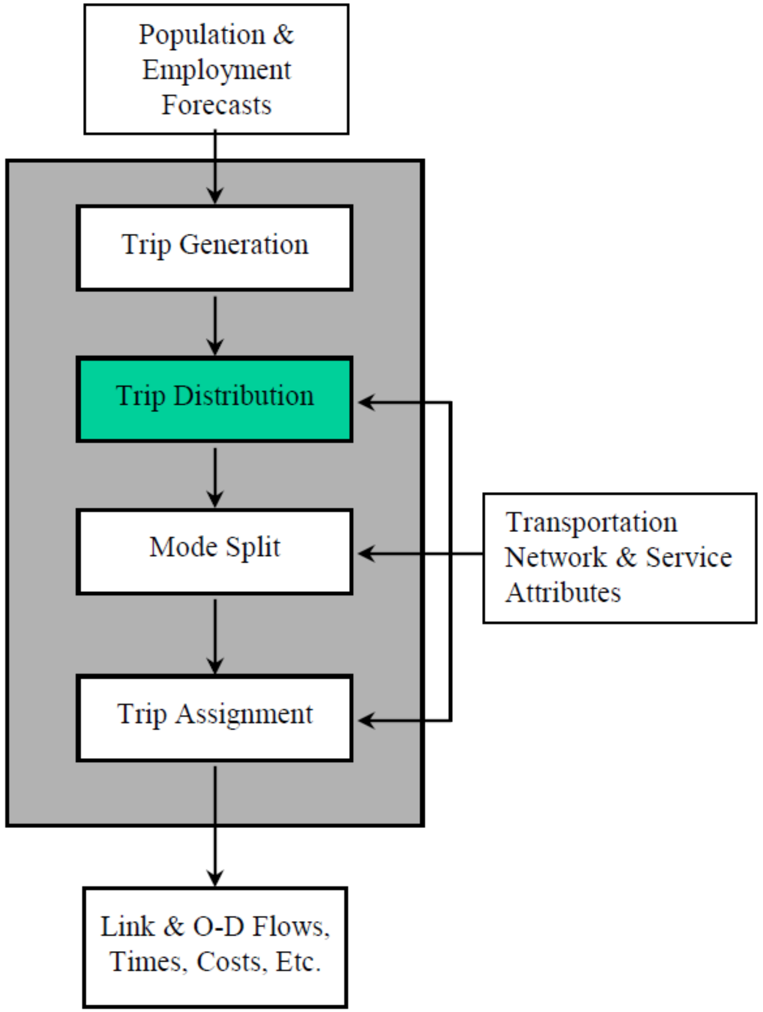
Trip Distribution

CIVE 461 Urban Transportation Planning
Supplemental Notes

Outline

1. Problem definition & terminology
2. Proportional flow theory
3. Growth factor model
4. Synthetic model:
   - Gravity model
   - Entropy maximizing model
5. Biproportional updating:
   - Growth factor model
   - Gravity model
6. Impedance function
7. Goodness of fit and other measures
8. Drawbacks



Problem Definition

- Trip generation gives **total trips** generated from each origin zone (O_i) and total trips attracted to each destination zone (D_j)

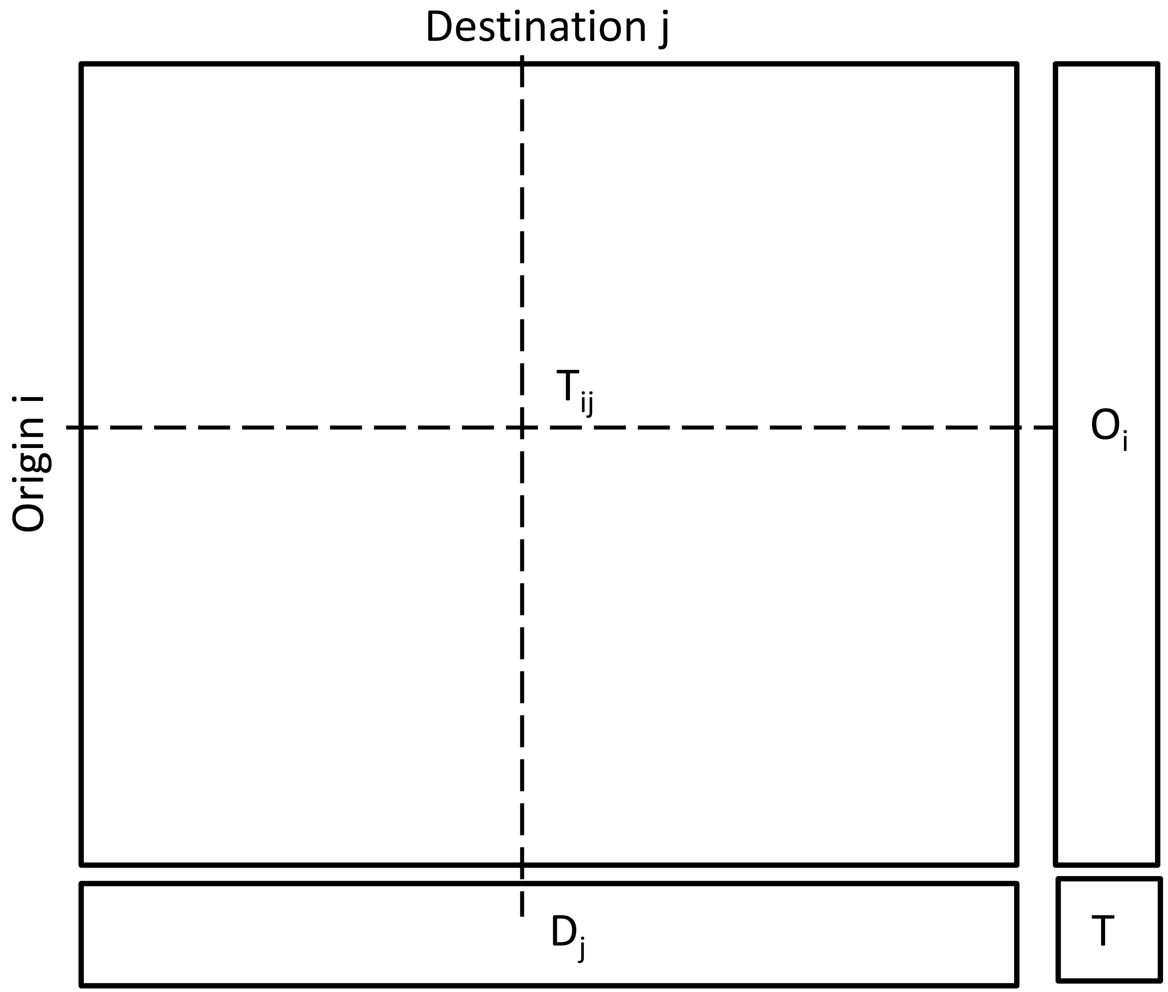
- OD intercept or home travel diaries can be used to determine trip patterns in terms of the combination of origin and destination pairs
  - Even with a large sample, we will have zone pairs that are **not sampled**, so we need a method that can generate trips between OD pairs

- Trip distribution stage estimates the "most likely" **origin-destination (O-D) trip (or flow) matrix**, given known origin and destination totals for each zone (plus any additional relevant & known information)

- We will focus here on **aggregate distribution methods**, but it is possible to consider distribution as a disaggregate choice using **discrete choice models** of destination choice (subsequent lectures)

Defining Terms

T_{ij} = total trips from origin i to destination j
O_i = total trips from origin i
D_j = total trips to destination j
T = total trips = $\sum_i O_i = \sum_j D_j = \sum_i \sum_j T_{ij}$



Defining Terms

	Destinations					
Origins	1	2	3	$\ldots j$	$\ldots z$	$\sum_i T_{ij}$
1	T_{11}	T_{12}	T_{13}		$\ldots T_{1z}$	O_1
2	T_{21}	T_{22}	T_{23}	$\ldots T_{2j}$	$\ldots T_{2z}$	O_2
3			T_{33}	$\ldots T_{3j}$	$\ldots T_{3z}$	O_3
$\vdots$						
i	T_{i1}	T_{i2}	T_{i3}	$\ldots T_{ij}$		
$\vdots$						
z	T_{z1}	T_{z2}	T_{z3}	$\ldots T_{zj}$	$\ldots T_{zz}$	O_z
$\sum_i T_{ij}$	D_1	D_2	D_3	$\ldots D_j$	$\ldots D_z$	$\sum_{ij} T_{ij} = T$

Trips from zone 2…

… to zone 3

Intrazonal trips

Defining Terms

- **Lower letter notation** (t_{ij}, o_i, d_i) denotes observations from a sample or from an earlier study

- We can further disaggregate trips by **person type** (n) and/or by **mode** (k)

$$T_{ij}^{kn} = trips\ from\ i\ to\ j\ by\ mode\ k\ and\ person\ type\ n$$

- Summation over sub- or superscripts may be indicated implicitly through omissions. E.g.,

$$T_{ij}^{n} = \sum_{k} T_{ij}^{kn}$$

Proportional Flow Theory

- It is **not unreasonable** to assume that the flow between i and j is proportional to the total flow out of i and into j:

$$T_{ij} \propto O_i D_j$$
$$T_{ij} = k O_i D_j$$

where k is a proportionality constant

- **Summing over i or j** and substituting into above relationship gives

$$\sum_i T_{ij} = \sum_i k O_i D_j \ or \ \sum_j T_{ij} = \sum_j k O_i D_j$$
$$D_j = k D_j T \ or \ O_i = k O_i T$$
$$k = \frac{1}{T}$$
$$\therefore T_{ij} = \frac{O_i D_j}{T}$$

Formal Trip Distribution Models

Possible methods/models to forecast future O-D distribution:

1. **Growth factor models**: simply updating base matrix
   - Uniform growth factor method
   - Singly constrained growth factor method
2. **Synthetic models**: use equations/models to predict future matrix
   - Gravity distribution model
   - Entropy maximizing model
3. **Biproportional updating** (also known as Fratar or Furness method)

Growth Factor Methods

- Use **base year O-D matrix** to update to **target year O-D distribution**: A non-parametric approach to trip distribution modeling
  - Base year O-D matrix needs to be accurate for an accurate forecast
  - Base year O-D patterns remain the same for future year. **Only cell values change**.

Uniform Growth Factor Model

Consider fixed growth rate (τ) for all O-D pairs

$$T_{ij} = \tau t_{ij}$$

Base trip matrix

	1	2	3	4	$\mathbf{o_i}$
1	5	50	100	200	**355**
2	50	5	100	300	**455**
3	50	100	5	100	**255**
4	100	200	250	20	**570**
$\mathbf{d_j}$	**205**	**355**	**455**	**620**	**t = 1635**

Consider 20% uniform growth – i.e., $\tau = 1.2$

Forecast trip matrix

	1	2	3	4	$\mathbf{O_i}$
1	6	60	120	240	**426**
2	60	6	120	360	**546**
3	60	120	6	120	**306**
4	120	240	300	24	**684**
$\mathbf{D_j}$	**246**	**426**	**546**	**744**	**T = 1962**

Constrained Growth Factor

- Logical constraints that any feasible trip matrix must satisfy

$$\sum_j \tau_i T_{ij} = O_i \;\; \forall i \in [1,..,N] \; (\textbf{row constraint})$$

$$\sum_i \tau_j T_{ij} = D_j \;\; \forall j \in [1,..,N] \; (\textbf{column constraint})$$

- If only one constraint is met, the corresponding model is called **singly constrained model**
- If both constraints are met, the corresponding model is called **doubly constrained model**
- Can define an origin-specific growth factor (τ_i) or a destination-specific growth factor (τ_j)

Singly Constrained Growth Factor

Base trip matrix

	1	2	3	4	o_i
1	5	50	100	200	**355**
2	50	5	100	300	**455**
3	50	100	5	100	**255**
4	100	200	250	20	**570**
d_j	**205**	**355**	**455**	**620**	**t = 1635**

Target O_i
400
460
400
702
T = 1962

- **Origin-specific growth factors** are then $\tau_i = \frac{O_i}{o_i} = [1.127, 1.011, 1.569, 1.232]$

Forecast trip matrix

	1	2	3	4	O_i	Target O_i
1	6	56	113	225	**400**	**400**
2	51	5	101	303	**460**	**460**
3	78	157	8	157	**400**	**400**
4	123	246	308	25	**702**	**702**
D_j	**258**	**464**	**530**	**710**	**T = 1962**	**T = 1962**

Growth Factor Methods: Pros & Cons

Pros:
- **Simple** to understand/estimate and directly use observed O-D matrix
- Good for **short-term planning** for changes in total demand with minimal changes in O-D distribution patterns

Cons:
- Requires an **accurate and full O-D table**: If any cell in base O-D matrix has zero value, it will be maintained into the future
- Not a reliable method if O-D patterns **expected to change** in forecasting year
- Does not consider trip length and its variation (**congestion delays**) due to changes in demand, network configuration, network performance, etc.
- Not a sensitive tool for policy analysis!

Synthetic Approach Gravity Model

- Rather than using base O-D table, synthetic approach uses **equations** to generate trip distribution table
- Synthetic approaches are **sensitive to O-D trip length**
- Basic formulation:

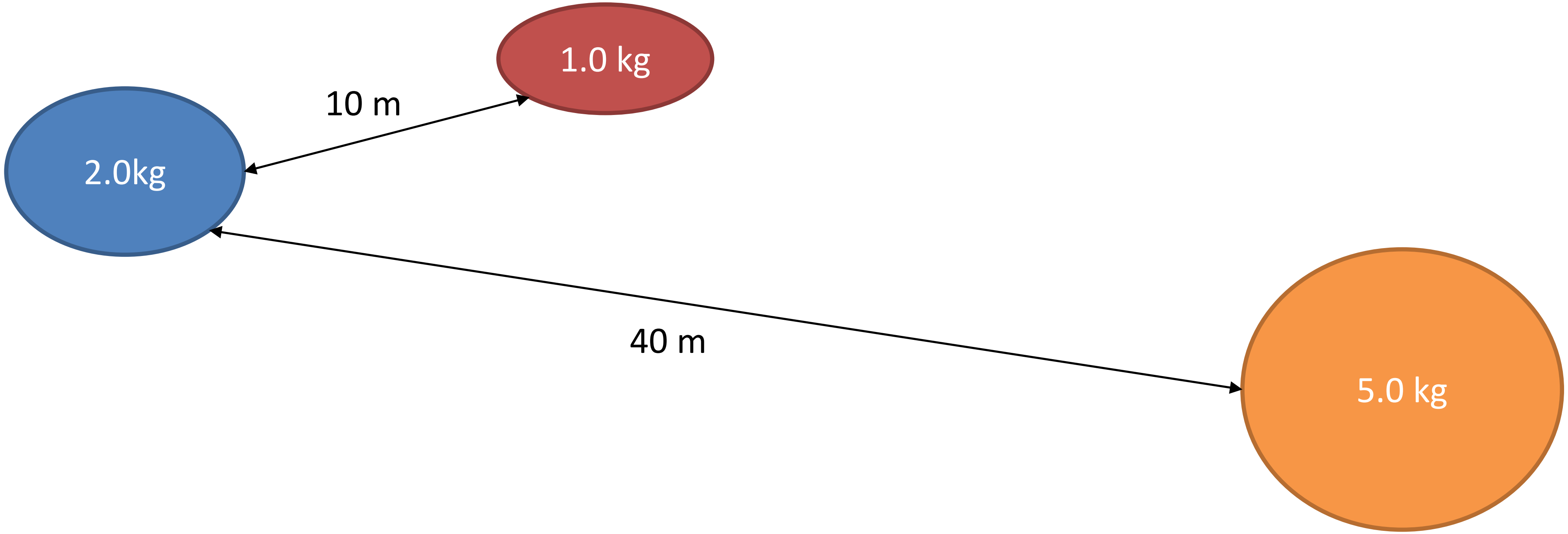
$$T_{ij} \propto O_i D_j f\left(c_{ij}\right)$$
$$T_{ij} = \alpha O_i D_j f_{ij}$$

Where α is a **proportionality constant**

- $f\left(c_{ij}\right)$ or f_{ij} is a function explaining **travel impedance/attraction** between origin i and destination j: $\frac{\partial f\left(c_{ij}\right)}{\partial c_{ij}} < 0$
  - Should capture the cost function for travel and the fact **increasing cost decreases attraction** between an O-D pair

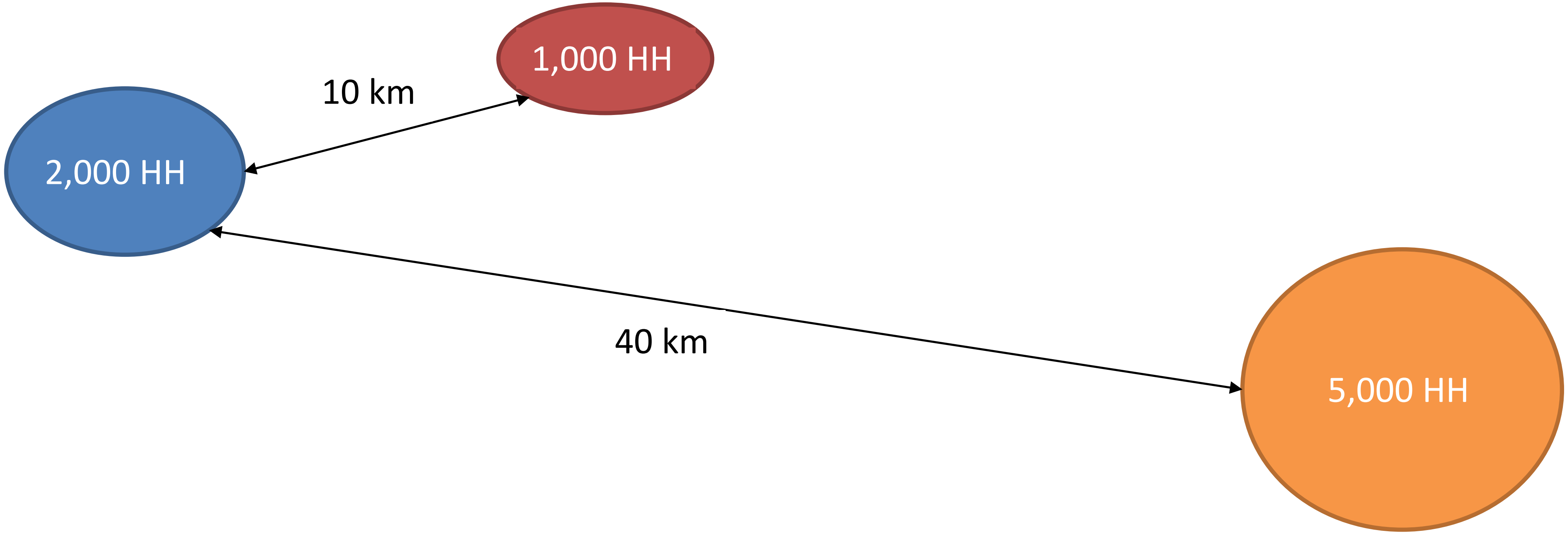
Gravity Model Analogy

$$F = \frac{Gm_1m_2}{r^2}$$



Gravity Model Analogy

$$F = \frac{Gm_1m_2}{r^2}$$



Singly-Constrained Gravity Model

Do not depend on index j

$$\sum_j T_{ij} = O_i = \sum_j \boxed{\alpha_i O_i} D_j f_{ij} \rightarrow \alpha_i = \frac{1}{\sum_j D_j f_{ij}} \quad (1)$$

$$\sum_i T_{ij} = D_j = \sum_i \boxed{\alpha_j} O_i \boxed{D_j} f_{ij} \rightarrow \alpha_j = \frac{1}{\sum_i O_i f_{ij}} \quad (2)$$

Do not depend on index i

- For an **origin-constrained gravity model**

$$T_{ij} = \frac{O_i D_j f_{ij}}{\sum_j D_j f_{ij}}$$

Weighted average as we saw in in-class work on statistics!

- For a **destination-constrained gravity model**

$$T_{ij} = \frac{O_i D_j f_{ij}}{\sum_i O_i f_{ij}}$$

Doubly-Constrained Gravity Model

- A doubly-constrained model requires **iteration between the origin- and destination-constrained models** until both converge – i.e., there are only **minimal cell value differences** between successive iterations

- First the origin-constrained gravity model

$$T_{ij} = \frac{O_i D_j^* f_{ij}}{\sum_j D_j f_{ij}}$$

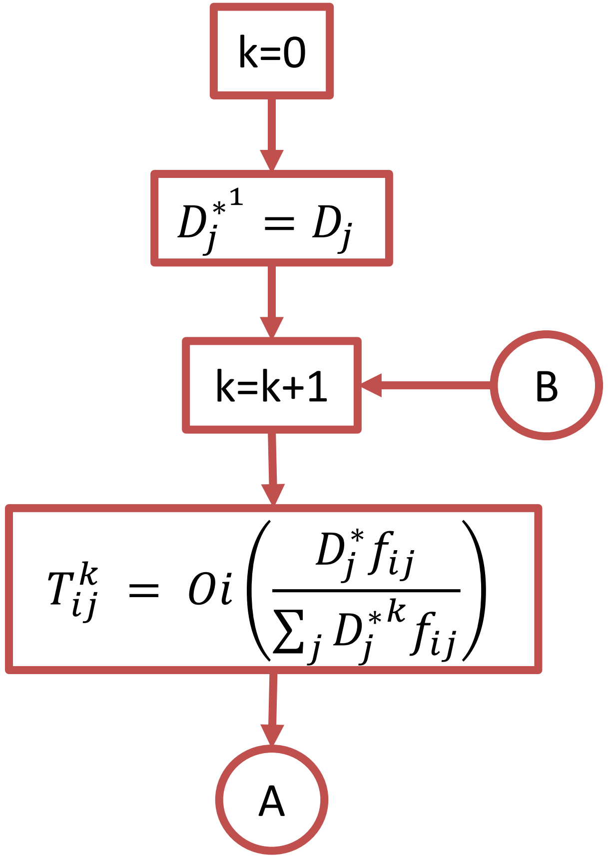
- Then the destination-constrained gravity model

$$T_{ij} = \frac{O_i^* D_j f_{ij}}{\sum_i O_i f_{ij}}$$

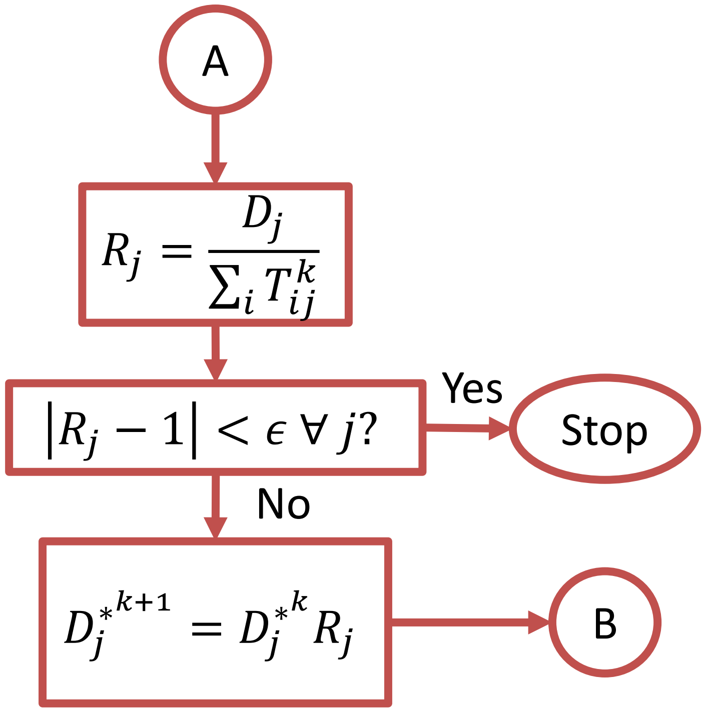
- D_j^* and O_i^* are intermediate column and row totals, respectively

- α_i and α_j are often labeled A and B, respectively

Doubly-Constrained Gravity Model Balancing Procedure



Doubly-Constrained Gravity Model Balancing Procedure Cont…



Doubly-Constrained Gravity Model Example

- A very simple future O-D trip matrix:

	1	2	O_i
1			**200**
2			**300**
D_j	**100**	**400**	**T = 500**

- Impedance function $\boxed{f_{ij} = 1/t_{ij}^2}$ Analogous to the inverse distance in Newton's Universal Law!

t_{ij}	1	2
1	5	15
2	12	6

Doubly-Constrained Gravity Model Example Solution

Future Year Trip Origins Destinations

O/D:	1	2	Total
1			200
2			300
Total	100	400	500

1

Impedance Function $f_{ij} = 1 / t_{ij}^2$

i \ j	1	2
1	0.04000	0.00444
2	0.00694	0.02778

Iteration k=1: ($D^{*1}_j = D_j$)

D^{*1}_j= 100 400

Denominator, Origin i=1 5.78 ← $\Sigma_{j'} D^{*k}_{j'} f_{1j'}$

Denominator, Origin i=2 11.81 ← $\Sigma_{j'} D^{*k}_{j'} f_{2j'}$

i \ j	1	2	Total
1	138.5	61.5	200.0
2	17.6	282.4	300.0
Total	156.1	343.9	500

← $T_{ij}^k = O_i D^{*k}_j f_{ij} / \Sigma_{j'} D^{*k}_{j'} f_{ij'}$

Rj= 0.641 1.163 ← $D_j / \Sigma_i T_{ij}^k$

|Rj-1|= 0.359 0.163

Doubly-Constrained Gravity Model Example Solution

Future Year Trip Origins Destinations

O/D:	1	2	Total
1			200
2			300
Total	100	400	500

2

Iteration k=2: (D*j(new)= D*j(old)*Rj)

D^{*2}_{j}=	64.1	465.3	
Denominator, Origin i=1			4.63
Denominator, Origin i=2			13.37

i \ j	1	2	Total
1	110.7	89.3	200.0
2	10.0	290.0	300.0
Total	120.7	379.3	500
Rj=	0.829	1.054	
|Rj-1|=	0.171	0.054	

3

Iteration k=3: (D*j(new)= D*j(old)*Rj)

D^{*3}_{j}=	53.1	490.6	
Denominator, Origin i=1			4.30
Denominator, Origin i=2			14.00

i \ j	1	2	Total
1	98.7	101.3	200.0
2	7.9	292.1	300.0
Total	106.6	393.4	500
Rj=	0.938	1.017	
|Rj-1|=	0.062	0.017	

4

Iteration k=4: (D*j(new)= D*j(old)*Rj)

D^{*4}_{j}=	49.8	498.8	
Denominator, Origin i=1			4.21
Denominator, Origin i=2			14.20

i \ j	1	2	Total
1	94.7	105.3	200.0
2	7.3	292.7	300.0
Total	102.0	398.0	500
Rj=	0.981	1.005	

Assume ε < = 2% : So, converged

Entropy Theory

- Consider a system with **a large number of distinct elements**
  - A full system description would require complete specification of its **micro states** – each individual traveler, their origin, destination, and mode
  - In many cases, it may be sufficient to work with a **meso state** specification – total trips between each origin and destination
  - There may be **numerous micro states** that produce the **same meso state**
  - **Macro state** – all trips on a particular link or total trips generated or attracted to each zone
- Entropy basis is that **all micro and meso states** consistent with a given macro state are **equally likely** to occur

Entropy Theory

- Micro states associated with meso state T_{ij} given by

$$W(T_{ij}) = \frac{T!}{\prod_{ij} T_{ij}!}$$

- If all micro states are equally likely, **what is the most likely meso state**?
  - One that results from most **micro states**

- We can define a **Lagrangian optimization problem** to maximize log(W) (producing a gravity model), but this would require technical background outside the undergraduate CEE curriculum…

Synthetic Model: Entropy Formulation

- Gravity model can be rewritten more compactly in the following form:

$$T_{ij} = A_i O_i B_j D_j f_{ij}$$

- We can then derive terms for A_i and B_j based on the **same conditions found in previous slides** – i.e., $\sum_j T_{ij} = O_i$ and $\sum_i T_{ij} = D_j$

$$A_i = \frac{1}{\sum_j B_j D_j f_{ij}}$$

$$B_j = \frac{1}{\sum_i A_i O_i f_{ij}}$$

- The same result can be found based on **information theory** – i.e., via entropy maximization instead of gravity assumption

Entropy Maximizing Model Balancing Procedure

k=0

$$B_j^{*0} = 1$$

k=k+1

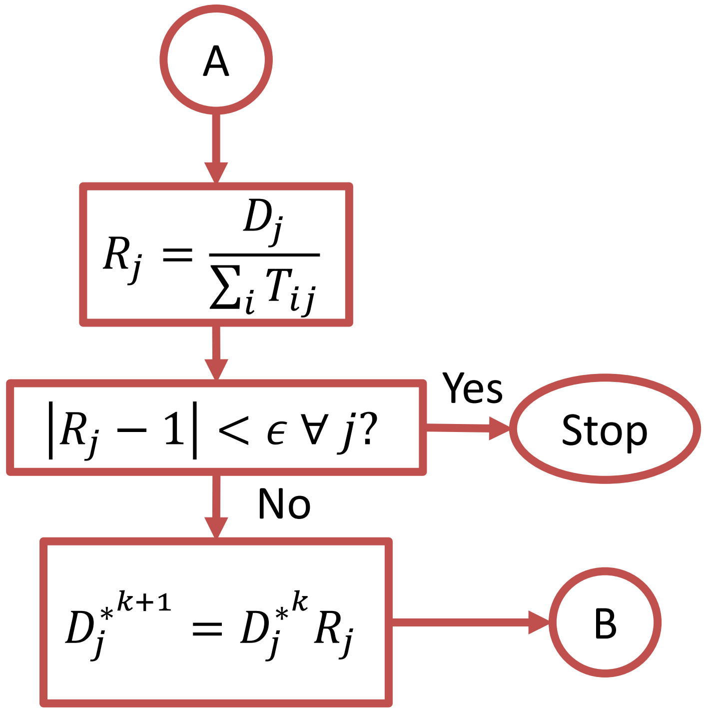
B

$$A_i^k = \frac{1}{\sum_j B_j^{k-1} D_j f_{ij}}; T_{ij}^k = A_i^k O_i B_j^{k-1} D_j f_{ij} = \frac{O_i B_j^{k-1} D_j f_{ij}}{\sum_j B_j^{k-1} D_j f_{ij}}$$

Update $B_j^k = \frac{1}{\sum_i A_i^k o_i f_{ij}}$

A

Entropy Maximizing Model Balancing Procedure Cont…



Doubly-Constrained Entropy Model Example Solution

Impedance Function $f_{ij} = 1 / t_{ij}^2$

i \ j	1	2
1	0.04000	0.00444
2	0.00694	0.02778

Future Year Trip Origins Destinations

O/D:	1	2	Total
1			200
2			300
Total	100	400	500

1

Iteration1 (Bj = 1)

Bj = 1 1

Denom A1 5.78

Denom A2 11.81

O/D:	1	2	Total	Ai
1	138.5	61.5	200.0	0.173077
2	17.6	282.4	300.0	0.084706
Total	156.1	343.9	500	

2

Iteration2

Bj = 0.64058 1.163158

Denom A1 4.63

Denom A2 13.37

O/D:	1	2	Total	Ai
1	110.7	89.3	200.0	0.215975
2	10.0	290.0	300.0	0.074801
Total	120.7	379.3	500	

3

Iteration3

Bj = 0.530887 1.226514

Denom A1 4.30

Denom A2 14.00

O/D:	1	2	Total	Ai
1	98.7	101.3	200.0	0.232341
2	7.9	292.1	300.0	0.071446
Total	106.6	393.4	500	

4

Iteration 4

Bj = 0.498114 1.247026

Denom A1 4.21

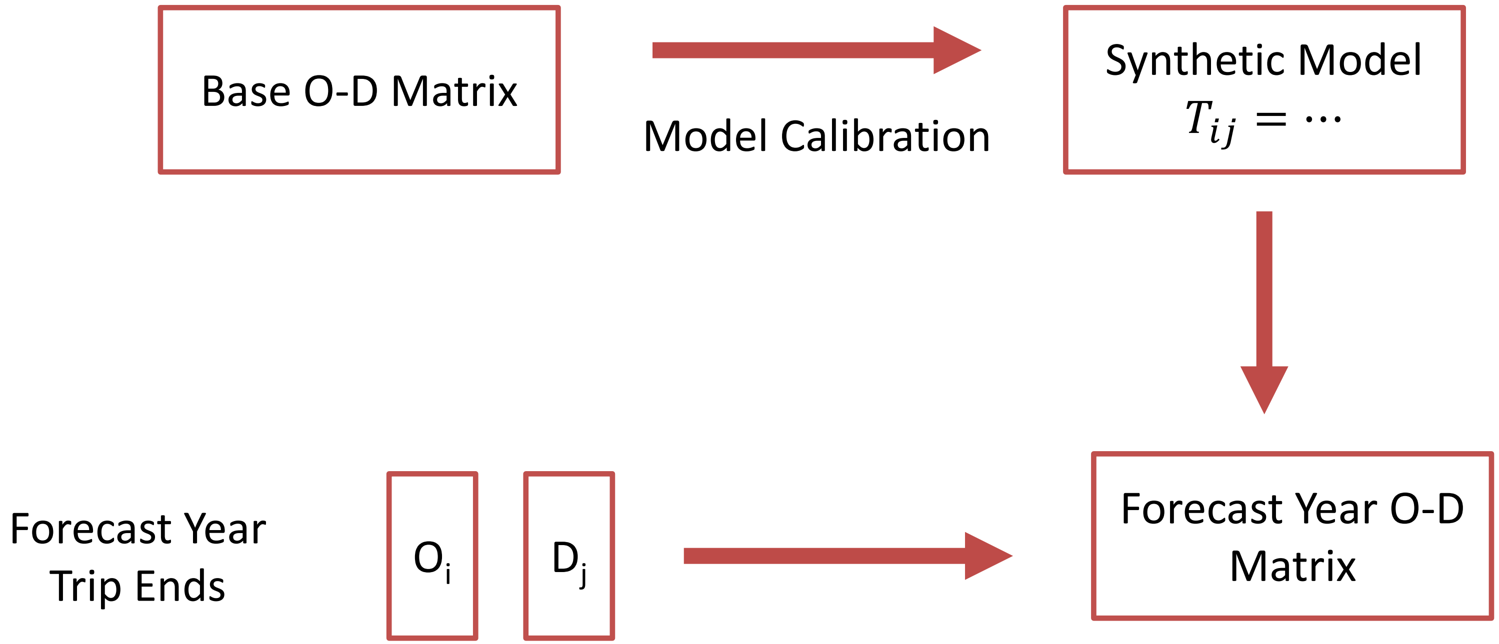
Denom A2 14.20

O/D:	1	2	Total	Ai
1	94.7	105.3	200.0	0.237564
2	7.3	292.7	300.0	0.070414
Total	102.0	398.0	500	

Rj= $\boxed{0.981 \quad 1.005}$

Assume ε < = 2% : So, converged

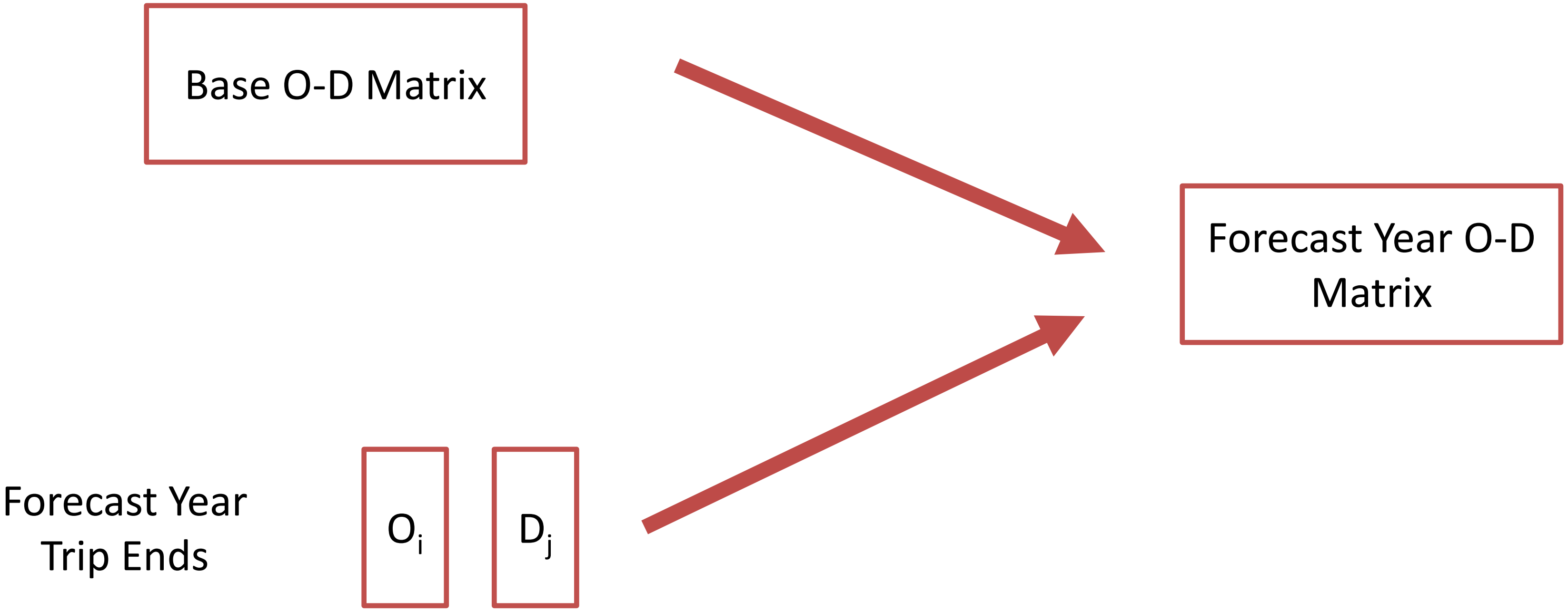
Synthetic Modeling Approach



Issues with Synthetic Models

- Synthetic models often contain **considerable errors**
- **Not surprising**: using a fairly simplistic model to represent very complex travel patterns
- Instead of converting observed trip matrix into a “synthetic” model, which can be used to predict future trips: **update base trip matrix directly**

Base Data Updating Approach



Biproportional Updating: Gravity Factor Method

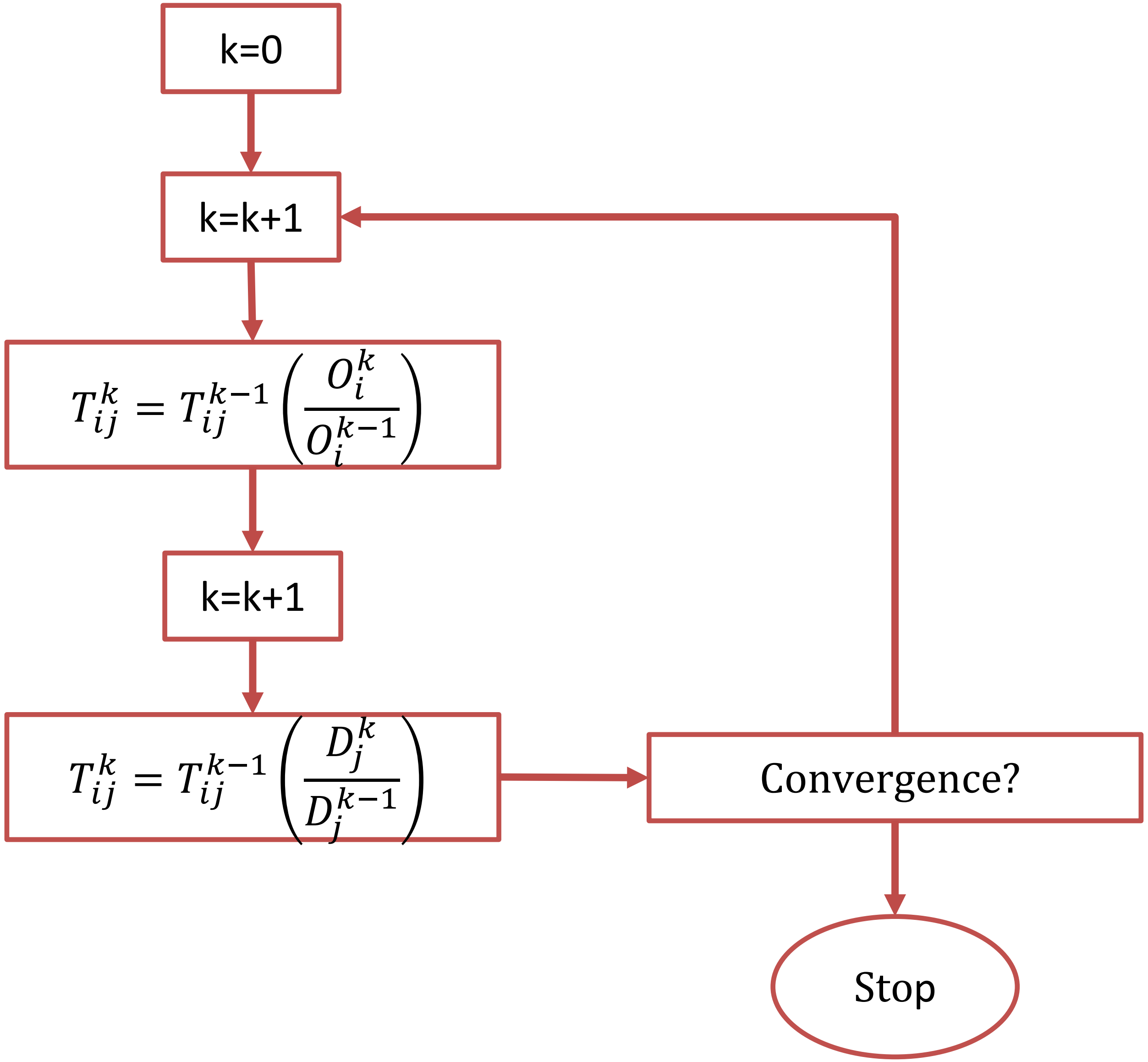
- Proportionally adjust observed base year trip matrix until it matches forecast year row & column totals

- Iterative procedure required to balance rows & columns

$$O_i^k = \sum_j T_{ij}^k \; for \; k^{th} \; iteration$$
$$T_{ij}^0 = Base \; year \; O - D \; flows/matrix$$
$$O_i^{new} = Forecast \; year \; row \; sum/generation$$

- Similar definitions for D_j^k and D_j^{new}

Balancing Algorithm



Biproportional Updating: Gravity Method

- Bi-proportional update for **Fratar procedure** is simply a doubly-constrained gravity model
- Base year trip matrix defines **impedance function**

$$T_{ij} = A_i O_i B_j D_j T_{ij}^0$$

$$T_{ij} = O_i(\alpha D_j) f_{ij} = O_i D_j^* T_{ij}^0$$

OR

$$T_{ij} = (\alpha O_i) D_j f_{ij} = O_i^* D_j T_{ij}^0$$

Doubly-Constrained Gravity Method Balancing Procedure

k=0

$D_j^{*^1} = D_j$

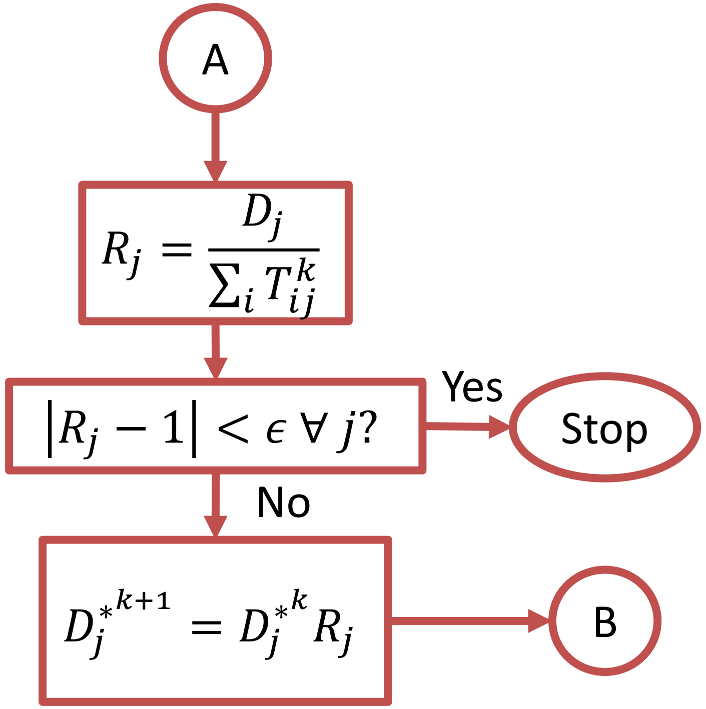
k=k+1

B

$$T_{ij}^k = \frac{O_i D_j^{*^k} T_{ij}^0}{\sum_j D_j^{*^k} T_{ij}^0}$$

A

Doubly-Constrained Gravity Method Balancing Procedure Cont…



Example: Biproportional Updating

Base Year Matrix

O/D	1	2	Total
1	60	90	150
2	30	220	250
Total	90	310	400

Future Year Trip Origin & Destination Totals

O/D	1	2	Total
1	?	?	200
2	?	?	300
Total	100	400	500

Solution: Biproportional Gravity Method

1

Iteration 1: (D*j = Dj)			
D*j=	100	400	
Denominator, Org=1			42000.00
Denominator, Org=2			91000.00
O/D:	1	2	Total
1	28.6	171.4	200.0
2	9.9	290.1	300.0
Total	38.5	461.5	500
Rj=	2.600	0.867	

2

Iteration 2: (D*j(new)= D*j(old)*Rj)			
D*j=	260	346.6667	
Denominator, Org=1			46800.00
Denominator, Org=2			84066.67
O/D:	1	2	Total
1	66.7	133.3	200.0
2	27.8	272.2	300.0
Total	94.5	405.5	500
Rj=	1.058	0.986	

3

Iteration 3: (D*j(new)= D*j(old)*Rj)			
D*j=	275.1273	341.9661	
Denominator, Org=1			47284.59
Denominator, Org=2			83486.36
O/D:	1	2	Total
1	69.8	130.2	200.0
2	29.7	270.3	300.0
Total	99.5	400.5	500
Rj=	1.005	0.999	

4

Iteration 4: (D*j(new)= D*j(old)*Rj)			
D*j=	276.5605	341.5236	
Denominator, Org=1			47330.76
Denominator, Org=2			83432.01
O/D:	1	2	Total
1	70.1	129.9	200.0
2	29.8	270.2	300.0
Total	100.0	400.0	500
Rj=	1.000	1.000	

Converged!

Which method is more appropriate?

- For **short time frames** - matrix updating usually more appropriate:
  - However, the bi-proportional updating is not a function of travel time
  - Trip distribution will be **insensitive to changes in travel time** due to congestion or infrastructure improvements
- For **long time frames** - better to use a **synthetic model** (e.g., gravity model, entropy maximizing model)
- Ex.: If HH travel survey runs **every 5 years**, what if a major employer changed its office location? Would the employees be able to move right away?
  - No! So this would result in **longer commute distances** in short-term (biproportional updating method will not give good prediction even in short time frame)
  - But eventually people would move closer, or change jobs, so the anomaly would fade away (long-term forecasting by synthetic model will be better).

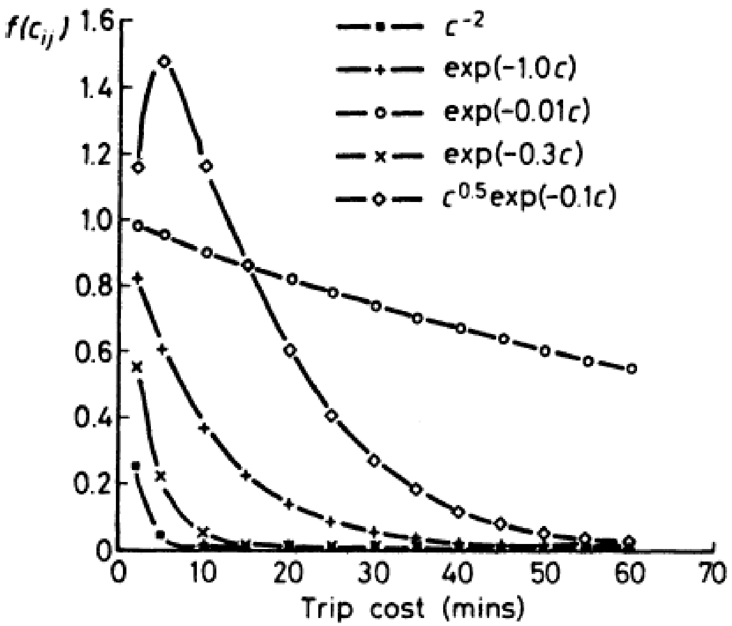
Impedance Functions

- β captures the **sensitivity of trips to travel time**
  - Note: β is capturing **current sensitivity** and keeping it constant for future forecasts is a drawback
  - β = 0 is the proportional flow model
  - β = negative infinity becomes a cost minimization model -> typical transportation problem
  - So $-\infty \leq \beta \leq 0$
- A good way to understand this type of relationship is to create a **quick plot** (see next slide)

Impedance Functions

Example impedance functions:

- $f\left(c_{ij}\right) = \exp\left(\beta c_{ij}\right), \beta < 0$ (exponential function)
- $f\left(c_{ij}\right) = c_{ij}^{n}, n < 0$ (power function)
- $f\left(c_{ij}\right) = \mathrm{c}_{\mathrm{ij}}^{\mathrm{n}}\exp\left(\beta c_{ij}\right), n, \beta < 0$ (combined function)
- n, β are parameters estimated from observed data separately from the trip distribution model



Understanding Impedance Functions

t_{ij}	1	2
1	5	15
2	12	6

Assume we have 2 trips starting from each i origin and apply a singly (origin)-constrained exponential gravity model

$\beta = 0$

T_{ij}	1	2
1	1	1
2	1	1

$\beta = -0.10$

T_{ij}	1	2
1	1.5	0.5
2	0.7	1.3

$\beta = -0.25$

T_{ij}	1	2
1	1.8	0.2
2	0.4	1.6

$\beta = -1.0$

T_{ij}	1	2
1	2	0
2	0	2

If we know the true trip distribution is as below, optimization (e.g., Excel solver) can be used to find β.

T_{ij}	1	2
1	1.16	0.84
2	0.91	1.09

Answer: -0.03142 (or $\pi/100$)

Impedance Function Parameter Estimation

- "Most likely" values of n, β are those that satisfy the equation:

$$\frac{\sum_i \sum_j t_{ij} T_{ij}}{T} = \frac{\sum_i \sum_j t_{ij} T_{ij}}{\sum_i \sum_j T_{ij}} = t_{avg} \; (1)$$

where t_{avg} is the observed travel time and the left-hand side of equation (1) is the average travel time predicted by the model

- This is the underlying assumption made when solving for β in the previous slide!

Gravity/Entropy Model Goodness-of-Fit Statistics

- How can we statistically measure how well our **model fits the data**?

$$T_{ij} = observed\ trips; T_{ij}^{*} = predicted\ trips$$

$$T = total\ trips; n = No.\ of\ zones, n^2 = n \times n\ (No.\ of\ trip\ distribution\ cells)$$

$$T_0 = T/n^2$$

$$R^2 = 1 - \left(\frac{\sum_i \sum_j \left(T_{ij} - T_{ij}^{*}\right)^2}{\sum_i \sum_j \left(T_{ij} - T_0\right)^2} \right)$$

$$\chi^2 = \sum_i \sum_j \frac{\left(T_{ij} - T_{ij}^{*}\right)^2}{T_{ij}^{*}}$$

$$Mean\ Absolute\ Error\ (MAE) = \sum_i \sum_j \left|T_{ij} - T_{ij}^{*}\right| / n^2$$

$$Normalized\ \phi = \sum_i \sum_j \left(\frac{T_{ij}}{T}\right) \left| \ln\left(\frac{T_{ij}}{T_{ij}^{*}}\right) \right|$$

Chi-squared Statistic

- Measures the hypothesis that observed and predicted matrices are the same (H_0)
- Estimated value less than theoretical value means we **fail to reject** null hypothesis (i.e., observed and predicted trip matrices represent same population)
- **Problem**: erroneous results if cell values < 5 trips because population too small

Example – R^2

Observed O-D Matrix

T_{ij} =

Origin Zone	Destination Zone 1	2	3	4	5	O_i^o
1	50	12	5	13	12	92
2	45	24	6	24	3	102
3	18	36	7	36	12	109
4	30	5	9	41	14	99
5	25	3	22	21	15	86
D_j^o	168	80	49	135	56	**488**

Predicted O-D Matrix

T_{ij}^* =

Origin Zone	Destination Zone 1	2	3	4	5	O_i^p
1	52	9	7	9	12	89
2	36	36	6	27	6	111
3	22	25	7	39	18	111
4	21	9	12	40	10	92
5	13	13	29	19	11	85
D_j^p	144	92	61	134	57	**488**

Average Trips Per Interchange, $T_o = T / n^2$ (where n is the number of matrix cells)

So To = 19.52

$(T_{ij} - T_{ij}^*)^2$ =

Origin Zone	Destination Zone 1	2	3	4	5
1	4	9	4	16	0
2	81	144	0	9	9
3	16	121	0	9	36
4	81	16	9	1	16
5	144	100	49	4	16
					894

$(T_{ij} - T_o)^2$ =

Origin Zone	Destination Zone 1	2	3	4	5
1	929	57	211	43	57
2	649	20	183	20	273
3	2	272	157	272	57
4	110	211	111	461	30
5	30	273	6	2	20
					4454

So R^2= 0.80

$$R^2 = 1 - \frac{\sum_i \sum_j \left(T_{ij} - T_{ij}^*\right)^2}{\sum_i \sum_j \left(T_{ij} - T_o\right)^2}$$

- Denominator: Sq. of observed differences from the simplest case (i.e., same trips in each cell)
- R^2 is a measure of the proportion of observed variation explained by the gravity model

Example – χ^2

Observed O-D Matrix

Tij =	Origin Zone	Destination Zone					
		1	2	3	4	5	Oi^o
	1	50	12	5	13	12	92
	2	45	24	6	24	3	102
	3	18	36	7	36	12	109
	4	30	5	9	41	14	99
	5	25	3	22	21	15	86
	Dj^o	168	80	49	135	56	**488**

Predicted O-D Matrix

Tij^* =	Origin Zone	Destination Zone					
		1	2	3	4	5	Oi^P
	1	52	9	7	9	12	89
	2	36	36	6	27	6	111
	3	22	25	7	39	18	111
	4	21	9	12	40	10	92
	5	13	13	29	19	11	85
	Dj^P	144	92	61	134	57	**488**

$(Tij - Tij^*)^2 / Tij^*$ =	Origin Zone	Destination Zone					
		1	2	3	4	5	
	1	0.08	1.00	0.57	1.78	0.00	
	2	2.25	4.00	0.00	0.33	1.50	
	3	0.73	4.84	0.00	0.23	2.00	
	4	3.86	1.78	0.75	0.03	1.60	
	5	11.08	7.69	1.69	0.21	1.45	
							49.44

Chi-Squared=	49.44138	
Degrees of Freedom	24	
Chi-Squared Value for 24 DF & 95% confidence =	36.41503	
	49.4>36.4	**REJECT**

- Validates whether two matrices represent the same population
- Since estimated value > theoretical (test) statistic, we reject the null hypothesis – i.e., two matrices are not the same

Example - MAE

Observed O-D Matrix

Tij =	Origin Zone	Destination Zone 1	2	3	4	5	Oi^o
	1	50	12	5	13	12	92
	2	45	24	6	24	3	102
	3	18	36	7	36	12	109
	4	30	5	9	41	14	99
	5	25	3	22	21	15	86
	Dj^o	168	80	49	135	56	**488**

Predicted O-D Matrix

Tij^* =	Origin Zone	Destination Zone 1	2	3	4	5	Oi^P
	1	52	9	7	9	12	89
	2	36	36	6	27	6	111
	3	22	25	7	39	18	111
	4	21	9	12	40	10	92
	5	13	13	29	19	11	85
	Dj^P	144	92	61	134	57	**488**

$$MAE = \sum_i \sum_j \left|T_{ij} - T_{ij}^*\right| / n^2$$

Origin Zone	Destination Zone 1	2	3	4	5	
1	2.00	3.00	2.00	4.00	0.00	
2	9.00	12.00	0.00	3.00	3.00	
3	4.00	11.00	0.00	3.00	6.00	
4	9.00	4.00	3.00	1.00	4.00	
5	12.00	10.00	7.00	2.00	4.00	
						118.00

MAE **4.72**

- Simplest measure of the average difference between observed and predicted matrix cells
- Lower is better

Example - $Normalized\ \phi$

Observed O-D Matrix

T_{ij} =

Origin Zone	Destination Zone 1	2	3	4	5	O_i^o
1	50	12	5	13	12	92
2	45	24	6	24	3	102
3	18	36	7	36	12	109
4	30	5	9	41	14	99
5	25	3	22	21	15	86
D_j^o	168	80	49	135	56	**488**

Predicted O-D Matrix

T_{ij}^* =

Origin Zone	Destination Zone 1	2	3	4	5	O_i^p
1	52	9	7	9	12	89
2	36	36	6	27	6	111
3	22	25	7	39	18	111
4	21	9	12	40	10	92
5	13	13	29	19	11	85
D_j^p	144	92	61	134	57	**488**

T_{ij}/T=

Origin Zone	Destination Zone 1	2	3	4	5	
1	0.10	0.02	0.01	0.03	0.02	
2	0.09	0.05	0.01	0.05	0.01	
3	0.04	0.07	0.01	0.07	0.02	
4	0.06	0.01	0.02	0.08	0.03	
5	0.05	0.01	0.05	0.04	0.03	
						1

$(T_{ij}/T)\text{abs}(\ln(T_{ij}/T_{ij}^*))$ =

Origin Zone	Destination Zone 1	2	3	4	5	
1	0.004019	0.007074	0.003447	0.009796	0	
2	0.020577	0.019941	0	0.005793	0.004261	
3	0.007402	0.0269	0	0.005905	0.00997	
4	0.021927	0.006022	0.005306	0.002075	0.009653	
5	0.0335	0.009014	0.012454	0.004307	0.009533	
						0.24

$\text{abs}(\ln(T_{ij}/T_{ij}^*))$

Origin Zone	Destination Zone 1	2	3	4	5	
1	0.04	0.29	0.34	0.37	0.00	
2	0.22	0.41	0.00	0.12	0.69	
3	0.20	0.36	0.00	0.08	0.41	
4	0.36	0.59	0.29	0.02	0.34	
5	0.65	1.47	0.28	0.10	0.31	
						7.92

Normalized Phi 0.24

- Based on information gain
- Larger value means larger error

Other Model Validation Methods

- Compare observed and precited **trip length frequency distribution** (TLFDs)
- Examine O-D residuals (perhaps on a super-zone basis)
- Examine predicted vs. observed **screenline** counts

Example: Trip Length Distribution

Observed 2000 O-D Matrix :

Origin Zone	Destination Zone 1	2	3	4	5
1	50.00	12.00	5.00	13.00	12.00
2	45.00	24.00	6.00	24.00	3.00
3	18.00	36.00	7.00	36.00	12.00
4	30.00	5.00	9.00	41.00	14.00
5	25.00	3.00	22.00	21.00	15.00
D_j^o	168.00	80.00	49.00	135.00	56.00

Predicted 2000 O-D Matrix :

Origin Zone	Destination Zone 1	2	3	4	5
1	52.00	9.00	7.00	9.00	12.00
2	36.00	36.00	6.00	27.00	6.00
3	22.00	25.00	7.00	39.00	18.00
4	21.00	9.00	12.00	40.00	10.00
5	13.00	13.00	29.00	19.00	11.00

Auto Travel Time (Minutes) :

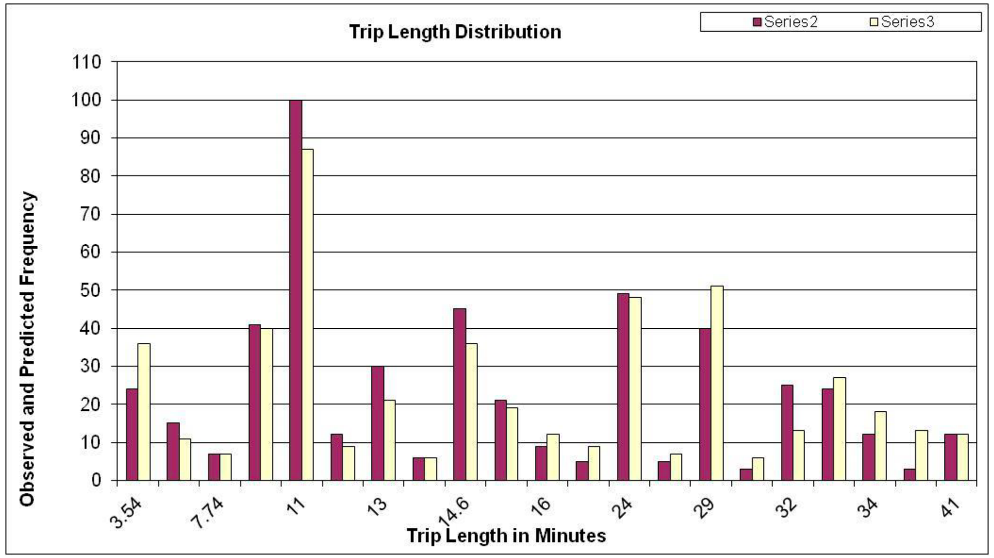
Origin Zone	Destination Zone 1	2	3	4	5
1	11.00	12.50	28.50	24.00	41.00
2	14.60	3.54	13.65	33.00	31.00
3	29.00	11.00	7.74	24.00	34.00
4	13.00	19.00	16.00	9.00	11.00
5	32.00	35.00	29.00	15.00	7.00

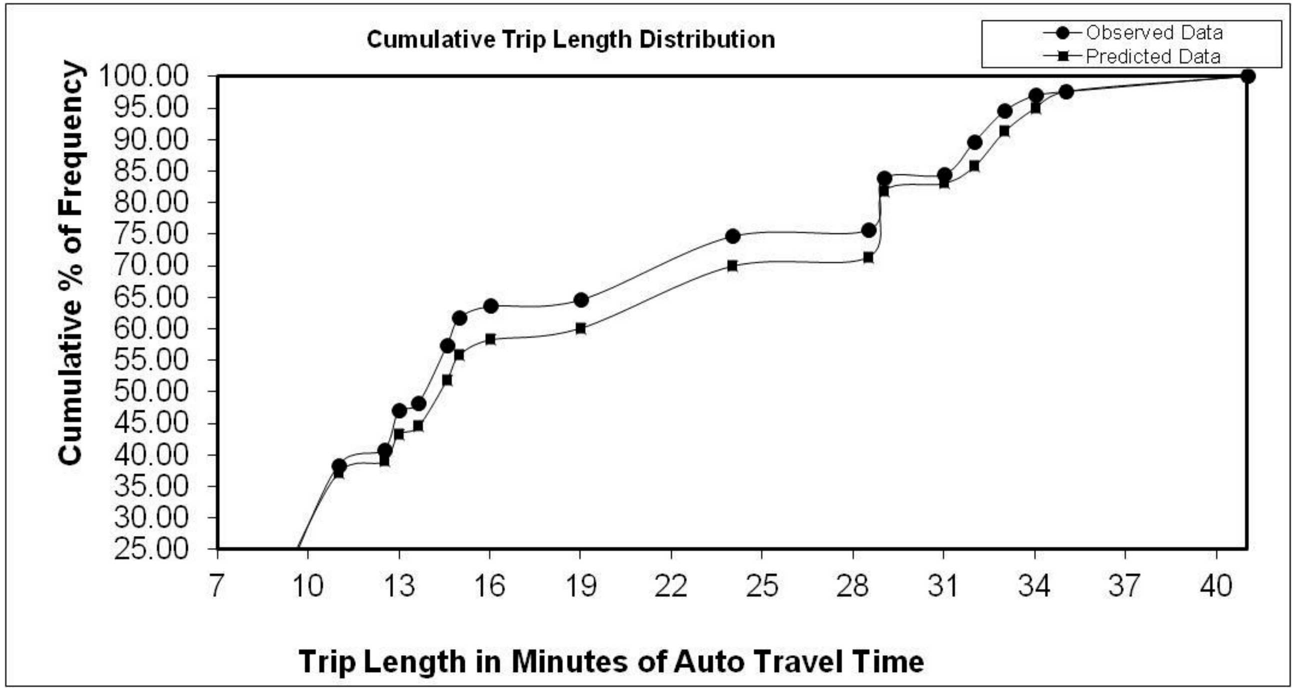
Trip Length In Minutes	Observed Frequency	Predicted Frequency
11	50	52
14.6	45	36
29	18	22
13	30	21
32	25	13
12.5	12	9
3.54	24	36
11	36	25
19	5	9
35	3	13
28.50	5.00	7.00
13.65	6.00	6.00
7.74	7.00	7.00
16.00	9.00	12.00
29.00	22.00	29.00
24	13	9
33	24	27
24	36	39
9	41	40
15	21	19
41	12	12
31	3	6
34	12	18
11	14	10
7	15	11

Ascending Order

Trip Length In Minutes	Observed Frequency	Predicted Frequency	%frequency-Obs	Cumulative Frequency-Obs	%frequency-Pred	Cumulative Frequency-Pred
3.54	24	36	4.92	4.92	7.38	7.38
7	15	11	3.07	7.99	2.25	9.63
7.74	7	7	1.43	9.43	1.43	11.07
9	41	40	8.40	17.83	8.20	19.26
11	100	87	20.49	38.32	17.83	37.09
12.5	12	9	2.46	40.78	1.84	38.93
13	30	21	6.15	46.93	4.30	43.24
13.65	6	6	1.23	48.16	1.23	44.47
14.6	45	36	9.22	57.38	7.38	51.84
15	21	19	4.30	61.68	3.89	55.74
16	9	12	1.84	63.52	2.46	58.20
19	5	9	1.02	64.55	1.84	60.04
24	49	48	10.04	74.59	9.84	69.88
28.5	5	7	1.02	75.61	1.43	71.31
29	40	51	8.20	83.81	10.45	81.76
31	3	6	0.61	84.43	1.23	82.99
32	25	13	5.12	89.55	2.66	85.66
33	24	27	4.92	94.47	5.53	91.19
34	12	18	2.46	96.93	3.69	94.88
35	3	13	0.61	97.54	2.66	97.54
41	12	12	2.46	100.00	2.46	100.00

Example: Trip Length Distribution





Trip Distribution Model Estimation Vs. Calibration

Estimation: Process of **finding model parameter values**, which cause the model to "best fit" observed data according to statistical procedure (e.g., regression)

Calibration: Process that occurs **post-estimation** (or in lieu of estimation) involving ad hoc adjustments to model parameters to "force" the model to "better fit" observed data

Gravity/Entropy Model Calibration: K-Factors

- Most common calibration procedure used for gravity models is k-factors:

$$T_{ij} = A_i O_i B_j D_j K_{ij} f_{ij}$$

$$A_i = \frac{1}{\sum_j B_j D_j K_{ij} F_{ij}}$$

$$B_j = \frac{1}{\sum_i A_i O_i K_{ij} F_{ij}}$$

- Choose K_{ij} to reduce difference between observed and predicted:
  - Screenline flows
  - TLFDs and/or
  - Key O-D pairs

Gravity/Entropy Model Calibration: K-Factors

- Rationale for K-factors is to capture systematic differences in spatial flows not explained by travel time (or other terms in gravity model)

  - E.g., residential zones with high concentrations of white collar workers are more likely to generate work trips to employment zones with high concentrations of white collar jobs than zones with predominantly blue collar jobs

  - Use K-factors to link “white collar” origin and destination zones

Drawbacks of Aggregate Trip Distribution Models

- Neither systematic- nor updating-type models capture idiosyncrasy of **destination location choice** by individuals because aggregate behavior into trip distribution matrices

- Lack of **policy sensitivity** – E.g., how would the trip distribution in Lincoln respond to a new employment center north of the city?