Decomposing Transit Demand: The effect of changes in competition on heterogeneous consumers

Yewon Kim   Olivia Natan

University of Chicago Booth School of Business

March 7, 2017

Outline

Background and Motivation

- The effects of new ride-hailing and ride-sharing services such as Uber and Lyft on driver welfare, rider welfare, and public transit use is of interest to policymakers and firms alike.
- Unlike the taxi industry, Uber and Lyft are able to price discriminate based on geography and demand conditions.
- Uber entered the Chicago market in 2011, and Lyft entered in 2013. However, the growth of users for these services was gradual. We propose studying a discrete change in their use as a result of a regulatory agreement.
- In late November 2015, the city of Chicago reached an agreement with both companies to allow them to operate at O'Hare and Midway airports. We intend to exploit this change in market structure to help identify the effects of competition on transit choices.

Research Question

How can we characterize demand for transit options from Chicago airports?

How does a change in market structure affect heterogeneous consumers' responses to prices?

Preview of Results

- Descriptive evidence supports the substitution of elastic demand away from taxi after rideshare service launch/ taxi fare increase
- Initial estimation results suggest that demand for taxis in the post-period has lower price sensitivity - suggesting selection

Model

To model demand for transit, we consider a discrete choice demand model across transit alternatives at Chicago Airports.
In building a discrete choice demand system, there are a number of effects we want to include when describing an individual's choice of transit options:

- Price - potential interaction with trip distance
- Time effects - day of week, month, (time of day)
- Weather
- Availability of alternatives

Describing the choice set

Individuals select an alternative to maximize indirect utility U_{ijdt}, where i is an individual, j is an alternative, d is an origin-destination pair, and t is time (here, day). Alternatives are selected from:

1. Taking the train
2. Taking a taxi
3. Taking an Uber (after November 2015)
4. Taking a Lyft (after November 2015)
5. Outside option (including driving, getting a ride, bus, and other)

The alternative set is conditional on airport. We take an individual's choice within a time period, and within an origin-destination pair. We don't directly observe destination for some options, so we collapse across destinations.

Utility Specification: Basic

Let U_{ijt} be the indirect utility consumer i gets from trip on alternative j on day t:

$$
\begin{aligned}
U_{idjt} = & \alpha_{jd} + \alpha_{r,j} R_t + \alpha_{w,j} W_t + \alpha_{month} M_t + \beta_{temp} F_t + \beta_{temp^2} F_t^2 \\
& + \beta_p \mathbb{E}[p_{djt}] + \beta_q \mathbb{E}[Q_{djt}] + \varepsilon_{ijt}
\end{aligned}
$$

- R_t and W_t: dummies for rain on day t and for a weekend or holiday
- M_t: month dummies
- F_t: average degrees on day t in Fahrenheit
- $\mathbb{E}[p_{djt}]$, $\mathbb{E}[Q_{djt}]$: expected price and expected travel time for a given origin-destination pair for an alternative on a day

We assume the utility shocks $\varepsilon_{idjt} \sim$ Type I extreme value i.i.d. across i, j, t, which are observed to consumers but not the econometrician.

Parameters of Interest

Our vector of parameters θ will be comprised of θ_{MDW} and θ_{ORD}. Within each airport's parameter vector, we estimate the following:

- α_j, the intercept for the alternative
- $\alpha_{r,j}$, the effect of rain on baseline utility
- $\alpha_{w,j}$, the effect of weekend or holiday on baseline utility
- α_{month}, the effect of each month on baseline utility over June
- β_{temp} and β_{temp^2}, the effects of temperature on utility
- $\beta_{p,j}$, the effect of price
- β_{time}, the effect of travel time

To solve the model given our data constraints, we use a normalization, where we let the baseline utilities of one option be set to 0. Since there are no price changes in train over the period, we set the utility of taking the train to be 0.

Solving the Model

Under these assumptions, we can represent the choice probabilities in the data as the typical multinomial logit expression (see McFadden 1973 for proof):

$$Pr(\text{i chooses j to go to d on t}) = \frac{exp(U_{idjt})}{\sum_k exp(U_{idkt})}$$

Since we have normalized one of the 2 alternatives in our data, this simplifies to a standard logit.

Data

- 3.7 million individual taxi trips data outbound from O'Hare and Midway airport from 2014 to 2016
- Daily CTA train ridership statistics from 2014 to 2016:
  Blue Line (O'Hare), Orange Line (Midway)
- Daily weather records

- Our goal is to construct a binary choice set (Taxi, Train) for each taxi taker and train taker given the data
- For each traveler, given his destination, we need to calculate estimated fare and travel time for both transit modes

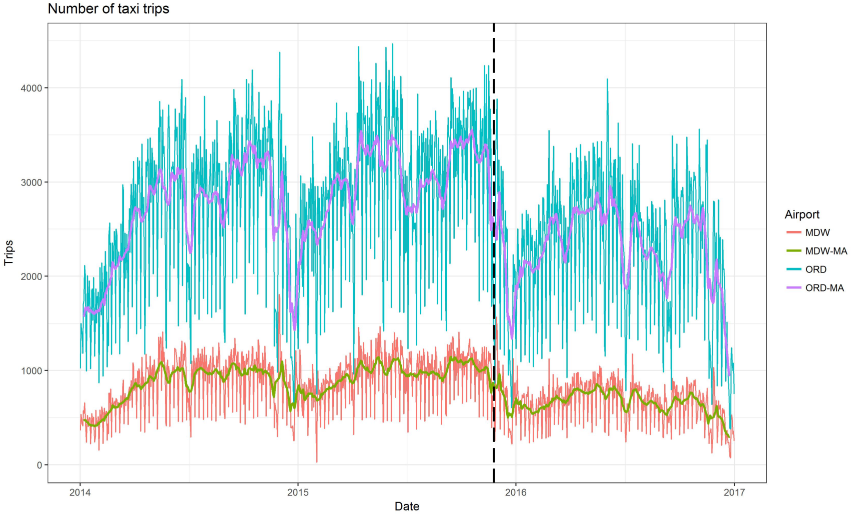
Data - Descriptive statistics

- Our hypothesis is that rideshare launch and taxi fare increase affected only those with elastic demand and made them substitute away from taxi, letting people with inelastic demand unaffected

- Elasticity of demand for taxi is determined by
  - Price sensitivity: Taxi fare
  - Sensitivity to non-monetary travel cost
    : Weather condition, Travel time, ...
  - Information availability
    : Information about public transit system, Access to Uber/Lyft, ...

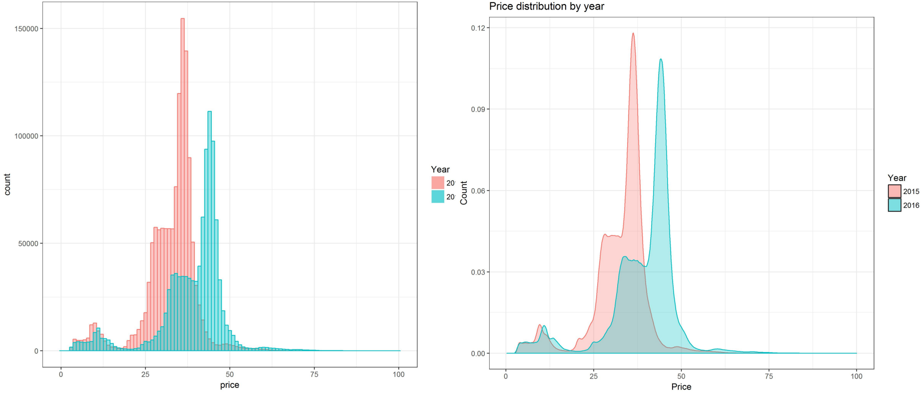
Data - Descriptive statistics

Demand for taxi has decreased after Jan 2016



Data - Descriptive statistics

Figure 1: Distribution of Fare

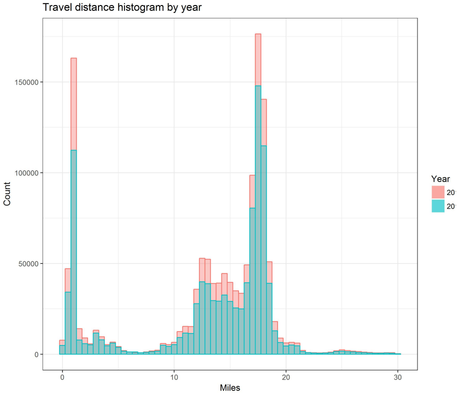


(a) Price histogram

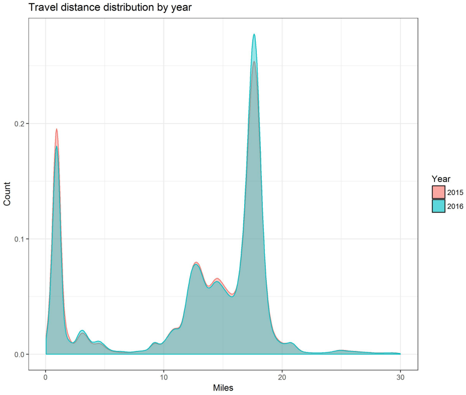
(b) Price density

Data - Descriptive statistics

Figure 3: Distribution of Trip Miles



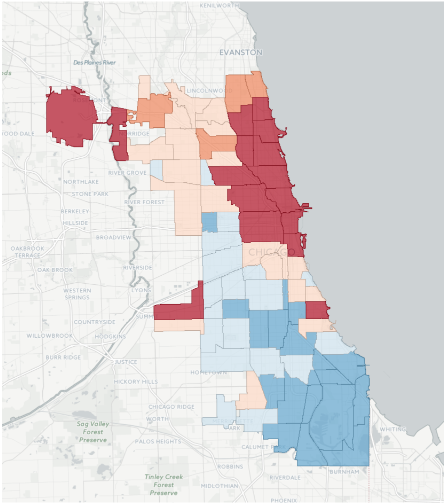
(a) Trip miles histogram



(b) Trip miles density

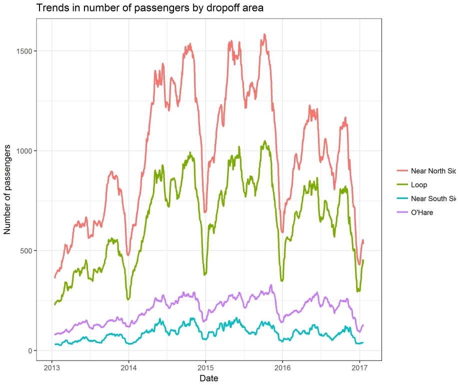
Data - Descriptive statistics

Figure 5: Heatmap by Drop-off Destinations

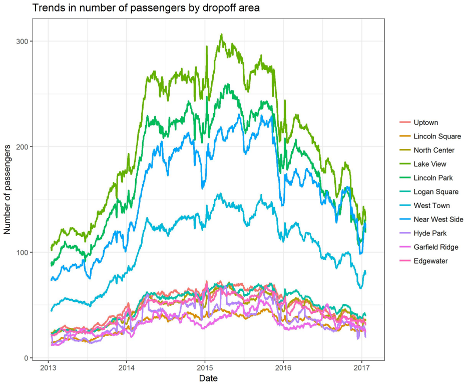


Data - Descriptive statistics

Figure 6: Trends in Taxi Demand by Destination



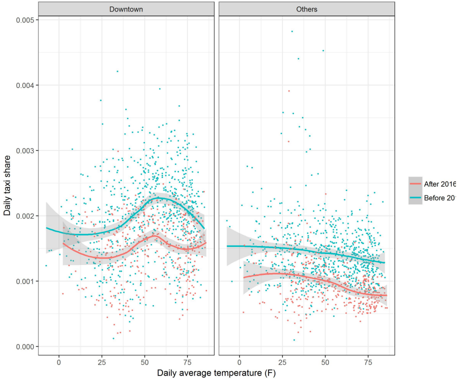
(a) Downtown



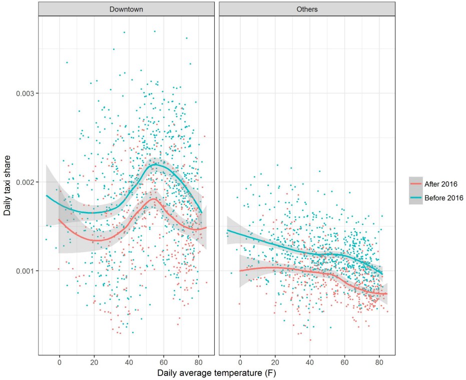
(b) Local

Data - Descriptive statistics

Figure 8: Estimated Daily Taxi Share and Daily Average Temperature



(a) Midway



(b) O'Hare

Data

Taxi trips: Variables

We have the following information for each taxi trip made:

- Pick-up/Drop-off location
  (Community area, geocode)
- Time stamp (date + hour)
- Total fare, Travel distance, Travel time

CTA ridership: Variables

We have the following information for daily ridership:

- Number of daily entries in O'Hare / Midway station
- No information about exits
  - No information about train takers' destinations
  - Becomes an issue because we need estimated taxi fare and estimated taxi travel time for train takers as well

Data - Taxi Takers

Variables for taxi alternative

- Estimated fare, distance, and travel time
  : Use daily average values given drop-off community area

Variables for train alternative

- Estimated fare
  : Use a constant train fare across destinations
- Estimated travel time
  : Use Google Map API to retrieve estimated public transit time

Data - Train Takers

Variables for taxi alternative

- Estimated fare
  : Use daily mean taxi fare (aggregated across destinations)
- Estimated travel time, distance
  : Should think about how to construct these variables

Variables for train alternative

- Estimated fare
  : Use a constant train fare across destinations
- Estimated travel time
  : Use daily mean public transit time from taxi takers' data

- Selection issues
- Search for more data

Estimation

Based on our previous assumptions, we can write the likelihood for a given individual's choice as:

$$\mathcal{L}_i(\theta) = \sum_{j=1}^{J} \mathbb{1}\{i \textit{ chose } j\} \frac{exp(U_{idjt})}{\sum_k exp(U_{idkt})}$$

The log likelihood is $\mathcal{L}\mathcal{L}(\theta) = \sum_{i=1}^{N} \mathcal{L}_i(\theta)$. We can use Maximum Likelihood to take this to the data.
We currently present the results for Midway data only.

Alternative Estimation Strategy - GMM

- Match choice probability (model moments) with market share data (data moments)
- Model moment:

$$Pr(\text{i chooses j} \mid \text{going to d on t}) = \frac{exp(U_{idjt})}{\sum_k exp(U_{idkt})}$$

- Data moment:

$$\frac{\text{N taxi rides}_t}{\text{N airport passengers}_t} = \prod_d \frac{\text{N taxi rides}_t | d_t}{\text{N passengers}_t | \text{d}} \cdot \frac{\text{N passengers}_t | d_t}{\text{N airport passengers}_t}$$

- Issue: data

Results: Naive Model

We begin by estimating a naive version of our model that omits price, month, and quality effects, supposing that baseline variation in utility and daily effects of weather and the day of the week are the only inputs to choice. We estimate all models only on Midway data.

Table 1: Estimation Results - Simplest Model, Midway

	Before Entry	After Entry
α_{dj}	-2.495757 (0.03752517)	-2.057149 (0.09931749)
α_{rj}	0.05104191 (0.003754602)	-0.02685534 (0.009124809)
α_{wj}	0.5005753 (0.003770825)	0.2655912 (0.007099897)
β_{temp}	0.008955185 (0.00178695)	0.002576567 (0.004106022)
β_{temp^2}	-0.0000785373 (0.00001466916)	-0.0001124353 (0.00002983485)
N	6,019,303	2,610,539
LL	-2,006,312	-801,444.5

Results: Naive Model with Month FEs

Table 2: Estimation Results - Simplest Model with Month FEs, Midway

	Before Entry	After Entry
α_j	-2.495828 (0.010347889)	-2.259852 (0.018167753)
α_{rj}	0.05095865 (0.002912903)	0.009464241 (0.004775116)
α_{wj}	0.5005634 (0.003107043)	0.2679429 (0.005588048)
β_{temp}	0.005616059 (NA)	-0.0004760779 (NA)
β_{temp^2}	-0.00004672049 (NA)	-0.00002300726 (NA)
N	6,019,303	2,610,539
LL	-2,020,383	-730,064.2

Results: Model, no FEs

Table 3: Estimation Results - Model with Price, Midway

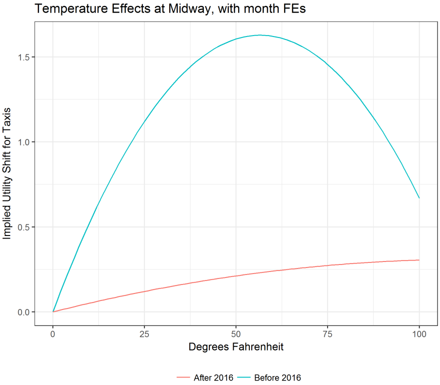
	Before Entry	After Entry
α_j	-2.7813129731 (0.031966309)	-2.489877 (NA)
α_{rj}	0.0343391815 (0.003367665)	0.04976002 (NA)
α_{wj}	0.3339328717 (0.003444115)	0.4963456 (NA)
β_{temp}	0.0132447591 (0.001354935)	0.001020395 (NA)
β_{temp^2}	-0.0001175519 (0.00001106557)	-0.00009932348 (NA)
β_p	0.0064451638 (0.0005918582)	0.009003366 (0.0005872499)
N	6,019,303	2,610,539
LL	-2,004,766	-802,544.9

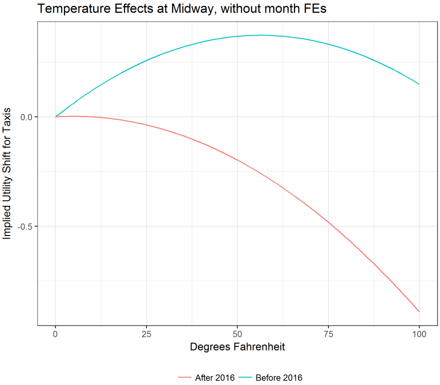
Results: Model with Month FEs

Table 4: Estimation Results - Model with Price and Month FEs, Midway

	Before Entry	After Entry
α_{dj}	-2.4884911605 (0.0223768077)	-2.490251086 (0.0384704269)
α_{rj}	0.0517281008 (0.0029604627)	0.050982751 (0.0045663254)
α_{wj}	0.4976793493 (0.0031339917)	0.496868292 (0.0051416978)
β_{temp}	0.0575152979 (NA)	0.005424446 (NA)
β_{temp^2}	-0.0005082469 (NA)	-0.00002374716 (NA)
β_p	-0.0295328034 (0.0005690839)	-0.001877242 (0.0007774288)
N	6,019,303	2,610,539
LL	-2,059,348	-830,945

Temperature Effects - Fitted Values of Coefficients





Interpreting results

- Caution - many local minima!
- Combination of MLE results and reduced form evidence suggests that variation across the year is driven by different mixes of consumers, but weather plays some role
- Month FEs don't appear to be particularly stable across specifications (not reported here for brevity).
- Strong evidence of selection out of the data into other modes - is it driven by price change or Uber entry?

Next Steps for Estimation

- Estimate model including quality measures - strong calibration of travel time variables
- Continue to search for stable starting values for O'Hare estimation
- Pursue alternative estimation strategy using aggregated moments
- Re-estimate model using Uber data if possible

Questions and comments?