IMPACT OF COMPETITION ON THE SCALE EFFECTS IN RIDESPLITTING: A CASE STUDY OF MANHATTAN

Hao Liu, Corresponding Author
The Pennsylvania State University
University Park, PA, United States
Email: hfl5376@psu.edu

Saipraneeth Devunuri
University of Illinois at Urbana-Champaign
Champaign, IL, United States
Email: sd37@illinois.edu

Xin Dong
The Pennsylvania State University
University Park, PA, United States
Email: xjd5036@psu.edu

Lewis Lehe
University of Illinois at Urbana-Champaign
Champaign, IL, United States
Email: lehe@illinois.edu

Vikash V. Gayah
The Pennsylvania State University
University Park, PA, United States
Email: gayah@engr.psu.edu

Word Count: 4429 words + 7 table(s) × 250 = 6179 words

Submission Date: April 11, 2024

ABSTRACT
Ridesplitting has become a popular type of shared mobility in major cities. A previous study revealed that the existence of two types of scale effects—economies of scale and increasing returns to scale—in ridesplitting service using trip records reported to the City of Chicago by transportation network companies (TNC). This paper confirms that the same scale effects are exhibited in Manhattan, NY: both for the entire ridesplitting system as a whole and for each individual TNC provider. Further, this paper investigates the influence of competition between TNCs on the overall efficiency of the ridesplitting system in Manhattan. The level of competition is quantified by the concept of entropy, and the efficiency is reflected by the matching rate and travel distance of shared trips. The results indicate that the matching rate decreases and the travel distance of shared trips increases as the intensity of competition rises, due to the lower demand received by each individual TNC. An estimate of the potential for improvement that might be gained due to collaboration between TNCs under a "best-case scenario" is derived using linear regression models. The results in this paper can be used by transportation policymakers to enhance the efficiency of the ridesplitting service by building collaboration between TNCs.

Keywords: Ridesplitting, Transportation Network Company (TNC), Scale effects, Competition

INTRODUCTION

The popularity of shared mobility—the shared use of vehicles by multiple people for different trips—has grown rapidly in the last decade (*1*). *Ridesplitting*, also called *pooled* service, is a shared ride-hailing service in which customers (riders) are willing to share vehicles with other customers, who have placed separate and uncoordinated requests for the same service (*2, 3*). A user willing to use this service sends a request, referred to as an *authorized* or *requested* shared trip, to a Transportation Network Company, and the service company uses matching algorithms and real-time demand data to match customers and drivers. Note an authorized shared trip will not always wind up being matched; trips serving a single customer regardless of whether it was authorized as a shared trip are still referred to as a single trip.

The upside of ridesplitting is a lower fare charged for individual customers. To encourage the use of this travel mode, some US cities and states (e.g., Chicago, New York City, Georgia, New Jersey) impose lower excise taxes on shared rides than on single rides (*4*). On the other hand, a shared trip can impose a longer travel distance, a longer travel time, and a higher travel uncertainty (*5*) due to the detour required for picking up and/or dropping off other customers. The tradeoff between these two aspects—the money costs and the travel efficiency—plays a crucial role in how much ridesplitting is used and thus whether this service can be operated in a financially sustainable way.

In the US, UberPool and Lyft Shared Rides—both of which were launched in August 2014 (*6*)—are the two most common companies providing ridesplitting, referred to as TNCs (Transportation Network Company). Note that Lyft discontinued their shared rides service in May 2023. All TNCs halted ridesplitting service in March 2020 due to the COVID-19 crisis, but the service returned to Chicago in June 2022, and Lyft and Uber pooling returned to New York City in August 2021 and June 2022, respectively. The demand has been gradually increasing towards the normal level. For example, the monthly number of authorized shared trips in the City of Chicago has increased from 21,484 in June 2022 to 214,689 in April 2023, and the corresponding number in NYC has increased from 51,238 in August 2021 to 75,213 in April 2023. However, when compared to pre-pandemic levels, the authorized shared trips are considerably lower. In January 2020, the city of Chicago had 1,499,012, and NYC had 3,694,815 authorized shared trips.

This study builds on the hypothesis that ridesplitting involves substantial *scale effects*: that as participation rises the average quality of matches rises. The matching rate of a ridesplitting system has been demonstrated to increase when the number of shared requests increases (*7–9*). This pattern was documented in the ridesplitting systems in New York City, San Francisco, Singapore, and Vienna (*10*). (*8*) claims that the joint effect of passenger demand and matching window can lead to a reduction in travel time for both single trips and shared trips. Another positive consequence of the increase in the matching rate is that the number of vehicles per hundred shared requests decreases (*11*). Moreover, recent studies identified that the increase in shared demand can also lead to an increase in the fleet size, which translates into a reduction in travel time resulting from the increase in the fleet size (*12*).

Most studies mentioned above were based on theoretical analysis and simulations, (*13*) confirmed the scale effects in ridesplitting data reported by Uber, Lyft, and Via to the city of Chicago in 2019. These companies have only offered ridesplitting in large markets, which suggests that there exists a threshold for the potential demand to make this service profitable. Specifically, that study unveils two economic features that exist in ridesplitting as the number of authorized shared trips during a time window increases:

1. the average detour distance decreases (known as *economies of scale* (*14*)).
   It can be easily seen that with the increase in the density of authorized shared trips, the probability of an authorized shared trip being matched with another request that shares close origins and destinations is higher. As a result, the average detour distance can be decreased.
2. the matching rate in ridesplitting rises (known as emphincreasing returns to scale (*15*)).
   This can be explained by a similar reason. If the detours required to match two authorized shared trips are too long to make the shared trip profitable, the customers are simply taken straight to their destinations as single trips. Therefore, a shorter detour distance that results from a rise in the density of ridesplitting requests can increase the matching rate.

More detailed discussions on these two economic characteristics can be found in (*13*). However, these scale effects have only been identified empirically for the city of Chicago; whether these findings are general—i.e., if they exist in other ridesplitting systems—is not clear.

Another shortcoming of (*13*) is that, in the Chicago dataset, some trip data were obscured or removed in order to protect customer privacy, which makes it challenging to precisely quantify the scale effects. For example, the start and end timestamps of all trips were rounded to the nearest 15-minute interval, and the pickup and drop-off locations are at the census tract level. Moreover, the TNC provider associated with each trip was also hidden. Compared to the Chicago dataset, the trip records reported to NYC (New York City), called "TLC" (Taxi and Limousine Commission), include more detailed information. Thus, the first question that can be answered by this dataset is whether the same scale effects exist for each individual TNC provider. More interestingly, how competition between TNC providers in the same region affects the scale effects can also be explored. (*16*) developed a modeling and simulation framework based on Macroscopic Fundamental Diagram (*17–19*) and showed that competition between TNCs can undermine short-term network mobility. The influence of competition on the economics of ridesplitting is also studied (*20–22*). However, these studies use theoretical modeling based on equilibrium conditions and do not reflect the impact on the aforementioned scale effects. In addition, the assumptions employed in these models might not be satisfied in reality, so it is critical to confirm the findings using empirical data (*23*). To this end, this paper investigates the impact of competition between TNC providers on the scale effects in ridesplitting using the "TLC" dataset in Manhattan in 2019.

The contributions of this study are as follows. This study (1) confirms the scale effects exist in the entire ridesplitting system in Manhattan and in each individual TNC provider; (2) demonstrates that competition between TNCs can diminish the system efficiency, i.e., reduce the matching rate and increase travel distance; (3) provides an estimate of the improvement in the scale effects considering collaboration between TNCs under a “best-case scenario” accordng to regression models developed from the “TLC" dataset.

The remainder of this paper is organized as follows. The following section describes the “TLC" dataset and the filters used to process the data. The next section presents the existence of scale effects in the entire ridesplitting system of Manhattan and in each individual TNC provider. This is followed by a section showing the impact of competition between TNCs on the overall efficiency of the ridesplitting system. The next section provides estimates on the improvement in matching rate and travel distance resulting from collaboration between TNCs using regression models. Finally, concluding remarks are provided.

DATA

The "TLC" dataset includes trip information from the six boroughs of NYC: Newark, Queens, Bronx, Manhattan, Staten Island, and Brooklyn. However, these boroughs are separated in space, and the travel pattern varies significantly between these boroughs. Since Manhattan has the highest trip density in both time and space, only trip data in Manhattan is considered in the rest of this paper to reduce the impact of this variation. In 2019, four TNCs operated in Manhattan: Juno, Lyft, Uber, and Via. Juno provides single rides only while the other three TNCs provide both single rides and ridesplitting services. The following information associated with each trip is used:

- Taxi zone ids for origins and destinations;
- Start timestamp including month, date, hour, and minute rounded to the nearest 15 min interval;
- Travel distance (mi);
- Travel time (min);
- Boolean for shared trip authorization; and,
- Boolean for matching of authorized shared trip matched.

The following filters were used to remove suspicious data:

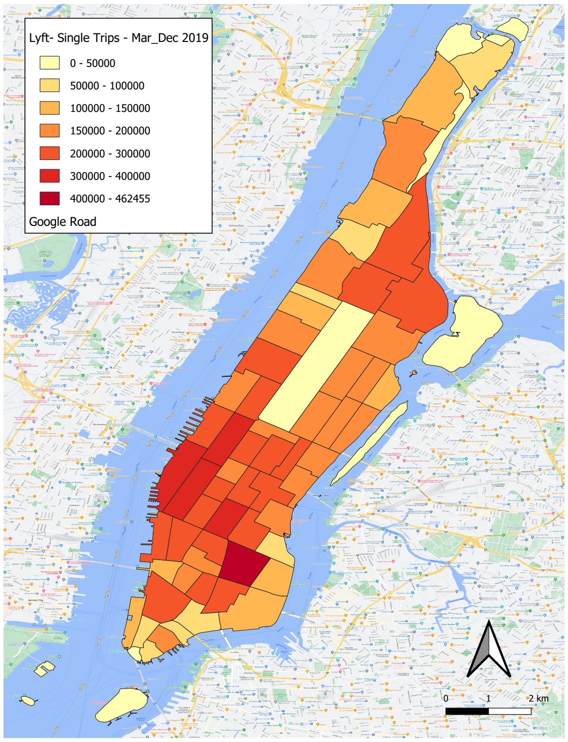
- Travel times of less than 2 minutes;
- Travel distances shorter than 0.1 miles;
- Missing origins or destinations

We also found some "matched" trips that were not "authorized" as shared trips. Specifically, 8.45% of Lyft trips in February 2019 possessed such attributes. The percentage for all other combined TNC and month is below 1. Therefore, the data in February 2019 was excluded from this study, and the rest of such records were fixed by changing them to authorized shared trips.

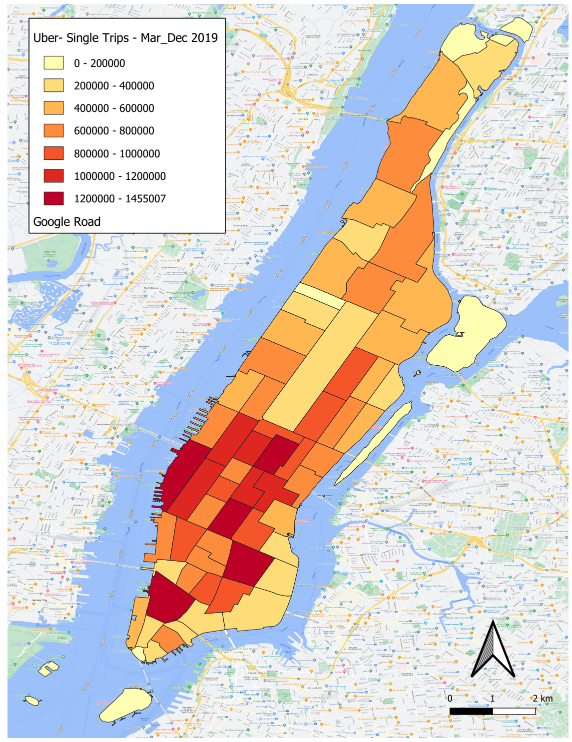
After applying the filters above, there are 67,176,764 trips remaining. Of these, 55,934,574 (83%) are single trips and 11,242,190 (17%) are (matched) shared trips. The distribution of average daily trips across the four TNC providers is shown in Table 1. It shows that the high imbalance exists in the proportions between single trips and shared trips served by each TNC provider. There are 73% of single trips served by Uber while the corresponding value for the shared trips is only 35%. Via provides 4% of single trips while it fulfills 44% percent of shared trip orders. Overall, the total number of rides provided by Uber is considerably higher than the other TNC providers. This distribution highlights the significance of the investigation on whether the scale effects exist in each individual TNC provider.

TABLE 1: Distribution of average daily trips

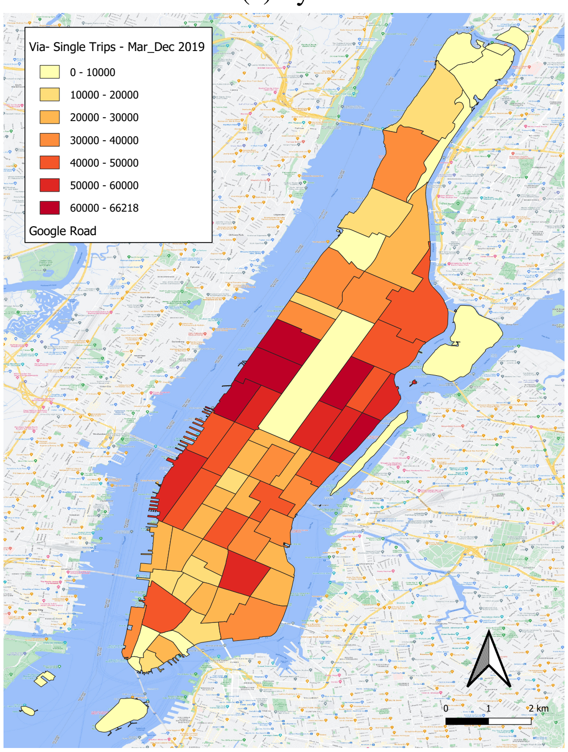
	Single trips		Shared trips (matched)		Total trips	
	Number	Percent	Number	Percent	Number	Percent
Lyft	37,538	20	7,792	21	45,330	21
Uber	134,497	73	12,818	35	147,315	67
Via	6,622	4	16,248	44	22,870	10
Juno	4,734	3	0	0	4,734	2
Total	183,391	100	36,858	100	220,249	100



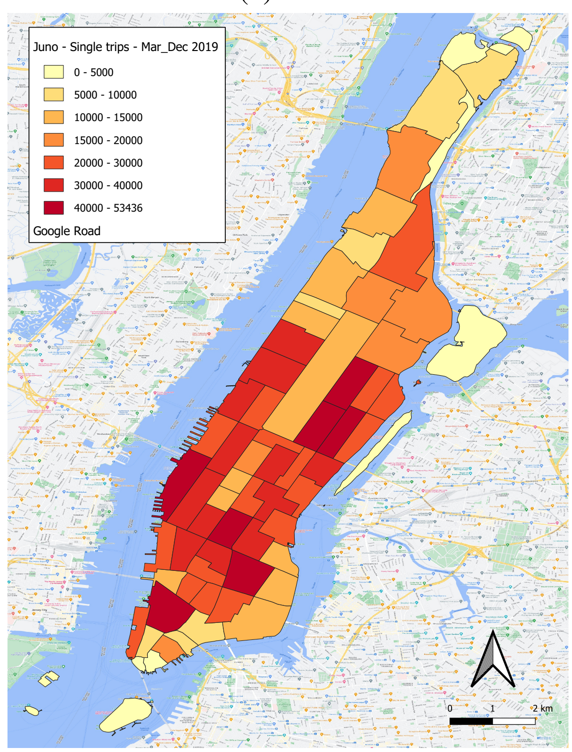
(a) Lyft



(b) Uber

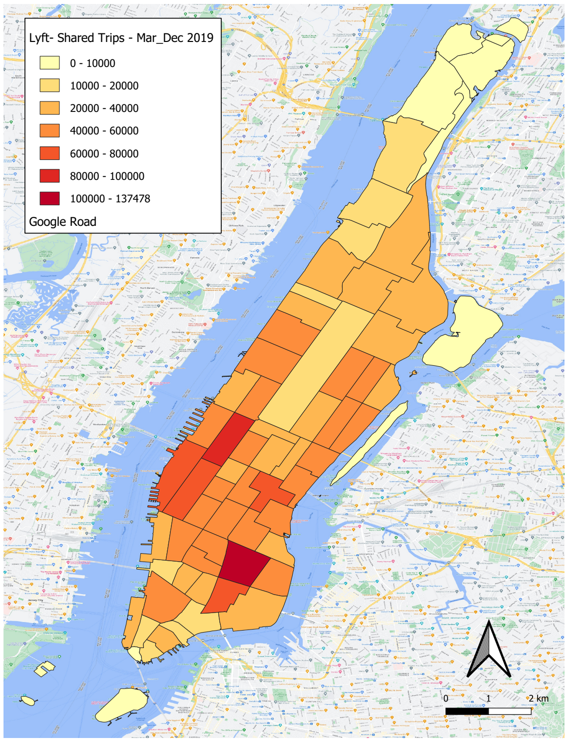


(c) Via

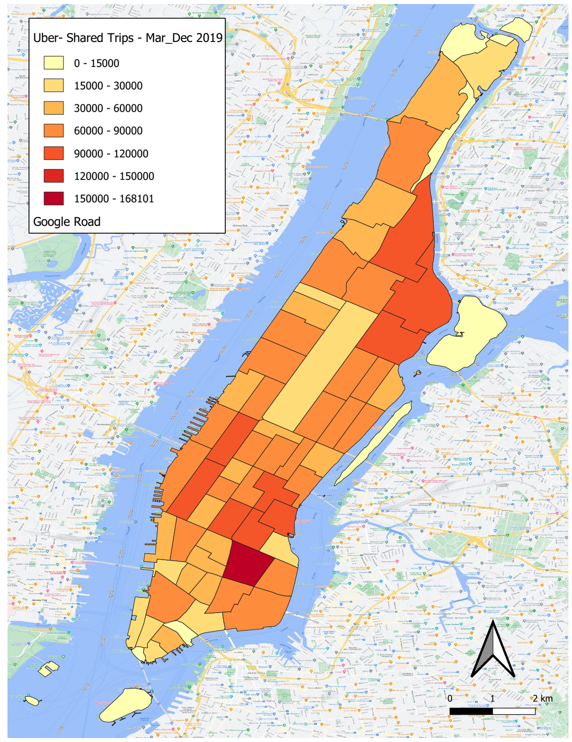


(d) Juno

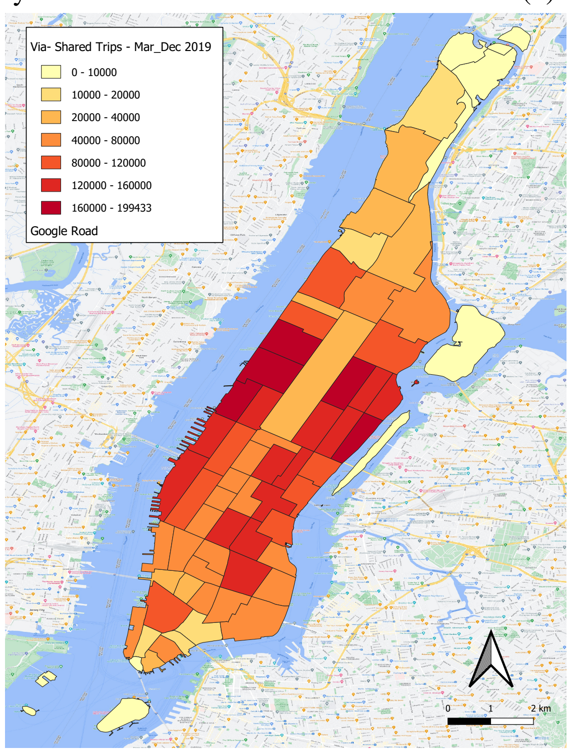
FIGURE 1: Distribution of origins of single trips.



(a) Lyft



(b) Uber



(c) Via

FIGURE 2: Distribution of origins of shared trips.

In addition, Figure 1 and Figure 2 show the distribution of the origins of single trips and shared trips for each TNC, respectively. It is evident that the TNC's vary by their shares of single/shared trips, so it is essential to explore scale effects at the level of individual TNC's.

SCALE EFFECTS

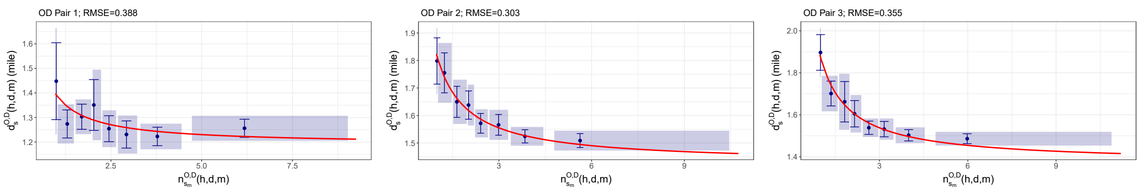
This section examines scale effects over the entire Borough of Manhattan and also at the level of each individual TNC's. To do so, the data were grouped based on hour (h), day of the week (d) and month (m), and the average number of (authorized shared/matched shared/single) trips in each hour-day-month combination was counted.

Economies of scale

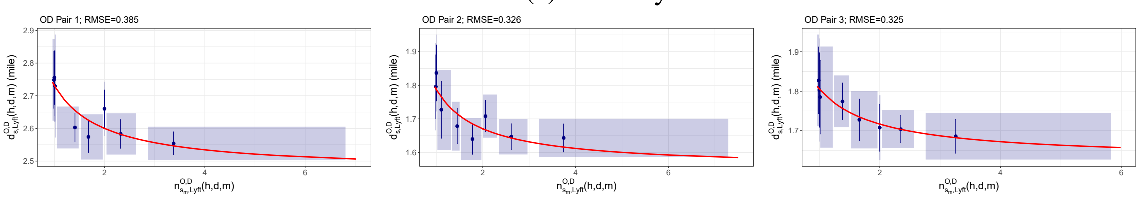
The detour distance is defined as the extra travel distance of shared trips resulting from picking up or dropping off other customers. It can be approximated as the difference between the travel distance of a shared ride and the travel distance of a single ride that have the same origin and destination and occur during the same time. The travel distance of the associated single rides is referred to as baseline travel distance. However, as mentioned before, since the timestamps were rounded to 15-min intervals, and the start and end locations of a trip are at a taxi zone level, the exact timestamps, origins, and destinations of trips, which are required to identify a precise estimate of the baseline travel distance, are not available in the dataset. To overcome this difficulty, we need to assume all shared trips that start in the same 15-min interval and connect the same origin and destination taxi zones have the same baseline travel distance. Consequently, the detour distance and actual travel distance of a certain OD pair have the same pattern. For simplicity, instead of the actual detour distance, the average travel distance over all shared trips between an origin-desination pair (O,D) in hour h, weekday d and month m, denoted by $d_{s,i}^{O,D}(h,d,m)$, is investigated.

There are 69 taxi zones in Manhattan. The influence of the average number of matched shared trips, denoted by $n_{s_m,i}^{O,D}(h,d,m)$, on the travel distance of matched shared trips in the first ten OD pairs with the highest number of trips of the entire system and each individual TNC were examined. A similar pattern exists in all OD pairs. Therefore, for simplicity, only three pairs of OD in each group are shown here in Figure 3. Since most data points have a relatively small number of matched shared trips, a dense cloud of points occurs in the classical scatter plot, which makes it impossible to visualize trends in the data. In addition, the dependent variable in the plots—the number of realized shared trips—is discrete, which makes it more difficult to observe the actual trend. Therefore, binned scatter plots are utilized (*24*) in Figure 3, in which an equal number of observations are assigned to "bins". The point and bar in each bin are the point estimates evaluated at the mean and the 95% confidence interval for the travel distance observed in each bin. The results show that for both the entire system and within each TNC provider, the travel distances of all examined OD pairs decrease as the number of matched shared trips increases.

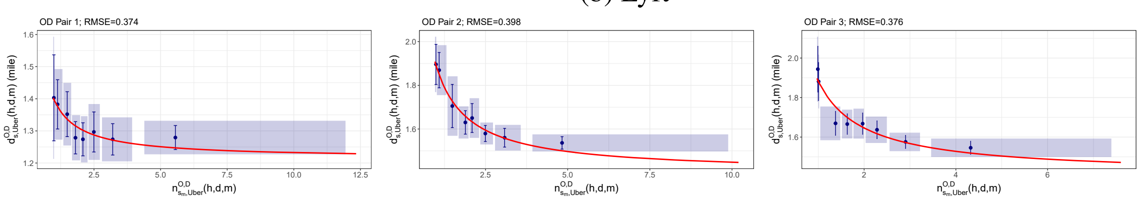
To test the strength and statistical significance of this effect, based on the visual patterns observed in Figure 3, a regression model was fitted. Note that the purpose of the regression model is to help to visualize the trend of the scatters instead of proposing any causation between the parameters. The form of the regression model is shown in Equation (1), and the regression results are provided in Table 2. The results confirm that the decreasing trend of travel distance with the increase in the number of matched shared trips is statistically significant for all examined OD pairs in individual TNCs and the entire system.



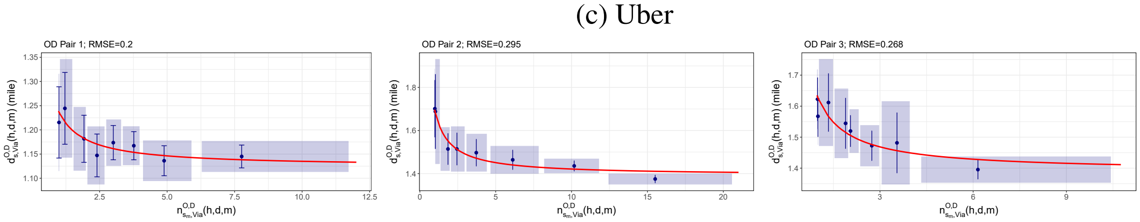
(a) Entire system



(b) Lyft



(c) Uber



(d) Via

FIGURE 3: Economics of scale: travel distance vs. matched shared trips.

$$d_{s,i}^{O,D}(h,d,m) = \beta_0 + \frac{\beta_1}{n_{s_m,i}^{O,D}(h,d,m)} \tag{1}$$

TABLE 2: Regression results for travel distance (mile)

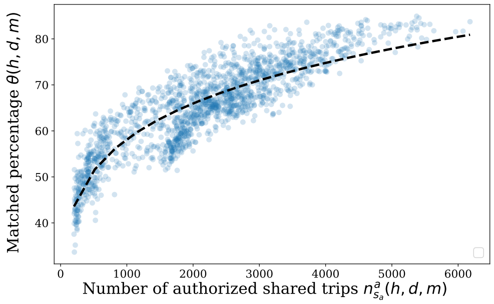
	Dependent variable: $d_{s,i}^{O,D}(h,d,m)$					
	Entire system			Lyft		
	OD 1	OD 2	OD 3	OD 1	OD 2	OD 3
$\frac{1}{n_{s_m,i}^{O,D}(h,d,m)}$	0.305***	0.450***	0.474***	0.360***	0.314***	0.288***
	(0.052)	(0.035)	(0.039)	(0.047)	(0.045)	(0.047)
Constant	0.104***	0.094***	0.095***	0.210***	0.205***	0.189***
	(0.030)	(0.021)	(0.022)	(0.035)	(0.032)	(0.035)
R^2	0.045	0.137	0.114	0.056	0.059	0.051
Adjusted R^2	0.044	0.136	0.113	0.055	0.058	0.049
	Uber			Via		
	OD 1	OD 2	OD 3	OD 1	OD 2	OD 3
$\frac{1}{n_{s_m,i}^{O,D}(h,d,m)}$	0.285***	0.461***	0.512***	0.180***	0.319***	0.227***
	(0.048)	(0.045)	(0.046)	(0.027)	(0.048)	(0.040)
Constant	0.129***	0.124***	0.080***	0.176***	0.190***	0.124***
	(0.030)	(0.028)	(0.031)	(0.015)	(0.028)	(0.026)
R^2	0.042	0.090	0.115	0.080	0.113	0.069
Adjusted R^2	0.040	0.089	0.114	0.078	0.110	0.067
Significance levels				*p<0.1; **p<0.05; ***p<0.01		

Increasing returns to scale

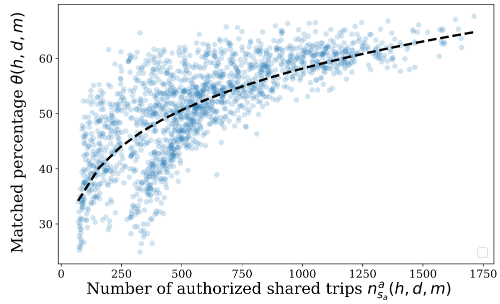
Let $n_{s_a,i}^{a}(h,d,m)$ and $n_{s_m,i}^{a}(h,d,m)$ denote the average number of authorized and matched shared trips in hour h, weekday d and month m for TNC i, respectively. The matched percentage for TNC i in the corresponding period can be expressed as:

$$\theta_i(h,d,m) = \frac{n_{s_m,i}^{a}(h,d,m)}{n_{s_a,i}^{a}(h,d,m)} \times 100 \quad \text{(\% matched shared trips).} \tag{2}$$

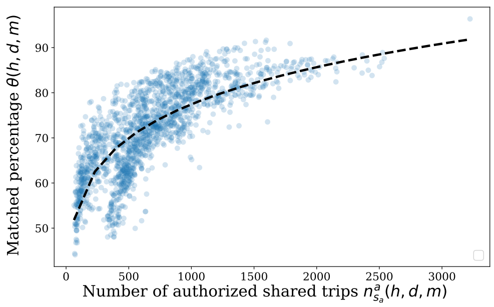
The matched percentage for the entire system can be obtained by replacing the numerator and denominator in Equation 2 with the sum of the corresponding values over all TNCs. Figure 4 shows that for all individual TNCs and the entire system, the matched percentage, $\theta_i(h,d,m)$, increases with the number of authorized shared trips $n_{s_a,i}^{a}(h,d,m)$. Similar to the travel distance, a regression model was fitted to confirm that this trend is statistically significant. The form of the model is shown in Equation (2), and the regression results are shown in Table 3. The results confirm that the increasing returns to scale are statistically significant for all TNCs and the entire system.



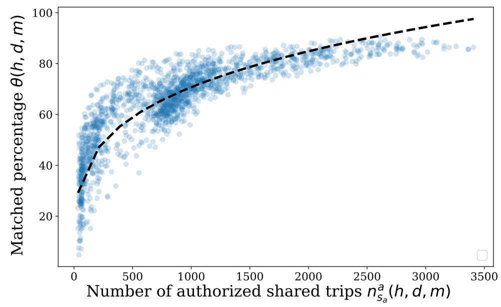
(a) Entire ridesplitting system



(b) Lyft



(c) Uber



(d) Via

FIGURE 4: Matching rate.

$$\log_{10}\theta_i(h,d,m) = \beta_0 + \beta_1 \log_{10} n^a_{s_a,i}(h,d,m) \tag{3}$$

TABLE 3: Regression results for matched percentage (%)

	Dependent variable:			
	Matched percentage $\log_{10}\theta_i(h,d,m)$			
	Entire system	Lyft	Uber	Via
$\log_{10} n^a_{s_a,i}(h,d,m)$	0.181^{***}	0.200^{***}	0.146^{***}	0.262^{***}
	(0.002)	(0.005)	(0.003)	(0.004)
Constant	1.222^{***}	1.165^{***}	1.452^{***}	1.063^{***}
	(0.007)	(0.013)	(0.008)	(0.012)
R^2	0.787	0.515	0.594	0.705
Adjusted R^2	0.787	0.515	0.593	0.705
Significance levels		$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$		

IMPACT OF COMPETITION

The previous section demonstrates that the two scale effects exist in all TNC providers and across the entire ridesplitting system as a whole in Manhattan. Since there are three TNC providers operating in this region, the demand for shared trips, which has a strong effect on travel distances and matching rates, received by each provider is by definition less than or equal to what would be realized if there was only a single TNC provider (or if the shared requests could be matched between different providers). This split of demand is expected to reduce the efficiency of the system, and the magnitude of this reduction is expected to depend on the intensity of competition.

Competition is reflected by the difference in the demand distribution across all three TNCs; competition is more intense when the distribution is more even. For example, the most intensely competitive scenario would occur when each of the three TNCs receives one third of the total shared requests. The least intense competition would occur when one TNC received all the shared requests; in this case, there would be no competition between services. The concept of entropy, which describes the disorder and randomness of system states, has been widely used to model diversity in transportation systems (*25–27*). In this paper, we use entropy to quantify the degree of competition between TNCs. The entropy for an OD pair during an hour can be expressed as:

$$E^{O,D}(t) = -\frac{\sum_{i=n}^{3} p_i^{O,D}(t)\ln p_i^{O,D}(t)}{\ln n} \tag{4}$$

where $p_i^{O,D}(t)$ is the ratio of shared requested between OD received by TNC i in hour t, and n is the number of TNCs, which is equal to three. The entropy increases with competition intensity. The entropy is equal to 1 when $p_i^{O,D}(t) = \frac{1}{3}$ for all i's and equal to 0 when there is no competition at all, i.e., only one TNC receives shared requests.

Different from the aggregation method in the previous section, we use data from 1-hour intervals without further aggregation to compute the variables in Equation (4). This is because the competition level may be misinterpreted by aggregation. Assume there are three hours in each of which there is only one distinct TNC receiving the same number of shared requests. In this example, no competition should be considered since each TNC receives requests in different periods. However, after aggregating these three hours together, the ratio of the shared requests of each TNC is $\frac{1}{3}$ and thus, the entropy becomes 1, which represents the most intensive scenario. We use the symbol t for the hourly intervals in Equation (4) to distinguish the notation for an hour in the aggregation level from the previous section.

The first ten OD pairs with the highest number of shared requests were analyzed. The influence of competition on travel distances and matching rates is shown in Figure 5 and Figure 6, respectively. For simplicity, only the first six OD pairs are shown, but all other OD pairs have a similar pattern. Note that the points with less than 10 shared requests in an hour were removed from the plots because the low number of shared requests can lead to a large randomness in entropy, travel distance and matched percentage. For example, let us consider two cases for the number of shared requests received by each TNC. In the first case, each TNC receives one request in an hour, and in the second case, each TNC receives 100 requests in an hour. The entropies of these two cases are equal; however, since the number of requests in the first case is so low that it is highly likely that these requests happen in different time windows, which makes it difficult to estimate the true competition. Therefore, it is assumed that the entropy can reflect competition more accurately when the shared demand is higher, and the points with low number of shared requests were removed. Consistent with expectations, the results show that as the entropy increases, the travel distance increases, and the matching rate decreases. Based on the observed pattern, the regression model, shown in Equation (5), was fitted for both travel distance and matched percentage. The results are shown in Table 4 and Table 5, respectively. All results are statistically significant.

$$\log y = \beta_0 + \beta_1 x \tag{5}$$

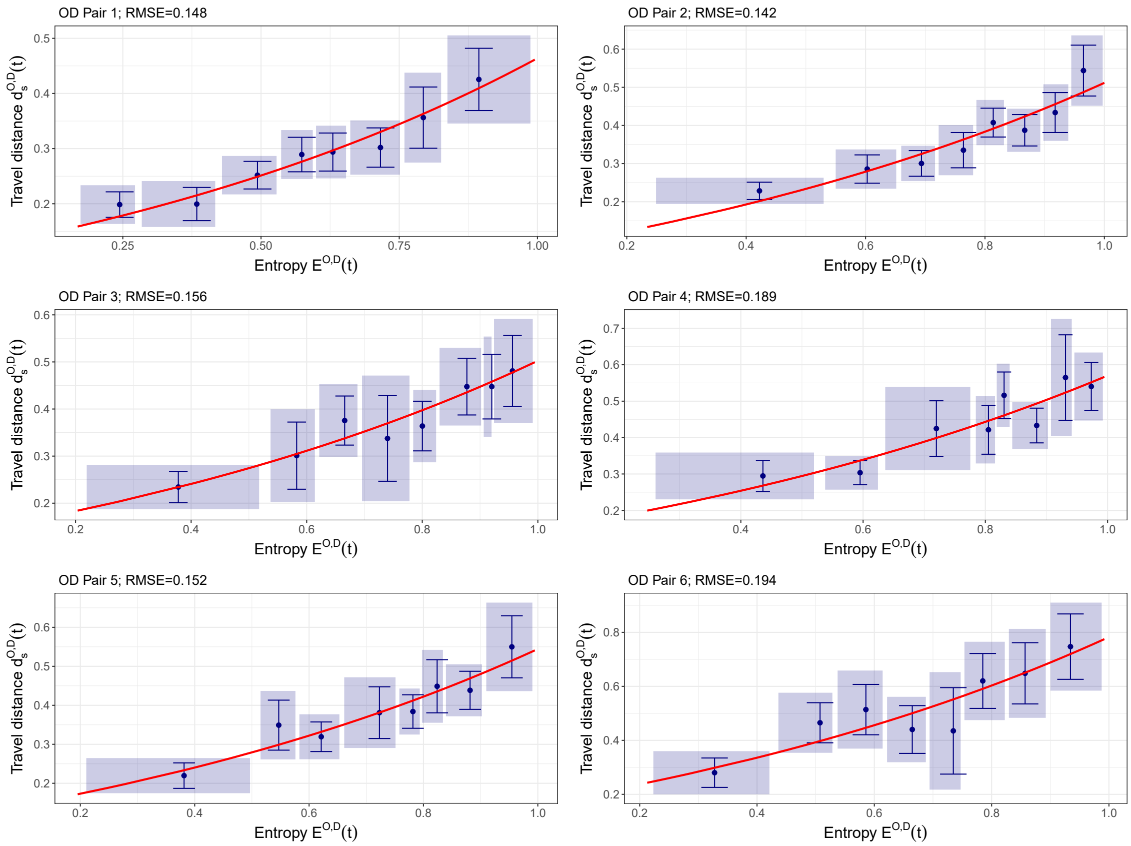


FIGURE 5: Impact of competition on travel distance

TABLE 4: Regression results for travel distance (mile)

	Dependent variable:					
	$d_s^{O,D}(t)$					
	OD 1	OD 2	OD 3	OD 4	OD 5	OD 6
$\exp(E^{O,D}(t))$	0.191***	0.260***	0.213***	0.259***	0.248***	0.360***
	(0.017)	(0.023)	(0.031)	(0.034)	(0.028)	(0.045)
Constant	−0.062**	−0.195***	−0.077	−0.133*	−0.130**	−0.200**
	(0.031)	(0.050)	(0.067)	(0.074)	(0.058)	(0.089)
R^2	0.208	0.287	0.199	0.193	0.263	0.353
Adjusted R^2	0.207	0.285	0.195	0.189	0.260	0.347
Significance levels						*p<0.1; **p<0.05; ***p<0.01

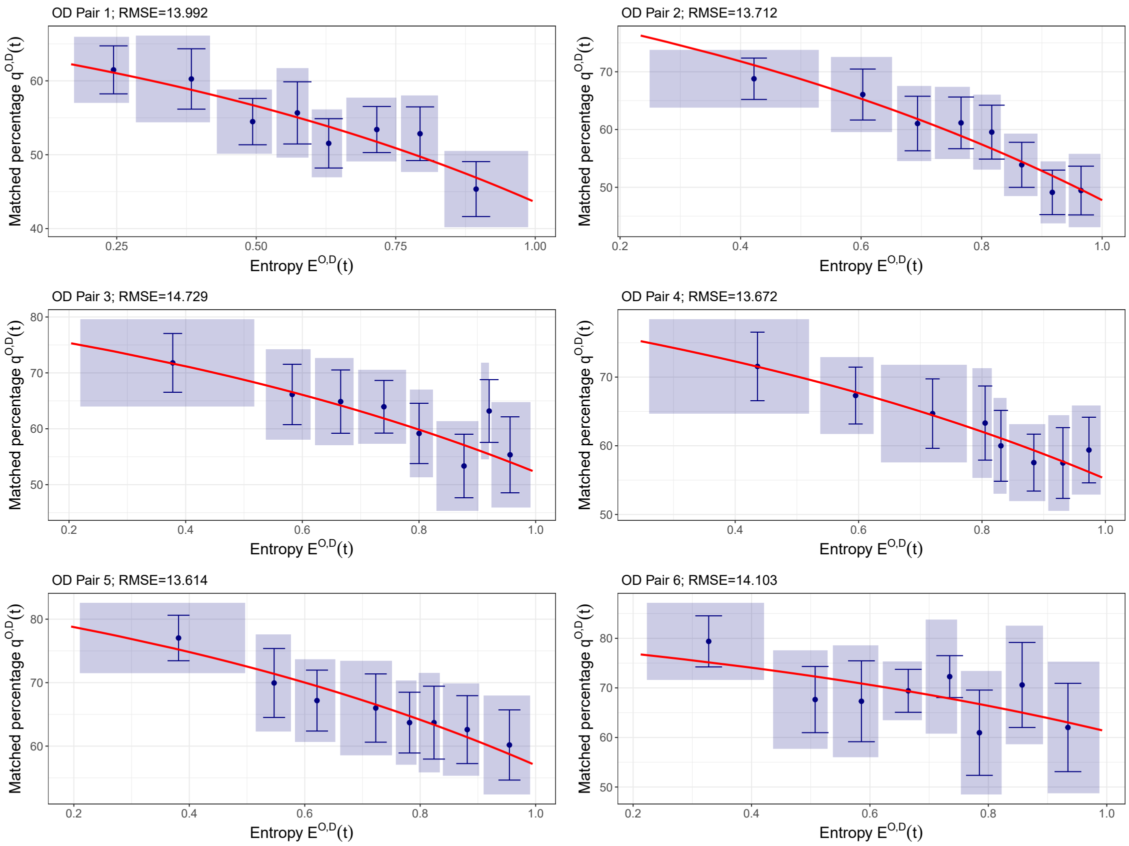


FIGURE 6: Impact of competition on matched percentage

TABLE 5: Regression results for matched percentage

	Dependent variable:					
	$\theta^{O,D}(t)$					
	OD 1	OD 2	OD 3	OD 4	OD 5	OD 6
$\exp(E^{O,D}(t))$	-11.930	-19.590	-15.450	-14.396	-15.180	-12.320
	(1.630)	(2.232)	(2.956)	(2.432)	(2.492)	(3.297)
	$***$	$***$	$***$	$***$	$***$	$***$
Constant	76.175***	101.023***	94.232***	94.117***	97.906***	93.569***
	(3.006)	(4.859)	(6.324)	(5.358)	(5.214)	(6.509)
R^2	0.100	0.198	0.128	0.124	0.142	0.106
Adjusted R^2	0.098	0.195	0.123	0.120	0.138	0.098
Significance levels				*p<0.1; **p<0.05; ***p<0.01		

IMPROVEMENT FROM COLLABORATION

The results from the previous section reveal that competition between TNCs can reduce the efficiency of a ridesplitting system in terms of the matched percentage and travel distances. Following that, this section develops regression models to estimate the potential improvement in efficiency as a result of a comprehensive collaboration between TNCs. This is defined as the scenario under which the real-time demand information from all TNCs are shared, and all authorized shared requests regardless of the associated TNC can be matched. Under this setting, the ridesplitting system operates as if there was only one TNC provider. Therefore, we regard this as a "best-case scenario".

It has been shown that both demand and the level of competition have a statistically significant impact on the matching rate and travel distance for shared trips. Therefore, these two parameters are considered independent variables in the regression models. Similar to previous sections, we selected the first six OD pairs with the highest number of shared authorizations and developed regression models for matched percentage and travel distance.

The regression model for the average travel distance of shared trips has the following form:

$$\log(d_s^{O,D}(t)) = \beta_0 + \beta_1 \frac{1}{n_{s_a}^{O,D}(t)} + \beta_2 E^{O,D}(t) \tag{6}$$

where $n_{s_a}^{O,D}(t)$ is the number of authorized shared trips in hour t.

An inverse logistic transformation was applied to the matching rate to ensure that the estimated matching rate is between 0 and 1. The regression model for the matching rate has the following form:

$$\log\left(\frac{\theta^{O,D}(t)}{1-\theta^{O,D}(t)}\right) = \beta_0 + \beta_1 \log\left(\frac{1}{n_{s_a}^{O,D}(t)}\right) + \beta_2 E^{O,D}(t) \tag{7}$$

The results of both fitted models are shown in Table 6 and Table 7, respectively. The results indicate that travel distance decreases and matching rate increases as the number of shared authorizations increases and entropy decreases.

TABLE 6: Regression results for travel distance (mile) with collaboration

	Dependent variable:					
	$\log(d_s^{O,D}(t))$					
	OD 1	OD 2	OD 3	OD 4	OD 5	OD 6
$\frac{1}{n_{sa}^{O,D}(t)}$	15.196***	14.124***	11.929***	13.474***	12.418***	17.104***
	(1.291)	(1.406)	(2.985)	(2.067)	(2.297)	(3.899)
$E^{O,D}(t)$	0.982***	1.505***	1.178***	1.347***	1.432***	1.420***
	(0.078)	(0.102)	(0.151)	(0.121)	(0.125)	(0.158)
Constant	−3.105***	−3.343***	−2.946***	−3.059***	−3.089***	−3.156***
	(0.103)	(0.136)	(0.265)	(0.187)	(0.210)	(0.330)
R^2	0.423	0.503	0.309	0.414	0.411	0.488
Adjusted R^2	0.421	0.500	0.301	0.410	0.406	0.479
Significance levels				*p<0.1; **p<0.05; ***p<0.01		

The estimates of improvement in both metrics resulting from the comprehensive collaboration were obtained from the fitted models when setting the entropy value equal to zero. This would

TABLE 7: Regression results for matching rate with collaboration

	Dependent variable:					
	$\log(\frac{\theta^{O,D}(t)}{1-\theta^{O,D}(t)})$					
	OD 1	OD 2	OD 3	OD 4	OD 5	OD 6
$\log(\frac{1}{n_{sa}^{O,D}(t)})$	-0.479^{***}	-0.555^{*}	-0.164	-0.777^{***}	-0.548	-1.530
	(0.171)	(0.295)	(0.424)	(0.280)	(0.445)	(1.155)
$E^{O,D}(t)$	-0.913^{***}	-2.010^{***}	-1.511^{***}	-1.468^{***}	-2.513^{***}	-1.386^{**}
	(0.152)	(0.303)	(0.281)	(0.233)	(0.334)	(0.616)
Constant	-0.504	0.487	1.232	-0.259	1.220	-1.822
	(0.467)	(0.789)	(1.096)	(0.743)	(1.142)	(2.957)
R^2	0.095	0.132	0.138	0.168	0.205	0.062
Adjusted R^2	0.091	0.126	0.129	0.162	0.198	0.046
Significance levels				*p<0.1; **p<0.05; ***p<0.01		

indicate the prediction when only one TNC provider receives shared requests. The results for travel distance and matched percentage are shown in Figure 7 and Figure 8, respectively. In both figures, the blue points are the original data, the red points are the results from the fitted models, and the green points are the results from collaboration (i.e., when the entropy is set to zero). The red and green curves are added to help visualize the trend.

The results reveal that the collaboration can improve the efficiency of the ridesplitting system significantly. According to the scale effects, the efficiency of the system can be improved by the increase in demand. Therefore, as expected, the improvement is more significant when the number of authorized shared trips is low. When the demand is relatively high, each individual TNC can receive enough requests to maintain a relatively high efficiency within their own service. Quantitatively, on average, the matched percentage can be improved by up to 25%, and the travel distance can be reduced by up to 0.3 miles. Quantitatively, the best improvement is seen at the lowest demand which improves the matched percentage by 20.9% and reduces travel distance by 0.25 miles on average. The worst improvement is observed at the highest demand with matched percentage increasing 13.65% and a travel distance reduction of 0.086 miles on average.

It is worth re-emphasizing that the estimates are for a "best-case scenario", which assumes a system with only one TNC providing the ridesplitting service. The actual improvement can highly depend on the policy and the matching algorithm. For example, a more reasonable scenario is that all TNCs first try to match the share requests within their own services, and they only send/accept requests that cannot be fulfilled by their own to other TNCs. This scenario is expected to generate a lower improvement than the values shown in this section.

CONCLUDING REMARKS

This paper confirms that the economies of scale and increasing returns to scale found in the ridesplitting system of Chicago (*13*) also exist in both the entire ridesplitting system of Manhattan and each individual TNC provider. This finding strengthens the generalization of the scale effects

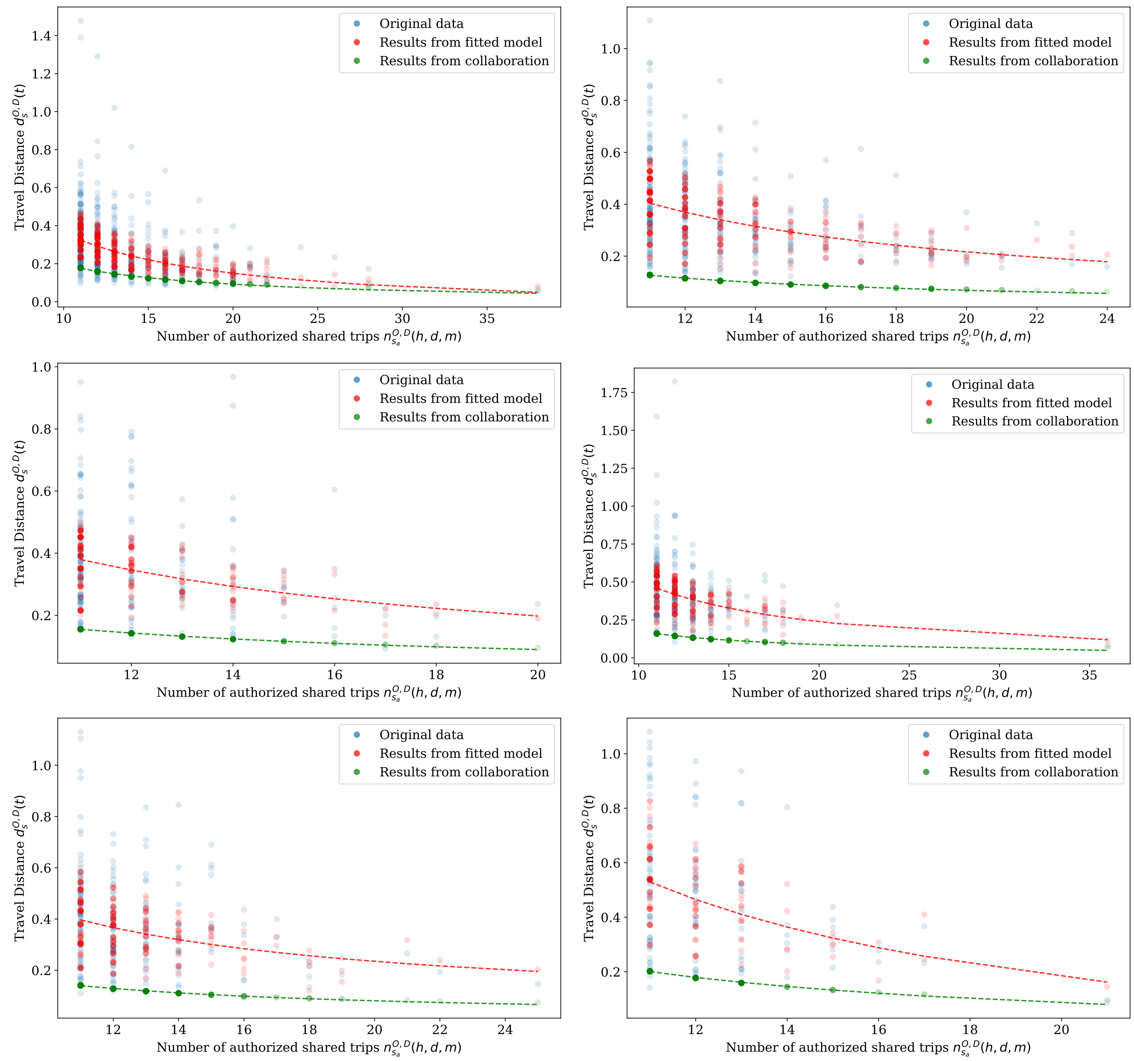


FIGURE 7: Improvement from collaboration on travel distance

in the ridesplitting service. In addition, TNC information associated with each trip is provided in the Manhattan “TLC" dataset. Using this information, this paper investigates the impact of competition, reflected by the concept of entropy, between TNCs on the scale effects. It is found that competition leads to a decrease in matched percentage and an increase in the travel distance for shared trips. When the entropy is higher, i.e., when competition is more intensive, the demand for shared trips received by each TNC tends to be lower, so the influence is more significant. Moreover, the improvement resulting from a “best-case scenario” for collaboration, under which all shared requests can be potentially matched regardless of the associated TNC, is estimated. The results manifest that the efficiency of the ridesplitting service can be improved by collaboration. The findings in this paper can be used by transportation agents and policymakers to improve the ridesplitting service through building and strengthening collaborations between TNCs.

Lyft discontinued its pooled service in May 2023, which is the most recent month that the trip data is currently available. After receiving more trip records in the future, it is essential to confirm if Lyft’s dropping from this service leads to the findings identified by this paper. Although the

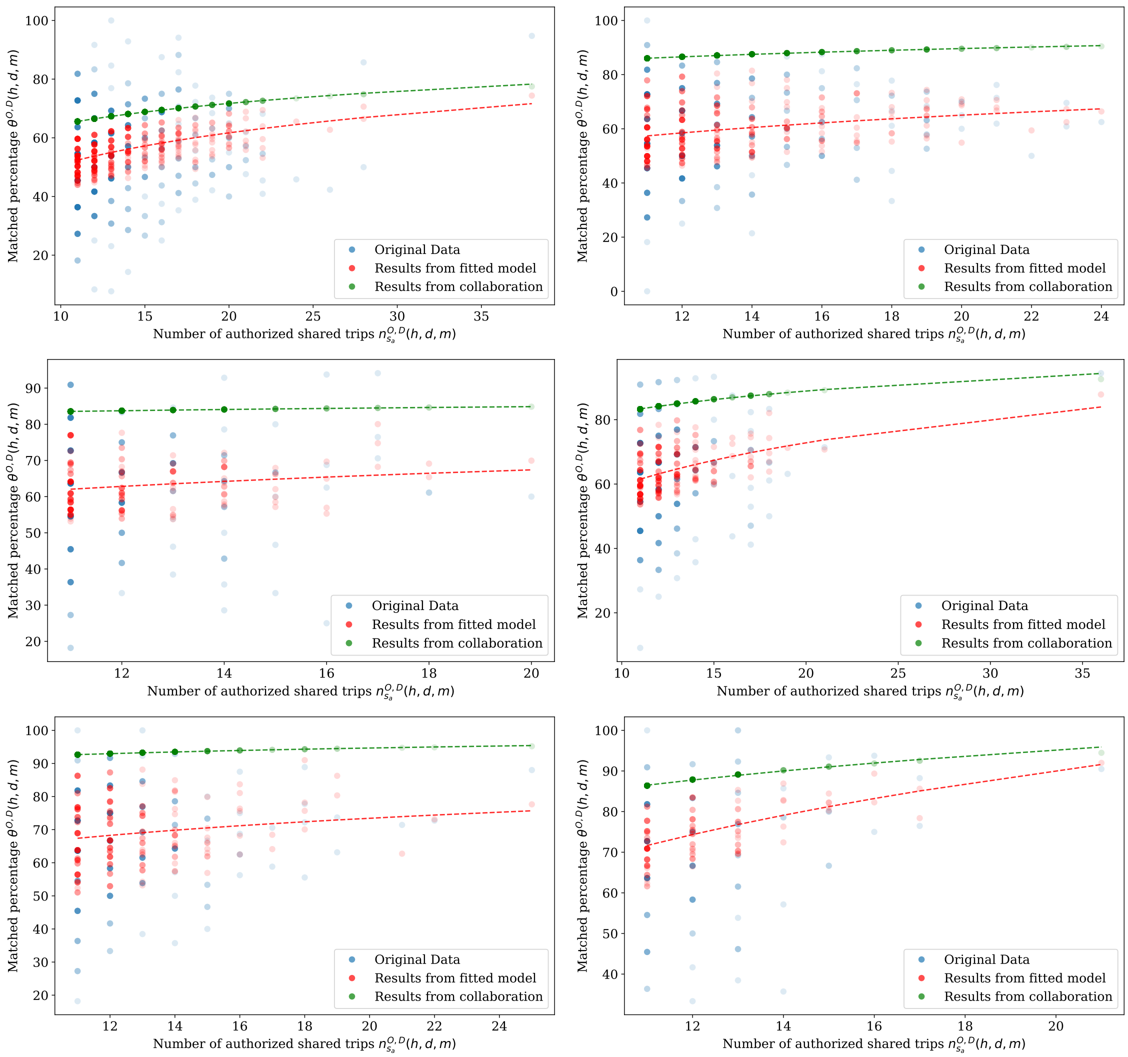


FIGURE 8: Improvement from collaboration on matched percentage

general findings about the influence of competition and collaboration on the ridesplitting service are interesting, the actual improvement highly depends on the matching algorithms and policies. Therefore, it is promising to propose a practical, sustainable, and beneficial policy for the realization of collaboration. Furthermore, it is interesting to investigate if similar findings exist during and after the re-introduction processing of ridesplitting from the COVID pandemic.

ACKNOWLEDGEMENTS

This research was supported by NSF Grants CMMI-2052337 and CMMI-2052512.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: H. Liu, S. Devunuri, L. Lehe, V. Gayah; data collection: H. Liu, S. Devunuri; analysis and interpretation of results: H. Liu, S. Devunuri, X. Dong, L. Lehe, V. Gayah; draft manuscript preparation: H. Liu, S. Devunuri, X. Dong, L. Lehe, V. Gayah. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Machado, C. A. S., N. P. M. de Salles Hue, F. T. Berssaneti, and J. A. Quintanilha, An overview of shared mobility. *Sustainability*, Vol. 10, No. 12, 2018, p. 4342.
2. SAE International, *Taxonomy and Definitions for Terms Related to Shared Mobility and Enabling Technologies*. SAE International, 2018.
3. Shaheen, S. and A. Cohen, Shared ride services in North America: definitions, impacts, and the future of pooling. *Transport Reviews*, Vol. 39, No. 4, 2019, pp. 427–442.
4. Lehe, L., S. Devunuri, J. Rondan, and A. Pandey, *Taxation of Ride-hailing*. Illinois Center for Transportation, Rantoul, 2021.
5. Li, W., Z. Pu, Y. Li, and X. J. Ban, Characterization of ridesplitting based on observed data: A case study of Chengdu, China. *Transportation Research Part C: Emerging Technologies*, Vol. 100, 2019, pp. 330–353.
6. Shaheen, S., *Shared mobility: the potential of ridehailing and pooling*. Springer, 2018.
7. Daganzo, C. F., Y. Ouyang, and H. Yang, Analysis of ride-sharing with service time and detour guarantees. *Transportation Research Part B: Methodological*, Vol. 140, 2020, pp. 130–150.
8. Ke, J., H. Yang, and Z. Zheng, On ride-pooling and traffic congestion. *Transportation Research Part B: Methodological*, Vol. 142, 2020, pp. 213–231.
9. Lehe, L., V. V. Gayah, and A. Pandey, Increasing returns to scale in carpool matching: Evidence from Scoop. *Transport findings*, 2021.
10. Tachet, R., O. Sagarra, P. Santi, G. Resta, M. Szell, S. H. Strogatz, and C. Ratti, Scaling law of urban ride sharing. *Scientific reports*, Vol. 7, No. 1, 2017, pp. 1–6.
11. Fielbaum, A., A. Tirachini, and J. Alonso-Mora, Economies and diseconomies of scale in on-demand ridepooling systems. *Economics of Transportation*, Vol. 34, 2023, p. 100313.
12. Kaddoura, I. and T. Schlenther, The impact of trip density on the fleet size and pooling rate of ride-hailing services: A simulation study. *Procedia Computer Science*, Vol. 184, 2021, pp. 674–679.
13. Liu, H., S. Devunuri, L. Lehe, and V. V. Gayah, Scale effects in ridesplitting: A case study of the City of Chicago. *Transportation Research Part A: Policy and Practice*, Vol. 173, 2023, p. 103690.
14. Silvestre, J., Economies and diseconomies of scale. *The New Palgrave: A Dictionary of Economics*, Vol. 2, 1987, pp. 80–84.
15. Vassilakis, S., Increasing Returns to Scale. In *The New Palgrave Dictionary of Economics*, Palgrave Macmillan UK, London, 1987, pp. 1–8.
16. Hryhoryeva, M. and L. Leclercq, Competition in Ride-Hailing Service Operations: Impacts on Travel Distances and Service Performance. *Transportation Research Record*, 2023, p. 03611981231171155.
17. Godfrey, J., The mechanism of a road network. *Traffic Engineering & Control*, Vol. 8, No. 8, 1969.
18. Daganzo, C. F., Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological*, Vol. 41, No. 1, 2007, pp. 49–62.
19. Geroliminis, N. and C. F. Daganzo, Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, Vol. 42, No. 9, 2008, pp. 759–770.

20. Zha, L., Y. Yin, and H. Yang, Economic analysis of ride-sourcing markets. *Transportation Research Part C: Emerging Technologies*, Vol. 71, 2016, pp. 249–266.
21. Jeitschko, T. D. and M. J. Tremblay, Platform competition with endogenous homing. *International Economic Review*, Vol. 61, No. 3, 2020, pp. 1281–1305.
22. Ni, L., C. Chen, X. C. Wang, and X. M. Chen, Modeling network equilibrium of competitive ride-sourcing market with heterogeneous transportation network companies. *Transportation Research Part C: Emerging Technologies*, Vol. 130, 2021, p. 103277.
23. Castillo, J. C. and S. Mathur, Matching and Network Effects in Ride-Hailing. In *AEA Papers and Proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 2023, Vol. 113, pp. 244–247.
24. Cattaneo, M. D., R. K. Crump, M. H. Farrell, and Y. Feng, Binscatter regressions. *arXiv preprint arXiv:1902.09615*, 2019.
25. Lin, J., Divergence measures based on the Shannon entropy. *IEEE Transactions on Information theory*, Vol. 37, No. 1, 1991, pp. 145–151.
26. Wang, D.-H., R.-H. Yao, and C. Jing, Entropy models of trip distribution. *Journal of Urban planning and development*, Vol. 132, No. 1, 2006, pp. 29–35.
27. Ren, Z., G. Fusco, N. Lownes, and J. Zhu, Entropy-based diversity quantification of multimodal transportation systems: Physical infrastructure perspective versus travel behavior perspective. *Journal of Urban Planning and Development*, Vol. 148, No. 3, 2022, p. 04022029.