

## Disruptive Change in the Taxi Business: The Case of Uber

Judd Cramer and Alan B. Krueger\*<sup>1</sup>

Occupational licensing has grown steadily in the U.S. since the 1950s, with nearly one third of private sector workers currently in jobs covered by occupational licensing requirements (Kleiner and Krueger (2010)). In many jurisdictions, taxi drivers are required to obtain an occupational license in order to transport passengers, and drivers are restricted from picking up passengers outside of the jurisdiction that issued their license. In addition, the number of taxi drivers is often limited by the number of medallions that are issued, and fares are often set by regulatory bodies. Although occupational licensing regulations can improve consumer safety and yield other benefits, they can also reduce the efficiency of the economy, raise costs for consumers, and lead to a misallocation of resources.

The innovation of ride sharing services, such as Uber and Lyft, which use internet-based mobile technology to match passengers and drivers, is providing unprecedented competition in the taxi industry. Weighted by hours worked, there were about half as many Uber and Lyft drivers as taxi and limo drivers operating in the U.S. at the end of 2015.<sup>2</sup> This paper examines

---

<sup>1</sup> \* Address/Affiliation for Cramer: Department of Economics, Princeton University, Princeton, NJ 08540, [cramer@princeton.edu](mailto:cramer@princeton.edu); Krueger: Department of Economics, Princeton University, Princeton, NJ 08540, [akrueger@princeton.edu](mailto:akrueger@princeton.edu). We are extremely grateful to Jason Dowlatabadi, Hank Farber, Jonathan Hall, Paul Joskow, Vincent Leah-Martin, Craig Leisy, and Eric Spiegelman for providing comments and/or data tabulations. We are solely responsible for the content and any errors. Krueger acknowledges that he has coauthored a paper that was commissioned by Uber in the past, although he has no ongoing relationship with the company.

<sup>2</sup> In 2015 there were around were nearly 500,000 taxi drivers and chauffeurs in the U.S. according to our tabulation of the Current Population Survey, and Uber and Lyft combined had nearly 500,000 active drivers. Uber drivers, however, work about half as many hours per week as taxi and limo drivers according to Hall and Krueger (2015).

the efficiency of the ride sharing service Uber by comparing the capacity utilization rate of UberX drivers to that of taxi drivers.

Capacity utilization is measured either by the fraction of time that drivers have a fare-paying passenger in the car or by the fraction of miles that drivers log in which a passenger is in the car. Because we are only able to obtain estimates of capacity utilization for taxis for a handful of major cities – Boston, Los Angeles, New York, San Francisco and Seattle – our estimates should be viewed as suggestive. Nonetheless, the results indicate that UberX drivers, on average, have a passenger in the car about half the time that they have their app turned on, and this average varies relatively little across cities, probably due to relatively elastic labor supply given the ease of entry and exit of Uber drivers at various times of the day. In contrast, taxi drivers have a passenger in the car an average of anywhere from 30 percent to 50 percent of the time they are working, depending on the city. Our results also point to higher productivity for UberX drivers than taxi drivers when the share of miles driven with a passenger in the car is used to measure capacity utilization. On average, the capacity utilization rate is 30 percent higher for UberX drivers than taxi drivers when measured by time, and 50 percent higher when measured by miles, although taxi data are not available to calculate both measures for the same set of cities.

Four factors likely contribute to the higher utilization rate of UberX drivers: 1) Uber's more efficient driver-passenger matching technology; 2) Uber's larger scale, which supports faster matches; 3) inefficient taxi regulations; and 4) Uber's flexible labor supply model and surge pricing, which more closely match supply with demand throughout the day.

## **I. Assembling Data on Capacity Utilization Rates**

Ideally, we would like to have data on the fraction of time in which taxi and Uber drivers have a fare-paying customer in their car each moment that they work. There is no single source of data for taxi drivers, however, so we must piece together information for cities where data are available. For New York City, we use micro-level daily data on anonymized taxi drivers' work hours and time with the meter running from the New York City Taxi and Limousine Commission (NYCTLC) for trips taken in 2013.<sup>3</sup> For San Francisco, Vincent Leah-Martin provided us with tabulations of similar micro-level data that he obtained from one midsized taxi fleet.<sup>4</sup> For Boston, the fraction of total hours worked that taxi drivers had a passenger in their car was reported in the Nelson/Nygaard (2013; Figure 4-1) report for the City of Boston for three days in 2013.<sup>5</sup> Information on miles driven by taxi cabs is not available for these cities.

For two cities, Seattle and Los Angeles, we have information on miles driven (total and with a passenger) aggregated across all taxi drivers. Aggregate revenue miles and aggregate miles driven by taxi drivers is available for 2013 and 2014 for Seattle from Soper (2015). For Los Angeles, comparable information at a monthly frequency from January 2009 to January 2015 is available from the Los Angeles Department of Transportation (LADOT).

There are a variety of ways to compute the capacity utilization rate. First consider a situation where we have access to individual-level data on  $N$  drivers' work hours in a given day, denoted  $H_i$ , and the number of hours in which they had a fare-paying passenger in the car,

---

<sup>3</sup> See Farber (2015) for a description of the data set.

<sup>4</sup> See Vincent Leah-Martin (2015) for further details on the data set. The data we report pertain to July, August, September and October of 2013.

<sup>5</sup> The data were from credit card terminal data, which record information for every trip, regardless of whether a credit card was used. The dates were January 9, April 11 and July 13. The sample of data for Uber drivers correspond to the same days of the week (and proximity to the Boston Marathon – i.e., the Thursday before the marathon) for those months in 2015: January 14, April 16 and July 11.

denoted  $h_i$ . We can compute the average fraction of time that a driver is working in which he or she has a passenger in the car, which we denote  $f^h$ :

$$(1) \quad f^h = \sum (h_i/H_i)/N = \sum f_i^h / N,$$

where  $f_i^h$  is  $h_i/H_i$ , the capacity utilization rate of driver  $i$  on the day in question.

Alternatively, in some instances data on passenger-fare hours aggregated across all drivers and total work hours of taxi drivers are available. In these cases, we compute the aggregate capacity utilization rate, denoted  $F^h$ :

$$(2) \quad F^h = \sum h_i / \sum H_i = \sum w_i f_i^h / N.$$

Notice that  $F^h$  is a weighted average of  $f_i^h$ , where the weights,  $w_i$ , are each driver's share of total work hours,  $H_i/\sum H_i$ . If drivers' hours do not vary much, or if driver hours and  $f_i^h$  are weakly correlated, then  $f^h$  and  $F^h$  will be similar.

To compute capacity utilization rates with respect to miles driven, as opposed to time, we simply replace  $h_i$  and  $H_i$  with miles driven while a passenger is in the car and total miles driven in the day, denoted  $m_i$  and  $M_i$ , respectively. The only information we could obtain on capacity utilization rates for miles driven for taxi drivers is of the F-type aggregate measure.

At our request, the Uber Research staff kindly provided us with statistics on  $f$  and  $F$  based on Uber's administrative database for Uber drivers in the five cities for which we were able to collect data on traditional taxi drivers. We focus on UberX drivers because that is the largest and fastest growing category of Uber drivers.<sup>6</sup> Work time  $H_i$  was defined as the total amount of time that a driver's app was on, while  $h_i$  was defined as the time in which a passenger was in the car.

With the Uber data, it is possible to calculate capacity utilization by either  $f$  or  $F$ , which is

---

<sup>6</sup> To be precise, the sample consists of UberX, UberXL, UberPool, and UberSelect drivers. We refer to all drivers in these service categories as UberX drivers. UberBlack drivers, who typically require a commercial driver's license, are excluded.

fortunate because daily work hours vary more across Uber drivers than they do across taxi drivers, who typically 7 or 8 hour shifts, or longer.

One difference between Uber drivers and Taxi drivers is that Uber drivers are not restricted from picking up passengers in one particular jurisdiction. The sample of UberX drivers in each city consisted of those who picked up at least one passenger in the city during the day, and those drivers were followed throughout the day regardless of where else they might have traveled.<sup>7</sup> As a practical matter, qualitatively similar results are obtained if the sample is limited to drivers whose first pickup was in the city. Because computing mileage driven is time intensive, a random sample of 2,000 drivers was selected for each city.<sup>8</sup>

Another issue concerns timing. One could argue that it is desirable to compare UberX and taxi drivers during the same period of time, or one could argue that it makes sense to compute the capacity utilization rate for taxis *before* Uber entered the market to assess the effect of taxi licensing and regulation, because the presence of Uber could have caused the productivity of taxi drivers to change. Regardless, as a practical matter we are limited by the data available. Due to lags in reporting, the taxi data are from an earlier year than the Uber data. The Uber data pertain to December 1, 2014 through December 1, 2015. For San Francisco the data were restricted to July through October 2015, to match the months of the taxi data, and for Boston the corresponding days of the year were selected to match the taxi data. The fact that the taxi data pertain to a period before Uber made significant inroads into the market likely raise the capacity

---

<sup>7</sup> One should also be aware that Uber drivers can simultaneously work for Lyft and other ride sharing services. Because Uber lacks information on whether UberX drivers are providing rides to customers through Lyft or other services, the Uber capacity utilization rate probably understates the actual rate that drivers achieve.

<sup>8</sup> More specifically, a day was defined as running from 4 AM to 4 AM, and a random sample of driver days was selected each period. The periods were selected from 2015 for the days and months corresponding to the available taxi data for Boston and San Francisco, or from December 1, 2014 to December 1, 2015 for the other cities.

utilization rate for taxis compared to Uber drivers, as the taxis had less competition for passengers at that time.

## II. Findings

Table 1 provides estimates of  $f^h$  and  $F^h$  for Uber in all five cities, three of which also have data for taxis. Figure 1 summarizes estimates of the mileage-based capacity utilization measure ( $F^m$ ) for Los Angeles and Seattle, the only two cities for which we have been able to obtain information on taxi drivers' miles.

**Table 1: Capacity Utilization Rate (Percent of Work Hours with a Passenger) for Taxi and UberX Drivers, Select Cities**

	<b>f</b>		<b>F</b>	
	<b>Taxi</b>	<b>UberX</b>	<b>Taxi</b>	<b>UberX</b>
Boston	NA	46.8%	32.0%	46.1%
Los Angeles	NA	51.7%	NA	50.3%
New York	48.3%	50.9%	49.5%	51.2%
San Francisco	38.4%	54.9%	38.5%	54.3%
Seattle	NA	43.5%	NA	43.6%

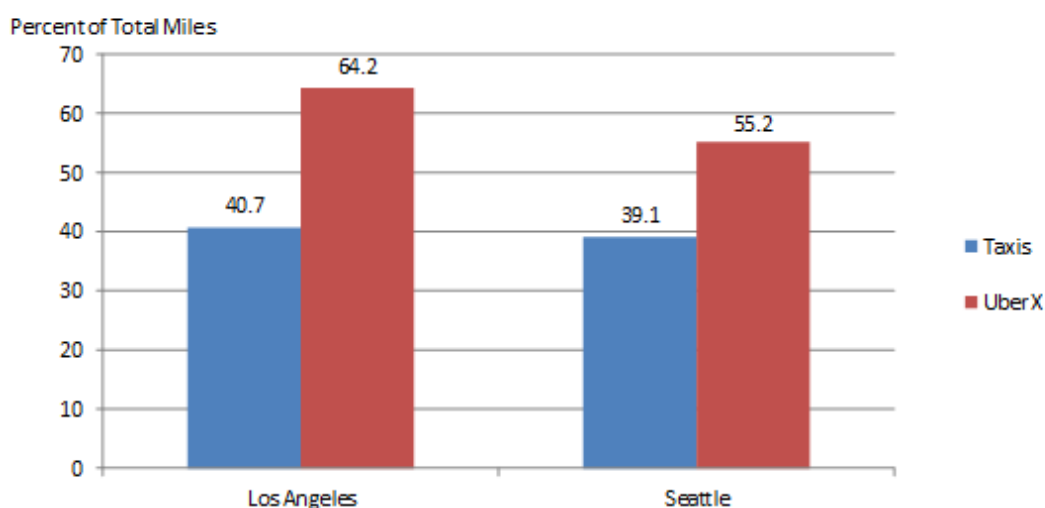
Sources: Uber Technologies, Inc.; authors' calculations using NYCTLC Microdata; Leah-Vincent; and Nelson/Nygaard (2013); see text for further details.

Notes:  $f$  is the average across drivers.  $F$  is ratio of aggregate passenger hours to aggregate hours worked. Boston taxi data are average of three days in 2013, New York taxi data are for 2013, and San Francisco taxi data are for July-October 2013; Uber data for Boston are for three corresponding days in 2015, Uber data for San Francisco are for July-October 2015, and Uber data for all other cities are for the 12 months ending December 1, 2015.

Regardless of the measure used, the results show a clear pattern: UberX drivers have a substantially higher capacity utilization rate than do taxi drivers in every city except New York, where the utilization rates are very similar. In Boston, the time-based capacity utilization rate  $F^h$

is 44 percent higher for UberX drivers than for taxi drivers, and in San Francisco it is 41 percent higher. Notice also that  $f^h$  and  $F^h$  are very similar where they both are available, consistent with there being little correlation between  $f_i$  and  $h_i$ . As a result, in San Francisco,  $f^h$  is 43 percent higher for UberX drivers than for taxi drivers, very close to the differential for  $F^H$ , and in New York both ratios are close to parity. Across the five cities, UberX drivers have a passenger in their car around half the time that they are working, whereas taxi drivers have a passenger in their car anywhere from 32 percent of the time in Boston to nearly half the time in New York City.

**Figure 1: Capacity Utilization Rate (Percent of Miles Driven with a Passenger) for Taxi and UberX Drivers in Los Angeles and Seattle**



Source: Uber Technologies, Inc.; LADOT; City of Seattle, Regulatory Compliance and Consumer Protection Division; Authors' calculations.  
 Notes: LA and Seattle are 2013-14 and Uber is the 12 months ending December 1, 2015; see text for further details.

The mileage-based capacity utilization rates ( $F^m$ ) tell a similar story.<sup>9</sup> In Los Angeles, taxi drivers have a passenger in the car for 40.7 percent of the miles they drive, while UberX

<sup>9</sup> For Los Angeles taxi drivers  $F^m$  is the average value of  $F^m$  taken over the 24 months of 2013 and 2014. For Seattle,  $F^m$  is the average of the 2013 and 2014 values.

drivers have a passenger in the car for 64.2 percent of their miles, resulting in a 58 percent higher capacity utilization rate for UberX drivers. In Seattle, UberX drivers achieve a 41 percent higher capacity utilization rate than taxis in terms of share of miles driven with a passenger in the car. Notice also that the capacity utilization rates are generally higher when measured by miles than hours. Across the five cities, for example, for UberX drivers the average of  $F^m$  is 61.0 percent and the average of  $F^h$  is 49.1 percent. (Unfortunately, no jurisdiction reports data that allow for the calculation of the capacity utilization rate in miles *and* in hours for taxi drivers, but looking across cities it appears that  $F^m$  is greater than  $F^h$  for taxis as well.) The mileage-based measure of the capacity utilization rate would be higher than the time-based measure if, for example, drivers arrive early to pick up some passengers and wait for them (without turning on the meter), or if drivers park or drive more slowly in the interval between dropping off a passenger and picking up a new one, or if drivers take breaks during their shifts that are counted as work hours.

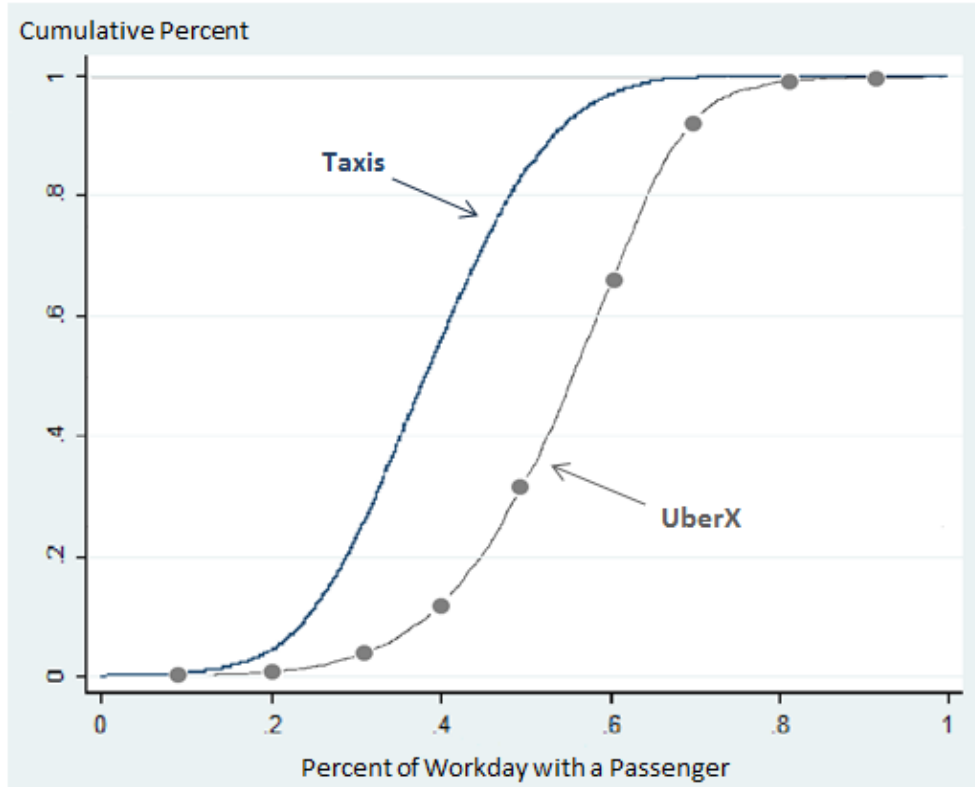
For taxis in Los Angeles and Seattle, we can look at variations in  $F^m$  over time. In Los Angeles the capacity utilization rate was relatively stable over time, only varying between 38.6 percent and 42.8 percent in the months between January 2009 and January 2015. In Seattle, the rate mostly trended upward from 40.7 percent in 2005 to 45.7 percent in 2013, before dropping to 32.6 percent in 2014, perhaps because of competition from Uber.

Lastly, Figure 2 presents the empirical cumulative distribution functions of  $f^h_i$  for taxi drivers and UberX drivers in San Francisco. Specifically, drivers are arrayed by the share of work hours they have a passenger in the car on the horizontal axis, and the percent falling below each value is shown on the vertical axis. The differences in the mean capacity utilization rates are not driven by a small number of drivers. At all percentiles, the UberX drivers have a higher capacity utilization rate than taxi drivers. Moreover, if we look at different time intervals of the



day, UberX drivers in San Francisco have a higher utilization rate at all time intervals than taxi drivers.

**Figure 2: Cumulative Distribution Function of Percent of Work Hours with a Passenger for Taxi and UberX Drivers in San Francisco**



Source: Data provided by Uber Technologies Inc. and Vincent Leah-Martin.  
Notes: Taxi data are for July-October 2013, Uber data are for July-October 2015.

### III. Discussion

There are several possible reasons why UberX drivers may achieve significantly higher capacity utilization rates than taxi drivers. First, Uber utilizes a more efficient driver-passenger matching technology based on mobile internet technology and smart phones than do taxis, which typically rely on a two-way radio dispatch system developed in the 1940s or sight-based street hailing. Second, in most cities Uber currently has more driver partners on the road than the largest taxi cab company. Apart from the technology, there are network efficiencies from scale,

as pure chance would likely result in an Uber driver being closer to a potential customer than a taxi driver from any particular company given the larger scale of Uber. Third, inefficient taxi licensing regulations can prevent taxi drivers who drop off a customer in a jurisdiction outside of the one that granted their license from picking up another customer in that location. Fourth, Uber's flexible labor supply model and surge pricing probably more closely matches supply with demand during peak demand hours and other hours of the day.

We cannot explore the importance of all of these factors, but we can explore aspects of some of them. First, for three cities -- New York, Seattle and LA -- we have capacity utilization rates for UberX drivers who worked at least 7 hours in the day. Because taxi drivers tend to work much longer shifts than UberX drivers, one possibility is that the longer work day reduces productivity, or the tendency to work during both slow and busy times of the day lowers the capacity utilization rate of taxi drivers. For UberX drivers, however, the capacity utilization rates were essentially identical for the drivers who worked at least 7 hours in the day as they were for drivers as a whole. This suggests that the exit and entry of UberX drivers during the course of the day equilibrates the market so that drivers achieve essentially the same utilization rate regardless of how long they work, or that longer shifts are not the central reason why taxi drivers have lower utilization rates than Uber drivers.<sup>10</sup>

Insofar as matching technology is concerned, Frechette, Lizzeri and Salz (2015) conducted an elaborate simulation exercise where they estimated a dynamic general equilibrium model of the taxi market in New York City in 2011-12, allowing for search frictions and endogenous driver entry and stopping decisions. In one counterfactual simulation, they changed

---

<sup>10</sup> The finding that hours and the capacity utilization rate are essentially uncorrelated is consistent with Hall and Krueger's (2015) finding that hours and revenue earned per hour are essentially uncorrelated for UberX drivers.

the matching technology and assumed that drivers knew the location of the closest passenger. Although this is not the same as switching to the Uber app, it gives a flavor for the potential role of more efficient technology for matching drivers and passengers. This policy was estimated to raise the fraction of work time with a passenger by 7.2 percent. Table 1 indicated that the capacity utilization rate is 5.3 percent or 3.5 percent higher for Uber than taxi drivers in New York. So these findings suggest that differences in driver-passenger matching technology can more than account for the minor difference in capacity utilization rate between taxi drivers and UberX drivers in New York City.

An important caveat, however, is that New York City is an apparent outlier in that the capacity utilization rates of taxi and UberX drivers are much more similar in New York than in other cities we have been able to examine. It is quite plausible that the high population density of New York City supports more efficient matching of taxis and passengers through street hailing than is the case in other cities. Indeed, our results suggest that New York is the only city where taxi and UberX drivers achieve a similar capacity utilization rate.

Regardless of the reasons for the higher capacity utilization rate of UberX compared to taxi drivers, our findings have implications for the efficiency of for-hire drivers. Averaging across the five cities with available data (and across the two measures), the capacity utilization rate is 38 percent higher for UberX drivers than for taxi drivers. Ignoring fixed costs, if fares are linear, this implies that UberX drivers could charge 28 percent ( $=1-1/1.38$ ) less than taxis and earn the same amount of revenue per hour. In Los Angeles, which exhibited the biggest difference in capacity utilization, fares could be 37 percent lower. It is also worth emphasizing that differences in utilization rates have implications for resources other than passengers and drivers. For example, for every mile that taxi drivers in Los Angeles drives with a passenger in

the car, they drive 1.46 miles without a passenger; the comparable figure for UberX drivers is 0.56 mile. This difference likely translates to greater traffic congestion and wasteful fuel consumption.

Lastly, our results bear on the literature on occupational licensing. Although occupational licensing can provide many benefits for consumers, workers and society, it could also reduce efficiency and distort markets. Occupational licensing has grown even in fields where there is little public safety or other societal benefit from licensing restrictions. Given that vested interests that benefit from occupational licensing (including the jurisdictions that collect licensing fees) have made it difficult to repeal occupational licensing, one way in which inefficient, unnecessary and counterproductive occupational licensing can be reduced is through disruptive change, such as brought about by a new technology.

## References

Farber, Henry S. “Why You Can’t Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers,” *The Quarterly Journal of Economics* 130(4): 1781-1824, 2015.

Guillaume Frechette, Alessandro Lizzeri, and Tobias Salz. “Frictions in a Competitive, Regulated Market: Evidence from Taxis,” New York University, August 10, 2015. Available at [http://www.econ.nyu.edu/user/lizzeri/draft\\_071015.pdf](http://www.econ.nyu.edu/user/lizzeri/draft_071015.pdf)

Jonathan Hall, and Alan Krueger. “An Analysis of the Labor Market for Uber’s Driver-Partners in the United States,” Princeton University Industrial Relations Section Working Paper No. 587, January 2015.

Leah-Martin, Vincent. “When to Quit: Narrow Bracketing and Reference Dependence in Taxi Drivers,” Working Paper University of California San Diego, 2015.

Kleiner, Morris, and Alan Krueger. “Analyzing the Extent and Influence of Occupational Licensing on the Labor Market. *Journal of Labor Economics* 31(Suppl. 1): S173-202, 2013.

Nelson/Nygaard (with Taxi Research Partners). “Taxi Consultant Report,” City of Boston, October 11, 2013. Available at <http://www.cityofboston.gov/mayor/pdfs/bostaxiconsultant.pdf>

Soper, Taylor. “The Uber effect: Seattle taxi industry revenue dipped 28% in past 2 years,” *GeekWire*, June 11, 2015. Available at <http://www.geekwire.com/2015/the-uber-effect-seattle-taxi-industry-revenue-dipped-28-in-past-two-years/>