	C	Sum
A	1	2	4	7
B	3	3	4	10
C	4	3	3	10
Sum	8	8	11	

New O
14
10
15

New D	12	15	12	

Compute the future matrix…

Uniform growth factor method: Example

Base year total trips = 27

Future year total trips = 39

Growth factor = 39/27 = 1.44

Future table

	A	B	C	Sum
A	1 × 1.44	2 × 1.44	4 × 1.44	10.08
B	3 × 1.44	3 × 1.44	4 × 1.44	14.4
C	4 × 1.44	3 × 1.44	3 × 1.44	14.4
Sum	11.52	11.52	15.84	

New O
14
10
15

New D	12	15	12	

We didn't do very well in capturing the future values

Singly Constrained Growth factor methods

- Origin-constrained growth factor method:
  - Future growth data is available on trips produced in each zone

- Destination-constrained growth factor method:
  - Future growth data is available on trips attracted in each zone

$T_{ij} = T_{ij}^0 \times \tau_i$ for origin – constrained factors

$T_{ij} = T_{ij}^0 \times \eta_j$ for destination – constrained factors

τ_i = growth factor applied to origin zone *i*

η_j = growth factor applied to destination zone *j*

Singly Constrained Growth factor methods

Origin-constrained growth factor method:

$$\tau_i = \frac{O_i}{O_i^0}$$

O_i = Future trips from origin zone *i*

O_i^0 = Base year trips from origin zone *i*

Base year	1	2
1	T_{11}^0	T_{12}^0
2	T_{21}^0	T_{22}^0

×

τ_1
τ_2

Future year	1	2
1	$T_{11}^0 \times \tau_1$	$T_{12}^0 \times \tau_1$
2	$T_{21}^0 \times \tau_2$	$T_{22}^0 \times \tau_2$

Singly Constrained Growth factor methods: Example

Base year matrix

	A	B	C	Sum
A	1	2	4	7
B	3	3	4	10
C	4	3	3	10
Sum	8	8	11	

New O
14
10
15

New D	12	15	12	

Compute the future matrix…

Singly Constrained Growth factor methods: Example

Origin-constrained growth factor method:

$$\tau_i = \frac{O_i}{O_i^0} \Rightarrow \tau_A = \frac{14}{7} = 2 \qquad \tau_B = \frac{10}{10} = 1 \qquad \tau_C = \frac{15}{10} = 1.5$$

τ
2
1
1.5

→

τ	A	B	C
A	2	2	2
B	1	1	1
C	1.5	1.5	1.5

Singly Constrained Growth factor methods: Example

Multiply element–wise

τ	A	B	C
A	2	2	2
B	1	1	1
C	1.5	1.5	1.5

×

	A	B	C
A	1	2	4
B	3	3	4
C	4	3	3

➡

	A	B	C	Sum
A	2	4	8	14
B	3	3	4	10
C	6	4.5	4.5	15
Sum	11	11.5	16.5	

Sums at destination do not match the expected future totals…

Singly Constrained Growth factor methods

Destination-constrained growth factor method:

$$\eta_j = \frac{D_j}{D_j^0}$$

D_j = Future trips from destination zone *j*

D_j^0 = Base year trips from destination zone *j*

Base year	1	2
1	T_{11}^0	T_{12}^0
2	T_{21}^0	T_{22}^0

$\times$

η_1	η_2

→

Future year	1	2
1	$T_{11}^0 \times \eta_1$	$T_{12}^0 \times \eta_2$
2	$T_{21}^0 \times \eta_1$	$T_{22}^0 \times \eta_2$

Doubly Constrained Growth factor methods

- Information available and reliable on both future trips originated and attracted to each zone
- Average growth factor method
- Furness method (Fratar in the USA)

Average growth factor

Average zonal growth factor:

$$P_{ij} = \frac{\tau_i + \eta_j}{2} = \frac{\frac{O_i}{O_i^0} + \frac{D_j}{D_j^0}}{2}$$

Base year	1	2
1	T_{11}^0	T_{12}^0
2	T_{21}^0	T_{22}^0

$\times$

P_{11}	P_{12}
P_{21}	P_{22}

$=$

Future year	1	2
1	$T_{11}^0 \times P_{11}$	$T_{12}^0 \times P_{12}$
2	$T_{21}^0 \times P_{21}$	$T_{22}^0 \times P_{22}$

Neither Os or Ds will match the expected future values

The Furness method

- A doubly constrained growth factor method
- Furness suggests updating the initial matrix by adjusting alternatively both the constraints to origins and destinations, until convergence is reached
- The final O-D matrix is obtained by iteratively adjusting the origins and destinations until both the new origins and destinations are close enough to the target values

The Furness method

$$T_{ij} = T_{ij}^0 P_{ij} = T_{ij}^0 \times \frac{O_i}{O_i^0} \times \frac{D_j}{D_j^0} \times A_i \times B_j = T_{ij}^0 \times a_i \times b_j$$

$$a_i = \frac{O_i}{O_i^0} \times A_i$$

$$b_j = \frac{D_j}{D_j^0} \times B_j$$

A_i and B_j are balancing factors

The Furness method

For a_i :

$$O_i = \sum_j T_{ij} = \sum_j \left(T_{ij}^0 \times a_i \times b_j\right) = a_i \times \sum_j \left(T_{ij}^0 \times b_j\right) \Longrightarrow$$

$$a_i = \frac{O_i}{\sum_j \left(T_{ij}^0 \times b_j\right)}$$

For b_j :

$$D_j = \sum_i T_{ij} = \sum_i \left(T_{ij}^0 \times a_i \times b_j\right) = b_j \times \sum_i \left(T_{ij}^0 \times a_i\right) \Longrightarrow$$

$$b_j = \frac{D_j}{\sum_i \left(T_{ij}^0 \times a_i\right)}$$

The Furness method: Algorithm

- We need the base year full matrix and future O_i and D_j values
- Iteration 0:
  - Calculate initial a_i and b_j as: $a_i = \frac{O_i}{O_i^0}$ and $b_i = \frac{D_j}{D_j^0}$
  - If all α_i and b_j values are within 0.95 – 1.05, STOP; Else, next iteration

- Iteration 1: origin constrained growth
  - Multiply matrix by α_i
  - Calculate new row and column totals and the new b_j
  - If all new b_j values are within 0.95 – 1.05, STOP; Else, next iteration

- Iteration 2: origin constrained growth
  - Multiply matrix by b_j
  - Calculate new row and column totals and the new α_i
  - If all new α_i values are within 0.95 – 1.05, STOP; Else, GOTO Iteration 1

The Furness method: Example

Base year matrix

	A	B	C	Sum
A	1	2	4	7
B	3	3	4	10
C	4	3	3	10
Sum	8	8	11	

New O
14
10
15

New D	12	15	12	

Use the Furness method to calculate the new matrix…

The Furness method: Example

Base year matrix

	A	B	C	O_i^0	O_i	a_i=O_i/O_i^0
A	1	2	4	7	14	2
B	3	3	4	10	10	1
C	4	3	3	10	15	1.5
D_j^0	8	8	11			
D_j	12	15	12			
b_j=D_j/D_j^0						

Step 1: adjust the origins

The Furness method: Example

Iteration 1:

	A	B	C	O_i^0	O_i	a_i=O_i/O_i^0
A	2	4	8	14	14	1
B	3	3	4	10	10	1
C	6	4.5	4.5	15	15	1
D_j^0	11	11.5	16.5			
D_j	12	15	12			
b_j=D_j/D_j^0	1.09	1.30	0.73			

Step 2: Compute b_j and check if between 0.95 – 1.05

The Furness method: Example

Iteration 2:

	A	B	C	O_i^0	O_i	$a_i = O_i / O_i^0$
A	2.18	5.22	5.82	13.22	14	1.06
B	3.27	3.91	2.91	10.09	10	0.99
C	6.55	5.87	3.27	15.69	15	0.96
D_j^0	12	15	12			
D_j	12	15	12			
$b_j = D_j / D_j^0$	0.99	1	1			

Step 4: Compute a_i and check if between 0.95 – 1.05

Step 3: adjust the destinations

The Furness method: Example

Back to Iteration 1:

	A	B	C	O_i^0	O_i	a_i=O_i/O_i^0
A	2.60	4.53	6.89	14	14	1
B	3.63	3.17	3.21	10	10	1
C	7.00	4.58	3.48	15	15	1
D_j^0	11.81	15.01	12.17			
D_j	12	15	12			
b_j=D_j/D_j^0	1.02	0.99	0.99			

Step 5: Adjust origins

… converged!

Step 6: Compute b_j and check if between 0.95 – 1.05

Growth Factor Methods

Advantages

- Simple and easy to understand and implement
- Make use of existing matrix and trip end (O or/and D) forecasts
- Suitable for short term planning
- Can use a ‘Uniform Growth Factor’ or ‘Zonal Growth Factors’

Disadvantages

- Needs lot of data (base year trip matrix)
- The results ‘replicate’ the patterns in the existing matrix – not suitable for long term planning
- Suffer from empty cell problem, i.e. if a cell is empty ($T_{ij} = 0$), growth factor methods cannot compute a new value to the empty cell
- Existing low volume cells also result in problematic outputs
- Cannot add new traffic zones
- Cannot account for changes in transport costs – limited in its ability to model changes in policy

Synthetic Distribution Models

- If no reliable previous matrix exists but there are reliable O_i and D_j
- Generating T_{ij} from O_i and D_j
- A common method is the gravity model:
  - Newton's Law of Gravity: the relative strength (i.e. gravity) between two objects is proportional to their masses and inversely proportional to the distance between the two
  - In trip distribution, the OD trips (due to attractions between two zones) are proportional to O_i and D_j and inversely proportional to travel cost (generalised cost) attributes (e.g. distance, time and cost)

The Generalised Cost (Reminder)

- Summarises everything about the travel experience from origin O to destination D
- A measure combining all the main cost attributes related to the disutility of a journey
- Typically a linear function of the attributes of the journey weighted by coefficients which attempt to represent their relative importance
- If different options available, each of them has a different generalised cost (even if monetary cost is the same)

The Generalised Cost form

$$c_{ij} = a_0 + a_1 t_1 + a_2 * t_2 + a_3 * t_3$$

where

c_{ij} : generalised cost from origin zone *i* to destination zone *j*

t_1 - t_3: factors affecting the perceived generalised cost from origin zone i to destination zone j (e.g. travel time, travel cost, level of comfort, etc.)

a_0 : constant (specific to the *ij* pair, travel mode used etc.)

a_1 - a_3: weights to be estimated (convert all factors to the same units)

The Gravity model

Newton's law

$$F_{1,2} = \frac{M_1 . M_2}{d_{1,2}^2}$$

where

$F_{1,2}$: the relative strength (i.e. gravity)

M_1, M_2: the masses

$d_{1,2}$: the distance between M_1 and M_2

The Gravity model: for trip distribution

The Gravity model

$$T_{ij} = kT_i . T_j . f(c_{ij})$$

where

T_{ij} : trips from origin zone i to destination zone j

k : the proportionality constant or balancing factor

c_{ij} : generalised cost from origin zone *i* to destination zone *j*

$f(c_{ij})$: the "deterrence function" representing the disincentive to travel

The Deterrence Function

Typical forms:

$$f(c_{ij}) = c_{ij}^{-n}$$ Power function

$$f(c_{ij}) = \exp(-\beta c_{ij})$$ Exponential function

$$f(c_{ij}) = c_{ij}^{-n} \exp(-\beta c_{ij})$$ Combined function

n, β: parameters to be estimated

The balancing factor k

- A single balancing factor may not be sufficient to ensure a match between the O and D totals
- Instead, we can use two balancing factors (and follow a Furness method approach)

$$T_{ij}=kT_iT_jf(c_{ij}) \longrightarrow T_{ij}=A_iD_jT_iT_jf(c_{ij})$$

- Origin constrained case:

$$A_i = 1/\sum D_j f(c_{ij}), \text{ all } B_j = 1$$

- Destination constrained case:

$$B_i = 1/\sum O_i f(c_{ij}), \text{ all } A_i = 1$$

- Origin and destination constrained case:

$$A_i = 1/\sum \left[B_jD_jf(c_{ij})\right], \text{ and } B_i = 1/\sum \left[A_iO_if(c_{ij})\right]$$

Example: uniform k

Steps:

1. $T_{ij}=kT_iT_jf(c_{ij})$

1. $\sum_{ij} T_{ij} = k\sum_{ij}\left[O_iD_jf(c_{ij})\right]$

2. $k = \frac{\sum_{ij} T_{ij}}{\sum_{ij}\left[O_iD_jf(c_{ij})\right]} = \frac{T}{\sum_{ij}\left[O_iD_jf(c_{ij})\right]}$

Example: uniform k

The empty OD matrix

	1	2	Totals
1	??	??	5
2	??	??	5
Totals	7	3	

The generalised cost matrix

c_{ij}	1	2
1	2	1
2	3	5

Example: uniform k

$$\sum_{ij} T_{ij} = k \sum_{ij} \left[O_i D_j f(c_{ij}) \right] =$$

$$k\left(O_1 D_1 c_{11}^{-1} + O_1 D_2 c_{12}^{-1} + O_2 D_1 c_{21}^{-1} + O_2 D_2 c_{22}^{-1}\right) =$$

$$k\left(5 \times 7 \times \frac{1}{2} + 5 \times 3 \times \frac{1}{1} + 5 \times 7 \times \frac{1}{3} + 5 \times 3 \times \frac{1}{5}\right) = k \times 47.17$$

$$k = \frac{\sum_{ij} T_{ij}}{\sum_{ij} \left[O_i D_j f(c_{ij}) \right]} = \frac{T}{\sum_{ij} \left[O_i D_j f(c_{ij}) \right]} = \frac{10}{47.17} = 0.21$$

The totals don't look great

We can now compute every T_{ij}

	1	**2**	**Totals**
1	17.5 ×0.21	15×0.21	6.83
2	11.67 ×0.21	3×0.21	3.08
Totals	6.125	3.78	

Example: singly constrained gravity model

Origin constrained gravity model

1. $T_{ij} = A_i O_i D_j f(c_{ij})$
2. $\sum_j T_{ij} = A_i O_i \sum_{ij} \left[D_j f(c_{ij}) \right]$ (sum over j)
3. $A_i = \frac{1}{\sum_j \left[D_j f(c_{ij}) \right]}$

Destination constrained gravity model

1. $T_{ij} = B_j O_i D_j f(c_{ij})$
2. $\sum_i T_{ij} = B_j D_j \sum_i \left[O_i f(c_{ij}) \right]$ (sum over i)
3. $B_j = \frac{1}{\sum_i \left[O_i f(c_{ij}) \right]}$

Example: origin constrained gravity model

The empty OD matrix

	1	**2**	**Totals**
1	??	??	5
2	??	??	5
Totals	7	3	

The generalised cost matrix

c_{ij}	**1**	**2**
1	2	1
2	3	5

Example: origin constrained gravity model

Solution: Step 1

$$A_1 = \frac{1}{\sum_j D_j \frac{1}{c_{1j}}} = \frac{1}{D_1 \times \frac{1}{c_{11}} + D_2 \times \frac{1}{c_{12}}} = \frac{1}{7 \times \frac{1}{2} + 3 \times \frac{1}{1}} = 0.154$$

$$A_2 = \frac{1}{\sum_j D_j \frac{1}{c_{2j}}} = \frac{1}{D_1 \times \frac{1}{c_{21}} + D_2 \times \frac{1}{c_{22}}} = \frac{1}{7 \times \frac{1}{3} + 3 \times \frac{1}{5}} = 0.34$$

Example: origin constrained gravity model

Solution: Step 2

$T_{11} = A_1 O_1 D_1 c_{11}^{-1} = 0.154 \times 5 \times 7 \times \frac{1}{2} = 2.695$

$T_{12} = A_1 O_1 D_2 c_{12}^{-1} = 0.154 \times 5 \times 3 \times \frac{1}{1} = 2.31$

$T_{21} = A_2 O_2 D_1 c_{21}^{-1} = 0.34 \times 5 \times 7 \times \frac{1}{3} = 3.97$

$T_{22} = A_2 O_2 D_2 c_{22}^{-1} = 0.34 \times 5 \times 3 \times \frac{1}{5} = 1.02$

	1	**2**	**O_i**	**A_i**
1	??	??	5	0.154
2	??	??	5	0.097
D_j	7	3		

	1	**2**	**O_i**	**A_i**
1	2.695	2.31	5	0.154
2	3.97	1.02	5	0.097
D_j	6.66	3.33		

Example: doubly constrained gravity model

The empty OD matrix

	1	**2**	**Totals**
1	??	??	5
2	??	??	5
Totals	7	3	

The generalised cost matrix

c_{ij}	**1**	**2**
1	2	1
2	3	5

Example: doubly constrained gravity model

Solution:

Step 1 (B_j=1):

$$A_1 = \frac{1}{\sum_j B_j D_j \frac{1}{c_{1j}}} = \frac{1}{B_1 \times D_1 \times \frac{1}{c_{11}} + B_2 \times D_2 \times \frac{1}{c_{12}}} = \frac{1}{1 \times 7 \times \frac{1}{2} + 1 \times 3 \times \frac{1}{1}} = 0.154$$

$$A_2 = \frac{1}{\sum_j B_j D_j \frac{1}{c_{2j}}} = \frac{1}{B_1 \times D_1 \times \frac{1}{c_{21}} + B_2 D_2 \times \frac{1}{c_{22}}} = \frac{1}{1 \times 7 \times \frac{1}{3} + 1 \times 3 \times \frac{1}{5}} = 0.34$$

Step 2: calculate B_j based on the new A_i

$$B_1 = \frac{1}{\sum_i A_i O_i \frac{1}{c_{i1}}} = \frac{1}{A_1 \times O_1 \times \frac{1}{c_{11}} + A_2 \times O_2 \times \frac{1}{c_{21}}} = \frac{1}{0.154 \times 5 \times \frac{1}{2} + 0.34 \times 5 \times \frac{1}{3}} = 1.05$$

$$B_2 = \frac{1}{\sum_i A_i O_i \frac{1}{c_{i2}}} = \frac{1}{A_1 \times O_1 \times \frac{1}{c_{12}} + A_2 \times O_2 \times \frac{1}{c_{22}}} = \frac{1}{0.154 \times 5 \times \frac{1}{1} + 0.34 \times 5 \times \frac{1}{5}} = 0.90$$

Example: doubly constrained gravity model

Solution:

Back to Step 1:

$$A_1 = \frac{1}{\sum_j B_j D_j \frac{1}{c_{1j}}} = \frac{1}{B_1 \times D_1 \times \frac{1}{c_{11}} + B_2 \times D_2 \times \frac{1}{c_{12}}} = \frac{1}{1.05 \times 7 \times \frac{1}{2} + 0.90 \times 3 \times \frac{1}{1}} = 0.157$$

$$A_2 = \frac{1}{\sum_j D_j \frac{1}{c_{2j}}} = \frac{1}{B_1 \times D_1 \times \frac{1}{c_{21}} + B_2 D_2 \times \frac{1}{c_{22}}} = \frac{1}{1.05 \times 7 \times \frac{1}{3} + 0.90 \times 3 \times \frac{1}{5}} = 0.334$$

Back to Step 2

$$B_1 = \frac{1}{\sum_i A_i O_i \frac{1}{c_{i1}}} = \frac{1}{A_1 \times O_1 \times \frac{1}{c_{11}} + A_2 \times O_2 \times \frac{1}{c_{21}}} = \frac{1}{0.157 \times 5 \times \frac{1}{2} + 0.334 \times 5 \times \frac{1}{3}} = 1.054$$

$$B_2 = \frac{1}{\sum_i A_i O_i \frac{1}{c_{i2}}} = \frac{1}{A_1 \times O_1 \times \frac{1}{c_{12}} + A_2 \times O_2 \times \frac{1}{c_{22}}} = \frac{1}{0.157 \times 5 \times \frac{1}{1} + 0.334 \times 5 \times \frac{1}{5}} = 0.894$$

Example: doubly constrained gravity model

Solution:

Back to Step 1:

$$A_1 = \frac{1}{\sum_j B_j D_j \frac{1}{c_{1j}}} = \frac{1}{B_1 \times D_1 \times \frac{1}{c_{11}} + B_2 \times D_2 \times \frac{1}{c_{12}}} = \frac{1}{1.054 \times 7 \times \frac{1}{2} + 0.894 \times 3 \times \frac{1}{1}} = 0.157$$

$$A_2 = \frac{1}{\sum_j D_j \frac{1}{c_{2j}}} = \frac{1}{B_1 \times D_1 \times \frac{1}{c_{21}} + B_2 D_2 \times \frac{1}{c_{22}}} = \frac{1}{1.054 \times 7 \times \frac{1}{3} + 0.894 \times 3 \times \frac{1}{5}} = 0.333$$

Back to Step 2

$$B_1 = \frac{1}{\sum_i A_i O_i \frac{1}{c_{i1}}} = \frac{1}{A_1 \times O_1 \times \frac{1}{c_{11}} + A_2 \times O_2 \times \frac{1}{c_{21}}} = \frac{1}{0.157 \times 5 \times \frac{1}{2} + 0.333 \times 5 \times \frac{1}{3}} = 1.055$$

$$B_2 = \frac{1}{\sum_i A_i O_i \frac{1}{c_{i2}}} = \frac{1}{A_1 \times O_1 \times \frac{1}{c_{12}} + A_2 \times O_2 \times \frac{1}{c_{22}}} = \frac{1}{0.157 \times 5 \times \frac{1}{1} + 0.333 \times 5 \times \frac{1}{5}} = 0.894$$

Example: doubly constrained gravity model

Solution:

Back to Step 1:

$$A_1 = \frac{1}{\sum_j B_j D_j \frac{1}{c_{1j}}} = \frac{1}{B_1 \times D_1 \times \frac{1}{c_{11}} + B_2 \times D_2 \times \frac{1}{c_{12}}} = \frac{1}{1.055 \times 7 \times \frac{1}{2} + 0.894 \times 3 \times \frac{1}{1}} = 0.157$$

$$A_2 = \frac{1}{\sum_j D_j \frac{1}{c_{2j}}} = \frac{1}{B_1 \times D_1 \times \frac{1}{c_{21}} + B_2 D_2 \times \frac{1}{c_{22}}} = \frac{1}{1.055 \times 7 \times \frac{1}{3} + 0.894 \times 3 \times \frac{1}{5}} = 0.334$$

We can assume the model converged!

Example: doubly constrained gravity model

Solution:

Step 3:

$$T_{11}=A_1B_1O_1D_1c_{11}^{-1} = 0.157 \times 1.055 \times 5 \times 7 \times \frac{1}{2} = 2.899$$

$$T_{12}=A_1B_2O_1D_2c_{12}^{-1} = 0.157 \times 0.894 \times 5 \times 3 \times \frac{1}{1} = 2.11$$

$$T_{21}=A_2B_1O_2D_1c_{21}^{-1} = 0.334 \times 1.055 \times 5 \times 7 \times \frac{1}{3} = 4.11$$

$$T_{22}=A_2B_2O_2D_2c_{22}^{-1} = 0.334 \times 0.894 \times 5 \times 3 \times \frac{1}{5} = 0.90$$

	1	**2**	$\mathbf{O_i}$	$\mathbf{A_i}$
1	??	??	5	0.157
2	??	??	5	0.334
$\mathbf{D_j}$	7	3		
$\mathbf{B_j}$	1.055	0.894		

	1	**2**	$\mathbf{O_i}$	$\mathbf{A_i}$
1	2.899	2.11	5	0.157
2	4.11	0.90	5	0.334
$\mathbf{D_j}$	7	3		
$\mathbf{B_j}$	1.055	0.894		

The estimated totals O_i and D_j match the expected numbers

Further considerations on the Gravity model

- The quality of our results is affected by the deterrence function
- The deterrence function has several parameters that need to be estimated e.g. n, β but we do not know the value of these parameters
- Usually, we evaluate the quality of a Gravity model by comparing the estimated outputs with an existing matrix
- As a goodness-of-fit indicator we use the observed trip length distribution (OTLD)
- The output of our model is the modelled trip length distribution (MTLD)
- Assumption: the same TLD will be maintained in the future
- There are techniques that allow the comparison between OTLD and MTLD by iteratively trying different values for the unknown parameters (n, β) until the best fit is found (e.g., Hyman method, Poisson model fitting)

Calibration of the deterrence function

- So far we assumed that the deterrence function is known
- However, the parameters of the deterrence function must be estimated
- Method 1: Hyman method
  - Easier to be implemented when the deterrence function only has one parameter
  - More complex for multiple parameters e.g. combined function
  - We calibrate the deterrence function while solving the gravity model
- Method 2: Poisson fitting
  - Requires a known OD matrix
  - No form for the deterrence function (it is assumed as a parameter)

The Hyman method

- The Hyman method is based on the following requirement for β (parameter of the deterrence function to be estimated

$$c(\beta) = \sum_{ij} \left[T_{ij}(\beta) c_{ij}\right] / T(\beta) = c^* = \sum_{ij} \left(N_{ij} C_{ij}\right) / \sum_{ij} N_{ij}$$

- c^* is the mean cost from the OTLD
- N_{ij} is the observed number of trips for each origin destination pair

The Hyman method: procedure

1. Initialise the first iteration (m=0) with $\beta_0 = 1/c^*$
2. Use β_0 and calculate a trip matrix using the gravity model
3. Obtain the mean modelled trip cost c_0 and update β as: $\beta_m = \beta_0 c_0 / c^*$ (only for m=0)
4. Using the latest value for β calculate a trip matrix using a standard gravity model
5. Obtain a new mean modelled trip cost (c_m) and compare it with c^*; if sufficiently closed then STOP. If not, go to step 6.
6. For m > 0, update the β of the next iteration as:

$$\beta_{m+1} = \frac{(MTL - MTL_{m-1})\beta_m - (MTL - MTL_m)\beta_{m-1}}{MTL_m - MTL_{m-1}}$$

7. Repeat steps 4 – 6 until convergence.

Poisson model fitting

Requirements

- Observed OD matrix
- Cost per i-j

Model formulation: $\hat{T}_{ij} = Q_i X_j F_k(c_{ij})$

- $\hat{T}_{ij}$ = estimated trips from origin i to destination j
- Q_i = production potential of zone i
- X_j = attraction potential of zone j
- $F_k(c_{ij})$ = deterrence function with respect to generalised cost – "willingness" to travel from i to j

Poisson model fitting

Poisson model probability

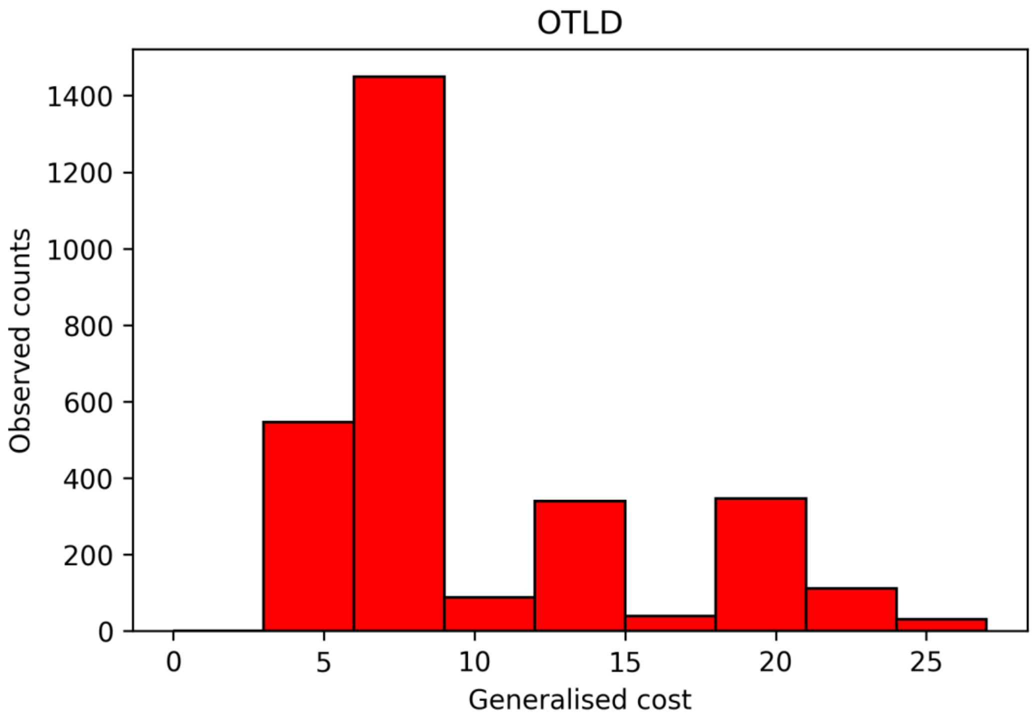
$$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \Longrightarrow P(T_{ij}) = \frac{e^{-\left(Q_i X_j F_k(c_{ij})\right)}\left(Q_i X_j F_k(c_{ij})\right)^{T_{ij}}}{T_{ij}!}$$

- where T_{ij} observed known trips from i to j
- Q_i, X_j, $F_k(c_{ij})$ are treated as unknown parameters to be estimated
- In every iteration we compute the LL $= \sum ln\left(P(T_{ij})\right)$
- We stop when LL stops improving

Poisson model fitting: Example

$$c_{ij} = \begin{pmatrix} 4 & 12 & 19 & 23 \\ 11 & 4 & 14 & 20 \\ 16 & 14 & 6 & 8 \\ 25 & 19 & 7 & 6 \end{pmatrix} \quad P_i = \begin{pmatrix} 601 \\ 691 \\ 601 \\ 1054 \end{pmatrix}$$

$$A_j = \begin{pmatrix} 390 & 601 & 752 & 1204 \end{pmatrix}$$



Poisson model estimator: Iteration 1

- We set the starting values of Q_i, X_j, $F_k(c_{ij})$ to 1 (hence their product results in unity for all i,j cells)

1	1	1	1	601
1	1	1	1	691
1	1	1	1	601
1	1	1	1	1054
390	601	752	1204	Sums

- Scale to productions ($Q_i = P_i/\hat{P}_i$):
  = 601/4
  = 691/4
  = 601/4
  = 1054/4

Poisson model estimator: Iteration 1

- Now we have a new version of the OD matrix…

150.25	150.25	150.25	150.25	601
172.75	172.75	172.75	172.75	691
150.25	150.25	150.25	150.25	601
263.5	263.5	263.5	263.5	1054
390	601	752	1204	Sums

- Scale to attractions ($X_j = A_j/\hat{A}_j$):
  = 390/736.75
  = 601/736.75
  = 752/736.75
  = 1204/736.75

Poisson model estimator: Iteration 1

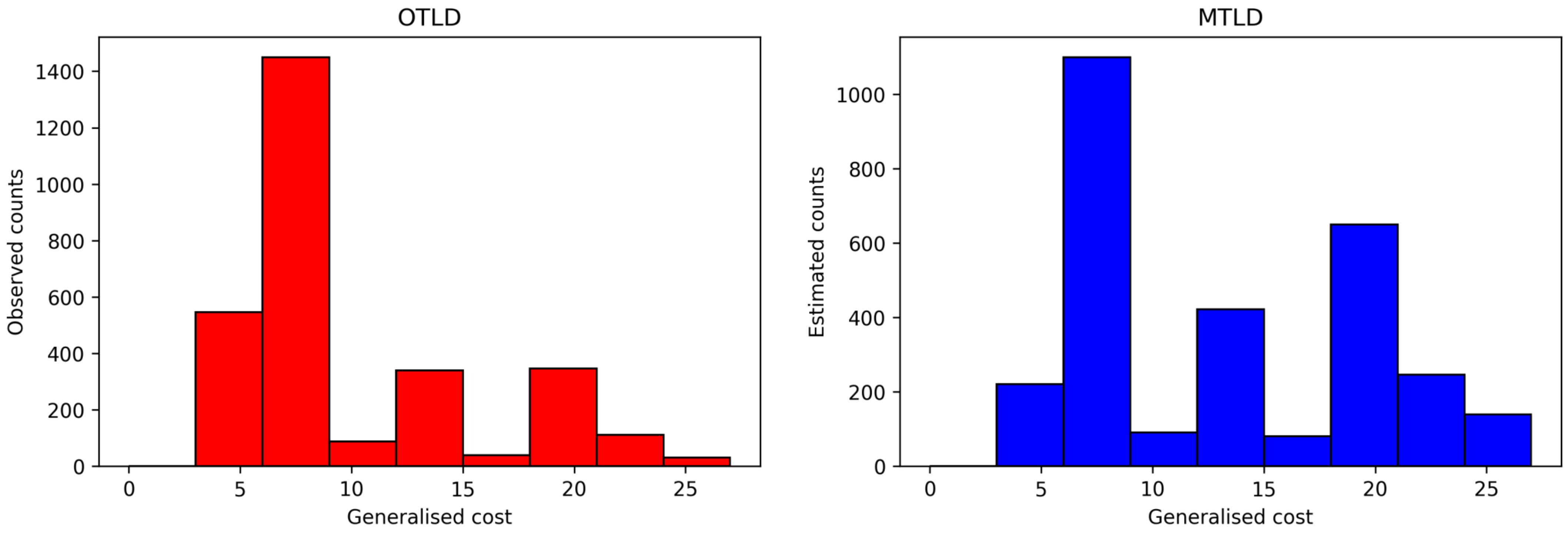
- Now we have a newer new version of the OD matrix (values rounded)…

80	123	153	246	601
91	141	176	282	691
80	123	153	246	601
139	215	269	431	1054
390	601	752	1204	Sums

- Scale to fit the observed distribution

Poisson model estimator: Iteration 1

- Scale to fit the observed distribution (take the ratio between observed and estimated counts for every bin)



The ratios ($F_k(c_{ij})$) are: 2.47, 1.32, 0.96, 0.80, 0.49, 0.53, 0.45, 0.22

Poisson model estimator: Iteration 1

- Now we must find in which bin each cell belongs and multiply with the respective factor from the previous slide
- We know this information from the generalised cost matrix of each ij pair e.g. if generalised cost is 5 then the factor is 2.47, if generalised cost is 10 then the factor is 0.96 etc.

80	123	153	246	601
91	141	176	282	691
80	123	153	246	601
139	215	269	431	1054
390	601	752	1204	Sums

Apply factors →

198	99	81	111	601
87	348	141	150	691
39	99	202	324	601
30	114	355	568	1054
390	601	752	1204	Sums

- Compute the LL and store the value
- Important! While running the loop, store the Q_i, X_j, $F_k(c_{ij})$ values of every iteration (it will become clear later why)

Poisson model estimator: Iteration 2 (and onward)

- Use the result table from iteration 1 as your starting point

198	99	81	111	601
87	348	141	150	691
39	99	202	324	601
30	114	355	568	1054
390	601	752	1204	Sums

- Repeat the whole process, same as iteration 1
- At an iteration n, if LL_n similar to LL_{n-1}, stop the loop
- The final values of Q_i, X_j, $F_k(c_{ij})$ are the products of each for all iterations.
- E.g. for n=10:
  - $F_k(c_{ij}) = F_k(c_{ij})_1 \times F_k(c_{ij})_2 \times F_k(c_{ij})_3 \times \ldots F_k(c_{ij})_{10}$

Trip distribution via discrete choice models

- Except for growth factors and synthetic methods, we can do the trip distribution step using discrete choice models…
- What is a discrete choice model?
- These are the models that we use for the modal split step (Step 3) so we will see next week!

Trip distribution summary

- Background and purpose of trip distribution
- From PA to OD
- Models:
  - Growth factor methods (Furness)
  - Synthetic methods (Gravity model)
- Calibration techniques (Hyman method, Poisson estimation)