THE POTENTIAL IMPACT OF RIDE HAILING ENTRY ON URBAN TRAFFIC CONGESTION

Garland Chow, Sauder School of Business, University of British Columbia[1]

Introduction

Ride hailing enabled by Transportation Network Company (TNCs) platforms such as UBER and Lyft have brought new competition into many passenger directed transportation markets. This has led to lower fares and improved service but it has also had impacts on traffic congestion. This paper seeks to identify the potential impact of the number of vehicles that would be added to roadways by the entry of ride hailing into the passenger directed market traditionally dominated by the taxi sector. Utilizing taxi trip data, a service model is developed that estimates the service level that can be achieved with a given number of vehicles for the City of Vancouver. The model is used to estimate the number of vehicles required to provide a given level of service and specifically for the very high levels of service that TNC companies seek to provide. Implications for public policy to manage the service-congestion trade-off are suggested.

Economics and Operational Features of Passenger Dircted Transportation

Ride hailing service is provided by individual drivers who are enabled by the digital platforms of Transportation Network Companies (TNC) such as UBER and Lyft. A key feature of ride hailing service has been improved service, often perceived to be a 5-minute waiting time, defined as the time from the commitment by the ride-hailing driver to pick up of the customer to actual pickup. The high level of service is achieved by having a very large supply of vehicles and drivers available to serve potential customers. In one TNC's words, ``"We do everything we can to get as many cars on the road system as possible". Ride hailing, despite the name, directly competes with dispatched taxi services. The customer contacts the ride-hailing driver via the TNC app or contacts the taxi company dispatch which in turn contacts the driver. Both the ride-hailing driver and the taxi driver accept the request for service, the former directly to the customer, and the latter through the dispatcher. In the taxi case, many taxi firms accept the request for a taxi and relay the request to driver or drivers afterwards; hence, there is more variability in taxi waiting time. The speed at which a vehicle is dispatch and reaches the customer for pickup depends on the supply of service vehicles relative to the demand for service, the geographic and temporal distribution of the demand origins and the demand destinations, the efficiency of repositioning service vehicle to serve the customers and the friction of moving from point to point in the road network.

---

[1] Presented at the 53rd Annual Meetings of the *Canadian Transportation Research Forum*, June 3-6, 2018 at Gatineau, Quebec

Impact of Higher Levels of Service on Congestion

The impact of ride hailing on congestion is the byproduct of increasing service. To provide a quicker response time to customers demanding service, there need to be more vehicles relative to demand and the vehicles need to be positioned near demand. Some argue that TNCs will reduce congestion by taking drivers off the road in their private vehicles. This is a factitious argument with respect to congestion; if a person does not drive their own vehicle, the TNC vehicle is on the road, so the actual miles driven stay the same. At the same time, the service vehicle continues to circulate and reposition in search of the next customer. TNCs are only beginning to develop shared and pooled driving services to reduce mileage.

There is controversy as to the net impact of ride hailing on congestion and the experience of other large metropolitan cities is just emerging. Several recent studies focus on the diversion of customers from public transit or from cycling and walking.

- Clewlow and Mishra (2017) conclude from a survey of seven major U.S. cities that ride-hailing is contributing to an increase in vehicle-km and a reduction in public transit use.
- The Boston Metropolitan Metropolitan Area Planning Council conducted a rider intercept survey of nearly 1,000 TNC passengers in fall 2017 around the Boston metro area and found more than 40 percent said they would have taken public transit if Uber or Lyft had not been available, and 12 percent said they would have walked or biked. Most cited speed and convenience as the main reasons for choosing an on-demand ride over buses or trains. (MBTA, 2017) "The responses to those questions provide strong evidence that TNCs are pulling from, not complementing, public transit,"(Bliss, 2018).
- Northeastern University researchers built a software script that scraped reams of vehicle location data. San Francisco County Transportation Authority staff then cleaned, modeled, and mapped that information to estimate where, when, and how many daily trips occur. Their analysis found that more than 170,000 vehicle trips are made by TNCs within city limits on a typical weekday, which is about 15 percent of all car trips, and 9 percent of all trips, across different modes. They also found that the vast majority of TNC trips are heavily concentrated in San Francisco's northeast quadrant, which is already the densest, most congested part of the city, as well as the area best served by public transit, bike lanes, and walkable streets. (Bliss, 2018).

Perhaps the best evidence of the impact of TNCs on congestion comes from New York where both taxis and TNCS are required to make trip logs, trip volumes and vehicle mileage readings available. Schaller (2017) analyzed this data and observes:

- "TNC growth has added nearly 50,000 vehicles … to the city's streets in just three years."
- "Much of this growth occurred in Midtown Manhattan and other already-congested areas of the city."
- "TNCs accounted for the addition of 600 million miles of vehicular travel to the city's roadway network over the past three years, after accounting for declines in yellow cab mileage and mileage in personal vehicles. The additional 600 million miles exceeds the total mileage driven by yellow cabs in Manhattan."

Schaller concludes that " Managing the impact of this growth on traffic congestion, vehicle emissions and traffic safety is thus a critical public policy challenge.``

As noted above, TNCs seek to provide a high level of service enabled in large part by having a very large supply of vehicles and drivers is available to serve potential customers. "For its part, Uber is desperate to avoid shortages, seeking instead to serve every customer quickly, ideally in five minutes or less." (Bliss 2018)

Both economics and service operations management principles recognize that there is a tradeoff between service capacity and waiting times (see for example, Laguna and Marklund, 2005, pp. 171 – 173) that companies must balance based on the cost of capacity and the costs of waiting times. The objective of the remainder of this paper is to provide insights on the relationship between the capacity (number of vehicles required) and waiting times for passengers to obtain service in passenger directed transportation.

Development of Total Service Level (TSL) Model

We investigate the impact of providing higher levels of service by estimating a service response model for taxi services. From the model we derive estimates of the number of taxi vehicles required to provide a given level of service. Subsequently we discuss the extent to which the estimates reflect the number of TNC vehicles which are required to provide the same level of service.

The four members of the Vancouver Taxi Association provided access to the dispatch records of all taxi trips originating in Vancouver and the Vancouver International Airport for 91 days in 2015. Over 2 million trips were processed and after deleting trips with incomplete or inconsistent data, the trips were classified as:

- Completed versus no shows and cancelled trips
- Dispatched and flag (street hail) trips
- Pre-booked and immediate (on demand) trips

Each trip record included:

- The time that a dispatched trip was booked, the create time.
- The time the dispatched trip was dispatched, the dispatch time.
- The time the customer trip began, the meter on time.
- The time the customer trip ended, the meter off time.

While we can measure the service for each dispatched trip, the relevant measure of service and productivity performance is over a time unit. Each of the 91 days in the data base was further disaggregated into each of the 24 hours in the day resulting in 2184 hourly periods used as the unit of analysis for the modeling of service, productivity, supply, demand and other relevant variable. The individual trip data is aggregated into the hour that each trip was created (the create time).

Service is defined as Total Service Response (TSR) for dispatched-immediate trips. TSR is the time it takes the taxi firm to provide a taxi to the customer after they have contacted the taxi company to order the taxi service and pickup. TSR is relevant to dispatched trips that are to be picked up as soon as possible or immediately. TSR is calculated as the difference between "meter on" and "create time of the trip". For example, the TSR provided could be 10 minutes meaning that the taxi firm has delivered a taxi to a customer so that the trip could start 10 minutes after the customer call was taken by the taxi company.

Service level is defined as TSRx, where x is the TSR in minutes that is provided or sought for an individual trip. TSR10 is TSR within 10 minutes. Service performance is defined as the percent of trips in a time period that meets the service level x. Our dependent variable is thus percent of trips provided within the TSRx or the Total Service Level (TSLx). In this paper we utilize TSL10 or percent of trips per hour where the customer has to wait no longer than 10 minutes to obtain service once the customer has contacted the taxi firm.

Based on a priori expectations and data availability, we explored the relationship between TSL10 and the following causal variables all measured as activity during each hour:

D – demand for service measured by all completed taxi trips, dispatched or flag. As D increases relative to the supply of taxis, customers must wait longer for the next available taxi.

S – number of taxis actively providing service during the hour. As S increases relative to the demand for taxis, customers wait less time for the next available taxi.

Service is fundamentally determined by the relationship between supply of taxi capacity and the demand for that capacity. In each hour, we measured the total completed trips (D) and the total active taxis (S) as measures of demand and supply respectively. These variables represent the supply and demand for all types of taxi service, dispatched-immediate, dispatched pre-booked and flag. We recognize that both are at best surrogate measures. Completed trips do not represent real taxi demand as it does not measure latent demand for taxi trips that did not occur if taxi service is inadequate or unavailable. Active taxis do not account for taxis parked or on call if demand warrants their utilization.

DS1- After model testing, we found that the ratio of D/S provided a single measure that reflected the impact of both influences on service and utilized that version of the model. Hence we expect that TSLx to be negatively related to DS1.

Avg.TripLengthDispatch and Avg.TripLengthFlag – This measures the time of each type of trip from pickup to dropoff. The longer the trip takes, the lower the capacity of a taxi to immediately service another customer. The longer trip time may be caused by a longer distance between the pickup and drop-off locations or road conditions such as congestion or circuity particular to the route. We expect ATL to be negatively related to TSLx.

FlagPer – Dispatched demand competes with flag demand for the same taxis. As the TSLx measures service to customers ordering dispatch service, when flag demand is higher relative to dispatch demand, fewer taxis are available to provide dispatch service. We expect that TSLx is negatively related to FlagPer.

PerZone1Zone2 – Measures the percent of trips originating in the downtown zone (Zone1) and the adjacent zone (zone2) which are observed to consistently account for the majority of the taxi trip originations at all hours. High traffic density or trip generation in a few concentrated areas enable taxi supply to position itself closer to a larger proportion of the demand and therefore reach the customer pickup point more quickly. TSL should be higher where traffic is concentrated in few zones.

CCNSPer – Measures the percent of total trips requested that were canceled or no shows. Canceled trips and no shows represent wasted service time where a taxi travels to a pickup point and waits for the customer without ultimately providing service. The taxi is not available to service other demand while in this status. Therefore as CCNSPer increases, TSLx should be lower. An alternate relationship is that CCNSPer is the result of poor service performance as customers cancel or no show when the TSR is too long. Thus TSL and CCNSPer are negatively related.

We also estimated the TSL models with the variables below and found them significant. In this paper, we utilize the results for the least complex model

- Weather and visibility conditions
- Dummy variables representing extreme peaking in specific locations
- Dummy variables representing shift changes
- Interaction variables between selected variables

All variables were transformed their log10 analog. The estimation of the model yielded the estimates found in Table 1. All coefficients had the a-priori expected signs and were significant at p > .001 or higher.

Table 1 – TSL Model Estimation Results

Variable	Coef.	SE	T Value
Avg.TripLengthDispatch	-0.371 ***	(0.015)	24.7
Avg.TripLengthFlag	-0.044 ***	(0.011)	4.0
CCNSPer	-0.058 ***	(0.003)	19.3
FlagPer	-0.104 ***	(0.006)	17.33
PerZone1Zone2	0.193 ***	(0.015)	12.87
DS1	-0.133 ***	(0.005)	26.6
Constant	0.794 ***	(0.036)	22.05
Observations	2,184		
R2	0.758		
Adjusted R2	0.757		
Residual Std. Error	0.041 (df = 2176)		
F Statistic	682.000*** (df = 7; 2176)		
***	p<0.01		

Empirical Analysis

The equation was utilized to generate fitted values for 10 different levels of actual demand and TSL10 (service performance performance) for the same TSR (service level). These included the 5 quintile break points, the absolute max and min and selected min and max hours during certain segments of the day listed in the first column and described in the second column of Table 2.

For each level of demand (case), the equation generated the TSL using the actual data for the hour in the case including Demand and varied the S, hence changing the DS1 ratio until a the TSL reach 95%. The estimated vehicles required to fulfill different levels of demand represented by the 10 cases for different TSL ranging from 80% to 95% are tabulated in Table 2 and illustrated in Figure 1. In addition in the last two rows, the ratio of D/S and its inverse S/D for Case 10 is calculate. The following is observed from both Table 2 and Figure 1:

- Focusing on the row for Case 10, Max Demand in order to achieve a TSL performance of 80%, 85%, 90% and 95%, there needs to be 1304, 2056, 3160, 4745 vehicles on the streets respectively. The maximum number of taxis licensed to provide service during this hour was 687 vehicles which would only produce a projected 73% TSL. The 4 taxi firms recognized this and had requested an additional 199 vehicles at the end of 2015 and were approved for 175 additional vehicles in 2017, bringing the total maximum fleet size to 862 during peak hours. This fleet size would provide a 75.7% TSL during this max demand hour. However this number of vehicles could provide more than enough vehicles to produce a TSL up to 85% for the majority of the hours and up to 95% for about half the hours in the typical day, as shown in the entries in Table 2. This illustrates the common tradeoff and dilemma between service capacity and waiting times in service industries where demand varies, the peak demand challenge and subsequent underutilization of capacity during peak periods.

- Higher levels of TSL for a given TSR (in this case TSR of 10 minutes) increase the number of vehicles needed . This was already illustrated for the Max demand hour. Similarly, in case 9, the max demand hour during the afternoon rush hour the number vehicles required are 465, 733, 1126 and 1691 to achieve a TSL performance of 80%, 85%, 90% and 95% respectively. Put another way, the number of vehicles required to provide 95% TSL within 10 minutes rather than 80% is 3.64 times higher or 1226.
- Productivity decreases as higher levels of service are sought. This is illustrated by the D/S and S/D ratios calculated for Case 11. At 80% TSL, there needs to be one vehicle for every 2.19 trips demanded (D/S) or conversely .45 vehicles per trip demanded (S/D). However for a 95% TSL, there needs to be one vehicle for every .60 trip demanded or conversely, 1.66 vehicles per trip demanded.
- The maximum demand hour (case 10) requires 4745 vehicles to provide a TSL of 95%. This case is during the weekend entertainment peak period whereby the demand is concentrated in the downtown entertainment districts. Congestion would be exacerbated by the concentration of the additional vehicles in this limited area.

Discussion and Implications

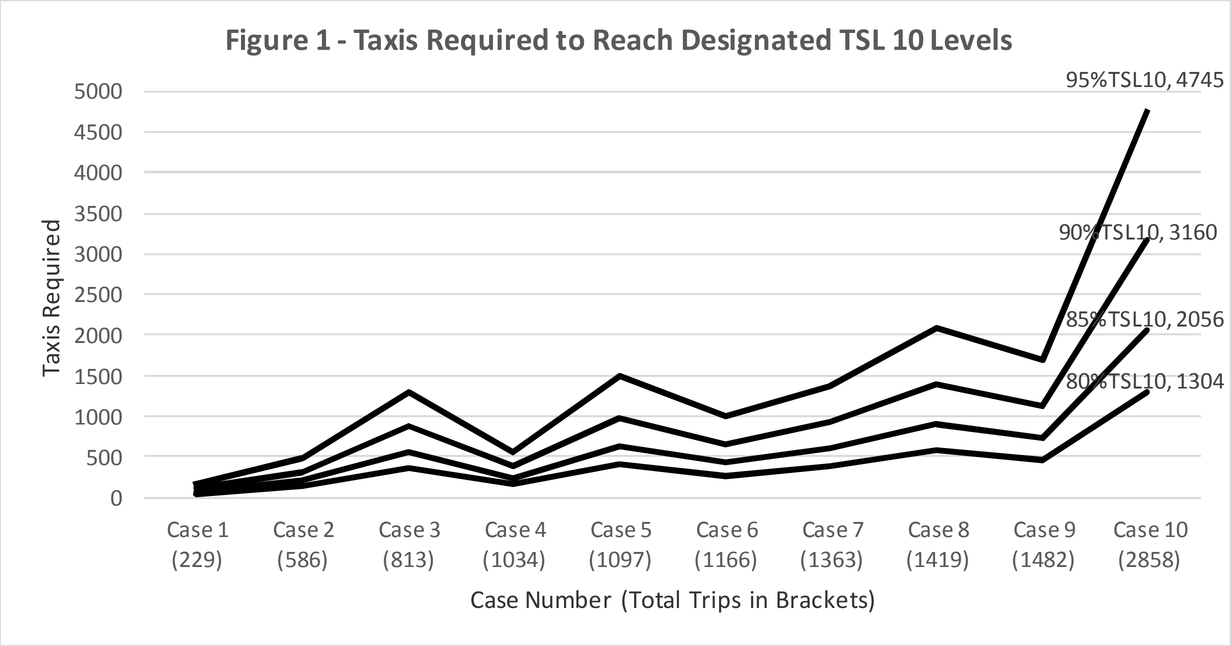
The entry of TNC and ride hailing competition into passenger directed transportation markets historically dominated by taxis is often praised for the improvement in service, specifically shorter waiting times that we have called TSR in this analysis. TSR of 10 minutes was expressly chosen to analyze as it is midway between a criteria once expressed by the British Columbia Passenger Transportation Board (PTB) that has regulatory jurisdiction over the taxi firms and their data which was used in this study. At various times, the PTB has stated an acceptable performance would be providing service within 15 minutes at least 92 percent of the time (PTB, 2012). While this service goal might not be met for the peak hours, it is met during the non-peak hours that generate the majority of the trips so that the aggregate performance of the Vancouver taxi sector has generally exceeded the PTB goal. At the other extreme, studies of TNCs such as UBER and stated company objectives indicate that waiting times of 5 minutes are sought by this large TNC. The estimation of vehicle requirements based on TSR10 are thus far below the estimated vehicle requirements for a TSR5. It is likely that a much larger increase in vehicles is required to meet that higher service level, that would congest roadways, increase the time for vehicle trips as well as repositioning, leading to even less productivity to meet demand. This could result in a cycle of increasing vehicle capacity to meet service expectations, leading to more congestion, leading eventually to gridlock. This is consistent with the New York City experience as analyzed by Schaller.

One solution is to implement congestion fees which would increase the cost of providing passenger directed transportation, reduce demand for both trips and high levels of service until equilibrium is reached in the market. Congestion charges are currently being considered for the Vancouver lower mainland but until there is certainty that they can and will be effectively implemented, this is not a near term solution.

Table 2 Taxis Required to Reach Designated TSL 10 Levels

Case (Completed trips)	Label	Demand (trips)	Taxis Needed for TSL			
			80%TSL10	85%TSL10	90%TSL10	95%TSL10
Case 1 (229)	Min	229	47	75	114	172
Case 2 (586)	Slowest Hour During Day	586	130	204	313	479
Case 3 (813)	20% point	813	359	567	871	1307
Case 4 (1034)	40%	1034	157	247	379	569
Case 5 (1097)	50% median	1097	408	643	988	1483
Case 6 (1166)	60%	1166	273	430	660	991
Case 7 (1363)	80%	1363	378	596	916	1375
Case 8 (1419)	Max-Peak Weekday Afternoon	1419	574	906	1391	2089
Case 9 (1482)	Max Peak Weekday Morning	1482	465	733	1126	1691
Case 10 (2858)	100% max	2858	1304	2056	3160	4745
D/S Case 10	N/A	N/A	2.19	1.39	.90	.60
S/D Case 10	N/A	N/A	.45	.719	1.10	1.66

The service levels provided by taxis in in the current environment in which they operate may not reflect accurately the efficiency in which TNC and ride hailing competition could provide service. The larger TNCs are known to employ advance data analytics such as machine learning to better forecast demand at finer temporal-spatial levels to optimize repositioning of their ride hailing service providers to be closer to their customers at the time of demand. Few, if any taxi firms have adopted these techniques, one barrier being the small scale or size of most taxi firms and therefore inability to invest in such innovation. A strategy for competing taxi firms is to merge or cooperate to obtain the benefits of size which go beyond spreading fixed costs. Each taxi firm operates a network but when combined, the network is denser and network economies arise. In today's taxi environment, a customer calling one taxi firm may have to wait longer for the taxi to arrive then if it had called a competing taxi firm which had a vehicle already positioned close to the customer. Preliminary research by this author using agent based simulation has found both increased productivity and higher TSL when the four taxi firms in Vancouver act as one entity and assign the closest vehicle, irrespective of the company, to the customer. While this will reduce the number of vehicles the taxi competitors need to provide a competitive level of service, the evidence from the New York City study still indicates considerable additional vehicle and vehicle miles traveled by the large TNC firms such as UBER and Lyft. Scale does not appear to be a sustainable solution.



The British Columbia Standing Committee heard testimony from multiple witness on the potential impact of ride hailing entry into B.C. Multiple witnesses cited concerns about the potential increase in congestion that may arise from unrestricted entry of TNCs and ride-hailing as well as a larger number of taxis seeking to compete against the TNCs. (BC Standing Parlimentary Committee, 2018). This paper has identified the tradeoff between increased service that will result from increased competition from ride hailing and the increase in vehicle capacity that could lead to congestion. This is typically unaccounted for in regulatory hearings on new entry and this has been the case in many jurisdictions that have already permitted TNCs like UBER and Lyft to enter without any entry limits. This suggests consideration of entry controls on both taxis and ride hailing providers to balance the benefits of faster service against the societal cost of congestion. This could be in the form of generalized vehicle restrictions such as the scheme employed in London that limits the number of vehicles entering specified traffic zones. A refinement is to regulate the maximum number of vehicles permited at a location/time dynamically. This is feasible to implement effectively due to the widespread use of GPS on taxi and ride hailing service vehicles (Chow 2017). At the very least data is needed to understand the nature of the congestion-service tradeoff. Jurisdictions should make data availability and reporting a condition of licensing by both taxis and ride hailing permitting the type of analysis conducted in New York. Jurisdictions need to immediately measure roadway congestion at key locations before allowing massive new entry in order to be able to measure congestion with and without ride hailing. Regulations should clearly define what information is to be reported, how it will be used, who can have access and protection of proprietary information to enable a regulatory framework that has the authority to restrict taxi and TNCs capacity in the future if the negative impacts such as congestion arise. This will support policy making that ensures a sustainable passenger transportation system with regards to both economic and environmental sustainability.

References available on request.