Measuring Mobility in the 21st Century: What Can We Learn From Mobile Device Location Data?

Guangchen Zhao, Ph.D. Candidate, Department of Civil and Environmental Engineering
Email: gczhao@umd.edu

Anna Alberini, Professor, College of Agriculture and Natural Resources
Email: aalberin@umd.edu

Cinzia Cirillo, Professor, Department of Civil and Environmental Engineering
Email: ccirillo@umd.edu
Travel Survey

- Historically, personal travel behavior has been tracked through national and local travel surveys
- Typically collect information from a sample of households about:
  - Their socio-demographic characteristics, including work status and work commute
  - The vehicles available to them
  - A diary of all travel undertaken by each member of the household in the last 24 hours
  - …
Mobile Device Location Data (MDLD)

- From cellphone tower, Global Positioning System (GPS), and location-based services (LBS), etc.

- Typically, one location sighting includes an
  - Anonymized device identifier (ID)
  - Latitude and longitude coordinates
  - Time stamps
  - Positioning accuracy
  - ...

[Image of a mobile device and icons representing various transportation modes connected by lines, symbolizing data flow and connectivity].
# Travel Survey vs. MDLD

<table>
<thead>
<tr>
<th>Pros</th>
<th>MDLD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detailed Information</td>
<td>Continuous Monitoring</td>
</tr>
<tr>
<td>Directly Reported</td>
<td>Large-Scale</td>
</tr>
<tr>
<td>Customization</td>
<td>Spatial Accuracy</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cons</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Conducted infrequently (every 5-8 years)</td>
<td>Limited Behavioral Context</td>
</tr>
<tr>
<td>Covers different households in each wave</td>
<td>Inferential Challenges</td>
</tr>
<tr>
<td>One-day diaries</td>
<td>Computationally expensive</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

- **Pros**
  - Detailed Information
  - Directly Reported
  - Customization
  - ...

- **Cons**
  - Conducted infrequently (every 5-8 years)
  - Covers different households in each wave
  - One-day diaries
  - ...

## Mobile Device Location Data (MDLD)

- **Pros**
  - Detailed Information
  - Directly Reported
  - Customization
  - ...
  - Continuous Monitoring
  - Large-Scale
  - Spatial Accuracy
  - ...

- **Cons**
  - Conducted infrequently (every 5-8 years)
  - Covers different households in each wave
  - One-day diaries
  - ...
  - Limited Behavioral Context
  - Inferential Challenges
  - Computationally expensive
  - ...

---

### Passenger OD data production flow chart for the Next Generation National Household Travel Survey (NextGen NHTS) OD Data Program
Research Questions

- How do MDLD Vehicle Miles Travelled (VMT) compare with VMT from high-quality, more traditional trip diaries/trip rosters?
- Can we form micro-level, longitudinal datasets using the MDLD?
- MDLD VMT...
  - Do they reflect fluctuations and local differences in gasoline prices?
  - Are they good predictors of station-level gasoline prices?
Study Area

DC + the “Middle Ring” counties

Commuting Patterns:

<table>
<thead>
<tr>
<th></th>
<th>2019</th>
<th>2022</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC</td>
<td>35.12%</td>
<td>34.96%</td>
</tr>
<tr>
<td>Middle Ring</td>
<td>61.21%</td>
<td>57.71%</td>
</tr>
<tr>
<td>Elsewhere or unknown</td>
<td>3.66%</td>
<td>7.33%</td>
</tr>
</tbody>
</table>

From Metropolitan Washington Council of Governments' State of the Commute (2019, 2022)
MDLD 2019 vs RTS Average Daily VMT

AVERAGE DAILY MILEAGE DRIVEN

MONTH

2019 MDLD
RTS
No government-run travel surveys were conducted at that time

With 2021-2022 MDLD:

- Can we form micro-level, longitudinal datasets using the MDLD?
- MDLD VMT…
  - Do they reflect fluctuations and local differences in gasoline prices?
  - Are they good predictors of station-level gasoline prices?
## Device Panel & Pseudo Panels

<table>
<thead>
<tr>
<th></th>
<th>Device Panel</th>
<th>Pseudo Panels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>#1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>#2</td>
</tr>
<tr>
<td><strong>Unit</strong></td>
<td><strong>Device</strong></td>
<td><strong>Census Block Group</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Census Tract</strong></td>
</tr>
<tr>
<td><strong>Average daily VMT in</strong></td>
<td><strong>By device</strong></td>
<td><strong>By all devices residing in</strong></td>
</tr>
<tr>
<td><strong>each week</strong></td>
<td></td>
<td><strong>the CBG</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>By all devices residing in</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>the CT</strong></td>
</tr>
<tr>
<td><strong>N. units</strong></td>
<td>82,866 (10% sample)</td>
<td>817</td>
</tr>
<tr>
<td></td>
<td></td>
<td>623</td>
</tr>
<tr>
<td><strong>Avg N. observations</strong></td>
<td>7</td>
<td>62</td>
</tr>
<tr>
<td><strong>(weeks) per unit</strong></td>
<td></td>
<td>68</td>
</tr>
<tr>
<td><strong>% of Obs with 0 VMT</strong></td>
<td>5%</td>
<td>0.2%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.02%</td>
</tr>
</tbody>
</table>
Gasoline Price Variation in Different States

![Graph showing the average gas price in different states over time. The states compared are Washington, DC, Maryland, and Virginia. The graph shows fluctuations in gas prices, with peaks and troughs throughout the dates from December 27, 2020, to March 27, 2023.](image-url)
People may choose to live where housing is more affordable, gas prices tend to be lower, and they drive more
Regression of Average Daily VMT from MDLD

- **Equation**

\[
VMT_{it} = \alpha_i + \gamma_t + \beta_1 \cdot GasPrice_{it} + \epsilon_{it}
\]

- \(VMT_{it}\) - the average daily vehicle miles travelled (VMT) by individual or residents in zone \(i\) at week \(t\)
- \(\alpha_i\) - the fixed effect of individual or zone \(i\)
- \(\gamma_t\) - time fixed effects
- \(\beta_1\) - the coefficient of \(GasPrice_{it}\)
- \(GasPrice_{it}\) - the average gasoline price near individual \(i\) or inside zone \(i\) at week \(t\)
- \(\epsilon_{it}\) - the error term
Regression of Average Daily Miles Driven from MDLD

- With local gasoline prices

<table>
<thead>
<tr>
<th></th>
<th>Device-level panel, Weekly</th>
<th>Device-level panel, incl. zeros, weekly</th>
<th>Block group averages, weekly</th>
<th>Census tract averages, weekly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coeff. on gasoline price</td>
<td>-0.6009 *** (0.1647)</td>
<td>-0.3255 * (0.1148)</td>
<td>-2.3679 *** (0.5727)</td>
<td>-2.4099 *** (0.6321)</td>
</tr>
<tr>
<td>Price Elasticity of daily VMT</td>
<td>-0.0790 (0.0217)</td>
<td>-0.0451 (0.0159)</td>
<td>-0.2679 (0.0649)</td>
<td>-0.2789 (0.0732)</td>
</tr>
<tr>
<td>Fixed effects</td>
<td>Device</td>
<td>Device</td>
<td>Block group</td>
<td>Census tract</td>
</tr>
<tr>
<td>Time fixed effects</td>
<td>Month, year</td>
<td>Month, year</td>
<td>Month, year</td>
<td>Month, year</td>
</tr>
<tr>
<td>R square</td>
<td>0.03</td>
<td>0.03</td>
<td>0.24</td>
<td>0.41</td>
</tr>
<tr>
<td>N. Observations</td>
<td>560,779</td>
<td>590,572</td>
<td>50,435</td>
<td>42,257</td>
</tr>
</tbody>
</table>
Regression of Gasoline Price at the Station Level

**Equation**

\[
\text{GasPrice}_{it} = \alpha_i + \gamma_t + \beta_1 \cdot \text{VMT}_{i,t-1} + \varepsilon_{it}
\]

- \(\text{GasPrice}_{it}\) - the average gasoline price of gasoline station \(i\) at week \(t\)
- \(\alpha_i\) - the fixed effect of gasoline station \(i\)
- \(\gamma_t\) - time fixed effects
- \(\beta_1\) - the coefficient of \(\text{VMT}_{i,t-1}\)
- \(\text{VMT}_{i,t-1}\) - the total vehicle miles travelled (VMT) at the previous week \(t - 1\) by individuals residing in the zone where the gasoline station \(i\) is located
- \(\varepsilon_{it}\) - the error term
## Gas station specific gasoline prices

<table>
<thead>
<tr>
<th></th>
<th>#1</th>
<th>#2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coeff. on total mileage driven in the previous week at Census tract-level, Weekly</td>
<td>-3.12E-08 (2.08E-08)</td>
<td>1.28E-07 *** (3.37E-08)</td>
</tr>
<tr>
<td>Coeff. on number of gas stations within 0.25 miles</td>
<td>-0.0172 *** (0.0011)</td>
<td>NA</td>
</tr>
<tr>
<td>Other independent variables that are statistically significant</td>
<td>Population density, Income distribution, Age distribution, Education-level distribution, Road-network density, ...</td>
<td>NA</td>
</tr>
<tr>
<td>Fixed effects</td>
<td>Brand, State</td>
<td>Station, Brand, State</td>
</tr>
<tr>
<td>Time fixed effects</td>
<td>Month, Year</td>
<td>Month, Year</td>
</tr>
<tr>
<td>R square</td>
<td>0.65</td>
<td>0.67</td>
</tr>
<tr>
<td>N. Observations</td>
<td>20,021</td>
<td>20,021</td>
</tr>
</tbody>
</table>
Conclusion

- Average Daily VMT from 2019 MDLD shows a similar monthly trend as VMT from RTS
- Different ways of forming longitudinal datasets with MDLD
  - Device-level panel, with shorter tracking period, and more units
  - Pseudo Panels, aggregated, with longer series, fewer units, and more stable
- MDLD VMT...
  - Average daily VMT negatively correlated with the price of gasoline, reflecting fluctuations and local differences in gasoline prices.
  - Total VMT help explain the price of gasoline in the area and thus serve as a good proxy for demand pressure on fuel prices.
Thank you!

Comments? Questions?

Guangchen Zhao, Ph.D. Candidate, Department of Civil and Environmental Engineering
Email: gczhao@umd.edu

Anna Alberini, Professor, College of Agriculture and Natural Resources
Email: aalberin@umd.edu

Cinzia Cirillo, Professor, Department of Civil and Environmental Engineering
Email: ccirillo@umd.edu