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

Yewon Kim† Olivia Natan‡

March 2017

(version 17.03.a)

Abstract

This paper considers how demand for transit from Chicago airports changes over time and across consumers. In particular, it identifies a combined effect of two experiments - the entry of new transit options and an increase in the price of taxi metered fare - on demand for taxis. We document how demand, as characterized by a discrete choice demand model, changes for taxis before and after the policy changes using trip-level taxi data and day-level public transit data from the City of Chicago. We find that consumers who remain in the market for taxis after the policy changes have less elastic demand (less price-sensitive and less sensitive to weather). Furthermore, we present descriptive evidence that shows a great amount of heterogeneity across consumers headed to different destinations, which implies that the heterogeneous effects of the policy changes can be more fully recovered using destination information.

keywords: Demand estimation, transportation economics, consumer heterogeneity.

*You can put thanks here, affiliation here, anything you want here.

†University of Chicago Booth School of Business, Department of Marketing, 5807 S Woodlawn Avenue, Chicago, IL 60637, yewonkim@chicagobooth.edu.

‡University of Chicago Booth School of Business, Department of Marketing, 5807 S Woodlawn Avenue, Chicago, IL 60637, olivia.natan@chicagobooth.edu.

1 Introduction

The effects of new ride-hailing and ride-sharing services such as Uber and Lyft on rider welfare and public transit use is of interest to policymakers and firms alike. Although there are rich documentations of taxi demand in transportation and economics literature (Gonzales, Yang, Morgul, and Ozbay 2014; Kamga, Yazici, and Singhal 2013; Qian and Ukkusuri 2015), the impact of ride-sharing service as a new transit option on demand for taxis has been less cleanly identified due to the gradual growth of users for these services. Furthermore, since the market for transit modes faces discrete changes in fare from time to time, understanding how consumers react to a change in price (here, taxi fare) given an entry of close substitutes (here, ride-sharing service) is important for policymakers to predict welfare changes in such repeated settings.

In this paper, we identify a combined effect of two discrete policy changes - the entry of rideshare services and an increase in taxi fare - on demand for taxis. On November 25, 2015, the city of Chicago reached an agreement with Uber and Lyft to allow them to operate at O'Hare and Midway airports, which happened a few years after they started operating in Chicago area.[1] Also, a 15% increase in taxi metered fare went into effect in January 2016. By looking at how transit choices for outbound trips from the two airports have changed after these policy changes, we intend to identify the effects of competition and price increase on transit choices.

With descriptive evidence and model estimates, we argue that the entry of ride-sharing services and taxi fare increase only affected consumers with elastic demand to substitute away from taxis, which resulted in a smaller market with higher ratio of consumers with inelastic demand. To understand this effect more deeply, we propose three determinants of demand elasticity for taxis - 1) sensitivity to price, 2) sensitivity to non-monetary travel cost (e.g., bad weather condition, travel time), 3) information availability (e.g., familiarity with public transit system, availability of Uber/Lyft applications) - and explain how the average level of these determinants of consumer demand elasticity have changed before and after policy changes. Both model estimates and descriptive statistics show that the average price sensitivity has decreased (in absolute value) after the two events, suggesting that highly price-sensitive consumers have substituted away from taxis. Furthermore, analysis of taxi trips based on their destinations (downtown areas versus other local areas) shows that there is a systematic difference between visitors' and local people's demand for taxis, which can be explained by differences in information availability or other sensitivity measures between these two groups of people. Overall, our results suggest that 1) the policy changes have lowered the demand for taxis significantly, and 2) there is a great amount of heterogeneity in

[1] Uber entered the Chicago market in 2011, and Lyft entered in 2013.

the effects of policy changes on consumers' transit choice based on their demand elasticities, which can be partially recovered by information on trip destinations.

Our paper proceeds as follows. In section 2, we describe the data sources being used, and we go through summary statistics. In section 3, we lay out the specification for our discrete choice demand model. Section 4 discusses estimation strategy and presents results, and we then discuss limitations and next steps for our approach.

2 Data

We use taxi and train ridership data outbound from the two airports in Chicago, O'Hare(ORD) and Midway(MDW), from 2014 to 2016. There are two major events that might have affected the airport passengers' demand for transit modes within this data window: 1) Uber and Lyft launched their rideshare service in each airport's pickup area in November 2015, and 2) there was a 15% increase in taxi fare in the City of Chicago, which became effective in January 2016. Because of the temporal proximity of the two events, here we do not separately identify the effect of rideshare service and taxi fare increase on the demand for taxi.[2]

The data we use include individual taxi trips data outbound from the airports, Chicago Transit Authority (CTA) ridership statistics, and daily weather information. More detailed information on each data set is in Appendix A-1.

2.1 Taxi Trips

We use about 3.7 million individual taxi trips data outbound from the two airports [3] from 2014 to 2016.

Figure 1 and Figure 2 show the change in number of taxi trips and the estimated market share of taxi from the two airports before and after the launch of rideshare service and taxi fare increase. Although the passenger volume coming through each airport has increased over time (Figure 12 in Appendix), the number of taxi trips outbound from the airports has decreased after the events, which is reflected in the estimated share (Figure 2) computed with daily passenger volume in each airport as denominator.

We argue that the launch of rideshare service and the increase in taxi fare only make people with elastic demand substitute away from taxi while keeping those with inelastic

[2] Since Uber started its service in Chicago area (excluding airport pickup service) in 2013, we expect to separately identify these two effects by comparing the demand for taxi trips outbound from the airports and the demand for other local taxi trips before and after 2016.

[3] Less than 2% of the total taxi trip data has been dropped from our data set due to the reporting error of either total fare or trip miles.

demand relatively unaffected. Elasticity of demand for taxi can be determined by multiple factors, including the following:

- Sensitivity to price,
- Sensitivity to non-monetary cost (e.g., weather conditions, travel time),
- Information availability (e.g., familiarity with pubic transit system, availability of Uber/Lyft app).

There are a number of factors that can affect these different categories of sensitivity. For example, expected taxi fare based on travel distance, bad weather condition including extreme temperature or rain, being a visitor or a local resident are candidates that might influence each of the three types of sensitivity respectively. In this section, we present detailed descriptive evidence that supports our claim.

Figure 3 and Figure 4 show how price and travel distance distribution of taxi trips has shifted over time. Shift in price distribution before and after 2016 is stark due to the 15% price increase. Decrease in the amount of taxi trips after the rideshare service launch and the taxi fare increase is even more visible in the histogram. As shown in the density plots, aggregate taxi demand after the two events has almost same travel distance distribution as before, except that the travel distance distribution after January 2016 has less mass in its left tail (less than 5 miles) and more mass in its peak, which implies that different levels of substitution might have occurred across different types of consumers based on their travel distance.

Major portion of taxi trips outbound from the airports are to downtown areas (Figure 5). Trips to downtown and trips to other local areas show systematic difference (Figure 6), which implies the significant effect of consumer demographics (mainly the effect of visitors-local residents ratio) on aggregate taxi demand. As expected, volume of taxi trips to downtown demonstrates much stronger seasonal pattern due to a high proportion of visitors than of locals. Therefore, in the seasons with high visitor volume (from April to September), the aggregate demand for taxi would be highly driven by visitors with relatively inelastic demand, whereas in the seasons with low visitor volume (from October to March), the demand would be mainly driven by local residents with more elastic demand.

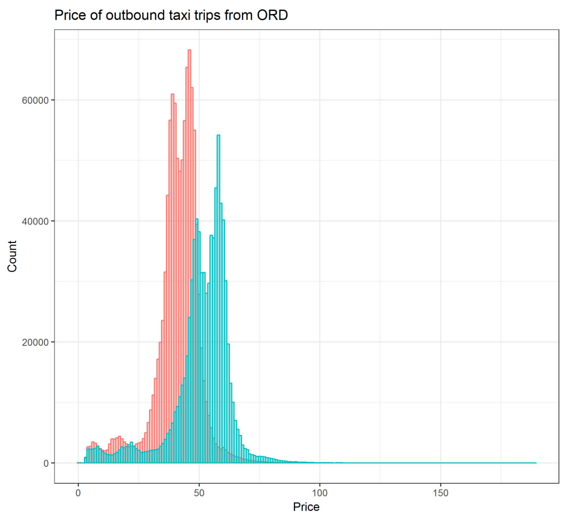
Figure 7 gives further descriptive evidence of the main (negative) effect of taxi fare increase and rideshare service launch on aggregate taxi demand. More importantly, they also indicate that the size of the effect varies across groups with different demand elasticity.

Figure 1: Number of Taxi Trips

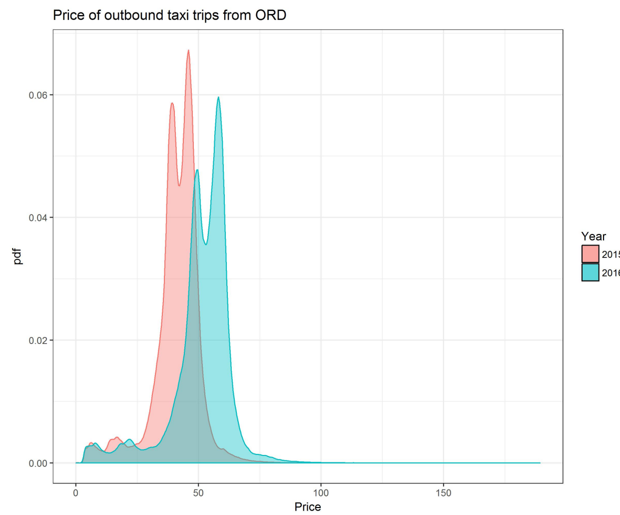
Figure 2: Estimated Market Share of Taxi

Figure 3: Distribution of Fare and Trip Miles, Outbound Trips from O'Hare

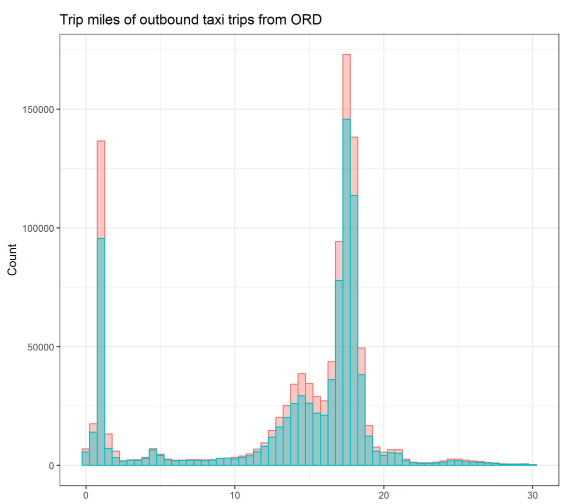
(a) Price histogram



(b) Price density



(c) Trip miles histogram



(d) Trip miles density

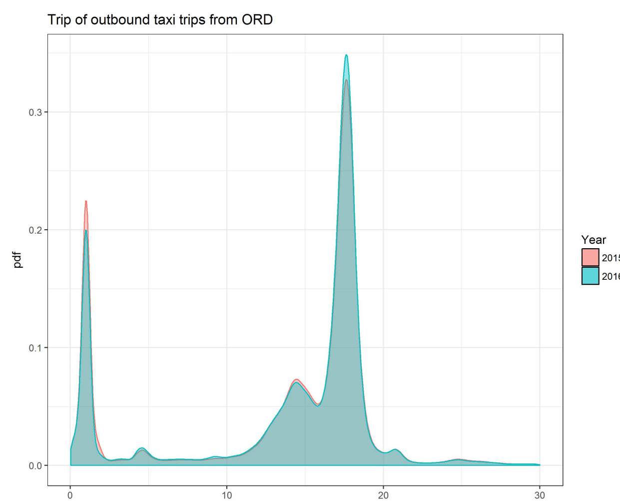
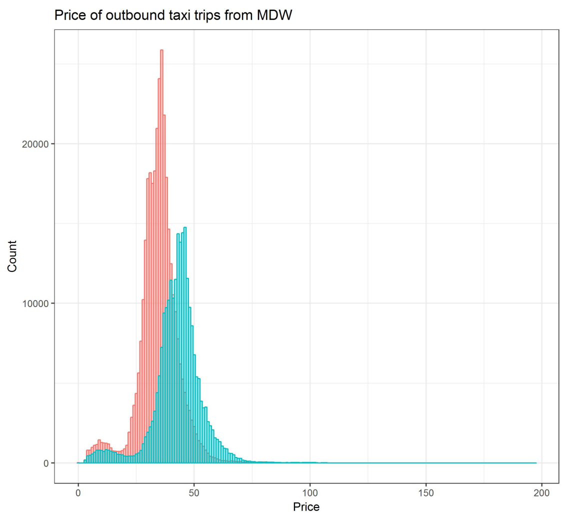
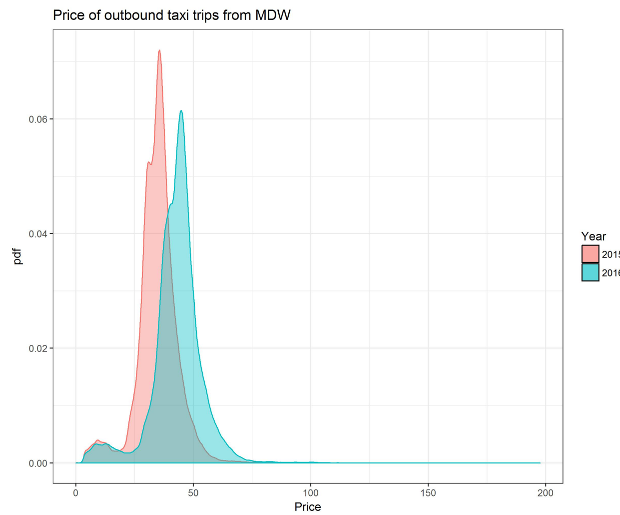


Figure 4: Distribution of Fare and Trip Miles, Outbound Trips from Midway

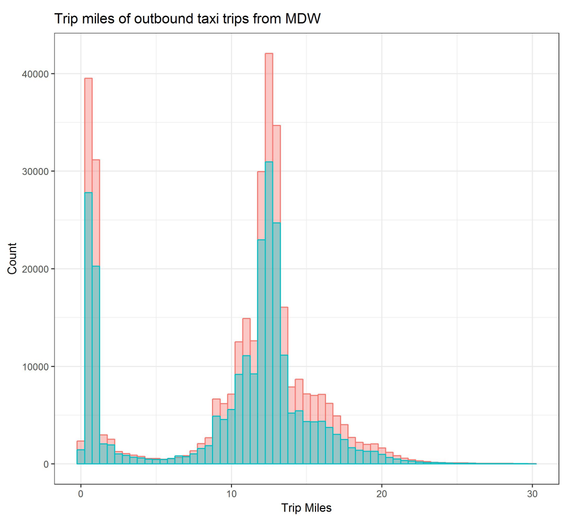
(a) Price histogram



(b) Price density



(c) Trip miles histogram



(d) Trip miles density

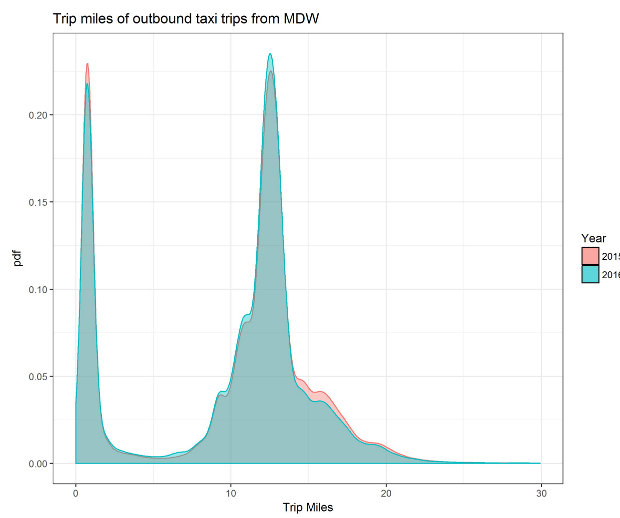


Figure 5: Destination of Taxi Trips Outbound from the Airports

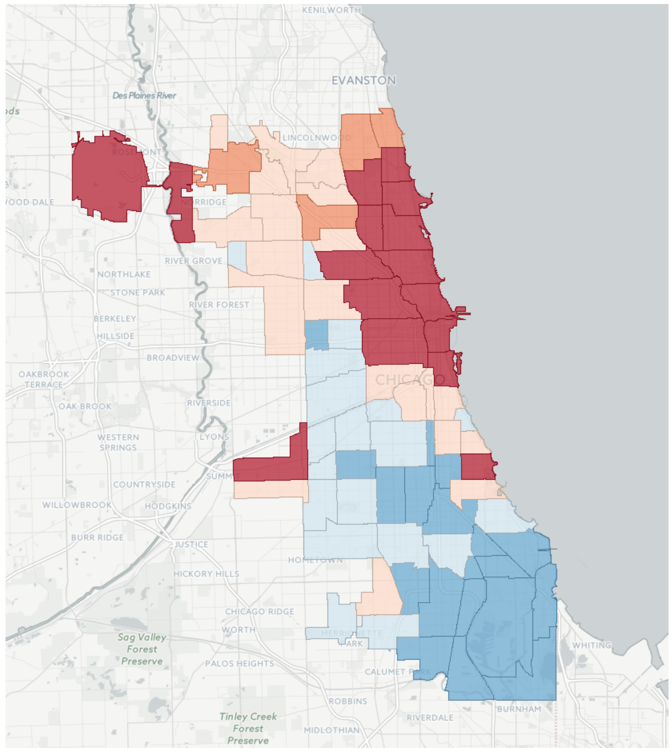
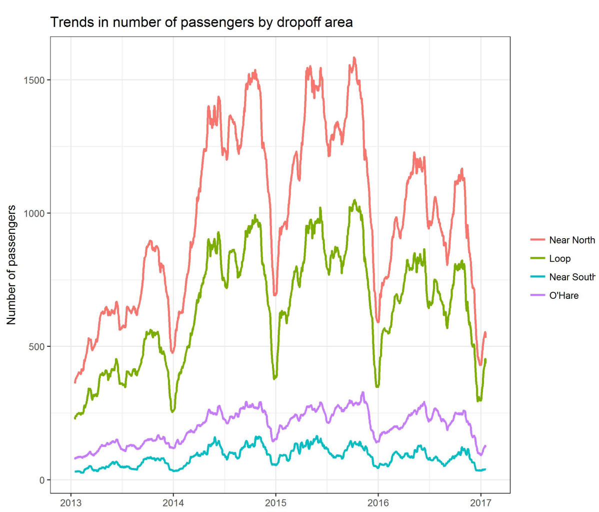
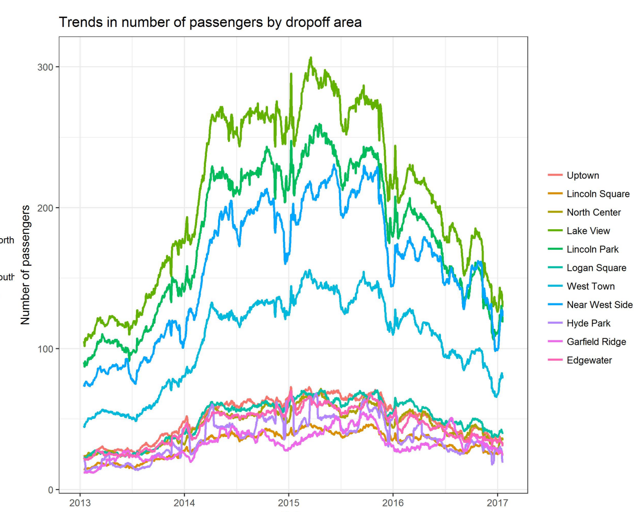


Figure 6: Comparison of Taxi Ridership Trends by Destination

(a) Visitor Destination



(b) Local Destination



One can expect that the relationship between taxi trip and temperature would be U-shaped, as people would have higher disutility from taking public transit when it is either too hot or too cold. Furthermore, as the demand for taxi would become more inelastic in the extreme weather, the taxi share in the extreme weather condition would have been less affected by the taxi fare increase, and only the demand in the mild weather condition that is relatively elastic might have substituted away from taxi after the policy changes.

At first glance, the taxi trips to downtown (panels on the left side in (a) and (b)) do not seem to support these arguments since they show an inverse-V pattern with their peak lying between 50 and 60F°. However, we argue that this potentially reflects a high portion of visitors with inelastic demand during the time period with such temperatures. The peak in Figure 7 corresponds to the temperature range in March to April or September to October, the time periods in which Chicago has the highest volume of visitors during a year (panel (a) in Figure 6). This suggests that the peak in Figure 7 implies the high number of visitors with inelastic demand who have either low price-sensitivity or low information availability and thus take taxis regardless of the weather conditions. Outside the temperature range around the peak, taxi share shows an expected U-shape in the extreme temperature. Morever, the gap between the market shares before and after the taxi fare increase is greater for mild temperatures, which is consistent with our hypothesis that elastic demand under the mild weather might have been affected more by the policy changes and thus substituted away from taxis. Trips to other local areas (panels on the right side in (a) and (b)) show a very different pattern; the estimated share almost linearly increases as the temperature drops.

Figure 7: Estimated Taxi Share by Daily Average Temperature after Taxi Fare Increase

(a) O'Hare **(b)** Midway

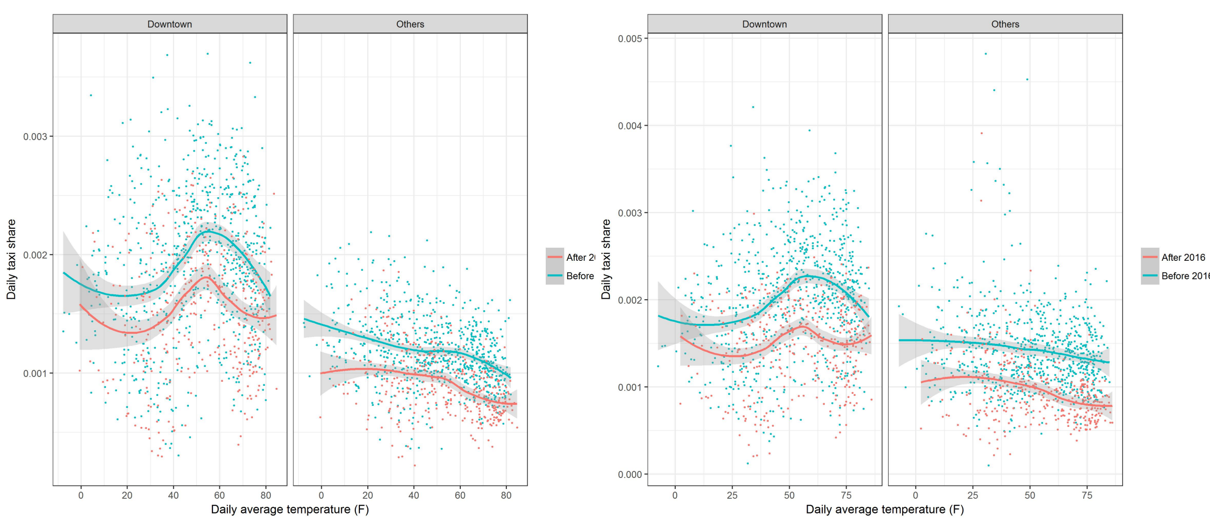
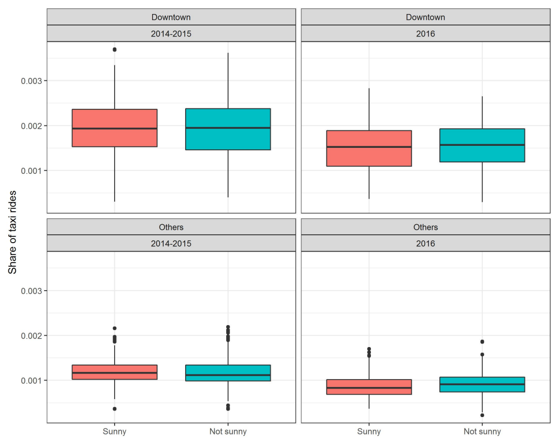


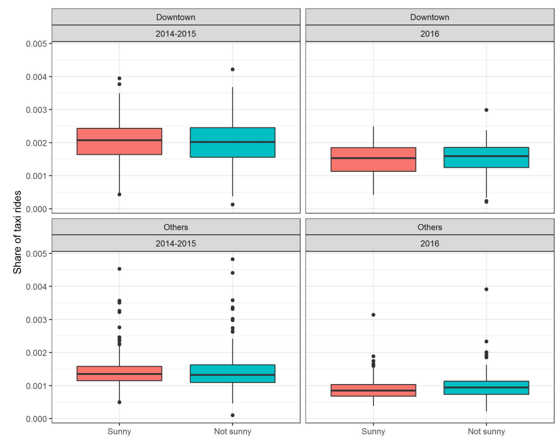
Figure 8 also adds subtle but favorable evidence to our claim. As explained above, we expect more decrease in taxi share in sunny days than in rainy or snowy days after the policy changes, which is consistent with what Figure 8 shows.

Figure 8: Estimated Taxi Share by Weather after Taxi Fare Increase

(a) O'Hare



(b) Midway



2.2 CTA Ridership

We use daily entries to the two airport stations (O'Hare station in Blue Line and Midway station in Orange Line). Figure 9 shows that the demand for CTA is much less affected by the two events. Insignificant change in the demand for CTA after the taxi fare increase can also be seen in Figure 10. Panel (b) in Figure 10 shows that the demand for public transit is lower in extreme temperatures, although the pattern is not significantly distinct.

2.3 Constructing Data Set for Estimation

Our goal is to construct a binary choice set (Taxi, Train) for each taxi taker and train taker given the data. To do so, we need to calculate estimated fare and travel time for both transit modes for each traveler given his destination. The data issue that prevents us from having this information is that the CTA ridership statistics only show the number of daily entries to O'Hare station and to Midway station, without any information about where these train passengers exit. Since we do not know the destinations of train takers, we use daily mean taxi fare (aggregated across all destinations) and daily mean taxi travel time (aggregated across all destinations) computed with taxi trips data as train takers' estimated fare and

Figure 9: Estimated Market Share of CTA

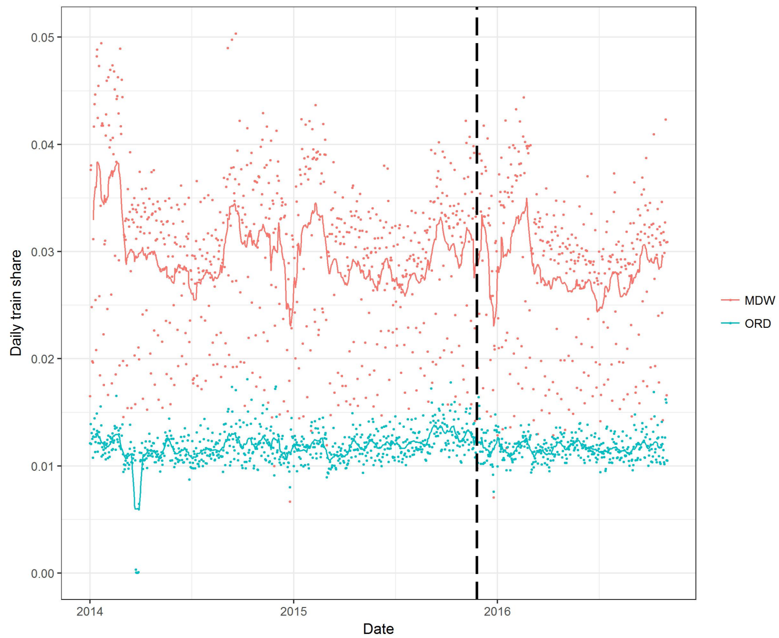
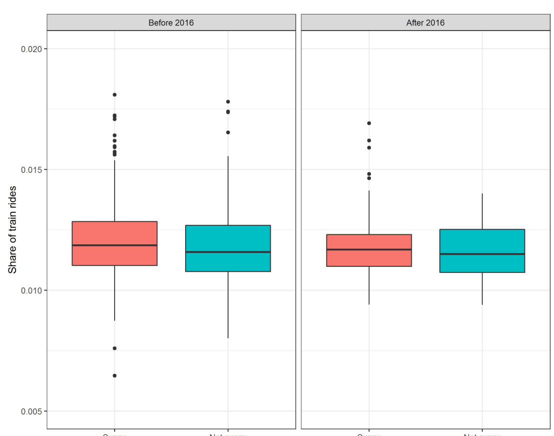


Figure 10: Estimated Market Share of CTA by Weather

(a) Weather



(b) Temperature

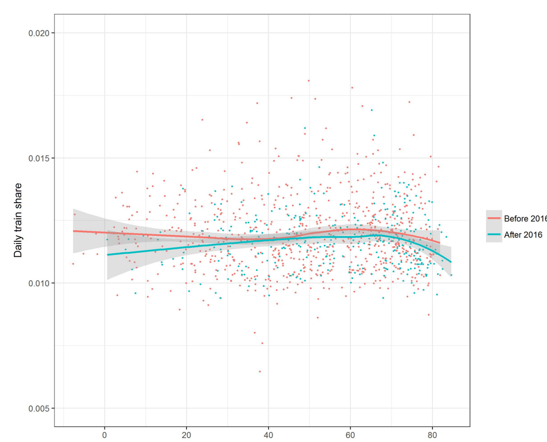


Table 1: Construction of Variables

	Alternative	Time Unit	Fare	Travel Time
Taxi takers	Taxi	Day	Daily averages given drop-off community area	
	Train	Day	Constant	Estimated travel time from Google Maps
Train takers	Taxi	Day	Daily averages from taxi trips data	
	Train	Day	Constant	Daily average train travel time from taxi data

travel time for taxi. This approach has a number of limitations including that it ignores selection issues (that the distribution of destinations, and thus of estimated fares and travel times, for taxi takers would be systematically different from those for train takers), but given the data set we find it as the best approach. Table 1 summarizes how the data set is constructed.

3 Model

Demand for transit options is posed here as a discrete choice demand model, following from a tradition of developing discrete choice demand for the express use of understanding transit choice (see Train 2009, Ben-Akiva and Lerman 1985). For a survey, see Chintagnta and Nair (2011). Consumers arriving at Chicago's O'Hare and Midway airports make a decision to use a single mode of transport to reach their actual destination (e.g. their home, a hotel, or other) when they arrive. We enumerate these choices as follows, conditional on airport arrival:

i. Taking the train (baseline)

ii. Taking a taxi

iii. Taking an Uber or Lyft (after Nov. 25, 2015)

iv. Other (Taking the bus, getting a ride, hotel shuttle, driving oneself)

We omit the last, unobserved option from our model. Getting a ride or driving is not a close substitute for taking the train or a taxi when arriving at the airport, as it requires planning well in advance. Bus ridership is very low at Midway Airport and only serves a small geographic area in the city, and no CTA buses serve O'Hare.[4] For these reasons, we primarily consider alternatives 1 through 3.

[4] For CTA route information, we access the CTA site at www.transitchicago.com. Maps and station guides indicate no bus availability at O'Hare.

Since we have a discrete point of entry for the ridesharing services, there are 2 options for constructing a model to understand demand before and after entry. We choose to use the same model in the pre- and post- periods on train and taxi data, as opposed to model choice among all the alternatives in the post period. This decision is driven by current data limitations.

Consumers choose an option to maximize expected utility U_{idjt}, where i is an individual, j is an alternative, and d is a destination. We model consumers' indirect expected utility as a function of the following product features. Each alternative has a baseline utility level α_{jd} for each destination (e.g. the baseline utility of a taxi between O'Hare and the Loop). In most of the models we estimate, we will restrict this baseline utility to be uniform across individuals (and thus across destinations).

Existing literature on taxi demand (see, for example, Kamga et. al. 2013) cites weather as a major shifter of demand. We include 2 sources of weather variation in the demand model. First, we add a shifter to the baseline utility for days when there is any rain. All else equal, we expect demand for taxi or Uber over the train to increase when it is raining. Second, we add average daily temperature (linearly and quadratically). We want to allow for the utility of a taxi or Uber over train to vary nonlinearly over the range of typical temperatures - intuitively, this should allow the train alternative to seem less attractive in very cold and very hot temperatures.

We include 2 time-determined fixed effects. Due to slightly different schedules and different traffic, we include a fixed utility shifter for weekend and holidays versus weekdays. As we will demonstrate in the Data section, the overall market size varies significantly over the course of the year due to fluctuations in the rates of travel of both locals and of incoming tourists or business travelers. We summarize this variation in the composition of travelers through fixed monthly shifters of the baseline utility for non-train options.

The final components of utility are price and product quality. Price enters linearly into the utility. Since fares for taxi and ride-hailing services vary based on the realized trip, we use an estimate of expected price for the trip. Quality of each alternative isn't directly observed and varies significantly across instances of the different alternatives even within the same day (e.g. a taxi driver could have varying levels of familiarity with the best routes, a train could experience delays). We will operationalize this as a proxy for convenience - expected travel time. In general, taxi and Uber will be much faster (except during heavy traffic) to reach many destinations in Chicago.

We specify the indirect utility as follows:

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

R_t is a dummy for rain on day t. W_t is a dummy for the day being a Saturday, Sunday, or holiday. M_t is a set of 11 dummies for each month (minus one, for identification). F_t is the average degrees on day t in Fahrenheit. $\mathbb{E}[p_{djt}]$ is expected price for a given origin-destination pair for an alternative on a given day. $\mathbb{E}[Q_{djt}]$ is the expected travel time for a given origin-destination pair for an alternative on a given day. We assume the utility shocks ε_{idjt}, which are observed to consumers but not the econometrician, are distributed Type I extreme value independently across i, j, t. Given these assumptions, the probability of a consumer choosing an alternative is expressed as a multinomial logit:

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

With 2 alternatives, we can simplify this choice to a standard logit. For a proof, see McFadden (1973).

This assumption about the nature of the random utility shocks is not without loss of generality, and it is not costless. Our parameterization of the problem may be wrong for a number of reasons. If the utility specification is wrong, estimation will be inconsistent. If our specification of the error process is wrong, then estimation will only recover the pseudo-true values.

Crucial to our model is the introduction of a new alternative after November 2015. Multinomial logit suffers from IIA - the independence of irrelevant alternatives. This suggests that introducing a new alternative will proportionally affect those who in the past chose each existing alternative. To address this, we can introduce a flexible form of heterogeneity to our utility specification. Suppose there exist S discrete segments of consumers with different preferences. We specify the choice probabilities as a discrete mixture over these classes with fixed probabilities across individuals:

$$U_{idjt|i\in s} = \alpha_{jd,s} + \alpha_{r,j,s}R_t + \alpha_{w,j,s}W_t + \alpha_{month,s}M_t + \beta_{temp,s}F_t + \beta_{temp^2,s}F_t^2 + \beta_{p,s}\mathbb{E}[p_{djt}] + \beta_{q,s}\mathbb{E}[Q_{djt}] + \varepsilon_{ijt}$$

This yields conditional choice probability within a segment:

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

This allows us to write down the joint likelihood of the data with respect to the parameters. The probability of segment memberships is π_s, which is bounded between 0 and 1 for all segments, and the sum of all probabilities is 1. Based on our previous assumptions, we can write the likelihood for a given individual's choice as:

$$\mathcal{L}_i(\theta) = \sum_{s=1}^{S} \pi_s L_i^s(\theta^s)$$

Where we have

$$L_i^s(\theta^s) = \frac{exp(U_{ijt}(\theta^s))}{\sum_k exp(U_{ikt}(\theta^s))}$$

and $\sum_{s=1}^{S} \pi_s = 1$. The log likelihood is $\mathcal{LL}(\theta) = \sum_{i=1}^{N} \mathcal{L}_i(\theta)$. There is one additional parameter that must be calibrated - the number of segments S. Additionally, to solve the model, we normalize the intercept of one alternative to be 0, which in our case is that of the train option. In fact, due to no price variation, we normalize the utility of the train option to be 0 in all versions of the model except for the travel time term. As such, the parameters we identify are relative effects for taxi over train.

4 Estimation

We estimate the model on the pre- and post-periods separately using Maximum Likelihood Estimation. To speed up computation, we can take the shortcut that the choice probabilities for each train rider on a given day are the same in our no-heterogeneous-effects model. For each day, we compute only one choice probability and multiply it by the number of riders that day, a shortcut enabled by the i.i.d. assumption. Minimization is performed using R's nonlinear minimization routine. As of the current version, we only include the specification without travel time.

We briefly describe the creation of each variable to match the model. Choice is constructed directly from the data source - if from taxi data, then the consumer chose taxi, else they chose train. The rain dummy is taken directly from the weather data (see Appendix A-1.5). Whether the day is a weekend or a holiday is pulled from the CTA day classification (see Appendix A-1.6). Month data is pulled from the date variable in both choice datasets. The temperature variable is the average daily temperature recorded at each airport's ground weather station, as described in Appendix A-1.5.

Constructing expected price is a little more in-depth. For taxi riders, we saw previously that the dropoff location is a major determinant of price (since fares are metered based on distance and idling time). The simplest estimator, then, is some sample mean over origin-destination drop-off pairs. We use daily price averaged by origin-destination pair to construct the expectation. For train riders, however, we can't observe destination, so we use daily mean across all destinations. This is problematic for a number of reasons, but it is the best estimator given the data. We might think that on average there is selection into outcomes based on the unobserved counterfactual price, i.e. $E[p_{it}|taxi] \neq E[p_{it}|train]$. This reflects different taxi-taking rates across destinations and across time. The distribution of destinations for train-takers is unlikely to be the same as the observed distribution of destinations for train-takers.

4.1 Estimation Results

We present the results of estimation iteratively. The first 2 specifications omit price effects and are intended primarily as a sanity check to describe baseline variation in choice shares. The second 2 specifications include price but not travel time. Indeed, the intercepts estimated throughout, which suggest the baseline utility difference is on average of the order -2.5, reflect the average choice probabilities. The results are reported (without month fixed effects where relevant) in Tables 2 and 3. Interestingly, the 'best fit' model for the post-period is the model without price but with month fixed effects. Moreover, the model with price in the post period without month dummies (Table 4) has even a small but positive coefficient on price. The most complete model, in Table 5, does find a small but negative coefficient on price in both the pre- and post-periods. This in part explains why the fit on the simple model with month effects is the best fitting in this nest. We expect that the objective function is fairly lumpy over price, since the expected price doesn't vary much even across days, so its predictive power in the post-period may be uninformative.

The before-period estimation performs best in the price-inclusive model without month effects (Table 4). However, this has a small but positive coefficient on price, which is not the direction we expect to see. This could come from the level of aggregation in price expectations, as mentioned before. The sign is corrected in the full model.

The overall variability in fit reflects the non-convexity of the objective function. In many cases, certain starting values could not reach the minima reported here. This was a larger problem for the O'Hare estimation, which is why those results are omitted. The results reported here are semi-stable, but the omitted month fixed effects were not at all stable across specifications, and highly contingent on initial values.

4.2 Discussion

Notably, the estimates for β_p differ between the before-entry estimation and the after-entry estimation. The estimated sensitivity to taxi price relative to train price declines (even if accounting for higher average prices in the post-period). This is fully in line with our

Table 2: Estimation Results - Simplest Model, Midway

	Before Entry	After Entry
α_j	-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

Table 3: 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 ***	-0.0004760779 ***
β_{temp^2}	-0.00004672049 ***	-0.00002300726 ***
N	6,019,303	2,610,539
LL	-2,020,383	-730,064.2

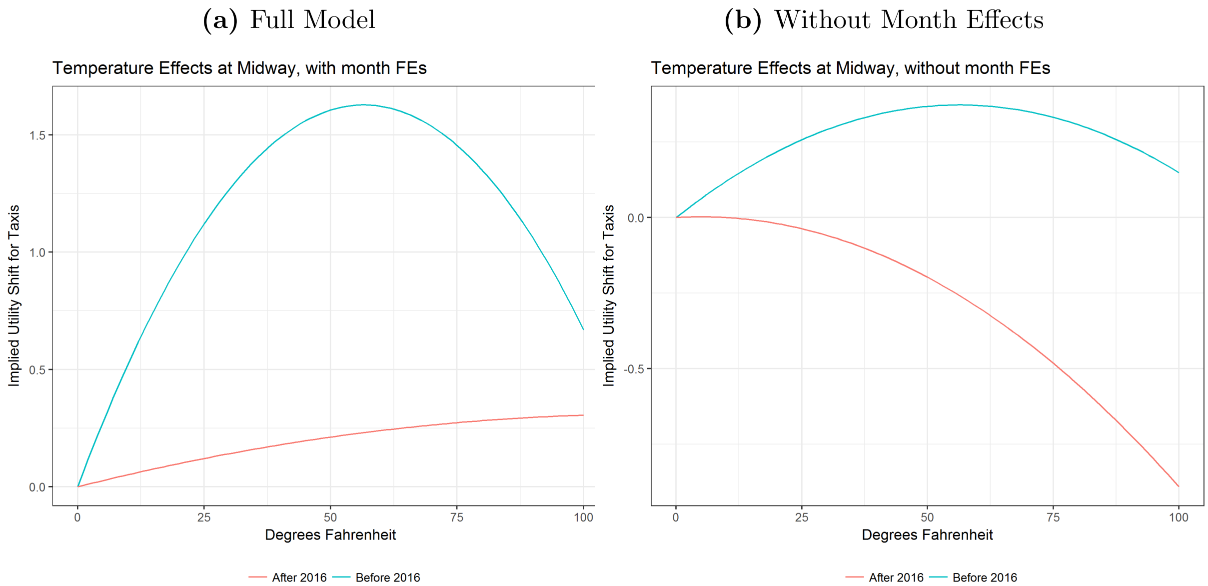
Table 4: Estimation Results - Model with Price, Midway

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

Table 5: 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 ***	0.005424446 ***
β_{temp^2}	-0.0005082469 ***	-0.00002374716 ***
β_p	-0.0295328034 (0.0005690839)	-0.001877242 (0.0007774288)
N	6,019,303	2,610,539
LL	-2,059,348	-830,945

Figure 11: Fitted Temperature Effects



expectations for how the market may be after the ride-hailing entry and the taxi meter fare increase. Additionally, sensitivity to temperature declined, as shown in Figure 11. The figure shows sample effects across a range of typical temperatures (see Figure 14) based on the estimates, suggesting shifts of the relative attractiveness of taxi trips at high and especially low temperatures. However, in the period after entry, the implied sensitivity to temperature has declined. These 2 pieces of evidence together are suggestive of selection out of the taxi market. We can't observe where these consumers are ending up, but it is likely that the decline in elastic demand is due to switching by the elastic consumers out of these 2 data sets and into alternatives (likely - Uber and Lyft). This can be the subject of further investigation should we acquire additional data sources. We would expect that much of this switching could be due to people with destinations we have classified as 'local', since they may be more likely to have access to ride-hailing apps or know they can be used at the airport.

4.3 Next steps for estimation

As mentioned in the results description, these results are in many cases highly sensitive to starting values for optimization. This may be in part due to the problem at hand, but we suspect the data aggregation contributes to the problem as well. If we can estimate or observe a distribution of destinations for consumers who took the train or a non-taxi alternative, then we can use more of the cross-sectional variation within a day to help identify the parameters.

Another next step will be to finish esimating the same models for O'Hare traveler data. In part due to the issues with aggregation we just discussed, the O'hare estimation was much less stable, since its larger volume lead to more aggregation.

Once the estimation of the 4 baseline models is complete for both airports, we can include a measure of travel time in the utility specification. As discussed above, this will be cast in terms of minutes saved by traveling via taxi, so that we can leave the normalized utility of the train option at 0. We predict that including travel time will yield larger (more negative) coefficients on price, since currently the model doesn't yield a benefit from taxi, and so must reduce how 'bad' the price is to match the data choice probabilities. However, it should be noted that we currently lack a means to estimate travel time via taxi for individual who take the train.

5 Conclusion

In this paper, we identify two policy changes in the supply of transit services from Chicago airports and use those changes to describe demand for transit alternatives. We estimate a discrete choice demand model at each airport across two alternatives, where we separately estimate the model on the pre- and post-policy change periods to directly compare the demand systems before and after the two changes.

We find evidence both from the estimation and from summary data analysis to suggest that consumers who remain in the market after the change have less elastic demand for taxis with respect to both price and weather characteristics. This could be suggestive of a change in which segments of consumers remain in the market for taxis when Uber is available. Based on the change in realized demand for taxis across different transit destinations, we hypothesize that those remaining in the market for taxis are visitors (i.e. non-local consumers).

The drawbacks of our approach come primarily from the limitations of the data on hand now. Without observing the destination of train takers, we are limited in constructing counterfactual utilities for the alternatives not selected. We solve this by aggregating across a given day, but this severely limits what we can include in our utility function and reduces stability of our estimates. Future directions for this research will include attempts to augment our data to better address this question.

References

Ben-Akiva, M. E., & Lerman, S. R. (1985). Discrete choice analysis: theory and application to travel demand (Vol. 9). MIT press.

Chintagunta, P. K., & Nair, H. S. (2011). Structural workshop paper - Discrete-choice models of consumer demand in marketing. Marketing Science, 30(6), 977-996.

Gonzales, E. J., Yang, J., Morgul, E. F., & Ozbay, K. (2014). Modeling taxi demand with GPS data from taxis and transit. MNTRC Report, 12-16.

Kamga, C., Yazici, M. A., & Singhal, A. (2013). Hailing in the Rain: Temporal and Weather-Related Variations in Taxi Ridership and Taxi Demand-Supply Equilibrium. In Transportation Research Board 92nd Annual Meeting (No. 13-3131).

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.

Qian, X., & Ukkusuri, S. V. (2015). Spatial variation of the urban taxi ridership using GPS data. Applied Geography, 59, 31-42.

Rayle, L., Shaheen, S., Chan, N., Dai, D., & Cervero, R. (2014). App-Based, On-Demand Ride Services: Comparing Taxi and Ridesourcing Trips and User Characteristics in San Francisco University of California Transportation Center (UCTC).

Train, K. E. (2009). Discrete choice methods with simulation. Cambridge university press.

Yang, C., & Gonzales, E. (2014). Modeling taxi trip demand by time of day in New York City. Transportation Research Record: Journal of the Transportation Research Board, (2429), 110-120.

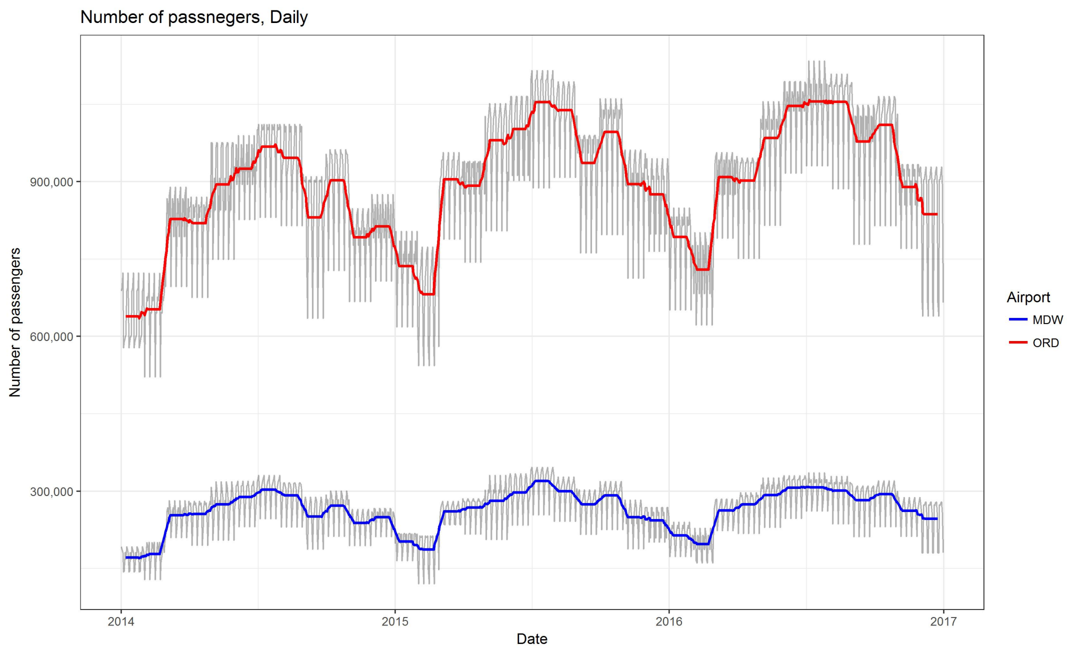
Appendix

A-1 Data

A-1.1 Market Size

Market size that has been used to compute market shares of each transit mode is computed based on monthly passenger volume[5] and monthly average of passengers in a given day of week in Chicago airports. Figure 12 shows the imputed daily passenger volume.

Figure 12: Daily Passenger Volume - O'Hare and Midway Airport



A-1.2 Taxi Trips

Less than 2% of the observations have been removed due to the inconsistency of fare, trip miles, and pick-up/drop-off locations. The measurement errors were not clustered in certain areas, so dropping off those observations did not change the data set in a significant way.

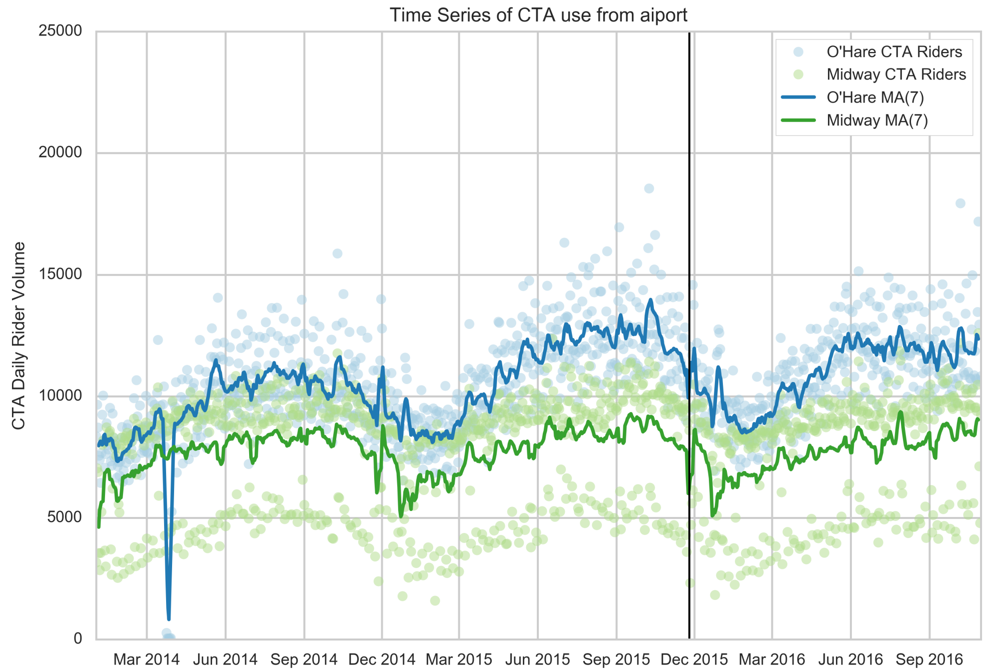
A-1.3 CTA Ridership

We use data from the City of Chicago's Open Data portal [6] on daily entries into O'Hare and Midway L stations. For each day at each airport, we observe the total number of riders to enter the station on that day and the day type (weekday, Saturday, or Sunday/Holiday, which are mutually exclusive categories).

[5] http://www.flychicago.com/OHare/EN/AboutUs/Facts/Pages/Air-Traffic-Data.aspx

[6] See: https://data.cityofchicago.org/Transportation/CTA-Ridership-L-Station-Entries-Daily-Totals/5neh-572f

Figure 13: Daily CTA Use by Airport



A-1.4 Market Size

- Number of passengers : how to impute daily statistics -

A-1.5 Weather Data

The data on weather is taken from the National Climactic Data Center's Integrated Surface Data. We use the Global Surface Summary of Day data to get daily weather summary data. We use data from 2 weather stations, one at each airport. The World Meteorological Organization station numbers are 725300 and 725340 for O'Hare and Midway, respectively. The three main measures we use are as follows. For rain, we use an indicator for any rain or drizzle throughout the day. For snow, we use an indicator for any snow, ice pellets, or hail throughout the day. For temperature, we use mean temperature in Fahrenheit.
We include summaries of the three main measures here over time to show strong seasonality in snow and temperature, but distributed rain throughout the year.

A-1.6 Other Variables

We grab the day of week classification from the CTA data, which includes whether the date is a weekend or holiday.

Figure 14: Average daily temperature

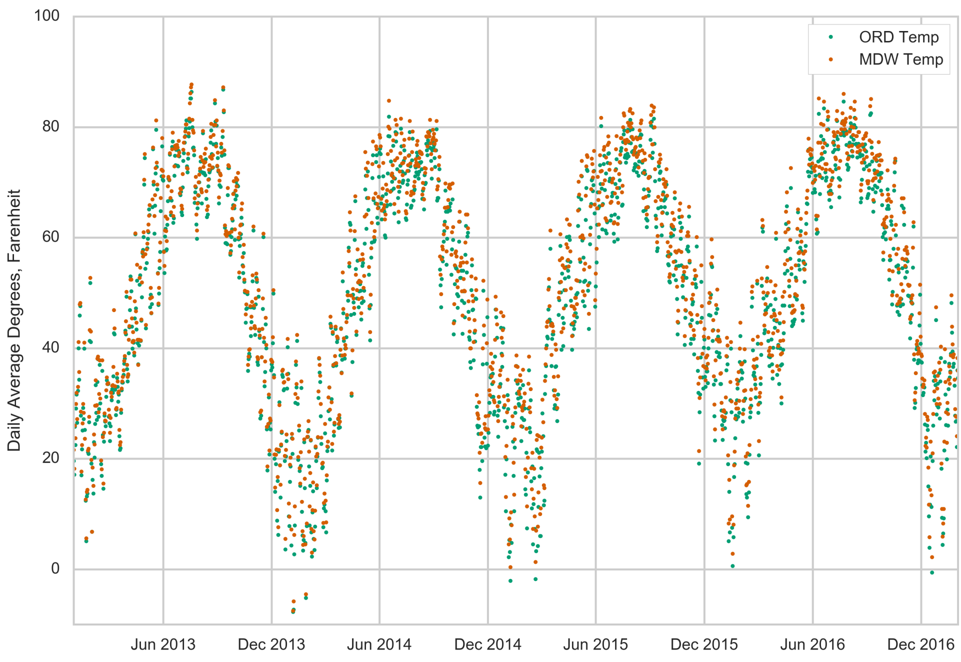


Figure 15: Daily Precipitation by Airport