Transportation Research Part C 79 (2017) 242–256

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

How can the taxi industry survive the tide of ridesourcing? Evidence from Shenzhen, China

Yu (Marco) Nie

Department of Civil and Environmental Engineering, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, United States

ARTICLE INFO

Article history:
Received 30 September 2016
Received in revised form 22 March 2017
Accepted 25 March 2017
Available online 2 April 2017

Keywords:
Ridesourcing
e-hailing
Street-hailing
Capacity utilization rate
Transportation network company

ABSTRACT

This paper aims to examine the impact of ridesourcing on the taxi industry and explore where, when and how taxis can compete more effectively. To this end a large taxi GPS trajectory data set collected in Shenzhen, China is mined and more than 2,700 taxis (or about 18% of all registered in the city) are tracked in a period of three years, from January 2013 to November 2015, when both e-hailing and ridesourcing were rapidly spreading in the city. The long sequence of GPS data points is first broken into separate "trips", each corresponding to a unique passenger state, an origin/destination zone, and a starting/ending time. By examining the trip statistics, we found that: (1) the taxi industry in Shenzhen has experienced a significant loss in its ridership that can be indisputably credited to the competition from ridesourcing. Yet, the evidence is also strong that the shock was relatively short and that the loss of the taxi industry had begun to stabilize since the second half of 2015; (2) taxis are found to compete more effectively with ridesourcing in peak period (6–10 AM, 5–8 PM) and in areas with high population density. (3) e-hailing helps lift the capacity utilization rate of taxis. Yet, the gains are generally modest except for the off-peak period, and excessive competition can lead to severely under-utilized capacities; and (4) ridesourcing worsens congestion for taxis in the city, but the impact was relatively mild. We conclude that a dedicated service fleet with exclusive street-hailing access will continue to co-exist with ridesourcing and that regulations are needed to ensure this market operate properly.

1. Introduction

The spectacular rise of *ridesourcing* is probably the most significant disruption occurred to the personal mobility market in decades. Offered mostly via mobile platforms, a ridesourcing service connects passengers to rides provided by private drivers using personal vehicles. The process that matches passengers and drivers on-line and in real-time is often called *e-hailing*, in contrast to traditional *street-hailing*. While those that offer ridesouring service - the likes of Uber, Lyft and Didi Chuxing, known as *Transportation Network Companies* (TNC) - are still locked in intense battles with each other around the world, it seems at least to some that collectively they have dug the grave for the once formidable rival: the traditional taxi industry (Oremus, 2016).

Mounting evidence suggests that the taxi industry has indeed suffered great losses in market share, revenue, workforce and asset. In Los Angeles, the annual number of taxi trips has plummeted from 8.4 million in 2013 to 6.0 million in 2015, a nearly 30% fall in less than three years (Nelson, 2016). The taxi industry in San Francisco, where both Uber and Lyft are headquartered, lost almost two thirds of its market share between 2012 and 2014 (Davidson, 2014). Not surprisingly, the Yellow

E-mail address: y-nie@northwestern.edu

http://dx.doi.org/10.1016/j.trc.2017.03.017
0968-090X/

Cab Inc. in San Francisco, the largest operator in the city, had to file bankruptcy protection in early 2016 due to "serious financial setbacks" (Corrigan, 2016). On the east coast of the US, the price for a yellow taxi medallion in New York City (NYC) has been cut in half since May 2013, when it was valued roughly at $1.32 million (Zuylen-Wood, 2015). Some taxi dispatchers, such as McGuinness Management, has seen almost half of their medallions become idle due to the lack of drivers, creating so-called "taxi graveyard" in the city (Whitford, 2015).[1]

Despite the gloomy picture, it may be premature to declare that the future of the taxi industry is all but doomed. Bershidsky (2015) observed that TNCs' success so far is due in large part to the aggressive pricing strategy that cannot be sustained in the long term.[2] At the end of the day, he argued, "the survival of licensed, regulated cabs is the only safeguard against" the potential monopoly of the winer emerged from the TNCs' hunger game. Steier (2015) noted that the demand for UberX in NYC may have peaked, based on the fact that only about 2,000 UberX drivers (out of more than 20,000 registered) were serving its CBD area between 7 AM and 7 PM from June through July 19 of 2015. He suggested that the ability to offer the old fashioned street-hailing is an important advantage held by traditional taxis, and predicted that the loss of the taxi industry should begin to level off. Newman (2016) observed that the pace of declining taxi ridership in NYC might have been slowed. The official data published by NYC's Taxi and Limousine Commissions (see Fig. 1) show that the taxi industry in the city has lost about 25% ridership since 2012.[3] Fig. 1 offers no compelling evidence for market stabilization, however, even though the first half of 2016 did bounce back a bit more from the second half of 2015, compared to a year ago (marked by arrows in the plot).

One naturally wonders what TNCs' expansion looks elsewhere in the world. More intriguing questions have to do with the underlying mechanisms that drive the interactions between the ridesourcing and taxi services. For example, where and when does the traditional taxi service compete more effectively against ridesourcing, and hence it may more easily retain market share in those market segments? To what extent can e-hailing applications help the taxi industry counteract the competitive edge of the TNCs?

To explore answers to these questions this paper mines a large taxi GPS trajectory data set collected in Shenzhen, China. We track more than 2700 taxis (or about 18% of all registered in the city) in a period of three years, from January 2013 to November 2015 and examine how various aspects of their operations are affected, temporally and spatially, by e-hailing and ridesourcing. Our results suggest that ridesourcing inflicted a disruption of similar scale as in NYC on the Shenzhen taxi industry, but it struck in a much shorter time period, with the average taxi ridership falling about 25% in less than a year. Interestingly, the taxi industry there has already begun to stabilize since the second half of 2015, according to our analysis. In general, the insights from our analysis agree that a dedicated taxi fleet equipped with exclusive right to street-hailing and e-haling should and will continue to exist, despite the strong competition from ridesourcing. In the near future, before autonomous driving wipes out human drivers, the personal mobility market is likely to benefit from a mixed supply model with both dedicated and part-time drivers. It is city managers' job to determine how to best regulate this market for the collective good of the society. While directly informing such policy making is beyond the scope of this paper, the empirical evidence and analysis presented herein could help guide the modeling process in due course.

In what follows, Section 2 briefly reviews related studies and Section 3 presents the details of the taxi data set and shows how it is processed to generate useful results. Sections 4–6 report findings from the data: Section 4 focuses on outputs measured by average hourly ridership and distance/time travelled, Section 5 discusses productivity and Section 6 examines spatial heterogeneity. The last section elaborates and analyzes the findings, and offers concluding remarks.

2. Literature

Due to its peculiar behavior,[4] the traditional taxi market has attracted ample attention from economists and transportation analysts since 1970s (see e.g. Douglas, 1972; De Vany, 1975; Beesley and Glaister, 1983; Arnott, 1996; Cairns and Liston-Heyes, 1996; Yang et al., 2002; Flores-Guri, 2003; Yang et al., 2005; Moore, 2006; Schaller, 2007; Yang et al., 2010). Most of these efforts focus on modeling market equilibrium and argue for or against the tight regulations imposed on these markets, such as entry control and price ceiling. Surprisingly, according to our search of literature, scholarly research on ridesourcing remains relatively scarce, despite the obsessive media coverage in recent years.

Santi et al. (2014) performed a simulation study based on a dataset of 150 million taxi trips in NYC, and found a large portion of the trips are routinely shareable. They show that with a modest increase in passenger discomfort (in terms of extra waiting and riding times), the cumulative vehicles miles travelled (VMT) could be reduced by 40% or more. A survey study conducted in San Francisco by Rayle et al. (2016) shows that at least half of the ridesourcing trip replace a non-taxi trip, indicating that the markets of two services have overlaps but also significant differences. They also found that ridesourcing consistently outperforms taxis in terms of waiting time. Hughes and MacKenzie (2016) tracked a UberX vehicle for two months through Uber's developer API, and generated a GPS trajectory sample with more than one million data points. After correlating the waiting time data with the socioeconomic variables in each zone, they found that (1) the waiting times for the

[1] For what is worth, this author was told similar stories by taxi drivers when traveling in Chengdu, China in the summer of 2016.

[2] For example, Uber lost more than $1 billion in the first half of 2016 largely because the pricing war it waged with competitors - see e.g. http://www.nytimes.com/2016/08/26/technology/how-uber-lost-more-than-1-billion-in-the-first-half-of-2016.html?_r=0.

[3] As a side but interesting note, the ridership in the second half of a year is always significantly lower than that in the first half, likely due to Uber's appeal to visitors during summer break and winter holidays.

[4] Specifically, the fact that the vacant taxi operating hours is both a blessing to the level of service and a waste for the service provider.

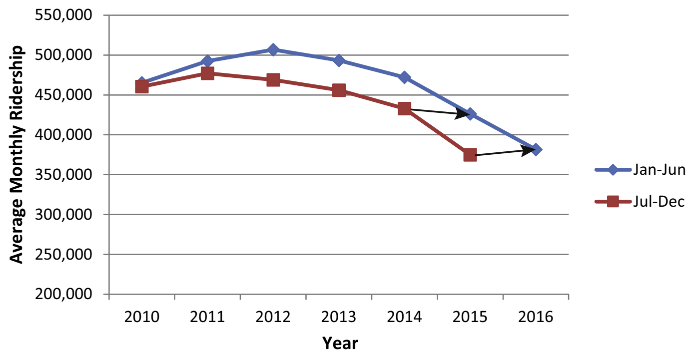


Fig. 1. Official taxi ridership data published by NYC's Taxi and Limousine Commissions. *Source:* http://www.nyc.gov/html/tlc/html/about/statistics.shtml..

UberX service are lower in denser urban areas; and (2) the percentage of minorities or low income residents in a zone does not seem to affect the waiting time negatively. Cramer and Krueger (2016) measured and compared the capacity utilization rate (the ratio between time/distance travelled with passenger and the total time/distance, a commonly used productivity index) of both UberX and taxi drivers in several cities. They concluded that UberX drivers' capacity utilization rate is about 30–50 percent high than taxi drivers when measured in both time and distance, thanks to (1) e-hailing; (2) Uber's scale; (3) Uber's flexible supply model including surge pricing; and (4) inefficient taxi regulations.

There have been even less efforts in the literature on modeling ridesourcing-taxi market. He and Shen (2015) extended the spatial taxi equilibrium model of Yang and Wong (1998) to analyze the impact of e-hailing on the regulated taxi market. Their analysis shows that e-hailing helps reduce taxi waiting time and increase the capacity utilization rate. Wang et al. (2016) modeled the taxi market with a single e-hailing platform using the classical market equilibrium approach, with a focus on examining the impacts of the platform's pricing strategies on the taxi market performance. Zha et al. (2016) conducted an economic analysis of a personal mobility market with only ridesourcing service, with the assumption that customers and drivers are matched through an exogenously defined function. While the model led to several interesting analytical results, it was not calibrated to real data, nor it considers the co-existence of ridesourcing and taxi.

To the best of our knowledge, few had characterized the impact of ridesourcing on the taxi industry, or modeled their interactions. After all, ridesourcing is a recent phenomenon, and, thanks to the private nature of TNCs, most researchers have rather limited access to detailed data about ridesourcing services. In order to bypass this data availability issue, we will focus on aggregated changes in the operational patterns of a fixed taxi fleet, as detailed in what follows.

3. Data

3.1. Raw COST data

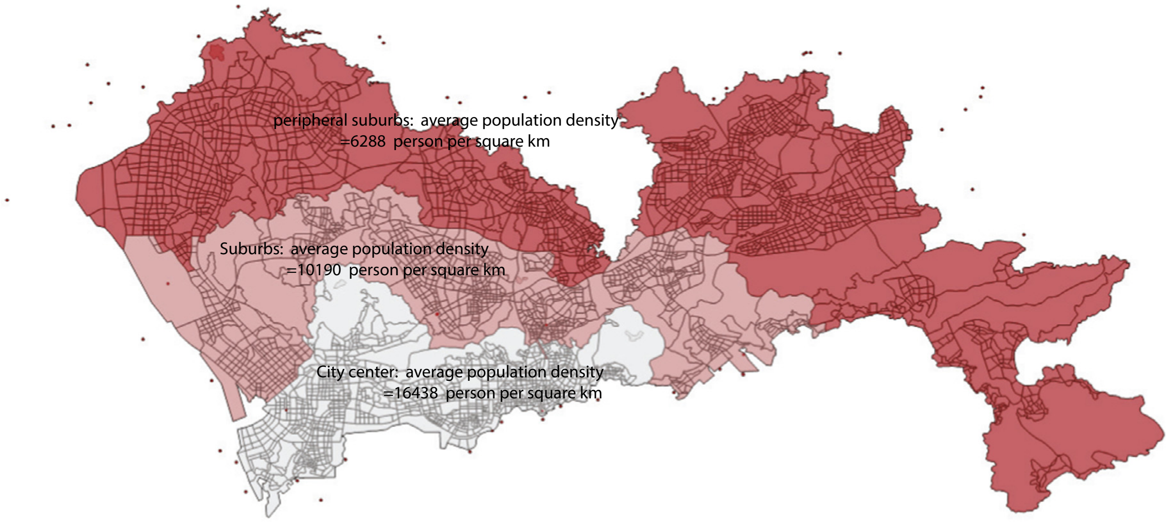
The taxi data set used in this study is a derived subset from the "City Of Shenzhen Taxi" (COST) database, which includes ten separate raw GPS trajectory data sets, each corresponding to a unique and continuous period between 2011 and 2015, as shown in Table 1.

The second column in Table 1 reports the period of time during which the data in each set were collected. Note that the data in Set 1 were collected in 31 consecutive days whereas those of all other sets were collected in seven consecutive days. For each set, the total number of all taxis that appear in the corresponding time period is given in the third column of the table. We note that each data set is nearly a full sample of all taxis registered in the city. The fourth and fifth columns in Table 1 give the total number of all GPS points and of GPS points with passengers, respectively. There are substantial variations in the sampling interval of the GPS trajectory in COST data. In general, it varies between 10 and 30 s; but it can be as long as 40–60 s for some periods. The average sampling interval is about 22 s in Set 1, and it gradually decreases and stabilizes around 12–14 s in later years.

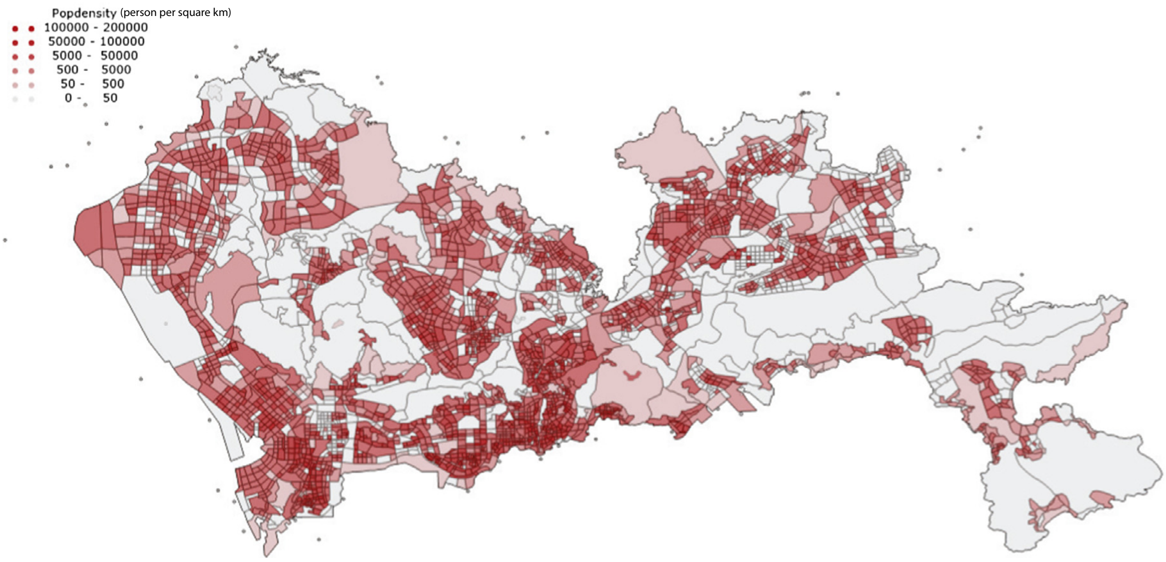
Along with the COST data, we also obtained the GIS data of the City of Shenzhen. According to Shenzhen Urban Transport Planning Center, the city is divided into 3561 small traffic analysis zones (TAZ) for the purpose of urban travel forecasting. The center classifies these zones into four types (see Fig. 2(a)): city center (average population density = 16,438 per/km^2), suburbs (average population density = 10,190 per/km^2), peripheral suburbs (average population density = 6288 per/km^2) and external zones (shown as dots in the plot). The population data is shared by the center as one of the TAZ attributes. It is unclear whether temporary workers are counted, but this detail need not concern us, because it is not expected to change the overall relative density of these areas. Fig. 2(b) visualizes the population density of TAZs using a coloring scheme.

Table 1
Description of COST database.

Set	Period	#Taxis	Total GPS points (million)	Total GPS points with passengers (million)
1	1/1/2011–1/31/2011	14171	658.71	242.98
2	1/15/2012–1/21/2012	13526	142.90	54.54
3	1/15/2013–1/21/2013	14514	311.74	135.83
4	1/15/2014–1/21/2014	14820	337.06	164.50
5	1/15/2015–1/21/2015	15726	420.19	194.13
6	3/25/2015–3/31/2015	15665	389.52	183.37
7	5/25/2015–5/31/2015	15414	369.50	174.33
8	7/25/2015–7/31/2015	15608	334.42	145.19
9	9/24/2015–9/30/2015	15564	358.16	151.78
10	11/24/2015–11/30/2015	15425	352.61	148.84



(a) Shenzhen traffic analysis zones (3561 in total, classified in four differen types)



(b) Illustration of TAZ population density

Fig. 2. Basic GIS data: the city of Shenzhen.

3.2. Trip identification

In each data set, the raw COST data gives a long sequence of GPS points, each with a time stamp and a passenger state (on or off). Trip identification aims to break this long sequence of data points into separate "trips". Each trip corresponds to a unique passenger state, an origin TAZ, a destination TAZ, a starting and an ending time. We emphasize that all GPS points associated with a trip must have the same passenger state. In the following, a trip with (without) a passenger will be called an occupied (vacant) trip.

The first criterion we use to break a GPS data sequence is to detect a switch in the passenger state. When two consecutive points in the sequence have different states, we assume that the previous trip ends and the new trip starts at the center of the straight line that connects the two points (see Fig. 3(a)).

The complexity arises when the elapsed time between two points are significantly longer than the average sampling interval (90 s is used as the threshold in our data processing). In some cases, such a long sampling interval simply means a temporal malfunction of the device (e.g. loss of GPS signals in urban areas), whereas in other cases, it could indicate that the taxi was out of service.

In this study, when the sampling interval is longer than 5 min, we assume that taxi was out of service. Accordingly, the segment between the two points would be excluded from consideration (see Fig. 3(d)). If the sampling interval ranges between 90 and 300 s, we check the distance between the two points in order to determine if the taxi is out of service. If the distance is very short (<50 feet), we conclude that the taxis did not move much, and hence the segment between the two points is discarded (see Fig. 3(c); otherwise, we assume that the GPS signal was just temporarily lost and hence the trip ends are still placed in the middle, as in Fig. 3(b).

When two consecutive points have the same passenger state, we break them only in the following cases:

1. If their difference in time stamp exceeds 5 min (see Fig. 4(a)), which means no data was recorded for more than 5 min.
2. If the time difference falls into the range between 90 and 300 s, the distance between the two points is shorter than 50 ft, and both points' passenger state is off (see Fig. 4(b)). In other words, if a vacant taxi is idle for a relatively long time during which no data was recorded, it is considered to be on break during that time period.

For the second case above, we require the passenger state to be "off" to break the two points into two separate trips. Dropping off one passenger and picking up another within five minutes does happen at hot spots, but the likelihood that it takes place concurrently with an unusually long sampling interval becomes quite small.

3.3. Study data set

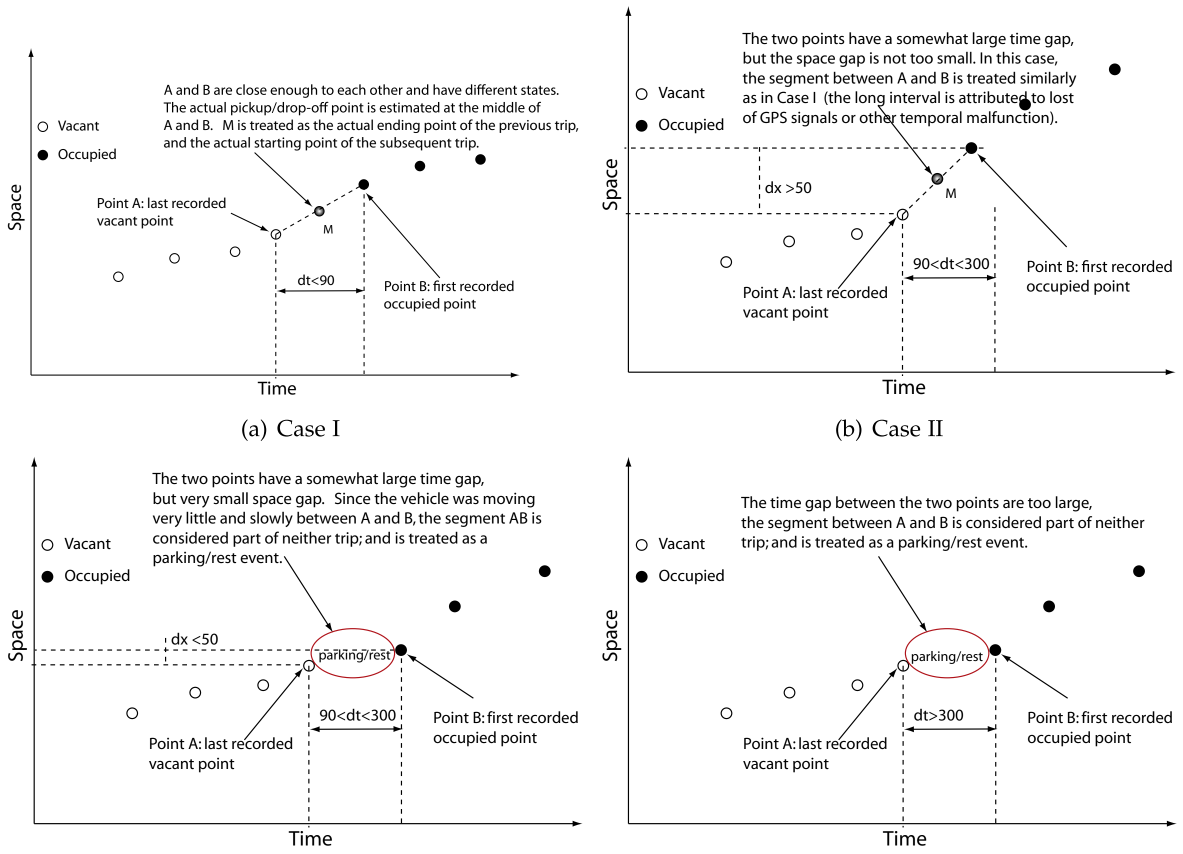
Since our focus is to examine the impact of e-hailing and ridesourcing, it is useful to first review the major events leading to the introduction of these technologies in the City of Shenzhen.

Didi Chuxing Inc., founded in Beijing in 2012 as an e-hailing platform (known as Didi Dache) exclusively serving taxi drivers, launched its service in Shenzhen in November 2012.[5] By October 2013, Didi Dache had seized more than half of the e-hailing market share, and by March 2014, it reportedly had been processing more than five million orders a day in China, with over 100 million users and one million drivers.[6] The year of 2014 also saw the intensifying competition between Didi Dache and its chief rival, Kuaidi Dache.[7] Around the time when Didi Dache and Kuaidi Dache began to dominate the e-hailing market in China, Uber, the best known ridesourcing platform, started its China operation and launched service in Shenzhen in November 2013.[8] In August 2014, Didi's own ridesourcing platform, called Didi Zhuanche, was launched, and since then it has grown explosively.[9] Didi Chuxing eventually merged with Kuaidi Dache in Feburary 2015, and with Uber China in August 2016.

Another relevant event is that, in December 2014, the city of Shenzhen banned private vehicles that are not registered in the city from driving within the city center (the restriction was later expanded to the entire city) during the rush hour (7–9 AM and 5:30–7:30 PM).[10]

From the above chain of events it is reasonable to expect that the effect of e-hailing would reach the taxi industry starting from 2013, and ridesourcing would begin to impact the market starting from the end of 2014. Therefore, Sets 3–10 in COST data (see Table 1) were chosen for this study, which together cover eight episodes between January 2013 and November 2015. Every taxi recorded in *all* eight data sets is included in the study except when taxi i does not satisfy the following qualification conditions for *any* of the eight data sets:

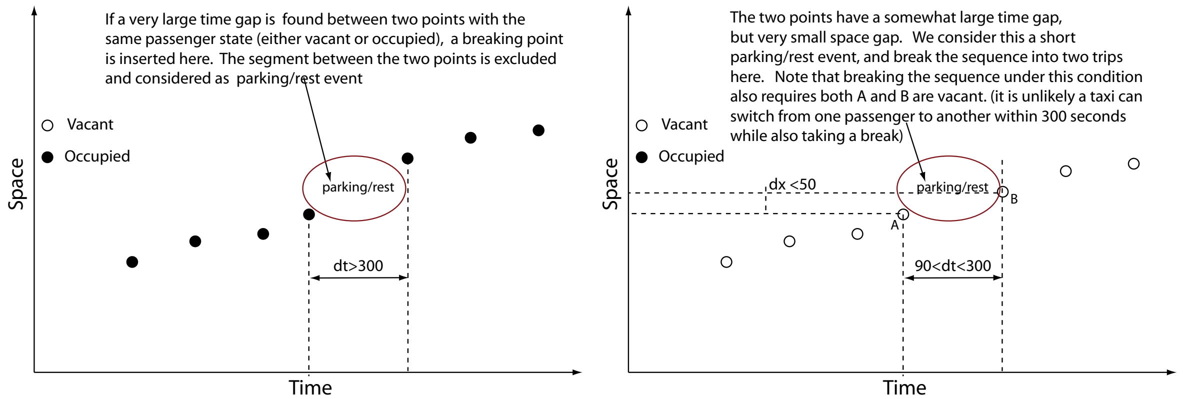
[5] http://business.sohu.com/20151118/n426841695.shtml, in Chinese.
[6] http://baike.baidu.com/view/10189534.htm, in Chinese.
[7] http://baike.baidu.com/view/10263782.htm.
[8] https://www.techinasia.com/uber-china-launches-in-shenzhen.
[9] The users of its smart phone application has increased five times between October 2014 and Feburary 2015, according to http://36kr.com/p/5042193.html, in Chinese.
[10] see http://www.bitenews.cn/szzc/4189.html.



(a) Case I (b) Case II

(c) Case III (d) Case IV

Fig. 3. Trip identification: breaking trips when the passenger state changes.



(a) Case I (b) Case II

Fig. 4. Trip identification: breaking trips when the passenger state does not change.

$$N_i^t > 10,000; \quad 0.05 < \frac{N_i^p}{N_i^t} < 0.95, \tag{1}$$

where N_i^t is the total number of GPS points recorded by taxi i, and N_i^p is the number of those GPS points whose passenger state is on. Essentially, a taxi is disqualified if it has logged too few GPS points, or has an unusually high or low occupancy (defined as the ratio between the number of recorded GPS points with passenger and the number of all recorded GPS points).

A taxi with too few GPS points may be not active for the recording period, hence not representative of the behavior of a "typical taxi". A very low/high occupancy indicates that the passenger state of these taxis might not be properly captured either due to the malfunction of the data collection device, or because the taxi driver disables it. After removing these outliers, the number of qualified taxis included in the study is 2781, roughly 18% of all taxis that register in the city in 2015.

Finally, to avoid the potential noises that weekend travel might create, we only consider data collected in three weekdays: Tuesday, Wednesday and Thursday.

4. Output analysis

We first examine how the overall outputs of the sample taxi fleet vary across the sampling periods. In order to capture the time-of-day effect, all data sets are further disaggregated based on three time periods: peak period (6 AM - 10 AM and 5 PM - 8 PM); mid-of-day period (10 AM - 5 PM); off-peak period (8 PM - 6 AM). For comparison, an all-day period is also considered.

4.1. Average hourly trips

Fig. 5 shows how the average number of hourly trips changes over the time in each of the four periods. The first observation from these plots is that on average the number of a taxi's vacant trips roughly equals that of its occupied trips. The only noticeable deviation between the two occurs in the off-peak period - a possible explanation is that taxi drivers tend to stop more frequently for rest at night when they are looking for business, and hence more vacant trips are identified by the proposed method (cf. Fig. 4(b)).

Fig. 5(d) shows that the average hourly occupied trips (AHOT) of the fleet peaked in January 2013 (at 2.13 trips/hour). It dropped about 5% to 2.03 trips/hour in January 2014, bounced back to 2.04 trips/hour in January 2015, and then began to fall rapidly until it appears to hit the bottom (1.63 trips/hour) in September 2015. The beginning of 2015 is evidently a tipping point, which correlates perfectly with the introduction of Didi Zhuanche (launched in October 2014), and the explosive increase in Didi's app users (between October 2014 and February 2015), due in part to the merge with Kuaidi Dache. Another tipping point occurs around September of 2015, when the sign of stabilization begins to emerge.

It is tempting to attribute the small decrease from 2013 to 2014 to the competition from Uber, which entered the service in Shenzhen in November 2013. Yet, given its relatively small market share in China,[11] it is unlikely that Uber would create a 5% market disruption in 2014. After all, even in the peak of ridesourcing expansion (the first half of 2015), the taxi industry in Shenzhen lost only 20% of its business. We will return to this issue at the end of the next subsection, when we examine other output indexes. For now, let us first corroborate the findings revealed in our sample with other data sources.

The official taxi ridership data reported by Shenzhen Government (see Fig. 6) reveal a remarkably similar pattern as shown in Fig. 5(d). Specifically, the taxi ridership (1) had a quick and dramatic drop starting from early 2015, (2) reached the lowest point between August and November in that year, and (3) began to turn around since then (in fact the ridership in May and July 2016 has returned to their respective levels in 2015). At its lowest point (October 2015), the ridership was about 24% lower compared to the value in January 2015, compared to 20% recorded in our sample.[12] Fig. 6 also shows that the taxi ridership is relatively stable in 2013 and 2014, even though e-hailing must have gained much ground during that period and ridesourcing has already risen on the horizon.

Having examined the overall pattern, let us have a closer look at different periods. Figs. 5(a)–(c) show that, prior to the first tipping point, the mid-of-day period has the highest AHOT of about 2.4 trips/hour, compared to 2.0 trips/hour in the peak period and about 1.7 trips/hour in the off-peak period. Clearly, ridesourcing impacts taxi ridership in the three time periods very differently.

It seems that the mid-of-day taxi market is also the most vulnerable to the aggressive competition from ridesourcing. In November 2015, this market segment lost almost 40% of its ridership (from 2.5 trips/hour in January 2015 to 1.5 trips/hour), and unlike the other two periods, with no sign of a rebound in sight (see Fig. 5(b)).[13] In contrast, while the peak period taxi market was hit hard in the first half of 2015 (the maximum loss is about 30%), it turned around quickly. By November 2015, it has recovered more than 60% of the lost ridership. In the off-peak period, the taxi industry kept its ridership nearly intact until July 2015. While there was a sharp drop in September,[14] it quickly recovered 50% of the loss two months later.

---

[11] A report shows that Uber only has about 10% market share in China in the first quarter of 2015, see http://www.forbes.com/sites/liyanchen/2015/09/09/uber-wants-to-conquer-the-world-but-these-companies-are-fighting-back-map/.

[12] Our sample does not have data in October, and so the lowest point was recorded in September 2015.

[13] It is interesting to note that the mid-of-day period also features quite dramatic fluctuations (with March, July and November 2015 being in the valley and May and September at the peak) compared to the other two periods. We hypothesis that each valley on the curve is related to a major promotion campaign of ridesourcing platforms. Although promotions from both Uber China and Didi Chuxing are too numerous and ubiquitous to track precisely, we note that Uber did have a major price cut (up to 40%) in March 2015 and then a promotion that would give qualified riders up to five free trips in the week following Thanksgiving (November 26, 2015)(see http://www.ifanr.com/621527, in Chinese.). Both events fit well with the above hypothesis. Moreover, the demand for ridesourcing is generally less elastic to price in the peak and off-peak periods than in the mid-of-day period. Intuitively, people travel on taxis more (as revealed in the data), have more options, and are less concerned about traffic congestion during the mid-of-day period. This can explain why the peak-time and off-peak travel are much less prone to the fluctuations experienced by the mid-of-day period.

[14] This drop may be an outlier because it is difficult to explain, noting taxis actually did well in other two periods in this month.

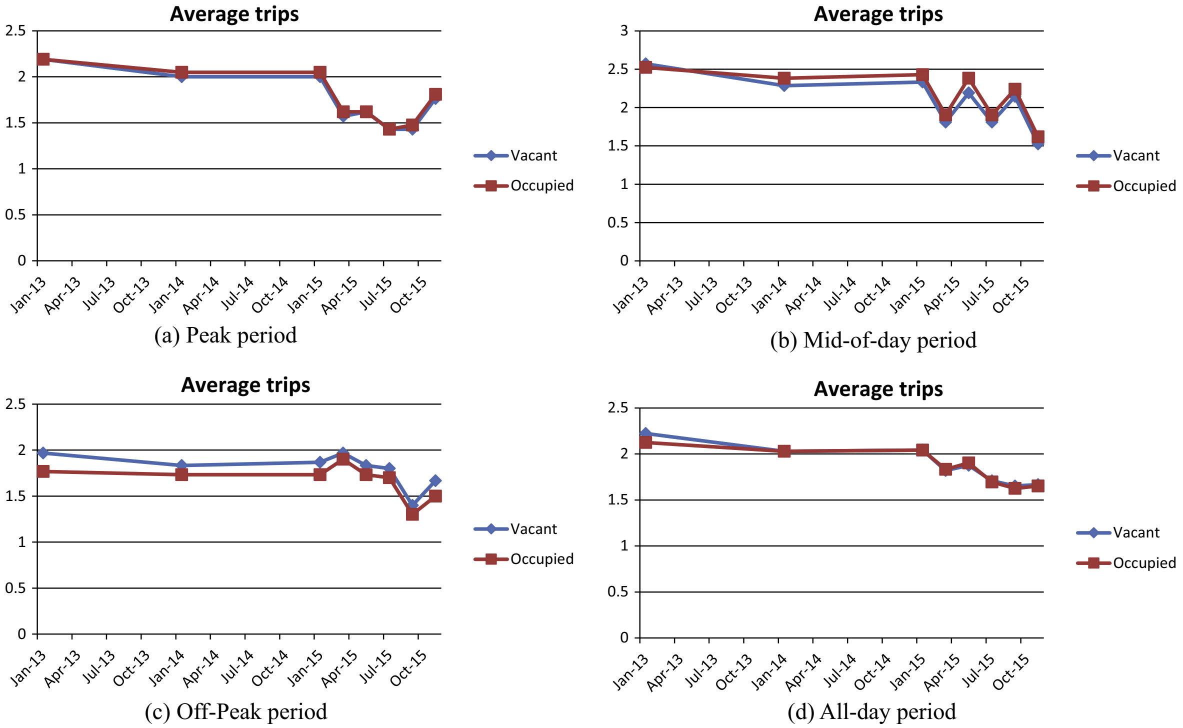


Fig. 5. Average hourly trips (weekday, size of the sample taxi fleet = 2781).

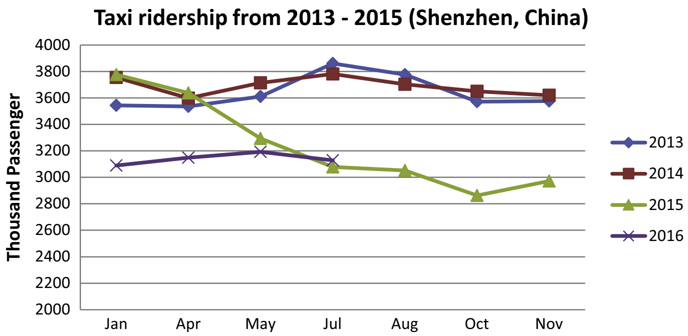


Fig. 6. Official taxi ridership data published by Shenzhen government agency. *Source:* http://www.sztb.gov.cn/pcjt/jbqk/yb/ in Chinese..

4.2. Average hourly distance and time travelled

We proceed to examine two other taxi output indexes: average hourly distance travelled (AHDT) and average hourly time travelled (AHTT), reported in Figs. 7 and 8 respectively. Overall, both AHDT and AHTT of occupied taxi trips follow a similar trend as AHOT (average hourly occupied trips), although there are some subtle differences. For the all day average, AHDT hit the lowest point in November 2015, about 24% lower than the value in January 2015, whereas AHTT dropped about 23% from January to November 2015. Interestingly, for the vacant trips, both indexes remain remarkably steady through the otherwise eventful year. On average, taxi drivers roughly drive about 4 miles and 22 min each hour looking for customers in 2015. Thus, while the competition from ridesourcing took away a large portion of their business, taxi drivers did not respond by substantially elevating search efforts. E-hailing, on the other hand, seems to help reduce the search effort - note that AHTT and AHDT of vacant trips were at about 5 miles and 26.5 min in January 2013, when Didi Chuxing was still in its infancy.

Differences can be found across the three periods. For example, the mid-of-day period has the highest AHDT of occupied trips (about 11 miles priori to the tipping point) whereas the off-peak has the lowest (about 8 miles priori to the tipping point). Also, in the off-peak period, drivers have to spend significantly more time (about 20–50% more) to search for customers than to actually drive them. This is not the case in the other two periods.

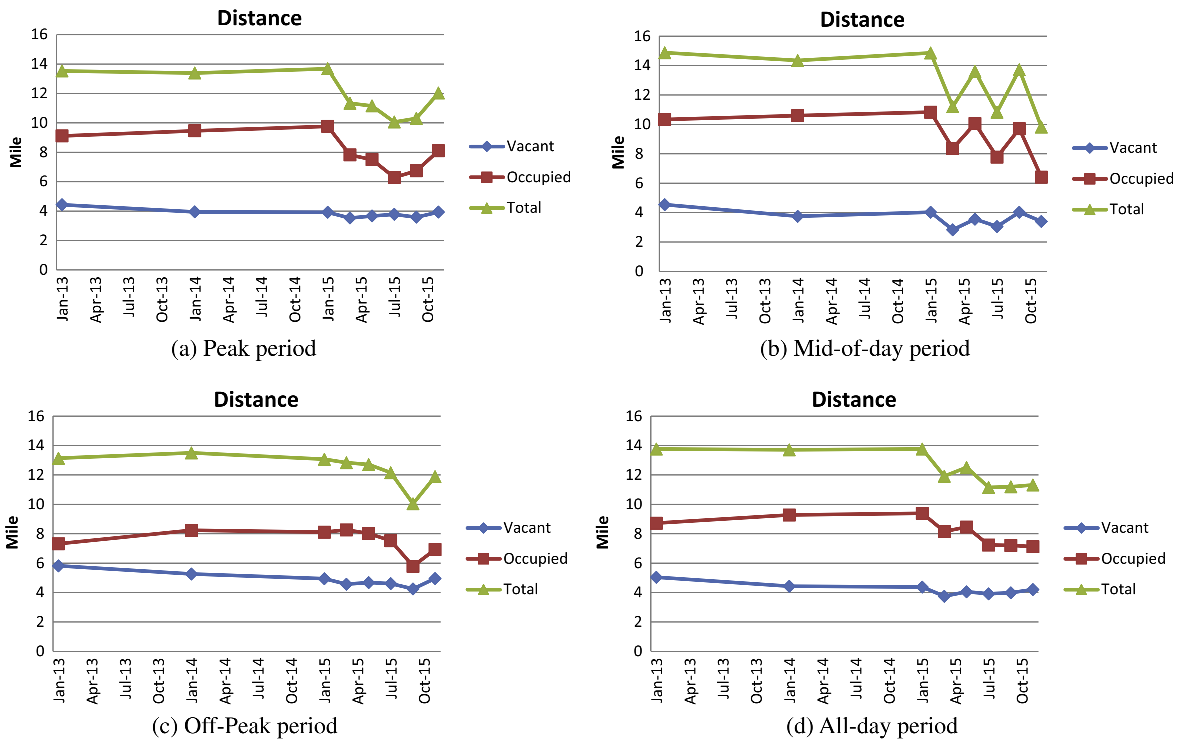


Fig. 7. Average hourly distance travelled (weekday, size of the sample taxi fleet = 2781).

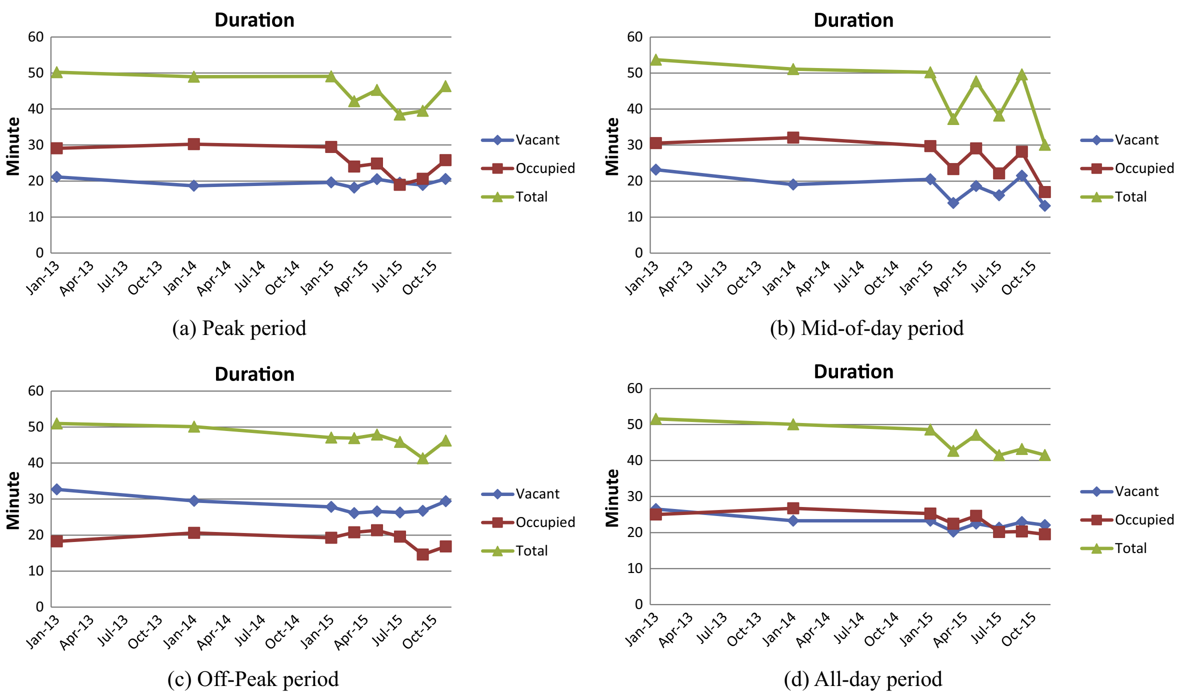


Fig. 8. Average hourly time travelled (weekday, size of the sample taxi fleet = 2781).

There are also similar patterns among the periods. For one thing, AHDT of vacant trips is only subject to minor variations across periods, especially in 2015. Somehow four miles is the average distance that taxis are willing to go in vain per hour in this city. Another phenomenon that is remarkably consistent across periods has to do with the change from 2013 to 2014. In all three periods, the data recorded gains on AHDT and AHTT of occupied trips, and losses on those of vacant trips. In other words, measured by distance travelled with passenger, the output of 2014 is better than that of 2013. This explains the abnormality pointed out before, which shows that AHOT decreases by 5% from 2013 to 2014. Taxis did seem to travel more profitable miles in 2014, but because on average these trips are longer, they ended up logging a slightly less number of trips. It seems reasonable to attribute the change from 2013 to 2014 to the positive effect of e-hailing, which not only increases the useful output, but also reduces the search cost.[15]

5. Productivity analysis

To measure the productivity of the taxi fleet, we employ two indexes: the average speed (defined as (total distance of all trips)/(total duration of all trips)) and the average capacity utilization rate by distance (CURD) and by time (CURT). CURD is defined as (AHDT by occupied trips)/(AHDT by all trips). The definition for CURT is slightly more complicated because the total working hours of a taxis driver may consist of not only the time spent on searching for customers and the time spent on serving them, but also the short break they take in between trips, which cannot be precisely measured using our method. To address this issue, we introduce an upper and lower bound for CURT, rather than presenting it as a unique value. Specifically

$$\text{CURTLB} = (\text{AHTT of occupied trips})/60 \tag{2}$$

$$\text{CURTUB} = (\text{AHTT of occupied trips})/(\text{AHTT of all trips}) \tag{3}$$

Note that 60 in the first equation above is the total number of minutes in an hour. Effectively, the lower bound assumes that drivers are always working in every minute of an hour. On the other hand, the upper bound assumes that drivers are always off when they are not driving. The true CURT should lie in between.

Fig. 9 reveals that the average speeds across all periods are significantly higher in January 2015 than January 2013 and 2014. This improvement is clearly related to the ban on the vehicles not registered in the city center during the peak periods, which was enacted in December 2014 (see http://www.bitenews.cn/szzc/4189.html.). Overall, the average speed of occupied trips in 2015 ranges between 23 and 25 miles per hour (mph) in the off-peak period, 20–22 mph in the mid-of-day period, and 18–20 mph in the peak period. For the vacant trips, as expected, the variations in the average speed is much smaller: it ranges between 9 and 12 in most cases, with the speed in the off-peak period being slightly lower. Thus, not only do taxis consume similar distance and time in searching for customers, they also do so traveling at a similar speed.

Compared to January, the average speeds in other months of 2015 are noticeably lower. The lowest recording is May, when the average speed is 20.6 mph, an 8% drop compared to January. The fact that the average speed profile perfectly correlates with the varying pattern of the taxi ridership indicates that ridesourcing was likely the cause for the elevated traffic congestion. We note that, however, the effect is relatively mild, with most recorded speeds within 5% from the highest point achieved in January. In fact, by November 2015, the average speed has risen to 21.9 mph, almost returning to the January level.

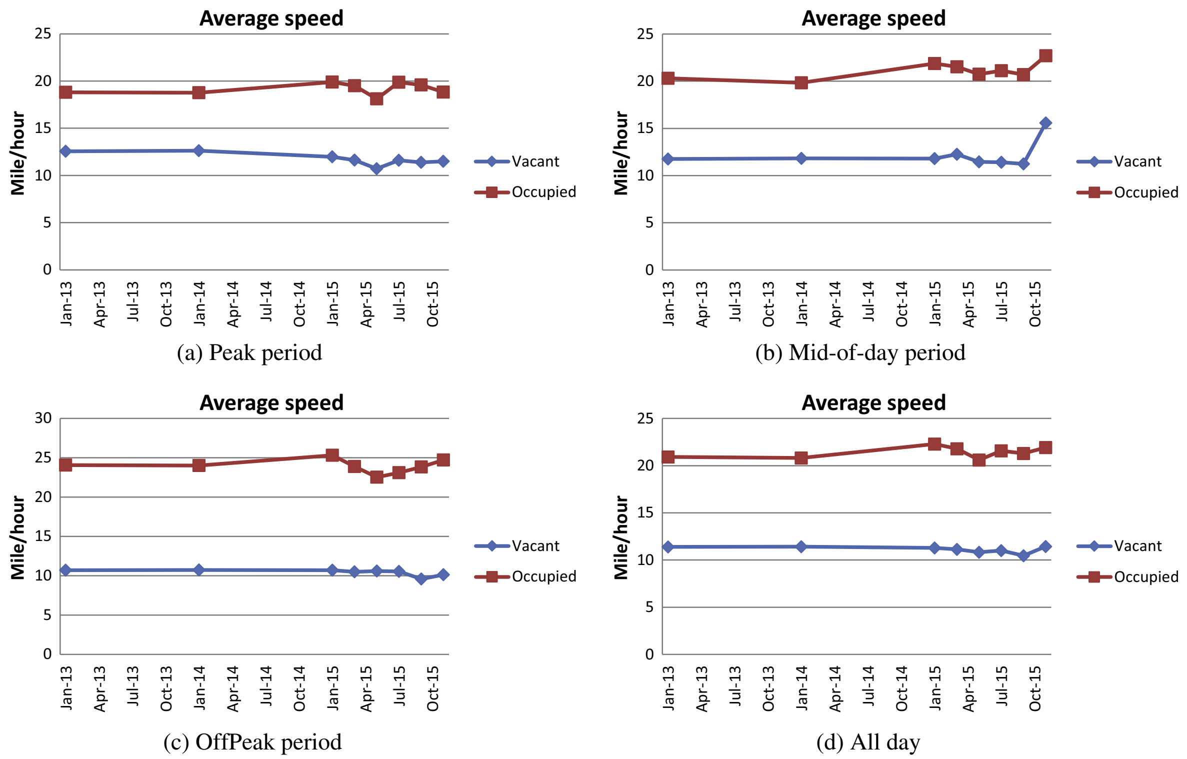
For cross validation, Fig. 10 plots the average weekday hourly speed using all occupied trips starting and ending within the city center, which is expected to be the more congested area of the three. In general, the sample captures the morning and evening rush hour patterns rather well. Importantly, the plot shows that, in city center, January has the highest average hourly speed between 9 AM - 12 PM (mid-of-day) and 8 PM - 11 PM (off-peak). Also, the congestion is clearly worse in July than in March.

Fig. 11 shows how the capacity utilization rate of the sample taxi fleet varies in the three years. Let us first focus on CURD. Averaged over the entire day (see Fig. 11(d)), CURD started at about 0.64 in 2013, peaked at 0.68 in early 2015, and gradually decreased until it arrived at 0.63 in November.[16] The mid-of-day period has the highest CURD (peaked at 0.75 in March 2015) whereas the off-peak has the lowest CURD (started from 0.56 in 2013, and peaked at 0.64 in March 2015).

Across all periods, the sample fleet has a consistent gain in CURD from January 2013 to January 2015. The largest increase, about 15%, occurs in the off-peak period. There seems no other logical explanation to this healthy growth in productivity (hence profitability) than the positive influence of e-hailing offered by TNCs. Yet, starting from January 2015, the upward trend of CURD was disrupted, likely due to the competition from ridesourcing. The peak period market was the first to be affected, followed by the mid-of-day and the off-peak periods. Of the three, the CURD in off-peak period is the least affected. It is also interesting to note that the mid-of-day period (and the off-peak period, to a lesser extent) is able to hold CURD steadily at a quite high level as late as July 2015, despite significant fluctuations in ridership (see Fig. 5(b)). In contrast, the varying pattern of CURD in the peak period correlates almost perfectly with that of the ridership. This finding reveals that taxi

[15] It is worth noting that the Chinese New Year was on January 31st in 2014. Also, the New Year Peak Travel Period officially started on January 16th, and the busiest travel days are 27th- 29th according to Baidu.com. The 2014 data used in this study was collected on 15th, 16th, and 21st. Thus, two thirds of the data were collected after the holiday travel has officially kicked in. One cannot rule out the possibility that the Chinese New Year travel might have affected the Shenzhen taxi market in January 2014, although gauging this effect is not trivial and beyond the scope of this paper.

[16] As a benchmark, Cramer and Krueger (2016) reported that taxis achieve a CURD of about 40% in both Los Angeles and Seattle, which is much lower than what is observed herein. The discrepancy may be caused by different operational characteristics of taxi services in US and China cities, and the different methods used to estimate CURD. It should be noted that Cramer and Krueger (2016) did not use detailed trajectory data in their computation.



(a) Peak period (b) Mid-of-day period

(c) OffPeak period (d) All day

Fig. 9. Average weekday speed for COST data set 3–10 (taxi sample size: 2781).

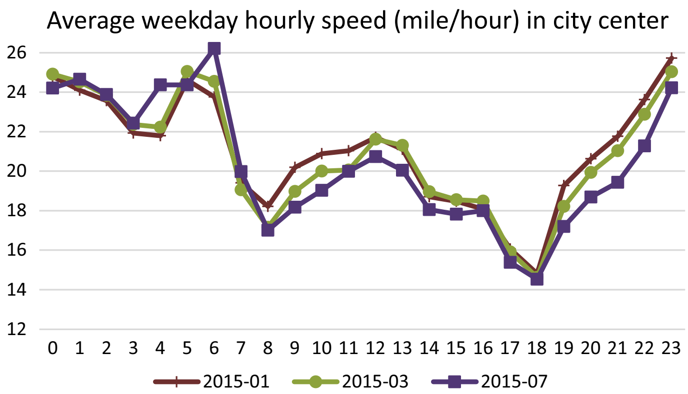
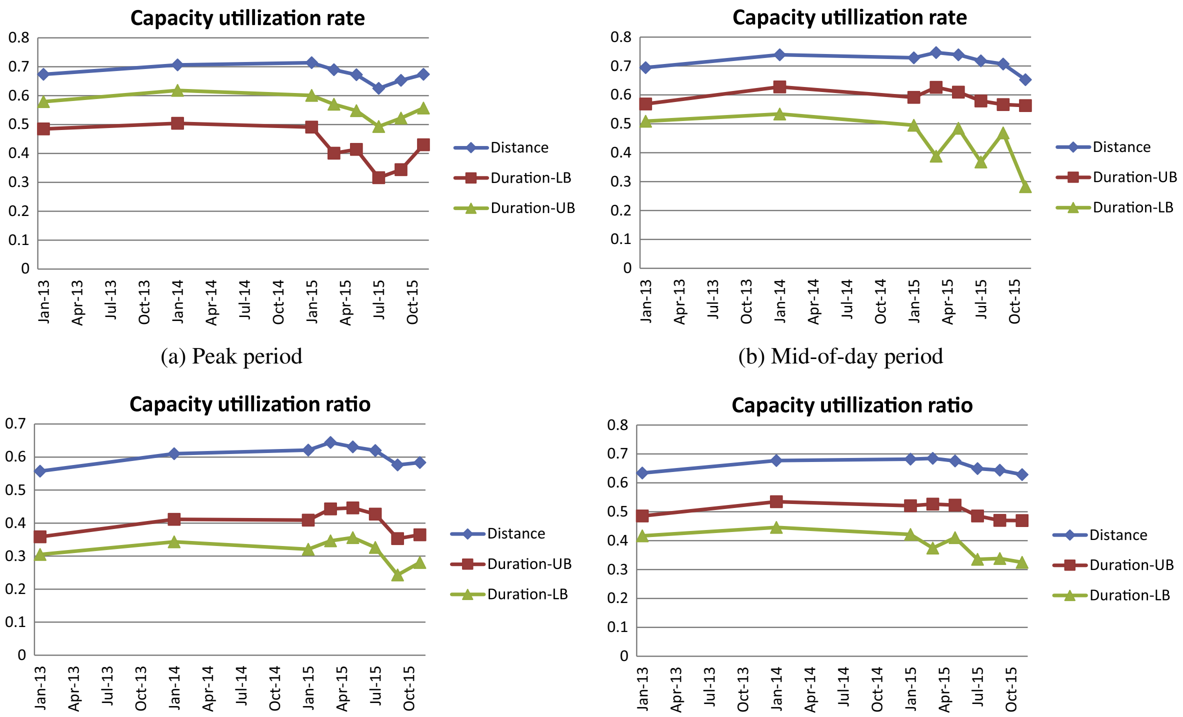


Fig. 10. Average weekday hourly speed for all occupied trips starting and ending within city center (taxi sample size: 2781).

drivers in the peak period is less able (or willing) to adjust its search effort in response to market conditions than those working in the other two periods (also see Figs. 7(a) and (b)). One possible explanation is that taxis drivers are less inclined to use e-hailing during rush hour.[17] As a result, they are not well informed of the market conditions.

For CURT, the profiles of its upper bound is quite similar to that of CURD in all periods, whereas that of its lower bound correlates with AHDT of occupied trips (as expected based on the definition). Generally, the utilization ratio by time is much lower than that by distance. For the all day average, its upper bound ranges between 0.47 and 0.52, compared to 0.32–0.5 reported in Cramer and Krueger (2016). Again, CURT is generally much lower in the off-peak period than in the other two periods.

[17] Based on the anecdotes of this author (including discussions with several taxi drivers in Chengdu, China), this is because they believe customers are easy to come by in the rush hour.



(a) Peak period

(b) Mid-of-day period

(c) OffPeak period

(d) All day

Fig. 11. Average weekday capacity utilization rate for COST data set 3–10 (taxi sample size: 2781).

6. Spatial analysis

We proceed to examine in this section if the impact of ridesourcing and e-hailing demonstrates any spatial heterogeneity. Given the data available, we focus on spatial distribution of population density. One may try to directly relate the relative change in taxis trips to the population density in each zone. Yet, this method does not yield any clear correlations between the two quantities. Instead, a plot of cumulative population influenced by a relative change in taxis trips is used to measure the spatial heterogeneity, as shown in Fig. 12. Each cumulative curve is constructed as follows.

1. All occupied trips are aggregated based on the four periods and their origin TAZ in each data set;
2. Using January 2015 as the benchmark (the month with the highest per-capita output of the year, as per the previous analysis), the *relative difference* of the total occupied trips in other data sets (July and November 2015 are selected for this analysis) are computed for each zone;
3. For all zones in city center (Area 1), count the number of people (in the unit of 1000 person) that experience a given relative change of the total occupied trips in both months.
4. Compute the cumulative population affected by a given relative change of taxi trips for Area 1.
5. Repeat Steps 3–4 above for Suburbs (Area 2).

In Fig. 12(d), the thick and thin curves represent November and July respectively, whereas the solid and dash curves represent Areas 1 and 2 respectively. The plots show that, while the November and July curves are fairly close in Area 1 (a sign of market stabilization in this area), the November curve is still further to the left of the July curve in Area 2. A leftward move by a curve on this plot suggests that more population experience negative changes of the total taxi trips (compared to the January 2015 level) that start from their home TAZ. As one may read from the plot, in July 2015, about 50 % of population in Area 2 experience more than 20% drop in the total taxi trips leaving their TAZ. This percentage becomes about 70% in November 2015, indicating that the decline in the taxi ridership had become more widely spread in Area 2. It is worth emphasizing that the trend in Area 2 is quite different from that seen with aggregated data (e.g. Fig. 5(d)), which highlights the importance of spatial heterogeneity.

Fig. 12(a) suggests that, in the peak period market, the taxi industry has a quite dramatic gain in November 2015 compared to July 2015, especially in Area 1. The percentage of the population subject to a more than 20% loss in taxi ridership has dropped from more than 80% in July to about 25% in November. In Area 2, there is also an improvement in November, though

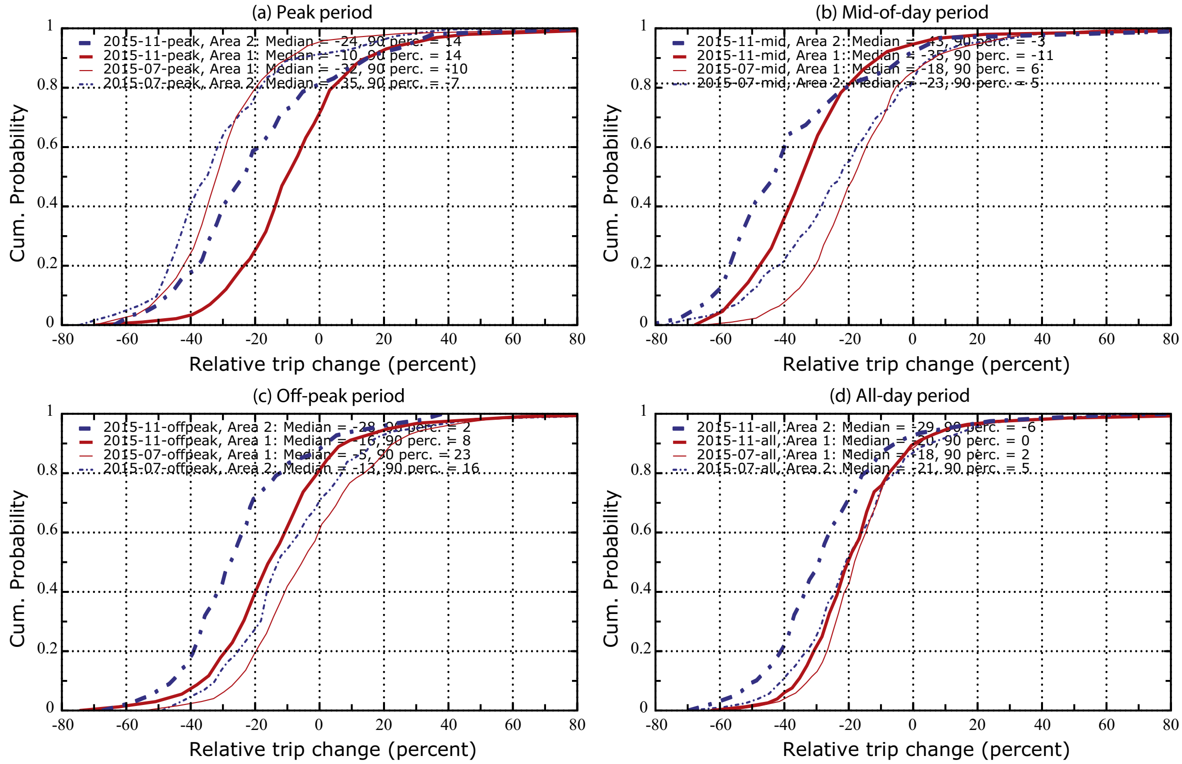


Fig. 12. Cumulative population affected by relative taxi trip change in different areas.

the gains are relatively modest. In the other two periods, the November curves for both areas continue to shift to the left side of their July counterparts. However, the improvement in the peak period from July to November seems enough to offset the losses in the other two periods in Area 1.

Overall, the spatial analysis indicates that taxis are relatively more competitive in the more densely populated areas and during the peak period.

7. Discussions

Several findings presented in the previous sections warrant further elaborations and discussions.

First, in Shenzhen, as in NYC, the taxi industry has experienced a significant loss in its ridership that can be indisputably credited to the competition from ridesourcing. Yet, the shock, although similar in intensity, was much shorter in Shenzhen (cf. Fig. 1). The evidence is strong that the high tide of ridesourcing has passed the city back in late 2015, and that the taxi industry had begun to stabilize since then. Why does the taxi industry in Shenzhen seem more resilient and adaptive? A possible explanation is that taxis in Shenzhen could compete more effectively because they, thanks to Didi Dache, had used to e-haling long before the arrival of ridesourcing. In contrast, taxi drivers in NYC had had no access to e-hailing until the summer of 2015, according to Garcia (2015).

The case of Shenzhen also suggests that a portion of the taxi market, possibly a majority, is still beyond the reach of a business model built on a fleet of mostly part-time amateur drivers. For one thing, e-hailing is not always more convenient and faster than street-hailing in areas where the density of vacant taxis is relatively high. The search cost in e-hailing seems negligible in theory, but not in reality. A customer has to pre-plan the trip (in order to minimize the waiting time), place the order using a smart phone, and (in most cases) talk to the driver over phone, sometimes repeatedly, explaining his/her exact location, which is not always easy when the person is unfamiliar with the place. All these actions take time and potentially money. Secondly, the matching algorithms of the TNCs are far from perfect. For example, a driver that is only 50 ft away from Customer A may be assigned to Customer B (who is one mile away), two seconds before customer A hits the "request" bottom on his/her phone. How is an algorithm supposed to foresee this and act accordingly? In addition, due to traffic congestion and other factors (e.g. drivers prefer not to take the route suggested by the platform), the predicted waiting times can be very inaccurate. Finally, the pitch of ridesourcing is to give drivers the flexibility to manage their work hours. As appealing as it sounds to potential workers, this rule limits the supply in the hours perceived as less profitable or less convenient (such as off-peak hours). To solve this last problem TNCs invented surge pricing. However, why should customers stick to ridesourcing if taxi stands out as a cheaper and more reliable option? The limited market potential of the current ridesourcing model is probably why Didi Chuxing is aggressively building its own dedicated fleet in China.[18] But such a move begs the question: how does the company distant itself from the boring image of an old-fashioned, less-regulated and market-monopolizing taxi firm?.

Second, taxis were found to compete more effectively with ridesourcing in peak period and in areas with higher population density. The detailed analysis presented in the previous sections consistently show that it is in the peak-time and populous city centers that the taxi industry has regained most of the lost ground in both ridership and productivity. First of all, due to traffic congestion, working in the peak period is more stressful and potentially less profitable, which may gradually drive away part-time workers and help relieve competition. Secondly, professional drivers have a competitive edge in rush hour because (1) they have the exclusive right for street-hailing, which may be preferred by business travelers in a hurry; and (2) they have more experience to navigate through heavy traffic whereas amateur drivers have to follow their GPS device. Finally, taxis tend to do better in more densely populated areas, likely because the exclusive street-hailing right works to their advantage in these areas. Essentially, high density areas attract more vacant taxis to visit, which makes street-hailing taxi a more appealing option.

Third, e-hailing helps lift the capacity utilization rate of taxis. Yet, the gains are generally modest (less than 10%), except for the off-peak period (8 PM to 6 AM) when the productivity of taxi drivers rises by as much as 15% at its peak. One explanation is that the productivity of taxis in Shenzhen is already high (compared to the statistiscs reported in Cramer and Krueger, 2016)) and so the room for improvement is limited. Also, the capacity utilization rate can be negatively affected by competition. While e-hailing may, to some extent, mitigate this impact by reducing the distance and time spent on searching customers, excessive competition can still lead to severely under-utilized capacities. This finding makes a strong case for regulations on the market served by both taxis and ridesourcing.

Fourth, ridesourcing did seem to impose some extra traffic congestion. At the peak of its expansion (the mid of 2015), we estimate that ridesourcing reduced the average travel speed by about 8%. If one read the data for the entire 2015, however, it is fair to say that the impact is not as dramatic as one might have imagined. Indeed, it is a little puzzling that throwing a whole ridesourcing fleet into a congested traffic network had not created a greater mess. To see why, we note that the amount of net traffic created by ridesourcing is probably not as significant as it sounds. If the induced demand for taxi and ridesourcing is excluded, all occupied trips only replace existing ones. For vacant trip, even though no data is available, it is safe to assume that ridesourcing would be at least as efficient as taxis - for taxis, the distance travelled on vacant trips is about 35% of that on all trips (see Fig. 11(d)). Hence, when taxis lost 25% of the travel distance paid by a passenger, the extra VMT contributed by vacant trips of the ridesourcing fleet is roughly $(25\% \times 65\% \times 0.35/0.65) \simeq 9\%$ of total VMT of the original taxi fleet. Assume that each private vehicle in Shenzhen travels about 30 miles a day, which is roughly 1/10 of a taxi's daily VMT in January 2015 (According to Fig. 7(d), the average hourly total distance is about 14 miles). There are 3.14 million private vehicles in Shenzhen in 2014 according Baidu (http://baike.baidu.com/view/1379561.htm), or 200 times the number of taxis. Thus, if we exclude the VMT by buses and trucks, the VMT contributed by taxis is about one 20th of the total VMT $(\simeq \frac{15{,}000\times300}{300{,}0000\times30})$, and hence, the increase created by ridesourcing would be roughly $9\%/20 = 0.45\%$. The actual increase is likely smaller given the conservative nature of the above estimate.

It seems safe to predict that the taxi industry is here to stay in the foreseeable future. Beyond e-hailing, economy of scale and aggressive pricing, ridesourcing does not seem to have other means at present to drive its expansion in the market. E-hailing is no longer the secret weapon that once glorifies the course of TNCs - it can be picked up by a taxi dispatcher that owns and operates its own fleet. Aggressive pricing, on the other hand, has proven at best a double-edged sword, as Uber's recent bitter defeat in China has vividly demonstrated. The scale of TNCs, which gives outside visitors a brand to stick to, is indeed an important competitive advantage. Even this lead is not that difficult to catch up, however, if a mobile platform, presumably operated by a third party, can unify taxi dispatchers around the world.[19] Such a platform can easily work within cities' existing regulatory structure, rather than against it, because it utilizes a dedicated and existing fleet. It can also improve the experience of street-hailing, a decisive advantage it holds against ridesourcing, by offering customers the amenities considered only available to e-hailing users, such as paying the fare on-line and rating drivers, all in real-time. An obvious solution may be allowing customers, as they board the taxi hailed off street, to open up an electronic transaction session similar to those seen on e-haling platforms, by e.g. scanning a QR code attached to the taxis or the driver's smart phone.

Finally, there is no question that ridesourcing has brought many positive changes to the taxi industry. It greatly improves the service quality by completely changing the play book and by tapping into a fresh reservoir of workforce. In the process, it has shaken the foundation of outdated institutions, regulations and policies that may have been a major source of inefficiency in this industry. Yet, as evidenced in this study, the revolution of ridesourcing is unlikely to eliminate the necessity of a dedicated service fleet, and for years to come we will continue to live in a world with both ridesourcing and (upgraded) taxis. How such a market can be properly managed and regulated is worth of further investigation. Indeed, right after the first version of this paper was completed in the summer of 2016, China enacted new regulations on ridesourcing services.[20] The long-term effect of such policies on the joint taxi/ridesourcing market is another interesting topic for follow-up studies.

[18] See e.g. http://tech.sina.com.cn/zl/post/detail/i/2016-04-21/pid_8506672.htm, in Chinese.
[19] Ironically, this was Didi Chuxing's business model at the beginning.
[20] http://finance.people.com.cn/n1/2016/0731/c1004-28598491.html.

Acknowledgement

The work was partially supported by National Science Foundation under the award numbers CMMI-1402911 and PFI:BIC-1534138. I wish to thank Mr. Jiandong Qiu from Shenzhen Urban Transport Planning Center for providing the COST data used in this study. The paper benefits greatly from comments offered by four anonymous reviewers, and from discussions with several colleagues, including Professors Song Yao and Hani Mahmassani from Northwestern University, Professor Yafeng Yin from University of Florida, and Professor Xiaobo Liu from Southwestern Jiaotong University. All remaining errors are my own.

References

Arnott, R., 1996. Taxi travel should be subsidized. J. Urban Econ. 40 (3), 316–333.
Beesley, M.E., Glaister, S., 1983. Information for regulating: the case of taxis. Econ. J. 93 (371), 594–615.
Bershidsky, L., 2015. Taxis can still survive uber. Bloomber https://www.bloomberg.com/view/articles/2015-10-26/regular-taxis-can-still-survive-uber.
Cairns, R.D., Liston-Heyes, C., 1996. Competition and regulation in the taxi industry. J. Public Econ. 59 (1), 1–15.
Corrigan, T., 2016. San franciscos biggest taxi operator seeks bankruptcy protection. Wall Street J. http://wwwwsj.com/articles/san-franciscos-biggest-taxi-operator-seeks-bankruptcy-protection-1453677177.
Cramer, J., Krueger, A.B., 2016. Disruptive change in the taxi business: the case of uber. Am. Econ. Rev. 106 (5), 177–182.
Davidson, J., 2014. Uber has pretty much destroyed regular taxis in san francisco. Time http://time.com/money/3397919/uber-taxis-san-francisco/.
De Vany, A.S., 1975. Capacity utilization under alternative regulatory restraints: an analysis of taxi markets. J. Polit. Econ., 83–94
Douglas, G.W., 1972. Price regulation and optimal service standards: the taxicab industry. J. Transport Econ. Policy, 116–127.
Flores-Guri, D., 2003. An economic analysis of regulated taxicab markets. Rev. Indust. Organiz. 23 (3), 255–266.
Garcia, A., 2015. Nyc yellow cabs get their own hailing app. CNN Money http://money.cnn.com/2015/08/28/news/arro-taxi-cab-app-nyc/.
He, F., Shen, Z.-J.M., 2015. Modeling taxi services with smartphone-based e-hailing applications. Transport. Res. Part C: Emerg. Technol. 58, 93–106.
Hughes, R., MacKenzie, D., 2016. Transportation network company wait times in greater seattle, and relationship to socioeconomic indicators. J. Transport Geogr. 56, 36–44.
Moore, A., 2006. t. balaker. do economists reach a conclusion on taxi deregulation. Econ. J. Watch. 3, 109–132.
Nelson, L.J., 2016. Uber and lyft have devastated l.a.'s taxi industry, city records show. LA Times. <http://www.latimes.com/local/lanow/la-me-ln-uber-lyft-taxis-la-20160413-story.html>.
Newman, R., 2016. Uber hasnt killed the taxi industry after all. Yahoo. <http://finance.yahoo.com/news/uber-hasn-t-killed-taxi-000000965.html>.
Oremus, W., 2016. The end of the taxi era. Slate. <http://www.slate.com/articles/technology/technology/2016/01/yellow_cab_in_san_francisco_is_just_the_beginning_uber_s_war_on_cabs_is.html>.
Rayle, L., Dai, D., Chan, N., Cervero, R., Shaheen, S., 2016. Just a better taxi? a survey-based comparison of taxis, transit, and ridesourcing services in san francisco. Transport Policy 45, 168–178.
Santi, P., Resta, G., Szell, M., Sobolevsky, S., Strogatz, S.H., Ratti, C., 2014. Quantifying the benefits of vehicle pooling with shareability networks. Proc. Natl. Acad. Sci. 111 (37), 13290–13294.
Schaller, B., 2007. Entry controls in taxi regulation: implications of us and canadian experience for taxi regulation and deregulation. Transport Policy 14 (6), 490–506.
Steier, F., 2015. Has uber hit a dead end in new york? taxis still hog the midtown. TheStreet.
Wang, X., He, F., Yang, H., Gao, H.O., 2016. Pricing strategies for a taxi-hailing platform. Transport. Res. Part E: Logist. Transport. Rev. 93, 212–231.
Whitford, E., 2015. Greenpoint's growing taxi graveyard. Gothamist. <http://gothamist.com/2015/08/21/why_yellow_cabs_are_taking_up_all_o.php>.
Yang, H., Fung, C., Wong, K., Wong, S., 2010. Nonlinear pricing of taxi services. Transport. Res. Part A: Policy Pract. 44 (5), 337–348.
Yang, H., Wong, S., 1998. A network model of urban taxi services. Transport. Res. Part B: Methodol. 32 (4), 235–246.
Yang, H., Wong, S.C., Wong, K., 2002. Demand–supply equilibrium of taxi services in a network under competition and regulation. Transport. Res. Part B: Methodol. 36 (9), 799–819.
Yang, H., Ye, M., Tang, W.H., Wong, S.C., 2005. Regulating taxi services in the presence of congestion externality. Transport. Res. Part A: Policy Pract. 39 (1), 17–40.
Zha, L., Yin, Y., Yang, H., 2016. Economic analysis of ride-sourcing markets. Transport. Res. Part C: Emerging Technol. 71, 249–266.
Zuylen-Wood, S.V., 2015. The struggles of new york city's taxi king. Bloomberg. <http://www.bloomberg.com/features/2015-taxi-medallion-king/>.