UNDERSTANDING HOW COMMUTERS USE (NEW) TRANSPORTATION SYSTEMS

KARAN GIROTRA

CORNELL TECH

ASHISH KABRA, KASHISH ARORA (INSEAD),
ELENA BELAVINA (BOOTH @CHICAGO),
FANYIN ZHENG, PU HE (COLUMBIA)

VELIB, TFL, EASYTAXI, OLA CABS
TODAY’S TALK

Decision Models

How to use archival data to estimate key input parameters for these models?
Illustrate use to support decisions (system design v/s operations)

Transportation Systems: Smart/New
Bike-Share, App-based Taxi Hailing Service, On-demand “public” transportation

General Principle/Template
Specific Projects
**System Intervention: A New Lane**

- Naïve Theoretical Prediction: Higher Capacity --- Lower Congestion

- Empirical Observation: Higher Capacity often has no effect on Congestion!

---

Substitution

- Alternate Routes
- Alternate Means
- Lifestyle Choices

Demand Function

System Use

Demand System

Consumer Choice Process
**System Intervention: A New Lane**

- Theoretical Prediction: Higher Capacity --- Lower Congestion

- Empirical Observation: Higher Capacity often Increases Congestion!
  - More People use the route!

Where are you more likely to add lanes? Supply is often Endogenous

**System Design Process**
SYSTEM INTERVENTION: A NEW LANE

► Theoretical Prediction: Higher Capacity --- Lower Congestion

► Empirical Observation: Higher Capacity often Increases Congestion!
  ► More People use the route!

Real Time Observation of Performance adds “noise” to data

Somewhat-Randomized Experiments

Accounting for this requires techniques to handle large data sets
**Estimating “Demand” for Transportation Systems**

Designing System Interventions Requires “Correct” understanding of *Consumer Behavior*

<table>
<thead>
<tr>
<th>Use</th>
<th>Operational Performance</th>
<th>Primitive</th>
<th>Evaluate Interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Customer Utility</td>
<td>Counterfactuals</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Travel</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Convenience</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>System Design Process</th>
<th>Consumer Choice Process</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Structural Estimation, Large Data Sets*
Some Recent Work


Closer or More Reliable Stations?

Kabra, A., K. Girotra and E. Belavina, “Designing Promotions to Scale Marketplaces”, 2017

Incentives and Achieving Scale


Density V/s Scope?


Where to offer “Shuttles”?
THE BIKE SHARE BUSINESS MODEL: PRODUCT AND REVENUE MODEL

Sturdy, Reliable Bikes

Convenient Parking Stations

Shared Consumption by Consumers

Sturdy, Reliable Bikes

- **Frame**: The frame is made of high-grade steel to ensure durability and safety.
- **Basket**: The basket is designed to accommodate personal belongings.
- **Saddle**: The saddle is ergonomic and comfortable for long rides.
- **Gears**: The bike has a Shimano 7-speed gear system for easy riding.
- **Rack**: The rear rack is designed to carry bags and other items.
- **Frame**: The frame is made of high-grade steel to ensure durability and safety.
- **Suspension Fork**: The suspension fork absorbs shocks for a smooth ride.
- **Tires**: The tires are made of durable rubber for grip and stability.

Convenient Parking Stations

- **Solar Power**: The station is powered by solar panels.
- **System Map**: The map is updated in real-time.
- **Pay Station**: The pay station accepts various payment methods.
- **Bicycle Dock**: Bikes can be rented and returned at any station.

Shared Consumption by Consumers

- **SIGN UP**: Daily, Monthly
  - Annual Plans
  - First 30 minutes free
- **SWIPE OUT**: Bike Availability Information
  - Automated Checkout
- **RIDE!**: City Bikes
  - Bike Lanes
- **DOCK**: Return at any station
  - Settle
**Velib**

Velib, Launched on June 15th, 2007
Signature Initiative of B. Delanoe

- **1800 Stations**
- **224,000 Subscribers**
- **Station/300 M**
- **20,000 Bikes**

**Velib** is Rented every Second

- **19 B Calories**
- **137 K Tonnes of CO₂**

<table>
<thead>
<tr>
<th>City</th>
<th>Bikes per inhabitant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paris</td>
<td>1 vélo / 97 habitants</td>
</tr>
<tr>
<td>Lyon</td>
<td>1 vélo / 121 habitants</td>
</tr>
<tr>
<td>Hangzou</td>
<td>1 vélo / 145 habitants</td>
</tr>
<tr>
<td>Barcelone</td>
<td>1 vélo / 270 habitants</td>
</tr>
<tr>
<td>Montreal</td>
<td>1 vélo / 300 habitants</td>
</tr>
<tr>
<td>Londres</td>
<td>1 vélo / 984 habitants</td>
</tr>
<tr>
<td>New-York</td>
<td>1 vélo / 8 336 habitants</td>
</tr>
</tbody>
</table>

**20 MM Trips/yr, 173 MM Total Trips**
**VELIB: HOW TO IMPROVE RIDERSHIP?**

- **Add More Bikes**
  - How Many?
  - Where?

- **Add More Stations**
  - How Many?
  - Where?

- **Re-allocate/Trans-ship Bikes**
  - How much?

- **Station Network**
SYSTEM DESIGN - OPERATIONAL PERFORMANCE AND CUSTOMER BEHAVIOR

**Facility Location**

**Sizing, Inventory Management, Transshipment, Pricing (RM)**

**Accessibility: Effect of Distance**

**Availability: Effect of Bike Availability**

Immediate (Substitution, Lost Sales)

Long-term (Customer Choices)

**Challenges in Using High Frequency Data**

**Illustrate Use in Making Better Operational Choices**
DATA

**SYSTEM DATA**

2-min Snapshot, 946 Stations,

4 Months: May-Aug, 2013

*Trips Originating | Bikes Available*

*Which stations are stocked out?*

*Cleaning: Reallocation, Broken Bikes*

---

**CITY DATA (DENSITY MODEL)**

*Metro locations and incoming traffic*

*Top tourist locations, number of visitors*

*Google places data on stores, grocery and supermarkets, government buildings, hotels, museums, movie theaters, etc.*

*Hourly Weather data on Temperature, Rainfall, Humidity and Wind speed.*

*Census Data*

---

<table>
<thead>
<tr>
<th>Number of Snapshots</th>
<th>Raw Data</th>
<th>59,710,574</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Removing Weekends</td>
<td>42,753,524</td>
</tr>
<tr>
<td></td>
<td>Removing Trans-shipments</td>
<td>42,005,052</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of Trips</th>
<th>Raw Data</th>
<th>4,504,145</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Removing Weekends</td>
<td>3,365,183</td>
</tr>
<tr>
<td></td>
<td>Removing Trans-shipments</td>
<td>3,251,787</td>
</tr>
</tbody>
</table>

---

*Orders of Magnitude larger data than typical structural choice model – Precise, Estimation Challenge*
A Consumer-Level Structural Model

\[ U_{ift} = \beta_0 + h(\beta_d; d(L_i, L_f)) + \xi_{f,t} + \gamma_{district(f),w(t)} + \gamma_{weather(t)} + \epsilon_{ift} \]

User utility from using a bike at station \( f \) at time \( t \)

Mean user utility

Distance of user \( i \) to station \( f \)

Station-Time Residual

Fixed effects at district and hour Level

Weather fixed effects

Error term Type-1 extreme value distributed

\[ U_{i0t} = \xi_{w,0} + \epsilon_{i0t} \]

Outside Option

Weather: Temperature, Humidity, Rain

where \( m(t) \) indicates month and \( w(t) \) indicates time window of time \( t \).
FROM USER UTILITY TO DEMAND

Probability of a user $i$ choosing station $f$ among **Available** stations

$$p_{ift}(L_i) = \frac{\exp(E[U_{ift}])}{1 + \sum_{g \in N(i)} \exp(E[U_{gft}])}$$

where,

$$E[U_{ift}] = \beta_0 + \beta_1 \cdot h(\beta_d, d(L_i, L_f)) + \gamma_{\text{district}(f),w(t)} + \gamma_{\text{weather}(t)} + \xi_{ft}$$

Aggregating across users (Location)

$$\lambda_{ft} = \int_{L_i} p_{ift}(L_i)P_t(L_i) dL_i$$

Density Model

$$P_t(L_i) = \alpha_0 + \tilde{\alpha}_{1,w(t)} \cdot \tilde{V}_{w(t)}(l) + \alpha_2 \cdot \text{serv_lvl}_{w(t)}(l) + \alpha_3 \cdot pd_{di}(l)$$

*The two effects of Bike Availability*
Estimation & Endogeneity: An Instrument

Distance Effect

- Cross-Section Variation + Longitudinal Variation (Stockouts)
- Alternately use only longitudinal variation

Long-term Service Level Effect

- Cross-Sectional Variation (Station x Time Window)
- Endogeneity: Unobserved Static Station Factors, Unobserved Demand Shocks
- Instrument: Incoming Demand in previous time window
  - Affects Service Level but not outgoing demand (baseline demand, catchment area is captured through density).

\[
E[U_{ift}] = \beta_0 + \beta_1 \cdot h(d(L_i, L_f)) + \gamma_{\text{distance}}(f)w(t) + \gamma_{\text{weather}}(t) + \xi_{ft}
\]

\[
p_{ift}(L_i) = \frac{\exp(E[U_{ift}])}{1 + \sum_{g \in N(i)} \exp(E[U_{gft}])}
\]

\[
\lambda_{ft} = \int_{L_i} p_{ift}(L_i)P_t(L_i)\,dL_i
\]

\[
P_t(L_i) = \alpha_0 + \tilde{a}_{1,w(t)} \cdot \tilde{V}_w(t)(l) + \alpha_2 \cdot \text{serv_lvl}_w(t)(l) + \alpha_3 \cdot pd_{d1}(l)
\]
**Estimation Challenge: A Transformation**

![Diagram showing bike availability at different times]

- **From Time Domain to Stockout-State Domain**
  - What changes from one time to another?
    - Time-Window + Month + Weather Fixed Effects, Choice-Set
  - What if we combined all data-points with same
    - Time-window, Month, Weather and Choice Set
  - Number of Choice Sets less than Number of Times— but still substantial

- **Use Local Choice Set.**
  - Develop a procedure for consistency of local choice sets.
  - Computations go from Trillions to Millions ~ 50 Hours (950 Stations)

- **Symbols:**
  - Green diamond: Bikes Available
  - Red diamond: Bikes Not Available

- **Accounting intervals:**
  - 18:52, 18:54, 18:56, 18:58
### Results

<table>
<thead>
<tr>
<th>Primary variables</th>
<th>Density Variables</th>
<th>Weather Variables</th>
<th>Wald Test (p-value)</th>
<th>Number of in-stock observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bike-Availability</td>
<td>Walking Distance (0-50mts)</td>
<td>Walking Distance (50-100mts)</td>
<td>Walking Distance (&gt;100mts)</td>
<td>Yes</td>
</tr>
<tr>
<td>0.024 (0.004)***</td>
<td>-3.057 (6.443)</td>
<td>-13.214 (5.024)***</td>
<td>-7.535 (0.967)***</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Marginal Effects</th>
<th>% Increase in Demand</th>
</tr>
</thead>
<tbody>
<tr>
<td>10% increase in Bike-Availability (Short-term)</td>
<td>9.937% (9.921% - 9.952%)</td>
</tr>
<tr>
<td>10% increase in Bike-Availability (Long-term)</td>
<td>11.795% (11.326% - 12.291%)</td>
</tr>
<tr>
<td>10% decrease in Walking Distance</td>
<td>14.109% (12.836% - 15.056%)</td>
</tr>
</tbody>
</table>

10% bike-availability increases system-use by 15.30%-16.13%, (9.56% Short-term)

On Stockout, 4.4% of commuters switch to neighborhood stations

Scaling all distances by 10% from current levels results in 10.46%-14.72% average increase in system-use.
**INTERPRETING EFFECTS**

Scaling all distances by 10% from current levels results in 10.46%-14.72% average increase in system-use.

10% bike-availability increases system-use by 15.30%-16.13%, (9.56% Short-term)

On Stockout, 4.4% of commuters switch to neighborhood stations
USE CASE 1: HOW TO ACHIEVE TARGET RIDERSHIP (ISO-DEMAND CURVES)?

Calibrated Simulation Tool (London) : Alternate (Station Networks, Operational Performance) <-> Ridership

Ridership <-> Scaled Network, Operational Performance

Illustrated at the Aggregate Level, Analysis available at Quartier, Time-Window Level
USE CASE 2: HOW TO IMPROVE SYSTEM (WITH MONEY)?

Costs Vary by Quartier
Assuming average benefits, also available by Quartier
Use Case 2: How to Improve System (By Adding Stations or Reallocations)?

Costs Vary by Quartier
Assuming average benefits, also available by Quartier
USE CASE 3: IMPROVED STATION NETWORK DESIGNS (WITH NO MONEY*)

**GIVEN THE SAME # OF BIKES, SAME OPERATIONAL PERFORMANCE**

**POOL INTO LARGE STATIONS**
**HIGHER BIKE AVAILABILITY**

**POOLING → BIKE AVAILABILITY → USAGE**

**SMALL STATIONS**
**CLOSER DISTANCES**

**DISTANCE → USAGE**

---

**System-Use Change (%)**

**Average System-Use**

**95% Confidence Interval**

**Status Quo**

**Too Small**

**Too Far Apart**

**29.4% HIGHER RIDERSHIP (290K)**
**IF STATION DESIGN INCORPORATES ESTIMATES**

---

**Station Density**

**SOME RECENT WORK**

**Closer or More Reliable Stations?**


**Density V/s Scope?**

Kabra, A., K. Girotra and E. Belavina, “Designing Promotions to Scale Marketplaces”, 2017

**Incentives and Achieving Scale**


**Where to offer “Shuttles”?**
**Spatially Differentiated Marketplaces**

Suppliers are independent agents
- Decide when and how much to work

Differentiated in Location

On Demand Services – Service Levels

Not today’s focus
Passenger vs Driver Incentives?
Short term and Long term effect

Prizes v/s Discounts
Effect of Incentives

Incentivize Passengers

- Direct effect

Direct effect

Higher Usage
More Passengers
EFFECT OF INCENTIVES

Incentivize Passengers
- Direct effect
- Cross-externality effect

Higher Usage
More Passengers

Direct effect

Cross-externality effect

Higher Usage
More Drivers

Cross-externality effect
**Scale Economies: Density Effect**

- **Less Passengers and Drivers**
- **More Passengers and Drivers**

More likely to have an available driver nearby a passenger
Probability of match higher at larger scale

Density Effect due to more Users in Spatially Differentiated Marketplaces
Effect of Incentives

Incentivize Passengers
- Direct effect
- Cross-externality effect
- Density effect

Higher Usage
More Passengers

Density effect
Cross-externality effect

Direct effect

Higher Usage
More Drivers
**Whom to Give Incentives To?**

Different effectiveness when giving incentives to passengers or drivers

- **Incentivize Passengers**
  - Direct effect
  - Cross-externality effect
  - Density effect

- **Incentivize Drivers**
  - Direct effect
  - Cross-externality effect
  - Density effect

**Comparison:** > ? <

**Density effect**

**Cross-externality effect**
System in week $w$

\[ \sum_{i \in \text{Active Passengers}_w} \text{Requests}_{i,w} \]

\[ \sum_{j \in \text{Active Drivers}_w} \text{Driver Availability}_{j,w} \]

Passenger, Driver and a matching equation (interlinked through service levels)
Passenger incentives more effective in short-term and Driver incentives more effective in long-term
Threshold incentives could be more effective than linear incentives.

Incentive Cost per trip increase

- Linear Incentive: 17.17 $/trip
- Threshold Incentives: 14.039 $/trip
- 13.092 $/trip
Are there substantial economies of scale

Doubling of Number of Passengers and Drivers

\[ x \text{ Passengers} + y \text{ Drivers} = z \text{ Trips} \]
\[ 2x \text{ Passengers} + 2y \text{ Drivers} = 2.5z \text{ Trips} \]

Substantial gains from density effect and positive externalities.
**SOME RECENT WORK**

**Closer or More Reliable Stations?**


**Density V/s Scope?**


**Incentives and Achieving Scale**

Kabra, A., K. Girotra and E. Belavina, “Designing Promotions to Scale Marketplaces”, 2017

**Where to offer “Shuttles”?**
BEYOND TRANSPORTATION

Myntra

CONTACT

Karan@Girotra.com
Girotra@Cornell.edu

LinkedIn, Facebook