Comparing OD matrices estimated from smart card data, large scale OD survey and household travel survey.

Oscar EGU – Dr. Patrick BONNEL
LAET – ENTPE

oscar.egu@entpe.fr

TRANSITDATA2019
5th International Workshop and Symposium
Table of contents

- Introduction
- Material and methods
- Results
- Conclusions
Background

• OD matrices: essential for transit planning

• Traditional way to construct matrices:
  → Active data collection such as surveys

• Advances in data collection + intelligent transport systems + passive sensing:
  → Passive data can be used to build OD matrices
Motivations

• Both type of data have their pros and cons
  – Active data: specifically design to answer a research question, rely on accuracy of people to report correctly their travel behavior, limited sample and costly ...
  – Passive data: larger sample and lower cost but lack of semantic, inference methods needed, near absence of sociodemographic ...

More heterogeneous corpus to estimate travel demand!
Is this corpus consistent?
Research goals

• **Main goal:**
  – Compare and evaluate the difference between three data sources to estimate public transit OD-trip matrices. Do they give the same results?

• **Secondary goals:**
  – Validate empirically the relevance of smart card data to estimate the full demand i.e. without neglecting non-interaction (fare evasion) and paper tickets
  – Compare two scaling methods for smart card
  – Detect potential bias and assess accuracy of active solicitation data
**Material and methods**

**Case study**

**Lyon network**
- metro (4 lines)
- tramway (5 lines)
- bus (> 100 lines)
- 1,3 million inhabitants

**Spatial aggregation:**
- 18 zones (matrices with 324 OD pairs)
Matrices construction

1. **Household travel survey**
   Geographically stratified random survey (1% - 2% of resident)
   Unweighted sample: N= 10 570 trips in public transit
Matrices construction

2. **Large scale origin destination survey**

Manually collected: face to face interview
Every 5 years for each line (total cost: 3.7 million €)
Bus: on board and exhaustive
Metro/tramway: at station for a sample of passengers (25-35 %)
For each passenger:
   - origin-destination on surveyed line
   - connection with other lines
   - origin-destination of full trip
   - socio-demo + other trip attributes
Matrices construction

3. **Automatic fare collection**
   Smart card (SCAFC) and paper tickets (TACF)

 Transfer identification: [Devillaine et al., 2012]

 Destination inference: [Trépanier et al., 2007] → 80.8% transactions with destination inferred

 Matrix scaling:

 1: origin of smart card and ticket + uniform non-interaction rate (21%) [Munizaga et al., 2012] → WOAFc: weighted on origin automatic fare collection data

 2: based on counted flow [Gordon et al., 2018] → WIAFC: weighted on itinerary automatic fare collection data
### 3. Automatic fare collection

Number of counted boarding and recorded transaction in million:

<table>
<thead>
<tr>
<th>APC</th>
<th>AFC</th>
<th>Smart card</th>
<th>Ticket</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.56</td>
<td>1.22</td>
<td>0.98</td>
<td>0.24</td>
</tr>
</tbody>
</table>

- Smart card penetration rate: $0.98/1.22 = 80\%$
- Fare non-interaction rate: $100 - (1.22/1.56) = 22\%$
- Smart card with inferred destination vs counted flow: $
  \frac{(0.8 \times 0.98)}{1.56} = 50\%$
Summary of the methodology

Automatic passenger counting
13-17 March 2017

Automatic fare collection
13-17 March 2017
- Data preparation
- Trip identification
  - SCAFC
  - TAFC

Origin destination survey
2012-2016
- Data aggregation
- Line survey deduplication

Household travel survey
2015
- PT trips extraction
- Trips geocoding

WIAFC
WOAFC
ODS
HTS
Data sets structure: volume of trips

- Similar volume for AFC weighted, OD-survey, but not for HTS

**Results**

<table>
<thead>
<tr>
<th></th>
<th>HTS</th>
<th>ODS</th>
<th>WOAFC</th>
<th>WIAFC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume</td>
<td>1,11</td>
<td>1,51</td>
<td>1,55</td>
<td>1,56</td>
</tr>
<tr>
<td>Trip-legs (million)</td>
<td>0.80</td>
<td>1.16</td>
<td>1.11</td>
<td>1.10</td>
</tr>
</tbody>
</table>

WOAFC: weighted on origin, WIAFC: weighted on itinerary
HTS: household travel survey, ODS: OD survey
Data sets structure: daily distribution

- **HTS**: concentration of trips during peak periods

SCAFSC: smart card, TAFC: paper ticket
WOAFC: weighted on origin, WIAFC: weighted on itinerary
HTS: household travel survey, ODS: OD survey
Data sets structure: trip-leg distribution

- ODS: smaller trips, HTS: higher share of two legs trip

SCAFC: smart card, HTS: household travel survey, ODS: OD survey
Matrices structure

• Varying trip distribution at the macro level:
  – WOAFC (weighted on origin): higher share of central trips
  – WIAFC (weighted on itinerary) and HTS (household travel survey): more transversal and peripheral trips
  – ODS (OD survey): balanced distribution

<table>
<thead>
<tr>
<th>Origin</th>
<th>Destination</th>
<th>WOAFC</th>
<th>WIAFC</th>
<th>ODS</th>
<th>HTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central area</td>
<td>Central area</td>
<td>63%</td>
<td>58%</td>
<td>61%</td>
<td>57%</td>
</tr>
<tr>
<td>Central area</td>
<td>Peripheral ring</td>
<td>13%</td>
<td>14%</td>
<td>13%</td>
<td>15%</td>
</tr>
<tr>
<td>Peripheral ring</td>
<td>Central area</td>
<td>15%</td>
<td>15%</td>
<td>12%</td>
<td>14%</td>
</tr>
<tr>
<td>Peripheral ring</td>
<td>Peripheral ring</td>
<td>9%</td>
<td>13%</td>
<td>14%</td>
<td>15%</td>
</tr>
</tbody>
</table>

WOAFC: weighted on origin, WIAFC: weighted on itinerary
HTS: household travel survey, ODS: OD survey
Matrices structure

WOAFC: weighted on origin, WIAFC: weighted on itinerary
HTS: household travel survey, ODS: OD survey
### Results

#### Matrices structure

\[
SMAPE = \frac{100\%}{n} \sum_{i=1}^{n} \frac{\frac{1}{2} |y_{1,i} - y_{2,i}|}{(y_{1,i} + y_{2,i})}
\]

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_{1,i} - y_{2,i})^2}
\]

<table>
<thead>
<tr>
<th></th>
<th>WOAFC</th>
<th>WIAFC</th>
<th>HTS</th>
<th>ODS</th>
</tr>
</thead>
<tbody>
<tr>
<td>WOAFC</td>
<td>-</td>
<td>1049</td>
<td>1741</td>
<td>1484</td>
</tr>
<tr>
<td>WIAFC</td>
<td><strong>13%</strong></td>
<td>-</td>
<td>1350</td>
<td><strong>959</strong></td>
</tr>
<tr>
<td>HTS</td>
<td>40%</td>
<td>39%</td>
<td>-</td>
<td>1656</td>
</tr>
<tr>
<td>ODS</td>
<td>20%</td>
<td>21%</td>
<td>40%</td>
<td>-</td>
</tr>
</tbody>
</table>

WOAFC: weighted on origin, WIAFC: weighted on itinerary
HTS: household travel survey, ODS: OD survey
Conclusions

• Automatic fare collection can be used to construct reliable trip OD matrices but proper **scaling** is critical (paper ticket + non interaction):
  – Better results if complemented with automatic passenger counting.

• Household travel survey **underestimate** the number of public transit trip in an average weekday:
  – Non resident are missing, under reporting of trips, sample issues?
  – Probably not the best source to estimate matrices? But still needed to calibrate attribute choice model? Analyze travel behavior with other mode than PT?

• Interesting **opportunities** in terms of data fusion and data collection cost optimization:
  – OD surveys don’t capture variability and are costly
  – AFC lack of socio demographic
  – ...