Measuring Commuting and Economic Activity inside Cities with Cell Phone Records

Gabriel Kreindler (Harvard University)
Yuhei Miyauchi (Boston University)

6th Urbanization and Poverty Reduction Research Conference

September 9th, 2019
Data on Economic Activity within Cities Valuable yet Scarce

- Detailed spatial data on firms, jobs, wages is important for policymakers and researchers.

- Useful for analyzing localized shocks within cities: floods, violence, industry-specific demand shocks, transportation policy, etc.

- However, such data is generally scarce.
Data on Economic Activity *within* Cities Valuable yet Scarce

- Detailed spatial data on firms, jobs, wages is important for policymakers and researchers.

- Useful for analyzing localized shocks within cities: floods, violence, industry-specific demand shocks, transportation policy, etc.

- However, such data is generally scarce.
Data on Economic Activity within Cities Valuable yet Scarce

- Detailed spatial data on firms, jobs, wages is important for policymakers and researchers.

- Useful for analyzing localized shocks within cities: floods, violence, industry-specific demand shocks, transportation policy, etc.

- However, such data is generally scarce.
Conventional data on firms (including informal sector) generally scarce

<table>
<thead>
<tr>
<th></th>
<th>N Countries</th>
<th>% Urban Pop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Sub-Saharan Africa Sample</td>
<td>27</td>
<td>100%</td>
</tr>
<tr>
<td>Have Economic Census ...</td>
<td>16</td>
<td>45%</td>
</tr>
<tr>
<td>... and Covers Informal Firms</td>
<td>11</td>
<td>25%</td>
</tr>
<tr>
<td>... and Wage Data</td>
<td>≤ 4</td>
<td>&lt; 9%</td>
</tr>
</tbody>
</table>
Conventional data on firms (including informal sector) generally scarce

<table>
<thead>
<tr>
<th>N Countries</th>
<th>% Urban Pop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Sub-Saharan Africa Sample</td>
<td>27</td>
</tr>
<tr>
<td>Have Economic Census ...</td>
<td>16</td>
</tr>
<tr>
<td>... and Covers Informal Firms</td>
<td>11</td>
</tr>
<tr>
<td>... and Wage Data</td>
<td>≤ 4</td>
</tr>
</tbody>
</table>
Conventional data on firms (including informal sector) generally scarce

<table>
<thead>
<tr>
<th>N Countries</th>
<th>% Urban Pop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Sub-Saharan Africa Sample</td>
<td>27</td>
</tr>
<tr>
<td>Have Economic Census ...</td>
<td>16</td>
</tr>
<tr>
<td>... and Covers Informal Firms</td>
<td>11</td>
</tr>
<tr>
<td>... and Wage Data</td>
<td>≤ 4</td>
</tr>
</tbody>
</table>
Conventional data on firms (including informal sector) generally scarce

<table>
<thead>
<tr>
<th>N Countries</th>
<th>% Urban Pop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Sub-Saharan Africa Sample</td>
<td>27</td>
</tr>
<tr>
<td>Have Economic Census ...</td>
<td>16</td>
</tr>
<tr>
<td>... and Covers Informal Firms</td>
<td>11</td>
</tr>
<tr>
<td>... and Wage Data</td>
<td>≤ 4</td>
</tr>
</tbody>
</table>
This Paper: Use Commuting Flows from Cell Phone Data to Infer Wages

Two facts:

1. Economic activity in cities intertwined with commuting behavior
2. Rich data on urban mobility increasingly available

1. Data: cell phone transactions in Colombo, Sri Lanka, and Dhaka, Bangladesh
   ▶ Construct and validate commuting flows

2. Method to recover labor productivity data from commuting patterns
   ▶ Based on gravity equation
   ▶ Unlike machine learning, no training data necessary
   ▶ We show validation results

3. Showcase application: impact of hartal strikes in Dhaka
This Paper: Use Commuting Flows from Cell Phone Data to Infer Wages

Two facts:
1. Economic activity in cities intertwined with commuting behavior
2. Rich data on urban mobility increasingly available

This paper:
1. **Data**: cell phone transactions in Colombo, Sri Lanka, and Dhaka, Bangladesh
   - Construct and validate **commuting flows**

2. **Method** to recover labor productivity data from commuting patterns
   - Based on gravity equation
   - Unlike machine learning, no training data necessary
   - We show validation results

3. Showcase **application**: impact of *hartal* strikes in Dhaka
This Paper: Use Commuting Flows from Cell Phone Data to Infer Wages

Two facts:
1. Economic activity in cities intertwined with commuting behavior
2. Rich data on urban mobility increasingly available

This paper:
1. **Data**: cell phone transactions in Colombo, Sri Lanka, and Dhaka, Bangladesh
   - Construct and validate **commuting flows**

2. **Method** to recover labor productivity data from commuting patterns
   - Based on gravity equation
   - Unlike machine learning, no training data necessary
   - We show validation results

3. Showcase **application**: impact of *hartal* strikes in Dhaka

3/18
This Paper: Use Commuting Flows from Cell Phone Data to Infer Wages

Two facts:
1. Economic activity in cities intertwined with commuting behavior
2. Rich data on urban mobility increasingly available

This paper:
1. **Data:** cell phone transactions in Colombo, Sri Lanka, and Dhaka, Bangladesh
   - Construct and validate **commuting flows**

2. **Method** to recover labor productivity data from commuting patterns
   - Based on gravity equation
   - Unlike machine learning, no training data necessary
   - We show validation results

3. Showcase **application**: impact of *hartal* strikes in Dhaka
Cell Phone Transaction Data (CDR) from Sri Lanka and Bangladesh

- Data from Dhaka and Colombo around 2013
  - 8 million anonymized user IDs
  - for each call: user ID, timestamp, cell phone tower location
  - no data on: gender, education, occupation, etc.

- Construct commuting flows by observing the same SIM card on the same day (morning and afternoon)
  - 440 million days with commuting information
  - Results robust to using “common” day and night places

- CDR commuting flows correlate well with survey commuting flows in Dhaka
Cell Phone Transaction Data (CDR) from Sri Lanka and Bangladesh

- Data from Dhaka and Colombo around 2013
  - 8 million anonymized user IDs
  - for each call: user ID, timestamp, cell phone tower location
  - no data on: gender, education, occupation, etc.

- Construct commuting flows by observing the same SIM card on the same day (morning and afternoon)
  - 440 million days with commuting information
  - Results robust to using “common” day and night places

- CDR commuting flows correlate well with survey commuting flows in Dhaka
Cell Phone Transaction Data (CDR) from Sri Lanka and Bangladesh

- Data from Dhaka and Colombo around 2013
  - 8 million anonymized user IDs
  - for each call: user ID, timestamp, cell phone tower location
  - no data on: gender, education, occupation, etc.

- Construct commuting flows by observing the same SIM card on the same day (morning and afternoon)
  - 440 million days with commuting information
  - Results robust to using “common” day and night places

- CDR commuting flows correlate well with survey commuting flows in Dhaka
Geographic Unit: Cell Phone Tower Voronoi Cells – Dhaka, Bangladesh
Geographic Unit: Cell Phone Tower Voronoi Cells – Colombo, Sri Lanka
Example: Commuting Flows from a Single Origin Tower (Colombo)
Commuting Flows from CDR vs Survey Data (Dhaka)

Commuting flows between pairs of survey wards

Dhaka
90 survey wards

- Cell data, log(mean())
- Survey data, log(mean())
- Bootstrapped CI

Log commuting flow vs Log Travel Time
The Logic of our Method

- Hypothesis: work destinations with high wages attract more workers, \textit{ceteris paribus}.

- Gravity equation: regress commuting flows on travel time and origin and destination factors
  - Estimate \textit{destination attractiveness}
  - Interpret as measure of wages

  - Procedure has nice theoretical properties
The Logic of our Method

- Hypothesis: work destinations with high wages attract more workers, *ceteris paribus*.

- Gravity equation: regress commuting flows on travel time and origin and destination factors
  - Estimate *destination attractiveness*
  - Interpret as measure of wages

  - Procedure has nice theoretical properties
The Logic of our Method

- Hypothesis: work destinations with high wages attract more workers, *ceteris paribus*.

- Gravity equation: regress commuting flows on travel time and origin and destination factors
  - Estimate *destination attractiveness*
  - Interpret as measure of wages

  - Procedure has nice theoretical properties
Estimated (smoothed) log Wages in Dhaka and Colombo

Figure 1: Estimated log Wages in Dhaka and Colombo
Validating Model-Predicted Income with Other Data Sources

- Model-predicted income is computed without “training” data
  - Only uses commuting behavior and Google Maps travel times

- Model: we know how income “moves” across the city
  - We compute income at workplace and at residential level

- Two validation exercises. Compare:
  1. Model workplace income and survey workplace income
  2. Model residential income and nighttime lights
Validating Model-Predicted Income with Other Data Sources

▶ Model-predicted income is computed without “training” data
  ▶ Only uses **commuting behavior** and **Google Maps travel times**

▶ Model: we know how income “moves” across the city
  ▶ We compute income at **workplace** and at **residential** level

▶ Two validation exercises. Compare:
  1. Model **workplace** income and survey workplace income
  2. Model **residential** income and nighttime lights
Validating Model-Predicted Income with Other Data Sources

- Model-predicted income is computed without “training” data
  - Only uses commuting behavior and Google Maps travel times

- Model: we know how income “moves” across the city
  - We compute income at workplace and at residential level

- Two validation exercises. Compare:
  1. Model workplace income and survey workplace income
  2. Model residential income and nighttime lights
Validation at *Workplace*: Model Income and Survey Income

(A) Model and Survey Workplace Income
Validation at Residential: Model Income and VIIRS Nighttime Data

- Nighttime satellite lights proxy of country GDP growth (Henderson et al 2010)
- Within cities, intuitively nightlights capture residential income
Discussion: How to Judge Model Predictive Performance?

- Model-predicted income consistently statistically significantly predictive of income from other sources

- However, predictive power for income from survey data is modest ($R^2 \approx 0.3$).
  
  1. Survey data itself not perfect
  2. Difficult prediction problem within cities
  3. In fact, machine learning approaches predict income with similar accuracy when looking at cities only (Blumenstock et al 2015, Jean et al 2016)
The Impact of Hartals in Dhaka

- Hartals are strikes in Bangladesh that involve partial shutdowns of urban transportation and businesses.

- 31 hartal days in 4 months in late 2013 in Dhaka (we use data from Ahsan and Iqbal, 2015)

- Objective: use rich commuting data and model predictions to estimate income losses due to hartal

- Accounting exercise:
  1. Measure income changes due to commuting changes.
  2. Assume that wages stay constant.
Hartal
Average Model-predicted Income is Lower on Hartal Days
Average Model-predicted Income is Lower on Hartal Days

![Graph showing income changes on different days.](image-url)
Commuters in Dhaka Travel and Earn Less on Hartal Days

- Commuters in Dhaka earn on average 4.4 to 4.8% less on hartal days compared to workdays
  - Effects much smaller compared to Fridays (20 to 45% lower predicted income)

- Effects driven primarily by the extensive margin, namely fewer trips

- Commuters with longer trips reduce trips relatively more
- Commuters working in high-income destinations reduce trips relatively more, controlling for trip duration
Commuters in Dhaka Travel and Earn Less on Hartal Days

- Commuters in Dhaka earn on average 4.4 to 4.8% less on hartal days compared to workdays
  - Effects much smaller compared to Fridays (20 to 45% lower predicted income)

- Effects driven primarily by the extensive margin, namely fewer trips

- Commuters with longer trips reduce trips relatively more
- Commuters working in high-income destinations reduce trips relatively more, controlling for trip duration
Commuters in Dhaka Travel and Earn Less on Hartal Days

- Commuters in Dhaka earn on average 4.4 to 4.8% less on hartal days compared to workdays
  - Effects much smaller compared to Fridays (20 to 45% lower predicted income)

- Effects driven primarily by the extensive margin, namely fewer trips

- Commuters with longer trips reduce trips relatively more

- Commuters working in high-income destinations reduce trips relatively more, controlling for trip duration
Conclusion: Using Big Data to Measure Revealed Preferences

- Used cell phone data to construct (a) commuting, and (b) detailed urban economic activity measures

- Income from this method predicts survey income and nighttime lights

- Potential applications: analyzing urban shocks localized in space and/or time

- Big data for revealed preferences: route choice (safety: Borker 2016), value of public services, payments (informal economic activity)
Thank you!
### Table C.1: Cell Phone Data Coverage at User-Day Level

<table>
<thead>
<tr>
<th></th>
<th>Dhaka, Bangladesh</th>
<th>Colombo, Sri Lanka</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Users in sample</td>
<td>$5.3 \times 10^6$</td>
<td>$3.0 \times 10^6$</td>
</tr>
<tr>
<td>(2) Days in sample</td>
<td>122</td>
<td>395</td>
</tr>
<tr>
<td>(3) All user-days possible = (1)×(2)</td>
<td>$6.5 \times 10^8$</td>
<td>$1.2 \times 10^9$</td>
</tr>
<tr>
<td>(4) User-days with data</td>
<td>$2.9 \times 10^8$</td>
<td></td>
</tr>
<tr>
<td>(5) User-days with data (5-10am)</td>
<td>$1.5 \times 10^8$</td>
<td></td>
</tr>
<tr>
<td>(6) User-days with data (10am-3pm)</td>
<td>$2.4 \times 10^8$</td>
<td></td>
</tr>
<tr>
<td>(7) User-days with data (5-10am and 10am-3pm)</td>
<td>$1.0 \times 10^8$</td>
<td>$3.4 \times 10^8$</td>
</tr>
<tr>
<td>(8) Coverage rate = (7)/(3)</td>
<td></td>
<td>16.1%</td>
</tr>
</tbody>
</table>