Dynamic Origin-Destination Estimation Using Smart Card Data: An Entropy Maximisation Approach

Abderrahman Ait Ali [a, b], Jonas Eliasson [b]

[a] The Swedish National Road and Transport Research Institute (VTI)

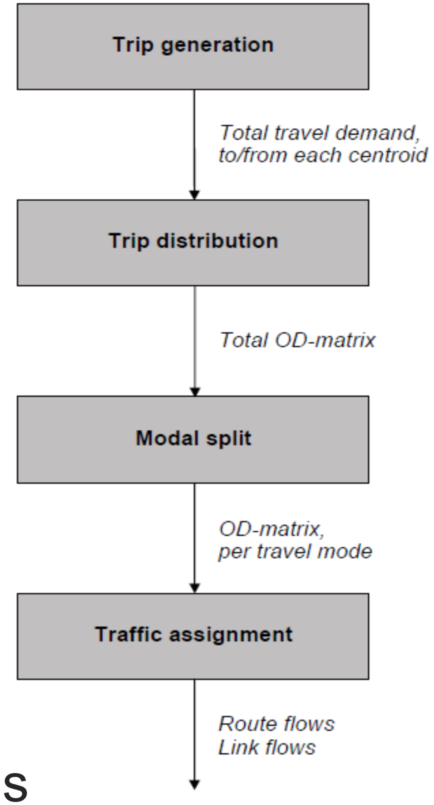
[b] Linköping University (LiU)

Outline

- Introduction
- Literature Review
  - OD estimation problems
  - OD estimation models
- Entropy Model
  - Basic model (BM)
  - Lagrangian relaxation
  - Symmetric assumption (SA)
  - Additional data (AD)
  - Calibration
- Case Study
  - Data
  - Results
  - Validation
- Conclusions and Future Works

Introduction

- Four step transport planning model
- Trip distribution ➔ OD matrix
- Different problems
  - Static vs dynamic OD (i.e. time-dependent)
  - Entry-only vs entry-exit systems
- Aim
  - Estimation of dynamic OD in entry-only systems



Literature Review – OD estimation problem

- Different problems
  - Static vs dynamic OD
  - Entry-only vs entry-exit systems
  - Zones and traffic (e.g. road vehicles, passengers)

- General formulation given target matrix ($\tilde{n}$) and link flows $\tilde{f}$

$$\min_{n,f} \alpha_n d_n(n, \tilde{n}) + \alpha_f d_f(f, \tilde{f})$$

vti

Literature Review – OD estimation models

- Growth (f) and Fratar ($f = f_i f_j$)

$$n_{ij} = f\,\tilde{n}_{ij}$$

- Gravity (c_{ij} as *deterrence function*)

$$n_{ij} = G_{ij}\frac{f_i f_j}{c_{ij}}$$

- Maximum Entropy

$$S_{entropy}(n) = \sum_{ijt}\left(n_{ij}^t \log\left(n_{ij}^t\right) - n_{ij}^t\right)$$

- Minimum Information

$$I(n,\tilde{n}) = \sum_{ijt} n_{ij}^t \log\left(\frac{n_{ij}^t}{\tilde{n}_{ij}^t}\right)$$

- Others: GLS, Bayesian, DCM, Kalman filters …

Entropy maximisation – Basic model (BM)

- Notations

Notation	Meaning
$\boldsymbol{T}$	set of time intervals over a working day
$\boldsymbol{S}$	set of all the nodes, i.e. train stations
$\boldsymbol{n_{ij}^t}$	number of trips from station i to $j \in S$ at time interval $t \in T$
$\boldsymbol{O_i^t}$	number of passengers entering station i at t

- Basic Model

$$\max_{n_{ij}^t \geq 0} \sum_{ijt} \left(n_{ij}^t \log\left(n_{ij}^t\right) - n_{ij}^t\right)$$

$$\sum_j n_{ij}^t = O_i^t \quad ; \forall i, t$$

Entropy maximisation – Lagrangian relaxation

- Lagrange function L

$$L(n,\lambda) = \sum_{ijt} (n_{ij}^t \log(n_{ij}^t) - n_{ij}^t) + \sum_{it} \lambda_{it} (\sum_j n_{ij}^t - O_i^t)$$

- KKT

$$\frac{\partial L}{\partial n_{ij}^t} = 0 \Rightarrow n_{ij}^t = e^{\lambda_{it}}$$

$$\sum_j n_{ij}^t = O_i^t \Rightarrow \sum_j e^{\lambda_{it}} = O_i^t \Rightarrow n_{ij}^t = \frac{O_i^t}{|S|}$$

vti

Entropy max – Symmetric Assumption (SA)

- Symmetric assumption, i.e. all those traveling from one station (to work/leisure) come back (home) to the same station.
- SA constraint

$$\sum_{it} n_{ij}^{t} = \sum_{t} O_{j}^{t} \quad ; \ \forall j.$$

- KKT

$$\frac{\partial L}{\partial n_{ij}^{t}} = 0 \ \Rightarrow n_{ij}^{t} = e^{\lambda_{it}} e^{\mu_{j}}$$

$$\begin{cases} \sum_{j} n_{ij}^{t} = O_{i}^{t} \\ \sum_{it} n_{ij}^{t} = \sum_{t} O_{j}^{t} \end{cases} \Rightarrow n_{ij}^{t} = O_{i}^{t} \frac{\sum_{\tau} O_{j}^{\tau}}{O}$$

vti

Entropy max – Additional data (AD)

- AD constraint, e.g. relative (d_{ij}) and average $(\bar{d})$ distances-pers.

$$\sum_{ijt} d_{ij} n_{ij}^t = \bar{d}$$

- KKT

$$\frac{\partial L}{\partial n_{ij}^t} = 0 \Rightarrow n_{ij}^t = e^{\lambda_{it}+\mu_j+\theta k_{ij}} = A_{it} e^{\theta d_{ij}+\mu_j}$$

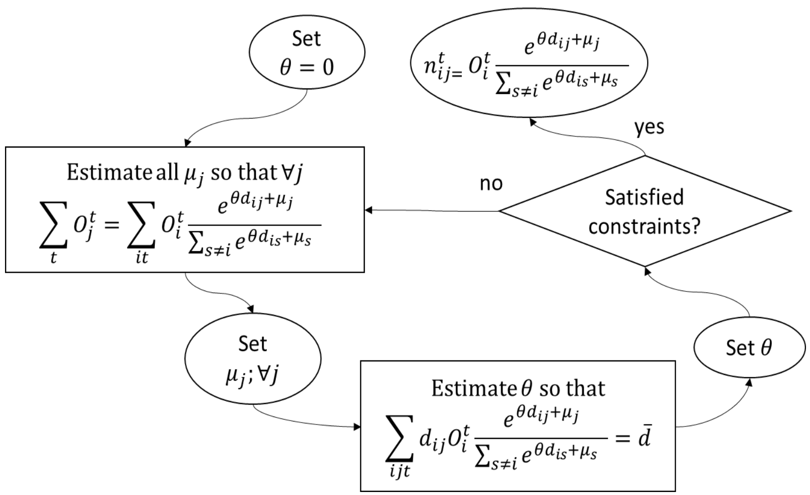
$$\sum_j n_{ij}^t = O_i^t \Rightarrow A_{it} \sum_j e^{\mu_j} e^{\theta d_{ij}} = O_i^t \Rightarrow n_{ij}^t = O_i^t \frac{e^{\theta d_{ij}+\mu_j}}{\sum_{s \neq i} e^{\theta d_{is}+\mu_s}}$$

- Generalisation (DCM), m pieces of AD

$$u_{ij} = K_j + \theta_1 k_{ij}^{(1)} + \cdots + \theta_m k_{ij}^{(m)} \Rightarrow p(j|i) = \frac{e^{u_{ij}}}{\sum_{s \neq i} e^{u_{is}}} \Rightarrow n_{ij}^t = O_i^t p(j|i)$$

Entropy max - Calibration

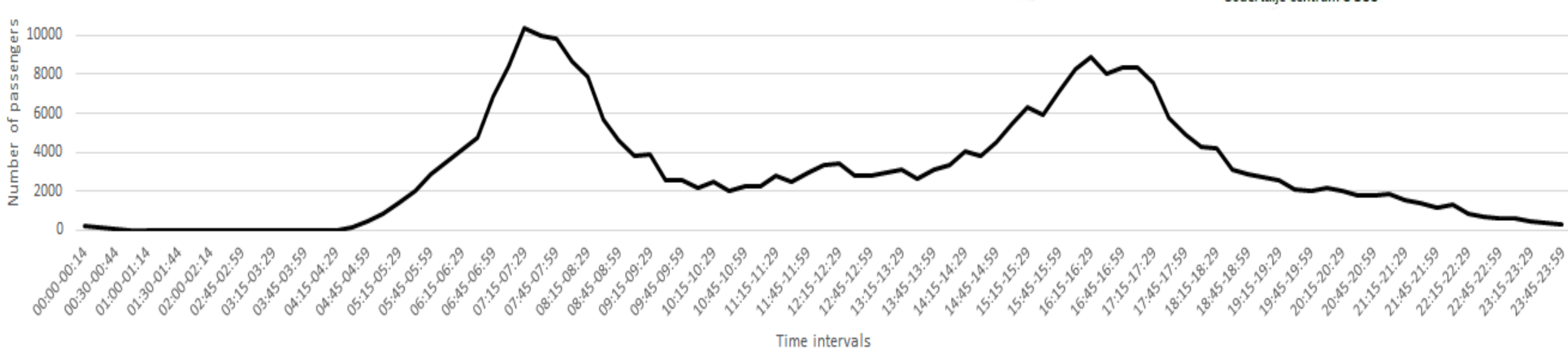
- This study, i.e. distances and person-km as AD
  - Iterative balancing (of additional constraints)
  - Until KKT satisfied (up to a certain tolerance)



- Generally
  - Iteratively balancing constraints (up to a tolerance)
  - Econometric methods (OLS, GLS, ..) but needs more data

Case Study - Data

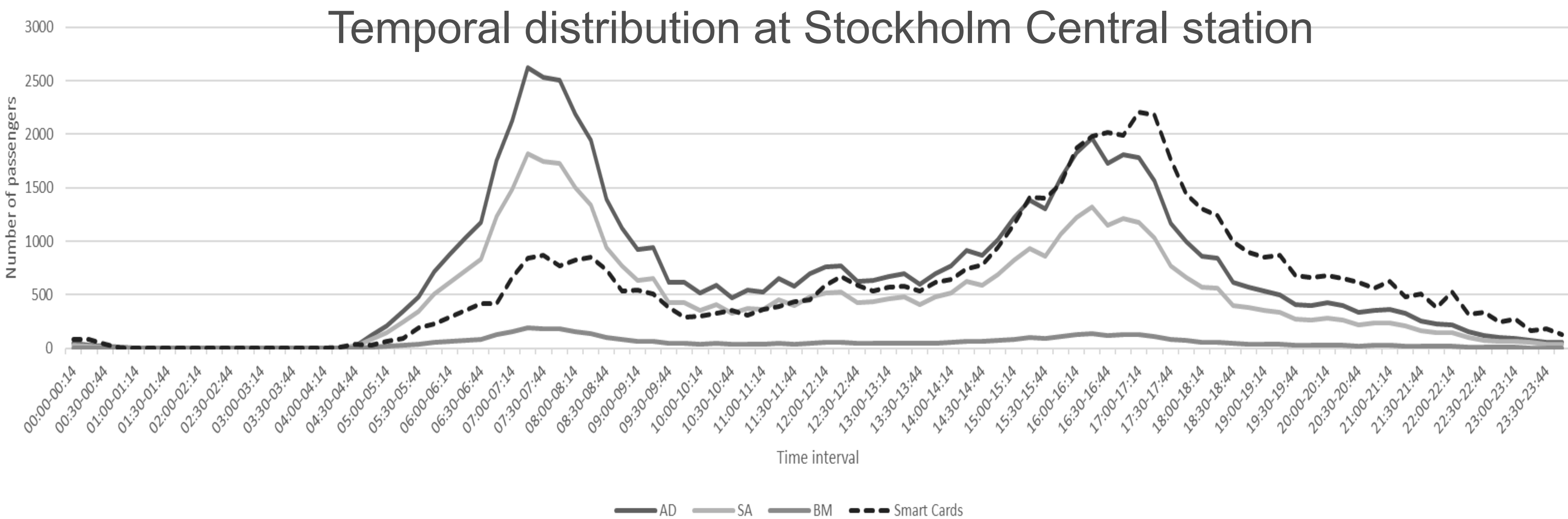
- Commuter train system (*pendeltåg*) in Stockholm
- Smart card (*SL Access*), normal working day (Sept 2015)



vti

Case Study - Results

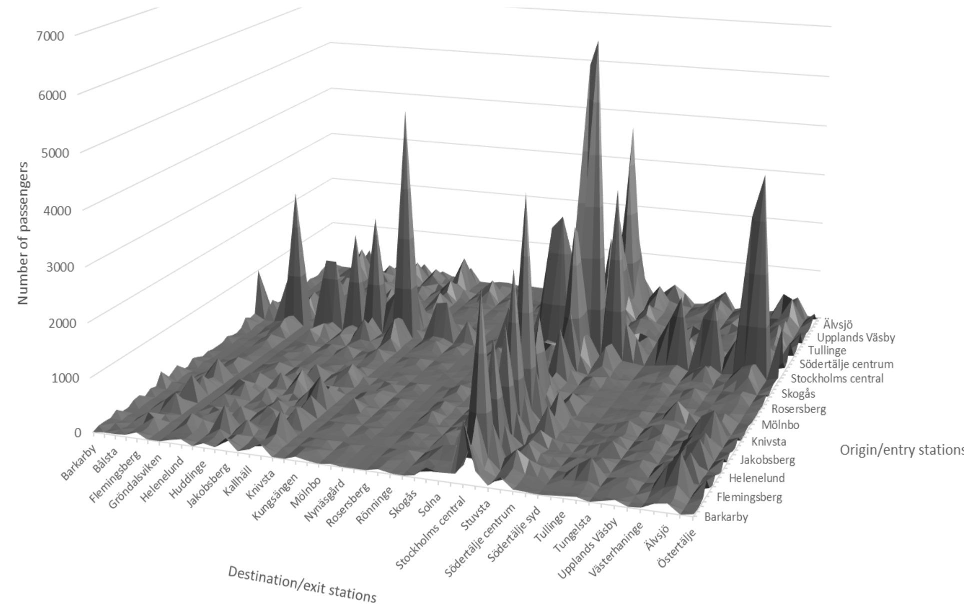
Variant	Characteristics
BM	Basic method, only smart card data.
SA	BM with the symmetric assumption (SA).
AD	SA with the additional data (AD): distances and total person-km.



- AD yields <u>higher</u> destination estimates than SA (and BM).
- <u>Higher</u> (estimated) exits than entries at morning peak and vice-versa at afternoon peak.

Case Study - Validation

- BM has worst estimates
- AD gives better estimates comparable to aggregated stats from SLL (2015)
- Daily OD matrix



	Total person-km (in 10^3)	Average travel distance (in km)	Total daily entry/exits at Central station
BM	10 194	34.8	56 000 / 47 380
SA	7 528	25.7	56 000 / 46 553
AD	5 761	19.7	56 000 / 58 314
SLL	**5 776**	**18.7**	**56 000 / 64 700**

Conclusions and Future Works

- Use of entropy max with entry-only smart card for OD estimation.
- Few additional assumptions/data can largely improve the destination estimation (OD matrix).
- Results can be validated with aggregated stats (from surveys, sensors, etc.)

- **Ideas for future works:**
  - Calibration with more data (econometric models).
  - Validation with sensor data from vehicle doors
  - Apply to other traffic (road vehicles, bus passengers

Thank you for your attention!

Questions?

abderrahman.ait.ali@vti.se