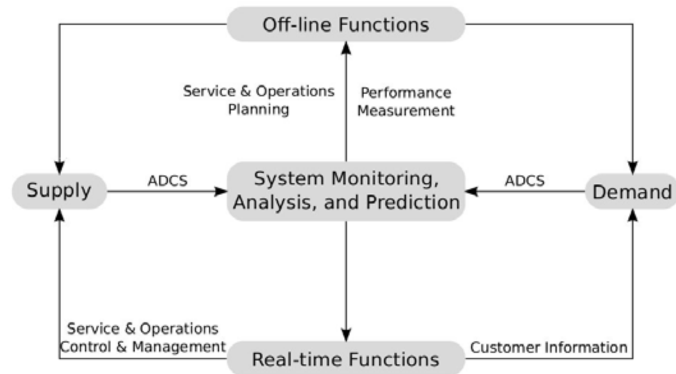
Origin, Destination, and Transfer Inference (ODX)

- Using automatically collected data: AFC, AVL, APC
- Infers destinations in open systems
- Infers transfers
- Only captures existing demand
- Does not make inferences for all fare transactions
  - only one tap
  - cash
  - fare evasion
  - trips on other modes
- Validated with surveys
- Needs to be scaled up to full demand

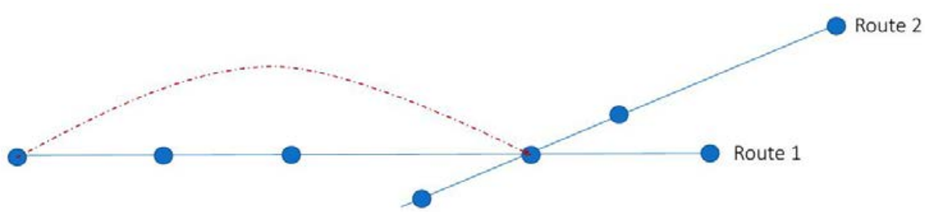
Key Automated Data Collection Systems

- Automatic Vehicle Location (AVL)
- Automatic Fare Collection (AFC)
- Automatic Passenger Counting (APC)



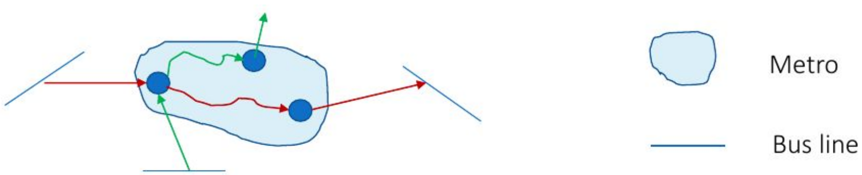
OD Matrix Estimation

Route Level



Network Level

Full Intermodal Journey Inference



Route Level OD Estimation with APC

APC provides "control totals"

TIME	BUS ID	ROUTE	TRIP ID	DIRECTION	Boardings	Alightings	STOP
9/12/2005 12:08:10 AM	6734	20	15065450	East	0	3	WASHINGTON + STATE
9/12/2005 12:00:04 AM	6734	20	15065450	East	1	1	MADISON + PEORIA
9/12/2005 12:03:29 AM	6734	20	15065450	East	0	0	WASHINGTON + CANAL
9/12/2005 5:37:19 AM	6729	20	15067244	East	0	1	WASHINGTON + LASALLE

Route #1 APC + seed matrix		Destination				
		Stop 1	Stop 2	Stop 3	Stop 4	Target on
Origin	Stop 1		25	10	2	40
	Stop 2			5	15	30
	Stop 3				9	20
	Stop 4					
	Target off	0	30	20	40	90

Iterative Proportional Fitting (IPF)

- Also known as biproportional fitting and matrix scaling
- Scales cell values of a sampled origin-destination matrix so that row and column sums equal marginal target values (counted boardings and alightings)
- If all values are strictly positive, IPF converges to a unique MLE solution
- Zeroes affect the solution

Initialization						
	A	B	C	D	Total Boardings	Target Boardings
A		1	1	1	3	40
B			1	1	2	30
C				1	1	20
D						
Total Alightings		1	2	3		
Target Alightings		30	20	40		90

Iterative Proportional Fitting (IPF)

Step 1							
	A	B	C	D	Total Boardings	Target Boardings	Factor
A		13.3	13.3	13.3	40	40	13.3
B			15	15	30	30	15.0
C				20	20	20	20.0
D							
Total Alightings		13.3	28.3	48.3			
Target Alightings		30	20	40			

Step 2						
	A	B	C	D	Total Boardings	Target Boardings
A		30	9.41	11	50.4	40
B			10.6	12.4	23.0	30
C				16.6	16.6	20
D						
Total Alightings		30	20	40		
Target Alightings		30	20	40		
Factor		2.3	0.7	0.8		

Iterative Proportional Fitting (IPF)

Step 3							
	A	B	C	D	Total Boardings	Target Boardings	Factor
A		23.8	7.46	8.75	40	40	0.8
B			13.8	16.2	30	30	1.3
C				20	20	20	1.2
D							
Total Alightings		23.8	21.3	44.9			
Target Alightings		30	20	40			

Step 4						
	A	B	C	D	Total Boardings	Target Boardings
A		30	7.02	7.79	44.8	40
B			13	14.4	27.4	30
C				17.8	17.8	20
D						
Total Alightings		30	20	40		
Target Alightings		30	20	40		
Factor		1.3	0.9	0.9		

Iterative Proportional Fitting (IPF)

Step 5							
	A	B	C	D	Total Boardings	Target Boardings	Factor
A		26.8	6.26	6.95	40	40	0.9
B			14.2	15.8	30	30	1.1
C				20	20	20	1.1
D							
Total Alightings		26.8	20.5	42.7			
Target Alightings		30	20	40			

Step 6						
	A	B	C	D	Total Boardings	Target Boardings
A		30	6.12	6.51	42.6	40
B			13.9	14.8	28.7	30
C				18.7	18.7	20
D						
Total Alightings		30	20	40		
Target Alightings		30	20	40		
Factor		1.1	1.0	0.9		

Iterative Proportional Fitting (IPF)

Step 7	A	B	C	D	Total Boardings	Target Boardings	Factor
A		28.2	5.74	6.11	40	40	0.9
B			14.5	15.5	30	30	1.0
C				20	20	20	1.1
D							
Total Alightings		28.2	20.3	41.6			
Target Alightings		30	20	40			

Step 8	A	B	C	D	Total Boardings	Target Boardings
A		30	5.66	5.88	41.5	40
B			14.3	14.9	29.2	30
C				19.2	19.2	20
D						
Total Alightings		30	20	40		
Target Alightings		30	20	40		
Factor		1.1	1.0	1.0		

Iterative Proportional Fitting (IPF)

Step 9	A	B	C	D	Total Boardings	Target Boardings	Factor
A		28.9	5.45	5.66	40	40	1.0
B			14.7	15.3	30	30	1.0
C				20	20	20	1.0
D							
Total Alightings		28.9	20.2	40.9			
Target Alightings		30	20	40			

Step 10	A	B	C	D	Total Boardings	Target Boardings
A		30	5.41	5.53	40.9	40
B			14.6	14.9	29.5	30
C				19.5	19.5	20
D						
Total Alightings		30	20	40		
Target Alightings		30	20	40		
Factor		1.0	1.0	1.0		

Iterative Proportional Fitting (IPF)

Step 11	A	B	C	D	Total Boardings	Target Boardings	Factor
A		29.3	5.28	5.4	40	40	1.0
B			14.8	15.2	30	30	1.0
C				20	20	20	1.0
D							
Total Alightings		29.3	20.1	40.6			
Target Alightings		30	20	40			

Step 12	A	B	C	D	Total Boardings	Target Boardings
A		30	5.25	5.33	40.6	40
B			14.7	15	29.7	30
C				19.7	19.7	20
D						
Total Alightings		30	20	40		
Target Alightings		30	20	40		
Factor		1.0	1.0	1.0		

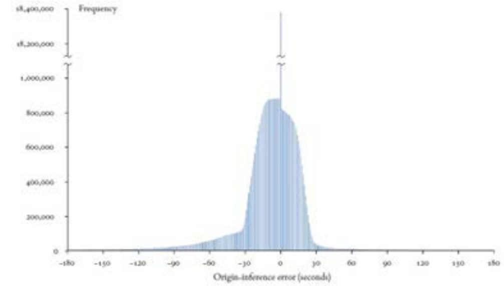
Route Level ODX with AFC and AVL

	TfL	MBTA	Seoul
AFC Rail	Closed	Open	Closed, Detailed, Including transfers
AFC Bus	Open	Open	
AVL	iBus	Announcements Heartbeat Time points	
Control totals	ETM (Buses)	APC (sample)	
	Gatelines (Rail stations)	some	

Origin Inference

Matching the AFC transactions with the AVL data to infer boarding stops

Origin Inference Results: London

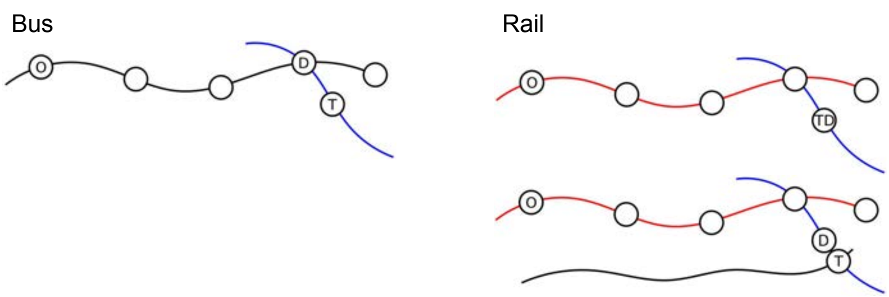


- 10 weekdays, 6.1 to 6.5 million Oyster bus boardings per day
- 96% of boarding locations inferred within ± 5 min
  - 96% within ± 2 min
  - 93% within ± 1 min
  - 28% between arrival and departure times
- 2.6% beyond ± 5 min.
- 1.4% not matched to iBus route or trip

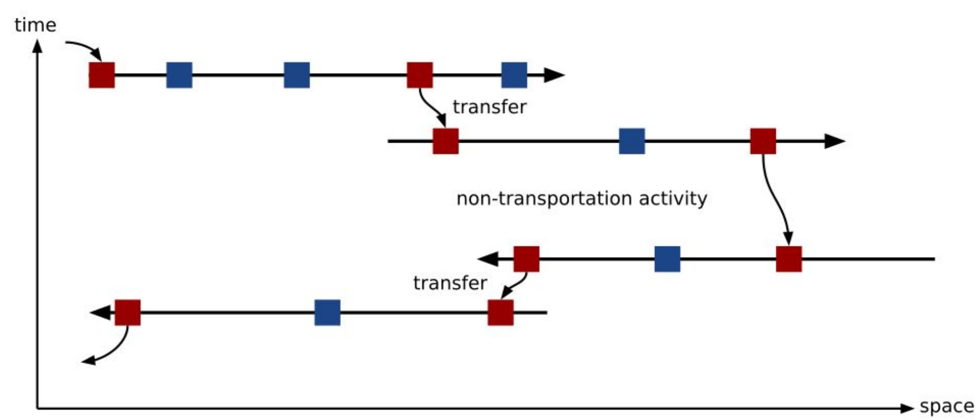
Destination Inference: Closest Stop

Key Assumptions

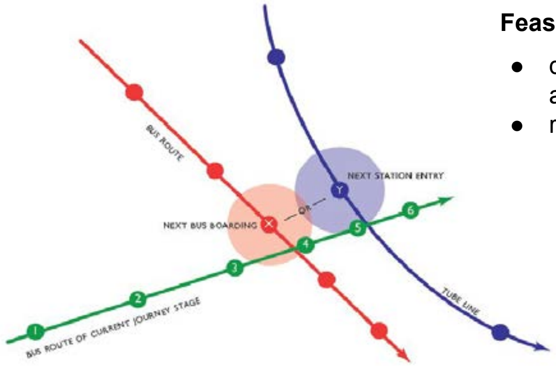
- The destination of many trip segments is close to the origin of the following trip segment.
- No intermediate private transportation mode trip segment
- Passengers will not walk a long distance
- Last trip of a day ends at the origin of the first trip of the day (symmetry assumption)



Destination Inference



Destination Inference

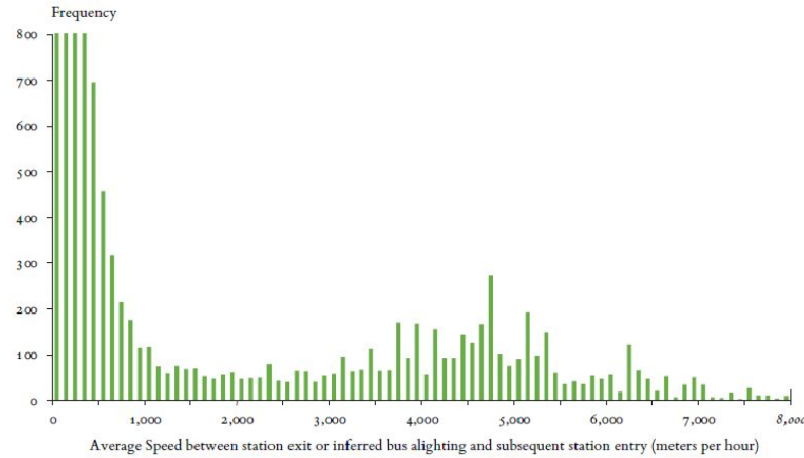


Feasibility Tests

- distance and time between alighting and next boarding stop
- relative location

Destination Inference Results: London

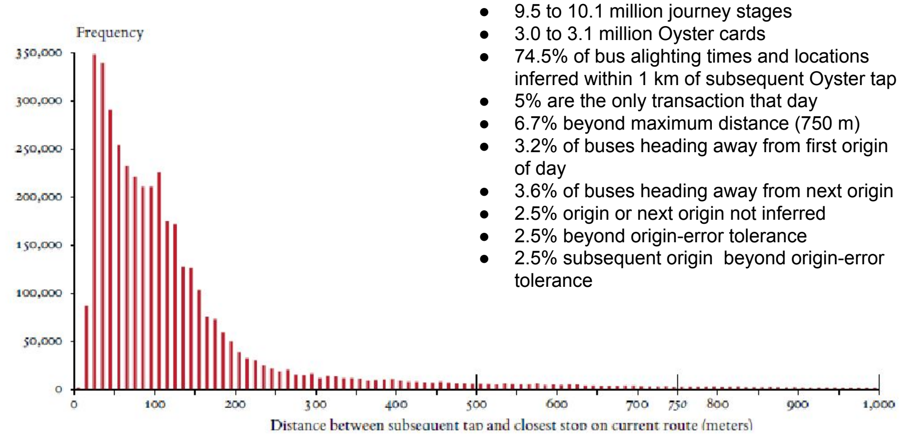
Destination inference: 74.6%



Destination Inference Results: London

Ten-weekday average: 6-10 and 13-17 June 2011

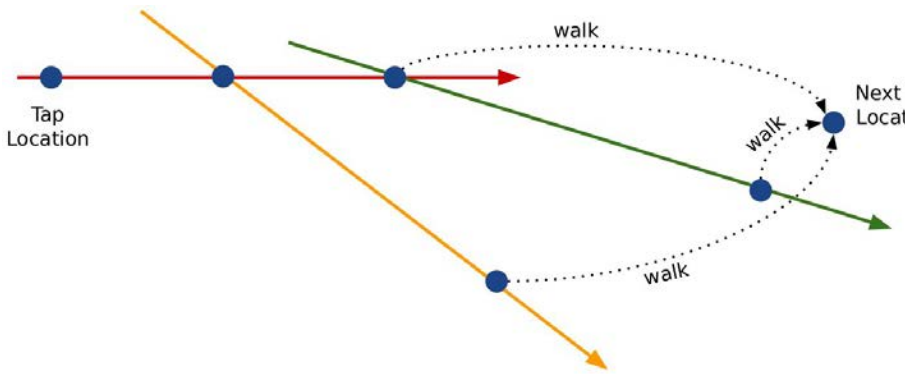
- 15.6 to 16.1 million Oyster transactions
- 9.5 to 10.1 million journey stages
- 3.0 to 3.1 million Oyster cards
- 74.5% of bus alighting times and locations inferred within 1 km of subsequent Oyster tap
- 5% are the only transaction that day
- 6.7% beyond maximum distance (750 m)
- 3.2% of buses heading away from first origin of day
- 3.6% of buses heading away from next origin
- 2.5% origin or next origin not inferred
- 2.5% beyond origin-error tolerance
- 2.5% subsequent origin beyond origin-error tolerance



Comparison to Other Sources

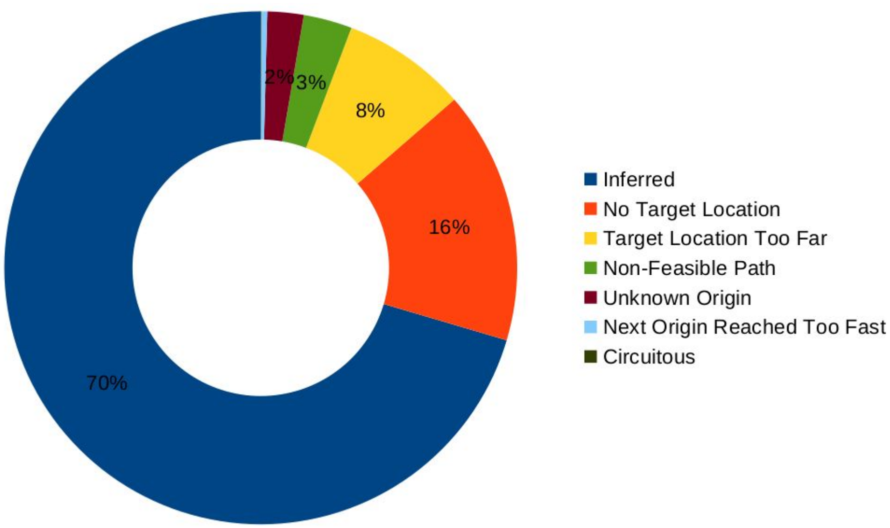
- Small stop-by-stop differences between estimated OD and results from the Bus OD Survey (BODS)
- BODS underestimated the ridership in peak periods and midday, especially when BODS survey return rates are low (50%-80%).
- Value for transportation planning

Destination Inference: Minimum Cost Path

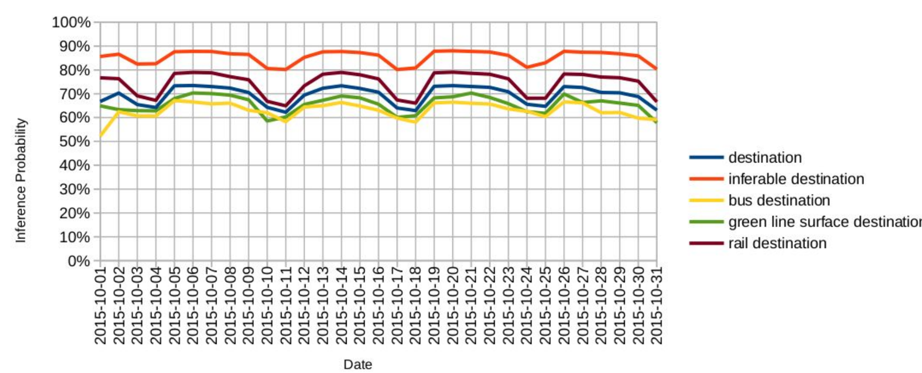


22

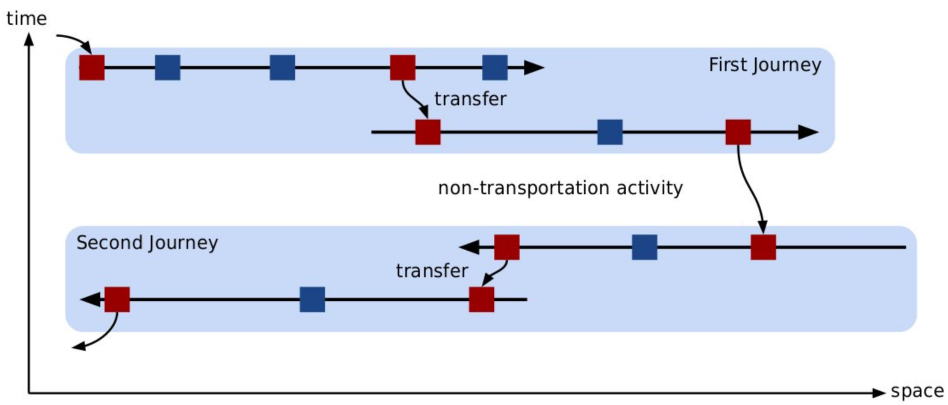
Destination Inference: MBTA



Inference Probability



Interchange (Transfer) Inference



Interchange (Transfer) Inference

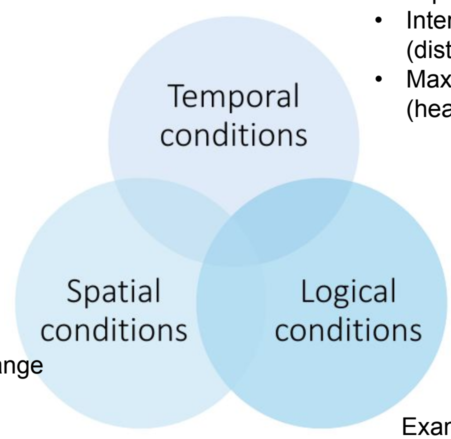
Journey stage: any portion of a rider's journey that is represented by a single Oyster bus record or by a rail entry/exit pair.

Interchange (Transfer): a transition between two consecutive journey stages that does not contain a trip-generating activity. Its primary purpose, rather, is to connect a previous stage's origin to a subsequent stage's destination.

Full journey: a sequential set of journey stages connected exclusively through interchanges.

Trip-Linking Assumptions

Examples:
- Interchange time (distance)
- Maximum bus wait time (headways)



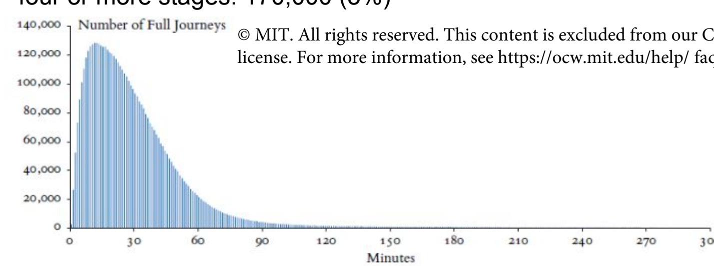
Examples:
- Maximum interchange distance
- Circuity between stages
- Ending journey near origin

Example:
- Must not continue at same station, same route, etc.

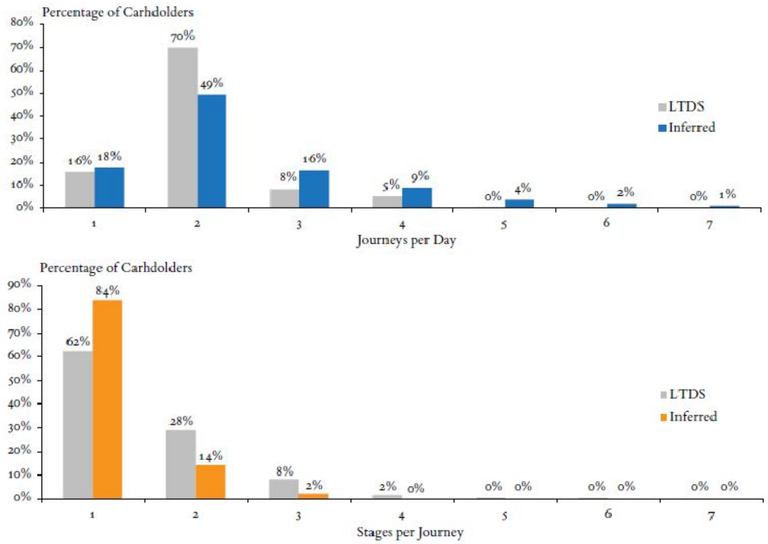
Interchange Inference Results

Ten-day average: 6-10 and 13-17 June 2011

- Link status inferred for 91% of journeys stages
  - link status could not be inferred for remaining 9% of stages: assumed not linked
- Stages per journey:
  - one stage: 4 million (66%)
  - two stages: 1.5 million (25%)
  - three stages: 400,000 (7%)
  - four or more stages: 170,000 (3%)



Comparison to Travel Surveys (LTDS)



Trip-Level Scaling

- AFC, AVL, and ODX give an OD matrix, but only for a sample of passenger trips
- APC gives full count of boardings and alightings
  - for all vehicles, a fraction of vehicles, or none
- Iterative Proportional Fitting (IPF) can be used to assign remaining destinations in probability
  - control totals are APC boardings and alightings minus ODX boardings and alightings

Trip-Level Scaling with Transfer Information and without APC

- The complete OD matrix $\mathbf{R}$ can be divided into an inferred part $\mathbf{I}$ and a missing part $\mathbf{M}$.

$$\mathbf{R} = \mathbf{I} + \mathbf{M}$$

- The missing part can be divided into trips with uninferred destinations $\mathbf{U}$ and trips not observed $\mathbf{N}$.

$$\mathbf{M} = \mathbf{U} + \mathbf{N}$$

- Therefore

$$\mathbf{R} = \mathbf{I} + \mathbf{U} + \mathbf{N}$$

and we want to estimate $\mathbf{R}$ as

$$\tilde{\mathbf{R}} = \mathbf{I} + \tilde{\mathbf{U}} + \tilde{\mathbf{N}}$$

Trip-Level Scaling with Transfer Information and without APC

- Trips followed by transfers may have different OD structure.
- Assume that all observed trips followed by a transfer have an inferred destination, i.e. no trips followed by a transfer in $\mathbf{U}$.
- Destinations of uninferred trips in $\mathbf{U}$ should be scaled excluding trips followed by a transfer.
- Let $\mathbf{u}$ be a vector of boardings with uninferred destinations.
- Let $\bar{\mathbf{L}}$ be a matrix of destination probability distributions for each origin for trips not followed by a transfer.
- The uninferred part $\mathbf{U}$ is estimated by

$$\tilde{\mathbf{U}} = \mathbf{u}\bar{\mathbf{L}}$$

Trip-Level Scaling with Transfer Information and without APC

- A portion of trips $\mathbf{N}$ is not observed.
  - Trips with uninferred origins
  - Trips without farebox interaction
- They can be estimated by combining ODX with passenger counts, e.g. APC data.
- Let $\bar{\mathbf{n}}$ be a vector of boarding scaling factors.
- Assuming destinations are distributed like observed trips,

$$\tilde{\mathbf{N}} = \bar{\mathbf{n}}\left(\mathbf{I} + \tilde{\mathbf{U}}\right)$$

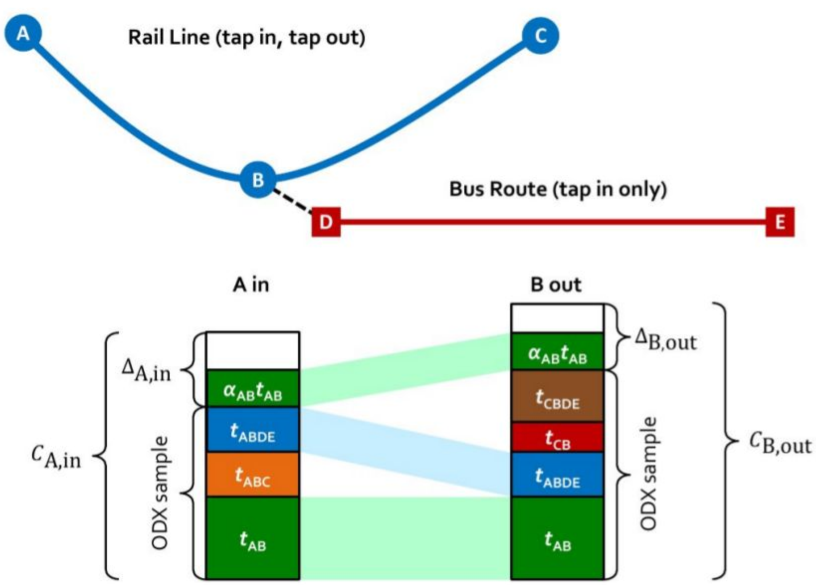
Trip-Level Scaling with Transfer Information and without APC

$$\begin{aligned}\tilde{\mathbf{R}} &= \mathbf{I} + \tilde{\mathbf{U}} + \tilde{\mathbf{N}} \\ &= \mathbf{I} + \tilde{\mathbf{U}} + \bar{\mathbf{n}}\left(\mathbf{I} + \tilde{\mathbf{U}}\right) \\ &= (\mathbf{1} + \bar{\mathbf{n}})\left(\mathbf{I} + \tilde{\mathbf{U}}\right) \\ &= (\mathbf{1} + \bar{\mathbf{n}})\left(\mathbf{I} + \mathbf{u}\bar{\mathbf{L}}\right)\end{aligned}$$

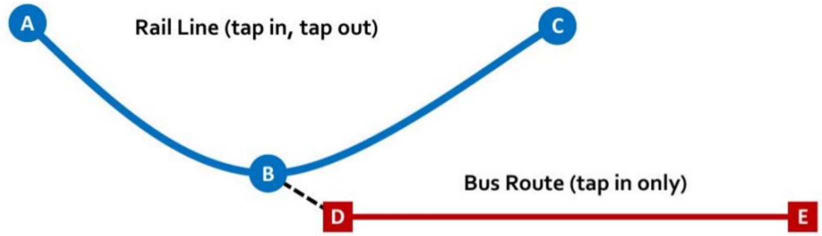
Journey Matrix Scaling

- Problem
  - Estimate expansion factors to scale Oyster-inferred full-journeys to represent non-Oyster and incompletely documented Oyster journeys.
- Challenges
  - Control totals available for stations, routes, but not itineraries
  - Large number of unique itineraries observed per day (if bus activity is aggregated to the route level)
  - Trillions of solutions can satisfy control totals
- Approach
  - Scale all full-journey itineraries to satisfy control totals

Journey Matrix Scaling



Journey Matrix Scaling



Count Node		Itinerary					
Station/Stop	Movement	AB	ABC	ABDE	CB	CBDE	DE
A	in	1	1	1	0	0	0
B	out	1	0	1	1	1	0
C	in	0	0	0	1	1	0
C	out	0	1	0	0	0	0
D	in	0	0	1	0	1	1

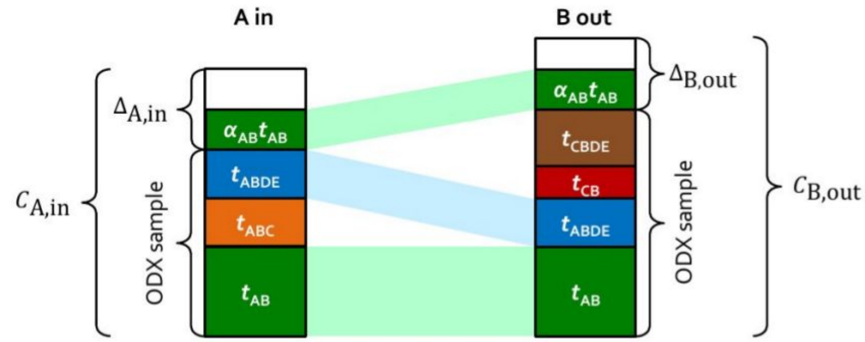
binary location-itinerary incidence matrix

B

Journey Matrix Scaling

$$T_i = (1+\alpha_i)t_i \quad \forall\, i \in I$$

$$\Delta_n = C_n - \sum_{i \in I} t_i b_{n,i} = \sum_{i \in I} t_i \alpha_i b_{n,i} \quad \forall\, n \in N$$



Journey Matrix Scaling

Initialize:

$$\boxed{\hat{\alpha}_i \leftarrow 1.0 \quad \forall\, i \in I}$$

Update:

$$\hat{\Delta}_n \leftarrow \sum_{i \in I} b_{n,i}\hat{\alpha}_i t_i \quad \forall\, n \in N$$

$$\hat{\alpha}_i \leftarrow \hat{\alpha}_i \frac{\sum_{n \in N} b_{n,i} \frac{\Delta_n}{\hat{\Delta}_n}}{\sum_{n \in N} b_{n,i}} \quad \forall\, i \in I$$

Count Location		Itinerary						Totals		
Station/Stop	Movement	AB	ABC	ABDE	CB	CBDE	DE	Δ_hat	Δ	Control
A	in	1	1	1	0	0	0	-	102	400
B	out	1	0	1	1	1	0	-	116	450
C	in	0	0	0	1	1	0	-	38	150
C	out	0	1	0	0	0	0	-	24	100
D	in	0	0	1	0	1	1	-	90	350
α		1.00	1.00	1.00	1.00	1.00	1.00			
αt		74	76	148	38	74	38			
(1+α)t		148	152	296	76	148	76			

Journey Matrix Scaling

Initialize:

$$\hat{\alpha}_i \leftarrow 1.0 \quad \forall\, i \in I$$

Update:

$$\boxed{\hat{\Delta}_n \leftarrow \sum_{i \in I} b_{n,i}\hat{\alpha}_i t_i \quad \forall\, n \in N}$$

$$\hat{\alpha}_i \leftarrow \hat{\alpha}_i \frac{\sum_{n \in N} b_{n,i} \frac{\Delta_n}{\hat{\Delta}_n}}{\sum_{n \in N} b_{n,i}} \quad \forall\, i \in I$$

Count Location		Itinerary						Totals		
Station/Stop	Movement	AB	ABC	ABDE	CB	CBDE	DE	Δ_hat	Δ	Control
A	in	1	1	1	0	0	0	298	102	400
B	out	1	0	1	1	1	0	334	116	450
C	in	0	0	0	1	1	0	112	38	150
C	out	0	1	0	0	0	0	76	24	100
D	in	0	0	1	0	1	1	260	90	350
α		1.00	1.00	1.00	1.00	1.00	1.00			
αt		74	76	148	38	74	38			
(1+α)t		148	152	296	76	148	76			

Journey Matrix Scaling

Initialize:

$$\hat{\alpha}_i \leftarrow 1.0 \quad \forall i \in I$$

Update:

$$\hat{\Delta}_n \leftarrow \sum_{i \in I} b_{n,i} \hat{\alpha}_i t_i \quad \forall n \in N$$

$$\boxed{\hat{\alpha}_i \leftarrow \hat{\alpha}_i \frac{\sum_{n \in N} b_{n,i} \frac{\Delta_n}{\hat{\Delta}_n}}{\sum_{n \in N} b_{n,i}} \quad \forall i \in I}$$

Count Location		Itinerary						Totals		
Station/Stop	Movement	AB	ABC	ABDE	CB	CBDE	DE	Δ_hat	Δ	Control
A	in	1	1	1	0	0	0	298	102	400
B	out	1	0	1	1	1	0	334	116	450
C	in	0	0	0	1	1	0	112	38	150
C	out	0	1	0	0	0	0	76	24	100
D	in	0	0	1	0	1	1	260	90	350
α		0.34	0.33	0.35	0.34	0.34	0.35			
αt		26	25	51	13	25	13			
(1+α)t		100	101	199	51	99	51			

Journey Matrix Scaling

Initialize:

$$\hat{\alpha}_i \leftarrow 1.0 \quad \forall i \in I$$

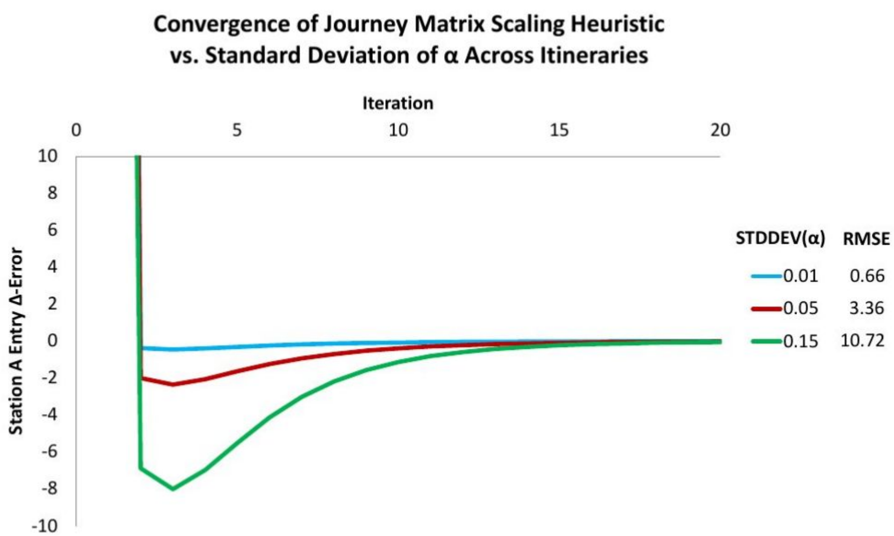
Update:

$$\hat{\Delta}_n \leftarrow \sum_{i \in I} b_{n,i} \hat{\alpha}_i t_i \quad \forall n \in N$$

$$\boxed{\hat{\alpha}_i \leftarrow \hat{\alpha}_i \frac{\sum_{n \in N} b_{n,i} \frac{\Delta_n}{\hat{\Delta}_n}}{\sum_{n \in N} b_{n,i}} \quad \forall i \in I}$$

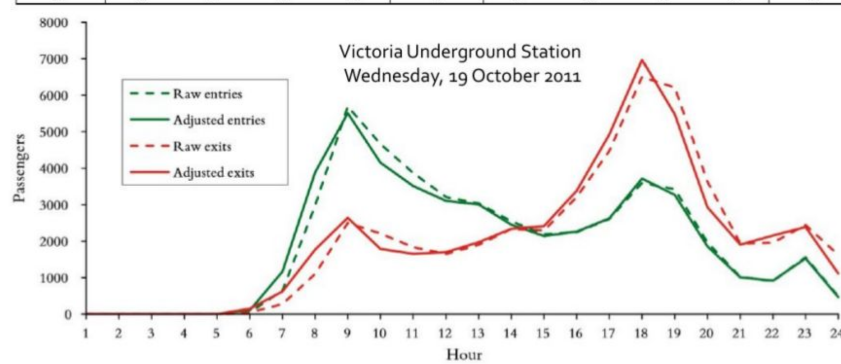
Count Location		Itinerary						Totals		
Station/Stop	Movement	AB	ABC	ABDE	CB	CBDE	DE	Δ_hat	Δ	Control
A	in	1	1	1	0	0	0	102	102	400
B	out	1	0	1	1	1	0	115	116	450
C	in	0	0	0	1	1	0	38	38	150
C	out	0	1	0	0	0	0	25	24	100
D	in	0	0	1	0	1	1	90	90	350
α		0.35	0.32	0.35	0.34	0.34	0.35			
αt		26	24	52	13	25	13			
(1+α)t		100	100	200	51	99	51			

Journey Matrix Scaling

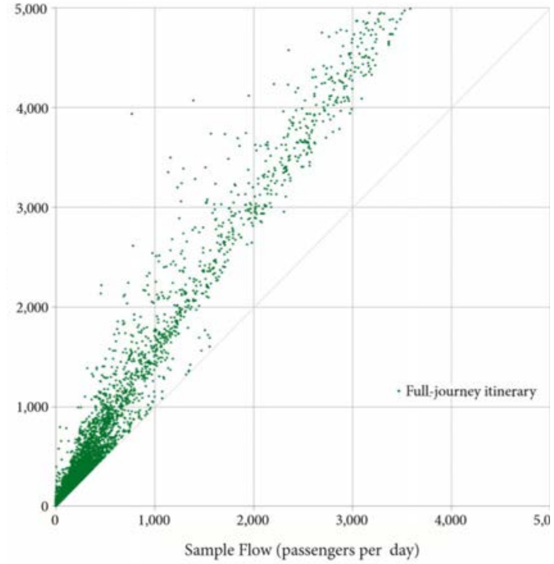


Journey Matrix Scaling

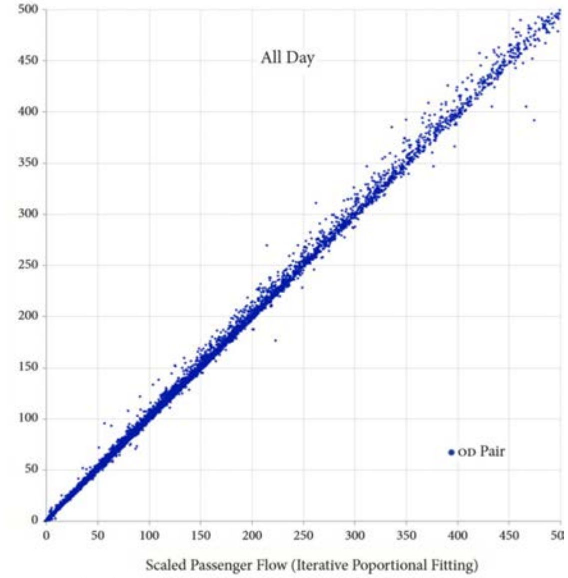
Recording Period	Period Offset 0	1	2	3	Unadjusted Count	Period Offset 0	1	2	3	Adjusted Count
5	100.0%	0.0%	0.0%	0.0%	1	1	0	0	0	22
6	66.7%	33.3%	0.0%	0.0%	59	39	20	0	0	199
7	64.7%	35.0%	0.3%	0.0%	444	287	155	1	0	...
8	59.4%	40.1%	0.5%	0.0%	803	477	322	4	0	...
9	47.7%	50.6%	1.7%	0.0%	803	383	406	14	0	...
...	...	...	...	...	...	...	...	...	...	...



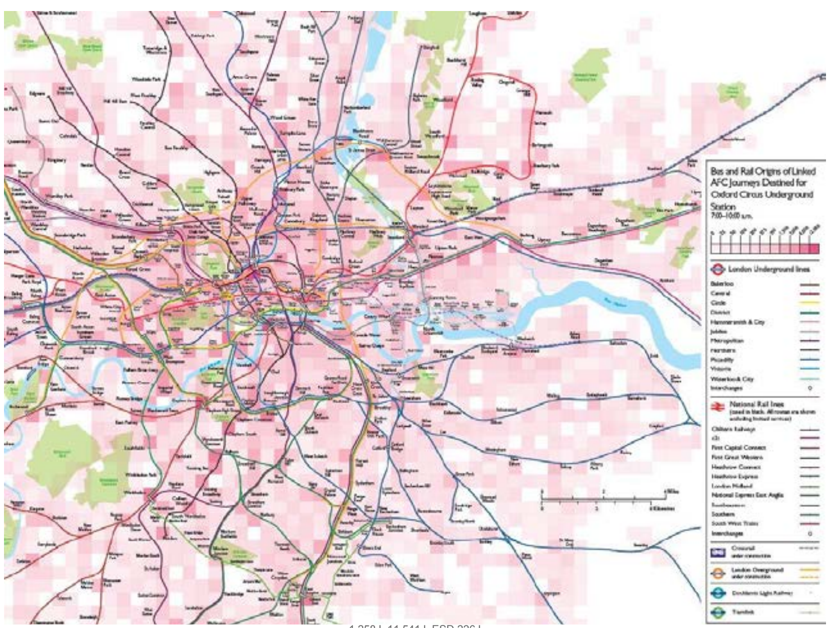
Scaling Factor Results



Journey Scaling vs. IPF (rail links only)



Full-Journey Scaling Results



References

- Gordon, Jason B. 2012. *Intermodal Passenger Flows on London's Public Transport Network: Automated Inference of Full Passenger Journeys Using Fare-Transaction and Vehicle-Location Data*. Master thesis, MIT.
- Gordon, Jason B., Koutsopoulos, Haris N., Wilson, Nigel H.M., Attanucci, J. 2013. *Automated Inference of Linked Transit Journeys in London Using Fare-Transaction and Vehicle-Location Data*. Transportation Research Record, 2343, pp. 17–24.
- Southwick, C.W. 2016. *Understanding Bus Passenger Crowding Through Origin Destination Inference*. Master thesis, MIT.
- Sánchez-Martínez, G.E. 2017. *Inference of Public Transportation Trip Destinations by Using Fare Transaction and Vehicle Location Data: Dynamic Programming Approach*. Transportation Research Record, 2652, pp. 1–7.
- Thesis, papers, poster, and visualization available at http://jaygordon.net