Do Digital Platforms Reduce Moral Hazard? The Case of Uber and Taxis *

Meng Liu † WUSTL and MIT

Erik Brynjolfsson ‡ MIT and NBER

Jason Dowlatabadi § Uber Technologies

February 1, 2019

Abstract

Digital platforms provide a variety of technology-enabled tools that enhance market transparency, such as real-time monitoring, ratings of buyers and sellers, and low-cost complaint channels. How do these innovations affect moral hazard and service quality? We investigate this problem by comparing driver routing choices and efficiency on a large digital platform, Uber, with traditional taxis. The identification is enabled by matching taxi and Uber trips at the origin-destination-time level so they are subject to the same underlying optimal route, by exploiting characteristics of the pricing schemes that differentially affect the incentives of taxi and Uber drivers in various circumstances, and by examining changes in behavior when drivers switch from taxis to Uber. We find that (1) taxi drivers route longer in distance than matched Uber drivers on metered airport routes by an average of 8%, with non-local passengers on airport routes experiencing even longer routing; (2) no such long routing is found for short trips in dense markets (e.g., within-Manhattan trips) or airport trips with a fixed fare; and (3) long routing in general leads to longer travel time, instead of saving passengers time. These findings are consistent with the platform tools reducing driver moral hazard, but not with competing explanations such as driver selection or differences in driver navigation technologies.

*We thank Keith Chen, Dean Eckles, Andrey Fradkin, Xiang Hui, John Horton, and Erina Ytsma, as well as seminar participants at MIT, Uber, AEA, Marketing Science, SICS, and NBER Summer Institute Industrial Organization Workshop and the NBER Digitization Workshop, CIST, FCC, and WISE for their valuable comments and suggestions. The MIT Initiative on the Digital Economy provided generous research support, and Uber provided essential data. Dowlatabadi is a current employee at Uber. The views expressed here are those of the authors and do not necessarily reflect those of Uber Technologies, Inc. All errors are ours.

† mengl@wustl.edu

‡ erikb@mit.edu

§ jasond@uber.com

1 Introduction

Digital platforms are growing rapidly, and so are their economic effects. Examples include large platforms such as Uber for ride-hailing and Airbnb for accommodations, as well as a growing number of smaller platforms such as ClassPass for fitness studios and Rover for dog-walking. Digital platforms are often designed to mitigate information asymmetry problems through the use of new technologies and incentive systems, such as ratings of buyers and sellers, real-time monitoring, and low-cost complaint channels. For example, 73.5% of New York City (NYC) UberX trips are rated by passengers and Uber fare adjustments are made for 1 in every 170 trips. In contrast, NYC taxi complaints are much more difficult to lodge and occur only 1 in every 6,356 trips.

One of the biggest barriers to market efficiency is asymmetric information, particularly moral hazard. Do digital platforms reduce moral hazard and improve service quality, compared to traditional settings? In this paper, we study this question by comparing a particularly successful and pervasive digital platform, Uber, with traditional taxis. Our findings will be of broad interest to economists because we document a significant effect of this digital platform in reducing moral hazard. This is essential for a better understanding of the nature of online-offline competition, welfare in the digital economy, and ultimately the potential for using technology and platform design to improve many other markets where moral hazard and asymmetric information is significant.

Specifically, we investigate driver detour, defined as the extra distance a driver adds to the fastest route. This is a measure of driver moral hazard in our context, and this type of strategic behavior is found prevalent among taxi drivers (Balafoutas et al. (2013), Rajgopal and White (2015), Balafoutas et al. (2017), and Liu et al. (2017)). In a hypothetical situation where a taxi driver and an Uber driver drive between the same two points at the same time, the difference in their routing decisions should reflect factors that affect the benefits and costs of detouring. To the extent that features such as shared GPS navigation, tech-aided monitoring, ratings, and digital feedback increase market transparency for passengers and therefore increase penalty of driver moral hazard, the Uber driver's routing is likely more efficient than that of the comparable taxi driver in situations with high moral hazard payoffs for both drivers.

A key challenge exists in identifying the effects of driver moral hazard — driver moral hazard is not directly observed. The inability to directly observe driver moral hazard is due to the lack of optimal routing benchmark at the time of the trip. For example, using a long-run average trip distance queried from routing engines such as Google Maps may underestimate the true real-time optimal route and overestimate the detour if there was a temporary road closure that required a longer route than indicated by the long-run average.

Therefore, one needs to construct valid comparison groups to infer opportunistic behavior by using detailed trip-level data of both taxis and Uber. We overcome this challenge by leveraging public taxi trip records and proprietary UberX data in NYC and matching taxi and Uber trips at a strict origin-destination-time level — both the pick-ups and drop-offs of a taxi trip and its matched Uber trip have to be from the same small area centered at a street intersection, on the same street, following in the same traffic direction, and less than or equal to 15 minutes apart. As a result, the drivers of the matched trips are subject to the same real-time optimal routes, even if these optimal routes are not directly observed.

These matched pairs of taxi and Uber trips then become our units of analysis. We explore the variation in the within-match taxi-Uber routing difference, across route types that represent different moral hazard incentives. We find that taxi drivers and Uber drivers share essentially the same driving distances when completing short trips that start and end in Manhattan; in fact, their routing behavior on this type of trips appears to be quite efficient when compared to a routing engine benchmark. However, when on airport trips where both taxi and Uber fares are metered in trip distance, taxi drivers on average route longer in distance relative to Uber drivers by 8%. Taxi drivers appear to route even longer in distance when the airport passenger is from outside the New York City area. However, for trips between Manhattan and JFK airport where taxi fare is a fixed amount while Uber fare is metered, no such taxi long routing is observed.

These empirical findings are consistent with our stylized model of driver moral hazard, where the driver decides whether and how much to detour and the speed of travel in order to maximize payoff. The key tension is a trade-off between the costs and benefits of the opportunistic behavior — detour increases driver earning from the current trip, but at the same time it increases penalty cost in terms of expected monetary and reputation cost of cheating behavior, as well as opportunity cost, in terms of expected forgone earnings (detouring usually prolongs travel time, reducing opportunities for additional trips). It then follows that drivers lack detour incentives when driving short trips in dense markets (e.g., within-Manhattan trips), because of low return due to short distance and high opportunity cost due to high demand at the drop-off location. Similarly, the detour incentive is essentially "shut down" when the airport fare is fixed. However, the detour incentive is greater on metered airport trips where the long distance rewards detour more, and drivers can exploit the information asymmetry further in the case of non-local passengers on these routes.

We explore several competing explanations and find that the data are not compatible with them. First, the observed moral hazard could be an artifact of increased GPS usage among Uber drivers, which may have improved their routing compared to taxi drivers. However, if GPS accounts for the 8% increase in taxi-Uber distance ratio for metered airport trips, the GPS effect[1] should be at least as salient for JFK trips, because

[1]By "GPS effect", we mean the effect due to technology-enhanced navigation. Note that this is different from the effect of GPS

JFK trips are significantly longer in distance than metered airport trips. Given that the taxi-Uber distance ratio is not significantly different from the Manhattan “no detour” benchmark, the GPS-enhanced navigation cannot convincingly explain the empirical patterns.

The second competing hypothesis we explore is whether taxi drivers possess superior routing information than GPS so that they route longer but save passengers time. We focus on a popular route between Midtown Manhattan and LaGuardia airport where we can identify the particular route drivers take, using information on bridge/tunnel tolls. We find that taxi drivers frequently choose the bridge that leads to the longest distance, these long routes on average result in longer travel times when compared to shorter routes taken by drivers completing similar trips at the same time, and this long-routing strategy is more seen in taxi drivers with more route-specific experience. Therefore, the findings are not consistent this competing hypothesis, but rather they lend more nuanced support to our main hypothesis of driver moral hazard.

Another competing explanation is driver selection, instead of moral hazard. On the intensive margin, it may be that strategic driver types select into profitable routes. We rule this out by controlling for driver fixed effects and finding no material changes in our results. On the extensive margin, taxi and Uber may represent different distributions of driver types. While we cannot directly observe and compare types, we indeed observe significant routing efficiency improvement after taxi drivers became Uber drivers, which indicates that drivers adapt to new market arrangements via behavioral updating.

Our findings shed light on the incentive devices and pricing schemes as the underlying mechanisms for the reduced strategic behavior at Uber. As such, our findings have implications for regulators and industry participants. For taxi regulatory agencies, our results provide support for the development and implementation of smart phone applications that handle functions such as taxi dispatching and matching with passengers, digital payment, and passenger monitoring. Also, it is important for taxi regulatory agencies to re-evaluate the current pricing scheme that rewards taxi cab speeding as well as the impacts of alternative pricing structures. For digital platforms such as Uber, our findings suggest an opportunity for machine-learning-based techniques to detect various types of driver opportunistic behavior, which may further enhance market transparency and trust building.

The rise of digital platforms has led to an enormous increase in transactions of services that were traditionally provided offline only, and it also presents new challenges and opportunities for technology-enabled market designs to improve market efficiency. The taxi industry offers a clean laboratory to study the relationship between technology and incentive design for two main reasons: on one hand, this is a highly competitive marketplace of a homogeneous, well-defined service (namely, transporting a passenger from

as a monitoring device for passengers.

one point to another); on the other hand, the rich spatial data allow us to make precise and valid comparisons between taxis and Uber, while such counterfactual groups can be difficult to form in other industries. As a result, evidence from this industry makes a strong and clear inference about the effect of digital platforms on moral hazard and service quality, which can help us better understand similar challenges in other industries and markets as well.

1.1 Literature and Contribution

Our paper is closely related to several strands of the literature. The first is on how technology, particularly information technology (IT), mitigates the agency problem in various settings (Tabarrok and Cowen (2015)). In the typical workplace, IT-enabled monitoring has been found to be productivity-enhancing through complementing performance pay (Aral et al. (2012), Bresnahan et al. (2002)), reducing employee shirking (Nagin et al. (2002)) or misconduct (Pierce et al. (2015)), and increasing standard process compliance (Staats et al. (2016)). In the context of trucking, Hubbard (2000) has found that on-board computers which facilitate monitoring of drivers increase productivity by improving both drivers' incentives and managers' resource allocation decisions. Duflo et al. (2012) have shown that incentive pay enabled by tech-aided monitoring can raise teachers' attendance rate and consequently student performance. Reimers et al. (2018) have found that insurance companies' monitoring technologies reduce driver moral hazard and fatal accidents. Sudhir and Talukdar (2015) have illustrated the role of IT in inducing business transparency by showing more corrupt businesses resist IT adoption more.

Besides the traditional settings, there are also studies on digital market designs that improve productivity by regulating agent incentives. Hui et al. (2016) have identified efficiency gains from eBay's buyer protection program as a result of reduced seller moral hazard and seller adverse selection. Klein et al. (2016) have shown that a change in eBay's policies that led to less biased buyer ratings of sellers also improved seller effort and quality without inducing sellers to exit the market. Gans et al. (2017) have evaluated the role of Twitter as a mechanism of consumer voice in disciplining firms for low quality. Liang et al. (2016) have found that IT-enabled monitoring mitigates moral hazard on an online labor platform.

While these aforementioned studies focus on technological improvements either within the offline or online setting, we are among the first to provide a direct online-offline comparison to study the relationship between technology, agent incentives, and quality provision. As many sectors are being digitized, empirical studies of how incentives and quality provision differ between online and offline markets become crucial for a better understanding of the nature of online-offline competition.

The second strand of literature our paper is related to is the literature on digital disruption and online-offline competition (Bakos (1997), Brown and Goolsbee (2002), Brynjolfsson et al. (2003), Brynjolfsson and Smith (2000), Forman et al. (2009), Overby and Forman (2014), among many others. See Goldfarb and Tucker (2017) for a review). In particular, this paper contributes to the studies of emerging tech-aided ride-hailing platforms. These platforms may reduce matching frictions between drivers and passengers (Buchholz (2015), Frechette et al. (2016)) with real-time technologies and dynamic pricing (Castillo et al. (2017), Hall et al. (2015)), as reflected in greater capacity utilization (Cramer and Krueger (2016)), as well as quick adjustments to market equilibrium (Hall et al. (2017)). Specifically, efficiency induced by dynamic pricing critically depends on consumer preferences and the tradeoff between wait time and price (Lam and Liu (2017)), and driver labor supply that responds to wage fluctuations (Chen and Sheldon (2016)). Consumers benefit from ride-hailing platforms extensively, because of surge pricing (Cohen et al. (2016)), reduced drunk driving (Greenwood and Wattal (2017)), and improved service quality due to safer driving (Athey and Knoepfle (2018)). Drivers also benefit from these platforms due to flexible work arrangement (Chen et al. (2017), Hall and Krueger (2015)) and commission schemes that allow for driving without a lease (Angrist et al. (2017)). We find that these technological and organizational features have important implications on driver incentives and quality provision, and thus add an important layer in the analysis of efficiency.

Finally, our findings resonate with empirical work on taxi driver opportunistic behavior. Balafoutas et al. (2013) have found that taxi drivers detour when passengers are less informed about the optimal routes or the local taxi fare structure. Liu et al. (2017) have identified non-local passengers from local passengers based on the destinations of trips originating at New York City's airports and have found that taxi drivers defraud non-locals more on LaGuardia trips that are metered, but not so on JFK flat-fare trips. Balafoutas et al. (2017) have shown that drivers may also defraud more when passengers explicitly state that their expenses will be reimbursed. Rajgopal and White (2015) point out the importance of regulatory restrictions on driver fraud, as they have found greater likelihood of driver fraud when dropping passengers off in areas where taxis are not allowed to pick up subsequent passengers. We contribute to this literature by examining how moral hazard can be mitigated by tech-aided ride-hailing platforms.

2 NYC Taxis vs. Uber: Market Design and Pricing

2.1 Taxi and Uber Market Design

The market design for the Uber platform differs significantly from that of taxis. First, GPS navigation is widely adopted and used by Uber drivers, while taxi drivers mainly navigate without GPS. The Uber app is designed in a way that GPS navigation is integral to both driver and passenger: When the driver picks up a passenger and starts the trip, Uber's built-in GPS automatically initiates, or the app switches to the preferred GPS that the driver has set up (e.g., Google Maps and Waze). Therefore, driver routing on Uber is more transparent for passengers relative to taxis.

Second, Uber implements a set of institutional design choices that aim at aligning driver incentives and facilitate monitoring by the passengers. These are absent or costly with taxis. With the Uber app, passengers can readily monitor driver routing in real time; passengers can either monitor the route on their own smartphone app, or look at the driver's app, since the driver's phone is usually mounted in a way that it is visible to passengers. This way, passengers can easily tell whether or not the driver is taking the route that is given by the GPS.

Uber uses a highly-visible rating system that is easy for users to update. After each ride, passengers are prompted to select a star rating, and therefore most passengers rate their drivers (73.5% for NYC UberX, January to June, 2016). Similar to other reputation systems that are subject to rating inflation (Filippas et al. (2018)), Uber driver ratings are highly concentrated with a mean of 4.74 out of 5 (see Figure 1a). Drivers with low ratings are constantly warned by Uber. Uber starts to consider deactivating a driver when the driver rating falls below a threshold (4.5 in NYC). Drivers appear to be very concerned about their ratings[2], and perhaps as a result, the actual deactivation risk is relatively low (about 3%).

Figure 1: Distance and Duration Ratios of Matched Taxi and Uber Trips

(a) NYC UberX Driver Rating Distribution

(b) Top UberX Fare Adjustment Reasons

Notes. Both plots are based on NYC UberX trip data, January to June, 2016. Figure (b) plots UberX fare adjustment reasons, conditional on a fare adjustment being made, for adjustments that account for 1% or more of the total.

In addition to monitoring and rating, verification and complaints can be made with little friction on Uber, thanks to electronic trip records. Passengers can revisit the historical trip summaries on their app to verify certain details. In the case of negative riding experiences, Uber passengers can easily file a complaint

[2]The qualitative study by Lee et al. (2015) states that "Drivers took their ratings seriously. High ratings such as 4.98 became a source of pride whereas a rating below 4.7 became a source of disappointment, frustration, and fear of losing their jobs."

through the app, and Uber handles the conflict resolution by evaluating the trip records. By contrast, taxi passengers in these situations can either call the Taxi and Limousine Commission (TLC) hotline or visit the TLC website, but the process is usually long and may require legal procedures. In 2016, taxi complaints were 1 in every 6,356 trips, whereas Uber fare adjustments were 1 in every 170 trips. Figure 1b lists the main reasons of fare adjustments, with the number one reason being "inefficient route".

2.2 Taxi and Uber Pricing

Pricing also differs between taxis and Uber. NYC taxi fares are set by the TLC[3]. Most routes are metered with a base fare of \$2.50 upon entry and \$0.50 for every $\frac{1}{5}$ miles traveled, plus taxes, fees, and tolls. A \$0.50 per-minute charge is applied in place of the per-mile charge when the traffic is slow (less than 12 miles per hour). The exception is that routes between Manhattan and JFK Airport are not metered; instead, a flat rate of \$52 applies. Some taxi drivers are medallion-owners who essentially run the business as an entrepreneur. Other drivers lease the medallions on a daily, weekly, or monthly basis, and they collect all the revenues minus gasoline and some vehicle maintenance costs. In both cases, drivers are residual claimants who are incentivized to maximize earnings.

Unlike the pricing of taxis, Uber's pricing schedule is consistent in both fast and slow traffic. The UberX base fare includes a fixed component of \$2.55, \$0.35 per minute of travel, and \$1.75 per mile of travel, plus taxes, fees, and tolls. On top of the base fare, passengers also need to pay the surge multiplier in effect at the time of request. For a 2-mile, 10-minute trip with a surge multiplier of 2, UberX costs $2\times(\$2.55+\$0.35\times10+\$1.75\times2) = \$19.10$, plus taxes, fees, and tolls. Unlike taxis' fixed fare on certain routes, all Uber routes in NYC are metered according to the same pricing formula. Uber drivers keep all trip earnings minus Uber's commission, which usually runs between 20% and 30%. Uber drivers who operate using their own cars are responsible for all operation-related expenses, such as insurance, maintenance, and gasoline. Many NYC Uber drivers instead rent a vehicle from fleet owners due to heavy TLC requirements such as commercial insurance.

3 Taxi-Uber Routing Difference Is Consistent with Moral hazard

3.1 A Simple Framework of Driver Routing Decisions

In this section, we describe a simple, stylized model of driver routing decisions and moral hazard, where the framework builds on Liu et al. (2017). The purpose of the model is to paint drivers' trade-offs and characterize situations of driver moral hazard, which will motivate the empirical analysis.

[3]Refer to the official language on the pricing rule: http://www.nyc.gov/html/tlc/html/passenger/taxicab_rate.shtml

A risk-neutral driver maximizes her payoff by choosing among alternative routes, where essentially the driver decides on the amount of detour (from the optimal passenger route) as well as driving speed. For a given route at a given time, let d_0 denote the distance of the optimal route given by a generic GPS that optimizes trip time. That is, any other route with a different distance than d_0 expects a longer travel time. Then the realized trip distance, d, is given by:

$$d = d_0(a + x + \epsilon), \tag{1}$$

where a represents the driver's ability, e.g., driver's knowledge of the streets and navigation skills, and $a \in [1, +\infty)$. Let x denote the amount of detour, where $x \in [0, +\infty)$. Let ϵ denote the random driver-route shock that affects routing efficiency, which is normally distributed with a mean 0 [4]. For simplicity, let the realized travel time, denoted by t, be linear in trip distance:

$$t = \gamma d_0(a + x + \epsilon) + y, \tag{2}$$

where γ measures how trip distance maps into trip time, and $\gamma \in (0, +\infty)$. Let y represent the extra travel time incurred when the driver drives at a different speed than the ongoing traffic: $y > 0$ when the driver drives relatively slow, and $y < 0$ when the driver drives relatively fast.

Let the metered fare be characterized by the base fare upon entry p_0, the per-mile rate p_d, and the per-minute rate p_t[5]. Let s denote the surge multiplier, where $s = 1$ for taxi trips, and $s \geq 1$ for Uber trips. Let q_e represent the probability of getting a subsequent passenger at the drop-off location and time if there was no detour. Let e_t denote the per-minute earning of the forgone trip.[6] Therefore, $q_e e_t(\gamma d_0 x + y)$ measures the earnings from forgone service minutes. Then the driver chooses the amount of detour (x) and the speed of driving (equivalent to y) to maximize the following expected payoff function:

$$\begin{aligned} \underset{x,y}{\text{Max}}\ \text{E}\ \{s[p_0 + p_d d_0(a + x + \epsilon) + p_t(\gamma d_0(a + x + \epsilon) + y)] \\ - f(x; d_0, \Theta) - g(y; d_0, \Theta) - q_e e_t(\gamma d_0 x + y)\}, \end{aligned} \tag{3}$$

where f is the penalty cost of detour, which can be viewed as the probability of getting caught times the

[4] It is possible for the realized trip distance to be shorter than the GPS-suggested distance d_0, when the random shock ϵ is sufficiently negative. This occurs, for example, when a road turn that is permitted during a certain time of the day shortens the route but is not captured by the GPS.

[5] Note that in NYC normal traffic, $p_t = 0$ for taxis.

[6] e_t can be thought of as $\frac{p_0 + p_d D_e + p_t T_e}{T_e}$, where D_e and T_e are the expected length and duration of the forgone trip, respectively.

monetary cost and/or reputation cost. The cost may be in the form of fines (taxis), lost tips (taxis)[7], low ratings (Uber), and refund to passengers (taxis and Uber). f is assumed twice differentiable in x, with a parameter set $\{d_0, \Theta\}$, and $f(x = 0) = 0$, $f_x > 0$, $f_{xx} > 0$. In addition, $s \in \Theta$, and $f_{xs} > 0$, meaning that the marginal detour penalty on Uber is greater when surge is greater[8]. Defined similarly as f, g is the expected monetary cost, reputation cost, or both associated with the travel speed: $g > 0$ for all y, i.e., both driving unnecessarily slow and unnecessarily fast relative to the traffic tend to be noticed and penalized by the passenger [9]; $g_y < 0$ for $y < 0$, $g_y > 0$ for $y > 0$, and $g_{yy} > 0$.

Taken together, the driver's problem in Equation 8 is to solve two trade-offs: One trade-off is between the monetary reward of detour and the opportunity cost of detour, where the opportunity cost consists of the expected detour penalty and the forgone payoff; the other similar trade-off applies to driving speed. Then, the first-order conditions and comparative statics lead to the following implications (see the online appendix for detailed derivations):

Implication 1: *Under mild assumptions, drivers tend to detour more on longer routes than on shorter routes.* Driver detour incentive is greater on longer routes because longer distance increases detour payoffs as a direct outcome of the pricing structure. This regularity holds as long as the demand at the drop-off location is not sufficiently high and the marginal detour penalty does not increase significantly with trip distance.

Implication 2: *Drivers detour more when the rider is a non-local passenger than when the rider is a local passenger.* This is because non-local passengers are less likely to notice the detour since they lack knowledge of the local traffic and road networks.

Implication 3: *Drivers detour more during high surge prices, under mild assumptions.* Driver detour incentive increases in surge price, provided that the increase in marginal detour payoff due to high surge dominates the increase in marginal detour penalty due to high surge.

Implication 4: *Drivers detour less (respectively, more) when the demand at the drop-off location is higher (respectively, lower).* This is a direct mapping of the payoff function that is strictly decreasing in the drop-off demand.

Implication 5: *Everything else held constant, taxi drivers have greater incentives than Uber drivers to drive faster than other traffic on the road.* This is intuitive because taxi driver time is not paid for according to the taxi meter rule while Uber driver time is.

[7] By the end of our sample period, Uber had not implemented the tip feature in the application.

[8] This is because (1) Uber passengers are more incentivized to monitor driver routing when surge is high, and detour may result in worse ratings than when surge is not in effect; (2) fare adjustments reflect surge multipliers.

[9] While it is true in some cases passengers prefer fast driving, speeding and weaving in and out of lanes are among the leading factors of traffic accidents.

Let c denote taxis and u denote Uber. Equation (1) requires that for a taxi driver and an Uber driver completing the same trip, the following equality must hold,

$$\frac{d^c - d^u}{d_0} = a^c - a^u + x^{c*} - x^{u*} + \epsilon^c - \epsilon^u. \tag{4}$$

That is, the (normalized) difference in taxi and Uber routing, for the same trip, is a function of driver skills, strategic detours, and driver-trip random shocks. Given Implications 1-4, the difference in detours is a function of various route characteristics that affect driver detour incentives. This will motivate our empirical set-up to test moral hazard, which will be made more clear after we discuss the data and the matching.

3.2 Data and Matching

Our data combine NYC taxi trip records and Uber's proprietary UberX trip records for two six-month periods: January to June, 2016, and July to December, 2013. Taxi trip records contain detailed information such as pick-up and drop-off time and GPS coordinates, trip distance and duration, and various fares and fees. The 2013 taxi data contain driver ID and medallion numbers, but the identifiers were subsequently removed by the TLC due to privacy concerns. UberX trip records contain similar information, plus extra information such as the surge multiplier, driver ID, driver total number of trips on Uber, rider total number of trips on Uber, driver lifetime rating, and driver and rider rating for each trip. These trip records are massive data sets: in 2016, average daily taxi ridership is about 350,000 trips and average daily UberX ridership is about 120,000 trips.

As suggested by Equation 1, inference of detour incentives needs to be built on a valid counterfactual construction of taxi and Uber. To that end, we conduct granular geographical matching of taxi and Uber trips such that the matched trips are subject to the same underlying optimal routing. In brief, we match an Uber trip and a taxi trip if they go from the same Point A to the same Point B, and begin at roughly the same time (i.e., same day *and* minutes apart). The matching process is detailed below:

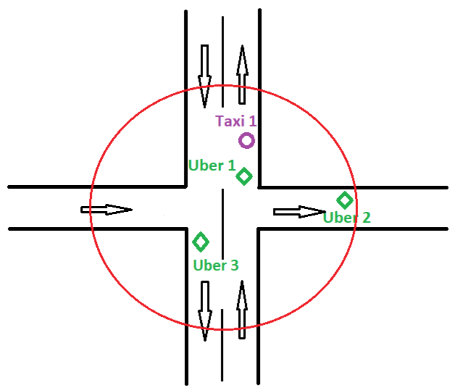
Step 1 (same street intersection): Because of the exceedingly high concentration of pick-ups and drop-offs around street intersections, we first define locations by dividing NYC into small Voronoi cells[10] (see Figure A1) centered at street intersections, where each street intersection is approximately 100 meters from its closest neighboring intersections. Using Figure 2 as an illustration, this means that we initially match Taxi 1, Uber 1, Uber 2, and Uber 3 in the circled area.

Step 2 (same street): We then restrict matched pick-ups to be on the same street,[11] because pick-ups on

[10]Roughly speaking, the Voronoi cell of one of pre-defined points (seeds) on a plane is the associated region that cover all points closer to that seed than to any other seed.

[11]The accuracy of GPS coordinates can be adversely affected by tall buildings in an urban area. Indeed, there are more cases

Figure 2: Pickup-dropoff-time Matching of Taxi and Uber Trips



different streets can be subject to different optimal routes even when they are going to the same destination. In Figure 2, this means that Taxi 1 will be matched with Uber 1 and Uber 3, but not with Uber 2.

Step 3 (same traffic direction): Following a similar logic as in Step 2, we further filter out matched pick-ups that follow different traffic directions of the same streets. Therefore, Taxi 1 and Uber 1 of Figure 2 remain in the matched sample.

We then apply the same filters (Steps 1-3) for drop-offs as well. For airport pick-ups and drop-offs, we match the trips based on the airport terminal.

Step 4 (real time): We further restrict matched trips to the ones that start within a short time window from each other so that they are subject to the same real-time traffic, road conditions, as well as other common factors. The time window for the main analysis is set at 15 minutes, and we apply other time windows (e.g., 5, 10, and 20 minutes) in the robustness checks.

When we use the 2016 taxi and Uber data, the matching process generates a sample of 173,770 pairs of matched trips. 70% of the matched pairs are airport trips, which is not surprising because the strict matching criteria make non-airport trips more difficult to match than airport trips. The vast majority (95%) of matched non-airport trips are trips that start *and* end in the most dense market of NYC — Manhattan Core, roughly the part of Manhattan below the north edge of Central Park. For reasons we will explain in the empirical section, we keep the Manhattan Core matched trips and drop other non-airport trips (e.g. trips between Brooklyn and Queens) from the sample, which leads to a 1.6% reduction of sample size. In addition, we drop matched pairs where the distance ratio (taxi distance divided by Uber distance, for the same matched pair) and the duration ratio (taxi duration divided by Uber duration, for the same matched pair) fall outside

where taxi and Uber pick-up and drop-off GPS pinpoints fall on a building instead of on the street in Midtown Manhattan than in other areas with less concentration of tall buildings. In these cases, we assign the trip to be on the street closest to its pinpoint.

of the range [0.5,1.5], in order to prevent extreme cases from affecting our results. This leads to a 2.5% drop in sample size. Finally, we discover that in the raw TLC taxi trip records, there are two taxi meter vendors with about equal shares, where Vendor 1 reports trip distance to the first decimal place and Vendor 2 to the second decimal place. A casual check of dozens of randomly selected short trips in Manhattan from Vendor 1 against their Google Maps shortest distances makes us believe that this meter vendor may have rounded down the actual trip distance. The rounding down may introduce bias to our estimates[12], which we provide evidence in the robustness section. Given this, we drop all matched pairs that involve Vendor 1 taxi trips.

Table 1: Summary Statistics (unit: a matched taxi-Uber pair)

Variable	**Mean**	**Std. Dev.**	**10th**	**Median**	**90th**
Taxi trip distance (miles)	8.76	5.30	1.13	9.69	16.66
Uber trip distance (miles)	8.56	5.35	1.16	9.21	16.80
Taxi distance / Uber distance	1.03	0.15	0.86	1.01	1.23
Taxi trip duration (minutes)	28.56	16.44	8.02	27.00	50.12
Uber trip duration (minutes)	30.39	17.34	8.93	28.68	53.17
Taxi duration / Uber duration	0.95	0.19	0.71	0.94	1.21
Airport	0.72	0.45	0	1	1
Metered airport	0.62	0.48	0	1	1
JFK (taxi fixed fare; Uber metered fare)	0.10	0.29	0	0	0
Non-local passenger	0.52	0.50	0	1	1
Surge multiplier	1.11	0.27	1	1	1.5
Surge (dummy; =1 if surge multiplier>1)	0.20	0.40	0	0	1
Uber driver total trips	2491.12	2010.00	358	2021	5311
Uber driver lifetime rating	4.75	0.09	4.63	4.76	4.85
Uber rider total trips	115.28	169.97	5	55	294
N	90,431				
No. of Uber drivers	23,484				

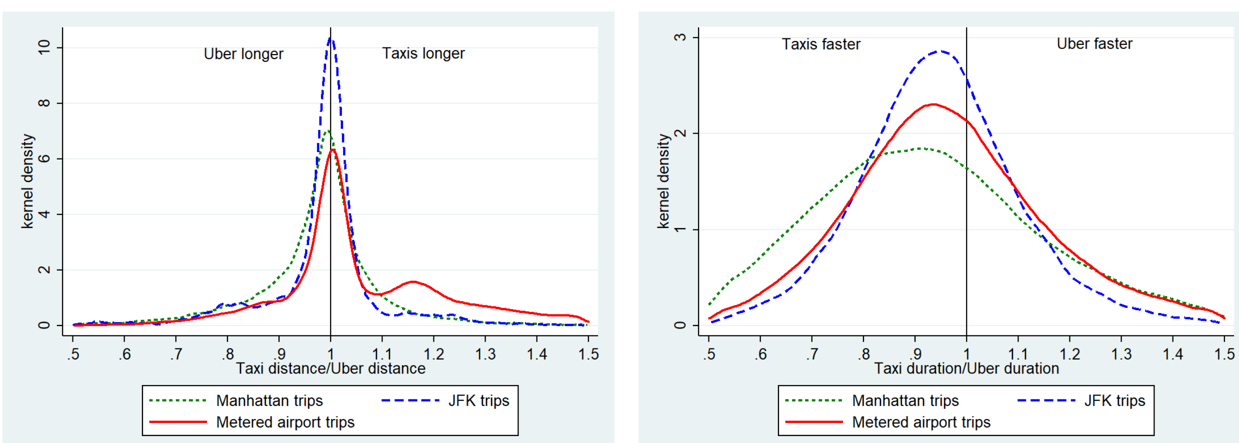
After above-mentioned sample restrictions, our sample contains 90,431 matched pairs, with 23,484 unique Uber drivers (about 54% of all UberX drivers in the same period). Table 1 summarizes the sample, where the unit of observation is a matched taxi-Uber pair. *Non-local passenger* takes the value of 1 if

[12]Consider a pair of matched Uber and taxi trips, where the Uber trip is 1 mile and the taxi trip is 0.95 miles, but the reported taxi trip length is rounded to 0.9 miles. Then the distance ratio would be 0.9 instead of 0.95, with a downward bias of -0.05. For a 9.95-mile taxi trip rounded to 9.9 miles with a matched 10-mile Uber trip, the downward bias is only -0.005 (0.99-0.995). Thus, the same amount of rounding error leads to proportionately greater downward bias on shorter routes.

the billing zip code of a given Uber passenger is outside of NYC, or if the billing zip code is missing, the passenger's city of most Uber trips is not NYC. The Uber surge multiplier is on average 1.11 and the frequency of surge (as opposed to the base fare) is about 20% in the matched sample. The sample is over-represented by airport trips, especially metered airport trips, compared to the population of taxi trips and Uber trips, due to the matching. There are three route types: (1) short, within-Manhattan routes, with an average route length of 1.67 miles; (2) routes between JFK and Manhattan where taxi fares are fixed and Uber fares are metered, with an average route length of 18.66 miles; and (3) all other airport routes where both taxi and Uber fares are metered, with an average route length of 10.10 miles, where 96.7% are LaGuardia trips, with the rest being Newark trips or trips between JFK and NYC outer boroughs. *For simplicity, from now on we refer to these route types as Manhattan trips, JFK trips, and metered airport trips*.

Figure 3a shows that taxi and Uber trips are similar in trip distance for Manhattan trips and JFK trips, as illustrated by the high concentration of the distance ratios around 1. However, taxi trips are significantly longer in distance than matched Uber trips for metered airport trips, as indicated by the second mode of the distribution around 1.15, as well as the fatter right tail. Figure 3b shows that taxis overall arrive faster than Uber. These patterns are consistent with our theory implications: (1) compared to Uber drivers, taxi drivers route longer on airport routes that are anecdotally more lucrative (Implication 1), and (2) taxi drivers in general drive at a greater average speed than Uber drivers (Implication 5). In the next section, we turn to formal tests of driver moral hazard.

Figure 3: Distance and Duration Ratios of Matched Taxi and Uber Trips



(a) Distance ratios for matched trips

(b) Duration ratios for matched trips

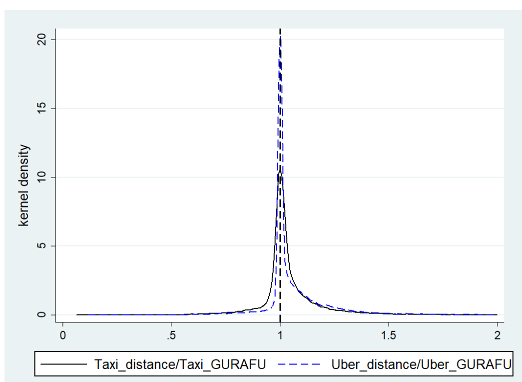
3.3 Empirical Strategy

Empirical Strategy and Baseline Specification:

As we discussed using Equation 4, the normalized taxi-Uber routing difference within a matched pair is a function of route characteristics that may deferentially affect taxi and Uber driver detour incentives. Given this, we combine the features of the matched sample and our theory implications to establish identification of moral hazard.

When driving short trips within Manhattan, both taxi and Uber drivers should have little or no incentive to detour, because of the low marginal detour payoff due to the short trip length (Implication 1) and the high opportunity cost of detouring when finding another ride at drop-off is easy (Implication 4). It is plausibly in drivers' best interest to take as many within-Manhattan trips as possible to exploit the proportionately larger fixed component of the fare, in stead of making individual trips longer. In fact, taxi-Uber distance ratios for Manhattan trips are on average 0.98[13], with a standard deviation of 0.12 and a median of 0.99. To add more empirical support for this assumption, we show in Figure 4 that both taxi and Uber trip distances closely concentrate on a measure of long-run average optimal routing given by GURAFU, Uber's internal routing engine, for Manhattan trips. Therefore, we consider Manhattan trips as the "no detour" benchmark.

Figure 4: Taxi and Uber Driver Routing Weighted by Uber's GURAFU



For metered airport trips, both taxi and Uber drivers' detour benefit increases because long distance rewards detouring (Implication 1). However, the detour penalty cost is higher for Uber drivers than for taxi drivers because of Uber market designs, which may reduce Uber drivers' detour incentives. If this taxi-Uber difference in penalty cost is non-trivial, it likely leads to a testable difference in routing choices — taxi

[13]The average is slightly below 1, which is plausibly due to the difference in pick-up and drop-off. Specifically, taxi drivers mainly cruise on major avenues and streets, while Uber drivers more often pick up and drop off passengers at their doorsteps. We show in Figure A2 that even after matching, taxi pick-ups (purple) are more concentrated on major avenues and streets, whereas Uber pick-ups (green) are more from cross-town streets with slower traffic. A similar distribution applies to matched drop-offs as well.

drivers may route longer than Uber drivers, producing a positive taxi-Uber distance ratio in these cases, when compared to the Manhattan "no detour" benchmark.

On the other hand, for JFK routes with fixed taxi fare, taxi driver detour incentive becomes absent, because detour will only increase cost while not increasing earning. For Uber drivers, the detour decision depends on the relative strength of detour benefit and cost in these cases. If Uber drivers do not detour, the taxi-Uber distance ratio for JFK trips will be similar to the "no detour" benchmark; otherwise, the taxi-Uber distance ratio will be less than the "no detour" benchmark.

Let r denote a given matched taxi-Uber pair. We specify and estimate the following empirical model:

$$\frac{d_r^c}{d_r^u} = \alpha_0 + \alpha_1 M\_Airport_r + \alpha_2 JFK_r + \phi_{t(r)} + \epsilon_r. \tag{5}$$

where $\frac{d_r^c}{d_r^u}$ is the taxi-Uber distance ratio for a given matched pair, $M\_Airport_r$ is the dummy for metered airport trips, JFK_r is the dummy for JFK trips, $\phi_{t(r)}$ represents time fixed effects of route r, and ϵ_r is the random shock. We construct the empirical model in a way that the taxi-Uber distance ratio is benchmarked at the Manhattan trips, such that if the Uber market designs create a binding penalty cost, we expect the distance ratio to be greater than the benchmark for metered airport trips (i.e., $\alpha_1 > 0$) and similar to the benchmark for JFK trips (i.e., $\alpha_2 = 0$).

Driver Fixed Effects and More Controls:

One endogeneity concern of our current empirical model is the possible correlation between route types and the unobserved random shock, which may produce biased coefficient estimates. For example, if strategic drivers consistently select into certain route types, then the observed effects can be an artifact of adverse selection instead of moral hazard. Noting this, we now discuss the institutional features of taxis and Uber that largely alleviate this endogeneity threat.

On one hand, for taxi drivers, the matching of passengers of certain destinations is close to randomly assigned, because (1) passengers do not select taxis as taxi cabs are *ex ante* homogeneous to passengers; (2) taxi refusal of passengers is heavily penalized by the TLC refusal law.[14] However, taxi drivers can indeed form expectations of passenger destinations and route profitability and develop their own geographical search strategies (Haggag et al. (2017), Zhang et al. (2016)), leading to a correlation between route characteristics and driver types. In this case, controlling for taxi-driver fixed effects is a good way to tease out the

[14]Per the TLC refusal law, "It is against the law to refuse a person based on race, disability, or a destination in New York City. A taxicab driver is required to drive a passenger to any destination in the five boroughs." Riders are encouraged to make a refusal complaint by calling 311. According to Haggag et al. (2017), "In 2009 the refusal punishment was \$200-\$350 for a first offense, \$350-\$500 and a possible 30-day license suspension for a second, and a mandatory license revocation for a third offense. The TLC received about 2,000 formal complaints per year in 2009 and 2010." While TLC strictly enforces the refusal law, anecdotal evidence exists that taxi drivers sometimes refuse passengers in spite of penalties.

bias. In the absence of taxi driver IDs in 2016, we demonstrate in Section 5 that driver selection appears to be insignificant when the same estimation is run on 2013 data where taxi-driver fixed effects are controlled for.

On the other hand, several features of the Uber platform limit the scope of endogeneity: (1) To Uber drivers, passenger assignment by the platform is virtually random by construction. Uber's matching of drivers and passengers is mainly based on spatial proximity and dispatching efficiency, and it gives little weight to driver and rider characteristics in the matching. (2) Uber drivers have the option to cancel trip requests, but cancellation of rides is costly. Once assigned a rider, the driver cannot see the rider's destination on the application until picking up that rider, which makes it difficult for drivers to "cherry pick" passengers before accepting a trip request. Moreover, frequent and suspicious ride cancellation is penalized on Uber, often in the form of warning, "time out", or even deactivation. In addition, it is difficult for a driver to form expectations on the next rider's profitability, making cancellation of the current ride risky and rare. Taken together, these institutional details suggest that the correlation between route types and unobserved driver-route shocks is at best limited. Nonetheless, to further reduce the potential bias, we control for Uber driver fixed effects in some specifications.

Another layer of variation that we can leverage to further provide evidence of moral hazard is whether the passenger is from the local area or not, as the theory indicates that drivers may be more likely to detour when driving non-local passengers due to information asymmetry (Implication 2). To the extent that Uber market designs reduce information asymmetry, the strategic routing inefficiency for non-local passengers should be more pronounced for taxi drivers than for Uber drivers, in situations where detouring is profitable. An ideal case to test this is to control for both taxi passenger and Uber passenger "localness". Yet without visibility to taxi passengers, we can only proxy for the "localness" of taxi passengers using information of the Uber passenger of the matched trip. We caution that the scope of measurement error of "non-local" should be smaller on airport routes than on non-airport routes, because it is more likely that the taxi rider and the Uber rider are either both locals or both non-locals when they head to/from the airport from/to the same specific place at the same time (e.g., a hotel).

Leveraging the above-mentioned sources of variation, our enhanced regression equation is the following,

$$\begin{aligned}\frac{d_r^c}{d_r^u} &= \alpha_0 + \alpha_1 M\_Airport_r + \alpha_2 JFK_r + \alpha_3 NonLocal_r \\ &+ \alpha_4 M\_Airport_r \times NonLocal_r + \alpha_5 JFK_r \times NonLocal_r + X_r\Omega + \eta_{i(r)} + \phi_{t(r)} + \epsilon_r.\end{aligned} \tag{6}$$

where $\eta_{i(r)}$ is the Uber driver fixed effects. We use α_3, α_4, and α_5 to detect additional taxi-Uber routing

Table 2: Taxi-Uber Routing Difference

D.V. = Taxi dist. / Uber dist.	**(1) Baseline**	**(2) Non-local**	**(3) More controls**	**(4) Driver FE**
M_Airport	0.080***	0.070***	0.070***	0.069***
	(0.002)	(0.002)	(0.002)	(0.002)
JFK	0.009***	0.007***	0.008***	0.006*
	(0.002)	(0.002)	(0.002)	(0.003)
NonLocal		-0.005**	-0.005**	-0.002
		(0.002)	(0.002)	(0.002)
M_Airport × NonLocal		0.019***	0.019***	0.016***
		(0.003)	(0.003)	(0.003)
JFK × NonLocal		0.005	0.004	0.002
		(0.003)	(0.003)	(0.004)
Log (Uber_driver_total_trips)			0.000	0.004**
			(0.000)	(0.002)
Uber_driver_rating			0.037***	
			(0.006)	
Log (Uber_rider_total_trips)			0.000	0.000
			(0.000)	(0.000)
Hour of week FE	Yes	Yes	Yes	Yes
Uber driver FE	No	No	No	Yes
N	90,431	90,431	90,431	90,431
R^2	0.069	0.070	0.071	0.371

Notes. For all specifications, standard errors are cluster-robust at the hour-of-week level; *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

difference when passengers are non-local, for Manhattan trips, metered airport trips, and JFK trips, respectively. In addition, we control for a set of Uber driver and Uber rider characteristics which may further explain the variation in the Uber-taxi routing difference (included in X_r with the associated parameters in Ω): (1) Uber driver experience is measured by the driver's total trips driven prior to the current trip, and it is expected to positively correlate with Uber driver routing efficiency if there is a learning-by-doing effect; (2) Uber driver routing efficiency is expected to positively correlate with Uber driver rating, as routing efficiency is an important metric in overall driver quality; (3) Uber rider experience on the platform, measured by the total number of trips completed, may also be relevant as more experienced riders may make the trip more efficient by better communicating with the driver, choosing a more efficient pick-up or drop-off location, and so on.

3.4 Results Are Consistent with Moral Hazard

Table 2 reports regression results of Equation 5 and Equation 6. In the baseline specification (1), we find that the taxi-Uber distance ratio for a metered airport trip is on average 8% larger than for a Manhattan trip, and this effect is statistically significant. The taxi-Uber distance ratio is about 0.9% larger for JFK trips than for

Manhattan trips, which is at odds with our conjecture. However, this slightly positive effect becomes smaller in size and loses statistical significance as more controls and fixed effects are included in the regression.

When we add "non-local" and its interaction with metered airport trips and JFK airport trips, we find that taxi drivers route additionally longer when the passenger is non-local than when the passenger is local (1.9%) when driving metered airport trips (Table 2 Specification (2)). In fact, this added variation across passengers splits the main effect of 8% in Specification (1) into 7% for local passengers and 8.9% (=7%+1.9%) for non-local passengers of metered airport trips. However, this heterogeneity is absent for JFK trips and only weak for Manhattan trips, given that the negative effect of the stand-alone *NonLocal* even loses statistical significance in the fixed effects model. These findings suggest that taxi drivers appear to be more strategic when information asymmetry is more severe (the case with non-local passengers) in situations with large detour benefit (i.e., metered airport trips), and Uber drivers do not exhibit such strategic responses compared to taxis in situations when taxi detour incentive is "shut down" and Uber detour benefit is large (i.e., JFK trips). These findings are in line with agency theory and moral hazard.

In Specification (3), we find no material changes to the estimates when adding a set of Uber driver and rider characteristics in the regression. We observe that Uber driver rating is positively correlated with the taxi-Uber distance ratio, and this correlation is expected as routing efficiency should be reflected in driver ratings. The estimated effects hardly change when we control for Uber driver fixed effects, as shown in Specification (4)[15]. Interestingly, the effect of Uber driver total trips becomes strong, suggesting a routing improvement due to accumulating driving experience (Cook et al. (2018) document a similar learning-by-doing effect among Uber drivers).

Overall, the regression results exhibit high consistence with our main hypothesis that Uber reduces driver moral hazard incentive, which is reflected in the relative routing efficiency with respect to taxis in situations where detouring is profitable. However, the results can also be consistent with other competing explanations, which we explore next.

4 Navigation-related Alternative Explanations

4.1 GPS effect or moral hazard?

One threat to the moral hazard story is that the increased GPS usage among Uber drivers may have improved their routing behavior compared to taxis drivers. GPS can be most effective in situations where real-time traffic information is valuable. Although NYC taxi drivers are well-known for their driving experience and sophistication, routing efficiency can still hurt without such information. For example, taxi drivers may

[15]Note that because Uber driver rating is driver-specific, it disappears in Specification (4) when Uber driver fixed effects are controlled for.

choose a longer route with longer but more certain travel time, as opposed to a shorter route with volatile traffic, as a practice of mean-variance trade-off in the absence of GPS.

Recall that we estimate an 8% increase in taxi-Uber distance ratio for metered airport trips from that of Manhattan trips. Thus, if GPS navigation accounts for the 8% increase in taxi-Uber distance ratio for metered airport trips, it should be at least as salient for JFK trips. This is because JFK trips are significantly longer in distance than metered airport trips and similar to LaGuardia trips (which are the vast majority of metered airport trips in our sample), JFK trips also involve bridges or tunnels to cross the river, except that JFK is further southeast in Queens than LaGuardia. That is, the scope for GPS navigation, if any, should be as large for JFK trips as for metered airport trips. Given that no such effect is observed on JFK trips, we find the GPS effect not consistent with the data.

That said, a more subtle threat is that Uber drivers strategically detour on JFK trips and this detour effect "cancels out" with the GPS effect. To explore whether this is compatible with the data, we introduce to the main regression analysis a new variation — Uber surge pricing, which affects Uber driver detour incentives (Implication 3). If the above-stated assumption is true, then we expect more Uber detours on JFK trips when surge is on than when surge is off. If, however, we do not observe much Uber driver behavioral change on JFK trips when surge is on, then there was probably not much of Uber detour to begin with (i.e., the penalty cost is binding), because if Uber drivers would detour with base fare, they would detour more when the surge is, say, 1.5 of base fare.

Two features of Uber surge multipliers enable an empirically-sound analysis: First, there is a nice variation in both the incidence of surge and the level of surge — in the 2016 matched sample, 20% of matched routes have effective surge pricing; conditional on surge, the surge multiplier is on average 1.54, or 54% increase in earning from the base fare. Second, surge can be considered random to individual Uber drivers, because surge multipliers are extremely volatile and difficult to predict[16], and it is not in the driver's best interest to "chase" the surge at the cost of forgone earnings from trip requests missed or declined.[17] In addition, the Uber app no longer shows the surge hot spots after driver's accepting a ride, which prevents drivers from canceling the current non-surge ride in order to get a high-surge ride. We propose and estimate

[16]As shown in Lam and Liu (2017), Uber surge multipliers are volatile and difficult to predict even after accounting for highly granular location-time fixed effects.

[17]Uber drivers commonly agree on this view, based on our conversations with Uber drivers in New York City and Boston, Massachusetts.

the following equation:

$$\frac{d_r^c}{d_r^u} = \beta_0 + \beta_1 M\_Airport_r + \beta_2 JFK_r + \beta_3 Surge_r \\ + \beta_4 M\_Airport_r \times Surge_r + \beta_5 JFK_r \times Surge_r + X_r\Omega + \eta_{i(r)} + \phi_{t(r)} + \epsilon_r. \tag{7}$$

where β_3, β_4, and β_5 identify surge-induced Uber driver detour for Manhattan trips, metered airport trips, and JFK trips, respectively. We use surge dummy (i.e., surge incidence) instead of surge multipliers for easier interpretation of the results. The Uber driver and rider characteristics as well as the effects of non-local passengers are included in X_r. Note that if surge induces more Uber detour, the coefficient estimate of surge for all route types should be negative, because Uber distance is in the denominator of the dependent variable.

Table 3 reports the regression results. The surge effect for Manhattan trip is positive, weak, and close to zero, suggesting that there is no additional Uber detour due to surge within Manhattan. This adds support to our earlier assumption that driver routing in Manhattan is the efficiency benchmark — even when surge is high, the benefit of detour is out-weighted by the penalty cost plus the opportunity cost in light of high demand. The surge effect for metered airports and JFK airports are negative but small in magnitude, indicating a limited scope for surge to incentivize extra routing. As more controls and fixed effects are added, the surge effect on JFK trips is absent. This small surge effect on JFK trips does not warrant a sizable GPS effect, because they have to be equal in size to cancel each other out as we discussed earlier. Therefore, it appears that there is at best a very limited GPS effect.

4.2 Distance vs. Time: Do Longer Routes Save Passengers Time?

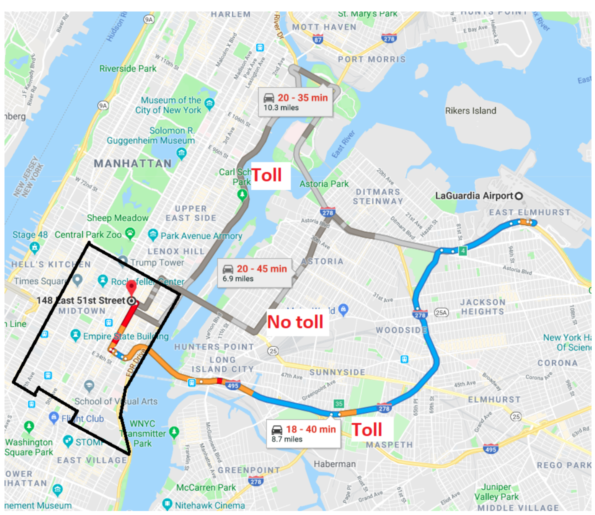
It is possible that taxi drivers in some cases possess superior routing information than the GPS. When this is the case, taxi drivers can save passengers time by taking a longer route. In this section, we show that the data are not fully compatible with this alternative explanation, but rather they lend more nuanced support to our main hypothesis of driver moral hazard. Specifically, we investigate taxi driver routing decisions via a case study and show that (1) among alternative airport routes, taxi drivers frequently choose the bridge that leads to the longest distance; (2) these long routes on average result in longer travel times when compared to shorter routes taken by drivers completing similar trips at the same time; and (3) this long-routing strategy is more seen in drivers with more route-specific experience.

Table 3: Taxi-Uber Routing Difference: The Role of Surge

D.V. = Taxi dist. / Uber dist.	**(1)**	**(2)**	**(3)**
M_Airport	0.083***	0.073***	0.072***
	(0.002)	(0.002)	(0.002)
JFK	0.010***	0.009***	0.007**
	(0.002)	(0.002)	(0.003)
Surge	0.003	0.003	0.004
	(0.002)	(0.002)	(0.002)
Surge × M_Airport	-0.016***	-0.016***	-0.018***
	(0.003)	(0.003)	(0.003)
Surge × JFK	-0.008*	-0.008*	-0.004
	(0.005)	(0.005)	(0.006)
NonLocal		-0.004**	-0.002
		(0.002)	(0.002)
M_Airport × NonLocal		0.019***	0.015***
		(0.003)	(0.003)
JFK × NonLocal		0.004	0.002
		(0.003)	(0.004)
Log (Uber_driver_total_trips)		0.000	0.004***
		(0.000)	(0.002)
Uber_driver_rating		0.037***	
		(0.006)	
Log (Uber_rider_total_trips)		0.000	0.000
		(0.000)	(0.000)
Hour of week FE	Yes	Yes	Yes
Uber driver FE	No	No	Yes
N	90,431	90,431	90,431
R^2	0.069	0.071	0.371

Notes. For all specifications, standard errors are cluster-robust at the hour-of-week level; *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Figure 5: Alternative Routes Between Midtown Manhattan and LaGuardia



The case study focuses on routes between LaGuardia Airport and an Midtown Manhattan area[18] (see Figure 5), for two reasons: First, this area has a high volume of taxi and Uber activities — approximately 37% of all LaGuardia taxi trips in 2016 either started or ended in this area; Second, the choice set of routes is relatively small and clear — the route in the middle via Queensboro Bridge is a toll-free route, usually the shortest and busiest among the three routes (we call this route S), whereas the other two routes have tolls of the same amount[19], with the top route via FDR Drive generally longer in distance (we call this route L) than the bottom route via the Midtown Tunnel (we call this route M). Taxi and Uber trip distance indeed exhibits a clear bimodal pattern for these two tolled routes (Figure A3), suggesting that the route choices are discrete. Therefore, using data on tolls as well as trip distance, we can identify, or at least proxy with good precision, which route drivers took.

A natural exercise here is to study how different routing choices affect trip duration by comparing taxi and Uber drivers on a matched route. If, say, a taxi driver who takes L finishes the trip later than an Uber driver who takes M, then that suggests inefficient routing and underlying moral hazard incentive. However, as we discussed before, taxi drivers generally driver faster than Uber drivers, in all cases, due to the taxi pricing rule that does not reward travel time. Thus, comparing taxi-Uber duration cannot cleanly infer moral hazard separately from the pricing effect. Therefore, we need to **match taxi drivers with taxi drivers** on the same route, and investigate whether taxis drivers save passengers time by choosing longer routes longer, where a taxi-taxi comparison can tease out the pricing effect on driving speed. We follow the same matching approach as we did for taxi and Uber trips. Since we focus on airport routes here, we relax the matching criteria to be just Step 1 (same street corner) and Step 4 (real time) with a time window of 30 minutes so that more trips can be matched without hurting precision much.

[18]This area consists of three NYC Neighborhood Tabulation Areas (NTAs): Midtown-Midtown South, Turtle Bay-East Midtown, and Murray Hill-Kips Bay.

[19]In 2016, the cash rate of the tolls was $8 for taxi-like vehicles. Taxi and Uber drivers normally pay the discounted rate of $5.54 by using an E-Z Pass.

Table 4: The Ratio of Focal Taxi Diver Time and Matched Taxi Driver Time, across Route Choice Combinations

		Focal taxi		
		S	M	L
Matched taxi	S	1.00 (0.0003)	0.90 (0.001)	0.94 (0.001)
	M		1.01 (0.0004)	1.05 (0.0006)
	L			1.00 (0.002)

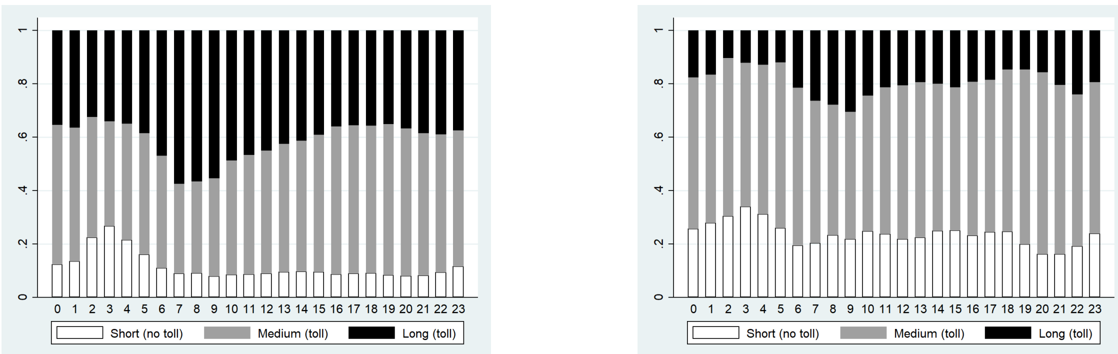
Notes. This table shows how focal taxi drivers' travel time compares with that of the matched taxi drivers across situations when they take the same or different routes. The number in each cell is the mean of focal-matched duration ratios for the corresponding route comparison group, and the associated number in parentheses is the standard error.

Table 4 shows how focal taxi drivers' travel time compares with that of the matched taxi drivers across situations when they take the same or different routes. For clean illustration, the cases SM (focal is S and matched is M), SL, and ML are not reported because they are symmetric to MS, LS, and LM, respectively. The number in each cell is the mean of focal-matched duration ratios for the corresponding route comparison group, and the associated number in parentheses is the standard error. The duration ratios in the diagonal cells are equal to or close to 1, meaning that when the focal taxi driver and the matched taxi driver take the same route (SS, MM, and LL), they spend about the same amount of time driving. When the focal driver takes M while the matched driver takes S (i.e., MS), the focal driver indeed performs better in travel time. In the case of LS, the focal driver still performs better in travel time, yet this advantage decreases from 10% in the case of MS to 6% (because the duration ratio increases from 0.90 to 0.94). In the case of LM, the focal driver on average needs 5% more time to complete the trip. Therefore, it appears that M is often the time-efficient route, which is in line with our casual checks with Google Maps.

In Figure 6, we use taxi and Uber raw trip records to plot the shares of alternative Midtown-LaGuardia routes taken by taxi and Uber drivers by hour. Clearly, taxi drivers are overall more likely to take L than Uber drivers and less likely to take M and S. Moreover, taxi drivers' greater tendency to take L is present for all hours of the day. Therefore, it appears that taxi drivers on average cost passengers more time, instead of saving passengers time, by choosing longer routes.

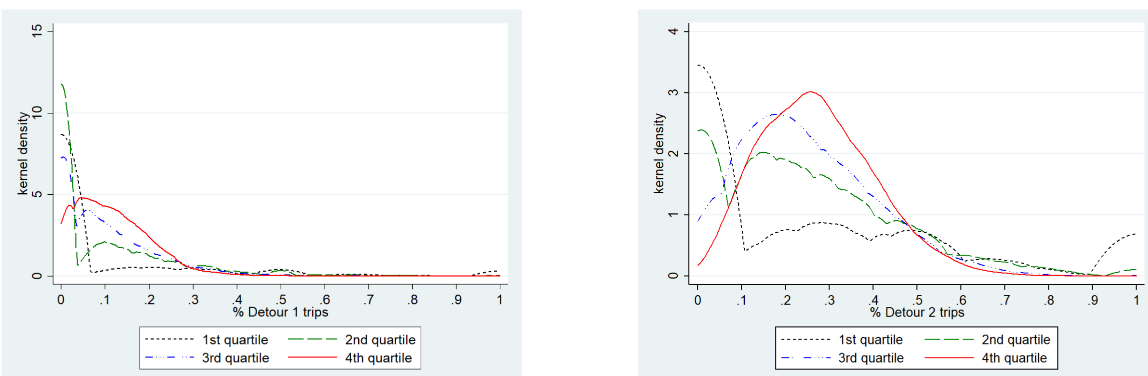
Finally, we provide evidence that the same-route experience is correlated with more detour. Using 2013 data with taxi driver ID, we split drivers into quartiles by their total number of trips between Midtown and LaGuardia. Figure 7a shows that drivers with more trips completed on this particular route (Midtown-LaGuardia) tend to detour more, as measured by *Detour 1* — the incidence when the focal taxi driver takes a longer route than the matched taxi driver (i.e., LM and MS), *and* the focal taxi driver arrives later than the matched driver. Figure 7b shows a more pronounced pattern with the definition *Detour 2*, where the focal driver both logs more distance and more time. Therefore, this evidence further provides support for moral hazard, as it shows that inefficient routing generally comes from experienced drivers who knowingly take the long route rather than from less-experienced drivers.

Figure 6: Shares of Alternative Midtown-LaGuardia Routes by Hour, Taxis vs. Uber



(a) Taxis (b) Uber

Figure 7: Drivers with More Same-route Experience Tend to Detour More



(a) Detour 1 (b) Detour 2

5 Moral Hazard vs. Adverse Selection

5.1 Selection on the Intensive Margin

As discussed in the identification, the unobserved driver types in the error term may correlate with route types. It is revealing that we see no significant changes in the coefficient estimates when Uber-driver fixed effects are controlled for in the main analysis. However, we need to explore whether the effects remain when taxi-driver fixed effects are accounted for.

Table 5: Taxi-Uber Routing Difference, 2013

D.V. = Taxi dist. / Uber dist.	**(1)**	**(2)**	**(3)**
M_Airport	0.124***	0.118***	0.114***
	(0.002)	(0.005)	(0.007)
JFK	0.077***	0.078***	0.068***
	(0.004)	(0.008)	(0.013)
NonLocal		-0.012***	-0.013***
		(0.003)	(0.004)
M_Airport × NonLocal		0.024***	0.025***
		(0.006)	(0.008)
JFK × NonLocal		0.017**	0.022
		(0.008)	(0.015)
Log (Uber_driver_total_trips)		-0.003*	-0.002
		(0.002)	(0.001)
Log (Uber_rider_total_trips)		0.000	0.001
		(0.001)	(0.001)
Hour of week FE	No	Yes	Yes
Taxi driver FE	No	No	Yes
N	23,774	23,774	23,774
R^2	0.123	0.258	0.591

Notes. The regression samples consist of matched pairs of taxi and Uber trips using 2013 data. For all specifications, standard errors are cluster-robust at the hour-of-week level; *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

We repeat the baseline regression analysis on the 2013 data, which contain taxi driver IDs. In the sample construction, we relax the matching criteria to only Step 1 (same street intersection) and Step 4 (real time) with a time window of 30 minutes, because following the original matching procedure would lead to a sample size too small for identification, due to the small market share of Uber in 2013. The final sample consists of 23,774 matches with 11,972 unique taxi drivers, 6,527 of which have more than one trip in the sample. Also because of the relaxed matching criteria, the taxi-Uber distance ratio for the Manhattan "no-detour" benchmark is on average 0.93, since the taxi distance has a downward measurement bias due to the taxi-Uber difference in pick-ups and drop-offs (refer to Figure A2). Thus, regressions that follow Equation 5 and Equation 6 should lead to an upward bias in coefficient estimates of *M_Airport* and *JFK*.

Table 5 reports the estimation results using the 2013 matched sample. We see that variables of interest remain strong and large, instead of being absorbed by fixed effects. Although the estimates of *M_Airport* and *JFK* are greater than their counterparts in the main analysis using 2016 matched sample, as we expected, their relative sizes are consistent with the moral hazard interpretation. In addition, the effects of *NonLocal* and its interactions with *M_Airport* and *JFK* are also broadly consistent with the estimates in the main analysis. Therefore, we find little evidence that driver selection into profitable routes is the major explanation of our results.

Table 6: Taxi-Uber Routing Difference: The Role of Switchers

	Switchers vs. 2013 Uber drivers		**Switchers vs. 2016 taxi drivers**	
	D.V. = Switcher dist. / Uber dist.		D.V. = Taxi dist. / Switcher dist.	
	(1)	**(2)**	**(3)**	**(4)**
M_Airport	0.120***	0.105***	0.054***	0.054***
	(0.005)	(0.019)	(0.002)	(0.004)
JFK	0.079***	0.078***	0.005	0.010*
	(0.011)	(0.029)	(0.004)	(0.006)
NonLocal		-0.016		0.001
		(0.011)		(0.004)
M_Airport $\times$ NonLocal		0.025		0.003
		(0.022)		(0.005)
JFK $\times$ NonLocal		0.012		-0.005
		(0.035)		(0.008)
Log (Uber_driver_total_trips)		-0.001		0.003
		(0.004)		(0.003)
Log (Uber_rider_total_trips)		0.000		-0.001
		(0.003)		(0.001)
Hour of week FE	Yes	Yes	Yes	Yes
Switcher FE	No	Yes	No	Yes
N	4,030	4,030	16,363	16,363
R^2	0.112	0.616	0.036	0.239

Notes. For all specifications, standard errors are cluster-robust at the hour-of-week level; *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

5.2 Selection on the Extensive Margin

Drivers of different types may select differently into being taxi and Uber drivers. If this was the case, then the observed moral hazard would be an artifact of the driver type distributions of taxis and Uber. Not being able to directly observe driver types, we cannot completely rule out this possibility. However, we shed light on the extent of behavioral change of former taxi drivers who switched to Uber between 2013 and 2016. Recall that we observe drivers' TLC driver IDs, for taxi drivers in the 2013 data and all Uber drivers. TLC driver IDs are issued by the TLC for all types of taxi-like services in NYC under the same system. Therefore, we are able to track the status (taxi or Uber) of a driver across time. Specifically, for a given taxi driver in the 2013 taxi data, if we observe the same TLC driver ID in the 2016 Uber data, then we identify the driver as one who switched from being a taxi driver to being an Uber driver at some point between 2013 and 2016. We refer to these drivers as switchers. Since we have to resort to regression analysis to test moral hazard, we only focus on 1,999 switchers who appear both in the 2013 matched sample and the 2016 matched sample.

We first show that the routing behavior of these switchers, who were taxi drivers in 2013, exhibits detour patterns when compared to that of Uber drivers in 2013. Table 6 (1) and (2) report the regression results using the 2013 matched sample of these 1,999 switchers and Uber drivers. Since this sample is a subset of the 2013 matched sample, it also suffers from the downward measurement error for Manhattan trips due to the relaxed matching criteria. As a result, we observe seemingly large point estimates of *M_Airport* and *JFK*. However, the rank order of distance ratios for metered airport trips, JFK trips, and Manhattan trips is preserved and significant, even with a small-sized sample that accounts for driver fixed effects.

In Table 6 (3) and (4) we perform the same regression analyses on the matched sample consisting of taxi drivers in 2016 and switchers, who were Uber drivers in 2016. The effect of metered airport dummy is strong and significant, suggesting that taxi drivers in 2016 route longer than switchers mainly on metered airport routes. Or, former taxi drivers appear to route more efficiently than current taxi drivers on metered airport routes. When the airport is JFK where taxi driver moral hazard incentive is "shut down", the taxi-Uber distance ratio is positive and only marginally different from the Manhattan "no detour" benchmark, providing little evidence that these switchers continued to detour in 2016. Therefore, evidence in Table 6 is consistent with behavioral updating of switchers, who used to detour as taxi drivers but abandoned the detour strategy after they joined Uber and adapted to the Uber environment.

We caution the readers that the behavior of the switchers may just reflect a common trend, for example, drivers may have become more honest over the years due to some unobserved factors. One good way to rule this out is by tracking taxi drivers who remained taxi drivers from 2013 to 2016; however, the exercise is not feasible due to missing taxi driver IDs in 2016. Nonetheless, we show in Figure A4 that 234 long-standing Uber drivers do not appear to have become more honest from 2013 to 2016, as suggested by their routing compared to taxi routing in the same time periods on JFK taxi flat-fare routes, where taxis are the "no detour" benchmark.

6 Other Robustness Checks

There are cases when drivers can make passengers better off by choosing longer routes that minimize tolls (see Figure A5 for an example). However, a necessary condition for this to be true is that taxi drivers take toll-free routes more often than Uber drivers. Figure A6 shows precisely the opposite: Across major neighborhoods of Manhattan, taxi drivers are *more* likely to take toll roads than Uber drivers do on their way to or from LaGuardia. Therefore, based on the data, we reject toll saving as the main explanation for taxi drivers' detouring.

In the main analysis of driver detour, we constrained the matched taxi and Uber trips to be 15 minutes apart. As a robustness check, we perform the same analysis using alternative time windows, namely 5, 10, 20, and 30 minutes. As shown in Table A1, the estimated effects are stable and consistent across time window lengths. To the extent that trips within a narrower time range are more likely to be subject to the same real-time traffic, and thus better approximate the experimental ideal, significant effects of similar size even when using a time window as short as 5 minutes greatly enhance our identification.

Recall that our main sample contains only taxi trips reported by Vendor 2, because Vendor 1's meter system appears to round down trip distance to the nearest first decimal place. Taxi trips may appear to be shorter because of the rounding, and the downward bias is greater in taxi-Uber distance ratios of shorter routes. This implies that the same regression analysis on Vendor 1 sample should yield an upward bias in the coefficient estimate of airport trips, instead of the opposite. In Table A2, we separately estimate the main regressions using Vendor 1 only, and then both Vendor 1 and Vendor 2. We compare these with the main regression results using only Vendor 2, which leads to an upward bias. We find this upward bias, instead of a downward one, consistent with our main findings.

7 Discussion

7.1 Mechanisms

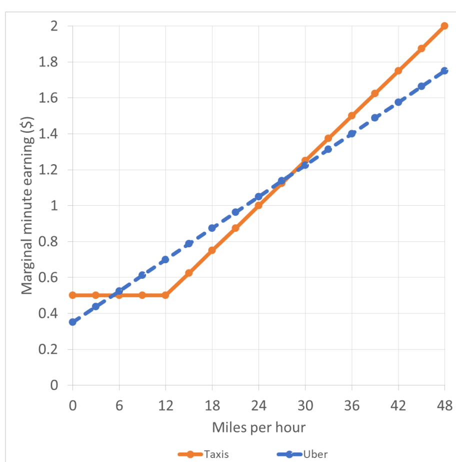
In this section, we discuss two mechanisms that account for the observed routing difference between taxi and Uber drivers. The first mechanism is the set of technology-enabled incentive devices implemented by the Uber platform but not by taxis. These incentive devices include tech-aided monitoring and verification, tech-enabled rider rating of drivers, and tech-aided conflict resolution. Each of these essentially makes the cost function of moral hazard "steeper" for Uber than for taxis. Our empirical results imply that these incentive devices (monitoring, rating, conflict resolution) enhance market transparency in most cases, as Uber drivers do not appear to detour on JFK airport routes, where the potential gain from detouring is large. That is, when the pricing rule rewards detouring, little or no evidence of detour points to a working (or even binding) incentive system.

One necessary condition for a working rating system to penalize strategic behavior is the negative correlation between passengers' trip ratings to the drivers and driver routing inefficiency. This negative correlation is robustly present in the data, as shown in Table A3. One interesting nuance is that conditional on the Uber ride is shorter than the matched taxi ride, the shorter the Uber trip is, the less likely the Uber driver will get a high rating. One likely reason for this is that passengers dislike off-GPS routing and thus give low ratings even when Uber drivers found a shorter route. It seems plausible that passengers cannot easily assess whether a driver's deviations from the prescribed GPS route are due to superior information used to shorten the route or an effort to extract a higher fare with a longer route. Therefore, tech-enhanced monitoring deters driver opportunistic behavior, yet on the other hand it may also create constraint for driver discretion.

Pricing is another important mechanism that predicts driver routing behavior. This is most clearly reflected in taxi driver routing efficiency on flat-fare JFK routes, where taxi drivers optimize routes as residual claimants. When taxi fares are metered as a two-part tariff, taxi drivers tend to detour on airport routes mainly because the variable part of the fare can justify the detour. In addition, taxi drivers' detour payoff is greater than that of Uber drivers, as the taxi pricing formula puts a larger weight on distance than Uber pricing does.

Furthermore, perhaps the pricing scheme is of first-order importance in explaining taxi driver travel speed. We document in the main analysis that taxi drivers, although detouring in some cases, in general drive at a faster speed than Uber drivers. This is expected because speeding is rewarded more for taxi drivers than for Uber drivers. In Figure 8, we compute driver marginal minute earnings at various traffic speeds, separately for taxis and Uber, following their pricing formulas. All the computations are net of the fixed component of the fare (i.e., \$2.5 for taxis and \$2.55 for Uber). An interesting divergence appears. For example, taxi-driver per-minute earning increases by 4 times from 12 miles per hour to 48 miles per hour (from \$0.5/min to \$2/min), while Uber-driver per-minute earning only increases by 2.5 times (from \$0.7/min to \$1.75/min). This difference stems from the weight given to trip distance in the pricing formula, where taxi distance is marginally more rewarding than that of Uber. One artifact of the taxi pricing schedule is that when traffic is flowing at about 12 miles per hour, a NYC taxi driver would earn a slightly higher fare by alternating between stopping and driving at 24 miles per hour.

Figure 8: Driver Marginal Minute Earning Across Travel Speeds



7.2 Mind vs. Machine

The data show that taxi drivers are more efficient at routing than Uber drivers on short, non-airport routes. For short routes within Manhattan Core, taxi trip distance is on average 98% that of Uber, and the difference is statistically significant at the 1% level. We have shown in A2 that this is largely due to taxi-Uber difference in exact pick-up and drop-off spots. Nonetheless, this also suggests that human navigation can often perform at least as well as the technology in dense markets.

One way that drivers can outperform GPS is by having more up-to-date information on the road networks and conditions, for example, temporary road closures, upcoming sporting events, or undocumented shortcuts. Another possibility is that experienced taxi drivers can suggest a better drop-off point than the exact address given by the passenger, based on the driver's extensive experience. For example, the driver might suggest dropping off the passenger on the opposite side of the street in order to avoid unnecessary travel. This is confirmed by an interview with Loai Yousef [20], an NYC Uber and Lyft driver, who stated, "Sometimes the Uber GPS map has mistakes. Sometimes it makes the driver do a U-turn to arrive at the exact address even though it would be easy for rider to just cross the street. Taxis drivers often drop passengers off a short distance from exact address."

NYC taxi driver expertise in routing should not be surprising, because as residual claimants, they are strongly motivated to learn the routes, optimize their routing, and take initiative when they can. In contrast, Uber drivers might not be as motivated to use their discretion, even in cases when they do possess better information than the GPS. The reason is that off-GPS routing might come across as suspicious behavior to the riders, which can result in bad ratings and complaints (as we witnessed in Table A3). Loai Yousef told us that "Uber passengers tend to want driver to go to the exact address even if it's wasteful." Therefore, the use of GPS, coupled with the monitoring and rating systems, can limit the incentives for human knowledge accumulation, as well as initiative and discretion.

8 Conclusion

In this paper, we study whether digital platforms affect moral hazard and service quality, when compared to traditional settings. We provide evidence from the taxi and Uber setting in the form of driver routing choices from identical start and end points. By analyzing trip-level data from NYC, we find that taxi drivers tend to detour more relative to Uber drivers on metered airport routes, particularly when the airport passenger is non-local. This long routing is not found for short, within-Manhattan trips or airport trips with a fixed fare. These findings are consistent with a model of driver moral hazard, where the Uber technology platform and pricing scheme reduce driver moral hazard behavior in situations where taxi moral hazard return is high. We have also explored alternative explanations but found none of them compatible with the data. That said, we also find evidence that the incentives for creating and using driver routing expertise may be reduced by the

[20]Loai Yousef was interviewed on July 9th, 2018.

Uber platform, relative to reliance on technologies such as GPS.

Digital platforms can make markets significantly more efficient by reducing information asymmetry, which has long been a key barrier to market efficiency. In the case of Uber, this is done in several ways, including the rating system, the easy complaint channel, and the highly salient GPS that enable both driver and passenger to see the same route. We identify sizable efficiency gains due to reduced agency problems, because detouring leads to welfare loss in the form of lost passenger time, which is estimated at 150 passenger hours per day[21]. In general, once the smart phone infrastructure is in place, these features can be rolled out at very low marginal cost.

There is growing body of research on the digital disruption and, in particular, the potential for digital platforms to mitigate moral hazard. The rise of Uber is a case example of the power of digital platforms and suggests that information asymmetry can be significantly mitigated by this type of technological advance. Our study provides models that can be applied to other settings facing the emerging challenges and opportunities created by the interaction of new technologies and incentive design.

References

Angrist, J. D., S. Caldwell, and J. V. Hall (2017). Uber vs. taxi: A drivers eye view. Technical report, National Bureau of Economic Research.

Aral, S., E. Brynjolfsson, and L. Wu (2012). Three-way complementarities: Performance pay, human resource analytics, and information technology. *Management Science 58*(5), 913–931.

Athey, Susan, J. C. C. and D. Knoepfle (2018). Service quality in the gig economy.

Bakos, J. Y. (1997). Reducing buyer search costs: Implications for electronic marketplaces. *Management science 43*(12), 1676–1692.

Balafoutas, L., A. Beck, R. Kerschbamer, and M. Sutter (2013). What drives taxi drivers? a field experiment on fraud in a market for credence goods. *Review of Economic Studies 80*(3), 876–891.

Balafoutas, L., R. Kerschbamer, and M. Sutter (2017). Second-degree moral hazard in a real-world credence goods market. *The Economic Journal 127*(599), 1–18.

Bresnahan, T. F., E. Brynjolfsson, and L. M. Hitt (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The Quarterly Journal of Economics 117*(1), 339–376.

Brown, J. R. and A. Goolsbee (2002). Does the internet make markets more competitive? evidence from the life insurance industry. *Journal of political economy 110*(3), 481–507.

Brynjolfsson, E., Y. Hu, and M. D. Smith (2003). Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science 49*(11), 1580–1596.

Brynjolfsson, E. and M. D. Smith (2000). Frictionless commerce? a comparison of internet and conventional retailers. *Management science 46*(4), 563–585.

Buchholz, N. (2015). Spatial equilibrium, search frictions and efficient regulation in the taxi industry. Technical report, University of Texas at Austin.

[21] Using the within-taxi matched sample, we estimate a 1.6% increase in travel time for the 8% increase in trip distance by taxi drivers on metered airport trips. For an average NYC metered airport trip that takes 32.7 minutes, the time waste is about 0.5 minutes. Given that there are 18,000 metered airport trips per day in NYC, a back-of-envelope calculation leads to an efficiency loss of 9,000 passenger minutes, or 150 passenger hours per day.

Castillo, J. C., D. Knoepfle, and G. Weyl (2017). Surge pricing solves the wild goose chase. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pp. 241–242. ACM.

Chen, M. K., J. A. Chevalier, P. E. Rossi, and E. Oehlsen (2017). The value of flexible work: Evidence from uber drivers. Technical report, National Bureau of Economic Research.

Chen, M. K. and M. Sheldon (2016). Dynamic pricing in a labor market: Surge pricing and flexible work on the uber platform. In *EC*, pp. 455.

Cohen, P., R. Hahn, J. Hall, S. Levitt, and R. Metcalfe (2016). Using big data to estimate consumer surplus: The case of uber. Technical report, National Bureau of Economic Research.

Cook, C., R. Diamond, J. Hall, J. A. List, P. Oyer, et al. (2018). The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. *Upubliceret paper. Tilgængelig på: https://web. stanford. edu/˜ diamondr/UberPayGap. pdf. Besøgt*, 26–04.

Cramer, J. and A. B. Krueger (2016). Disruptive change in the taxi business: The case of uber. *American Economic Review 106*(5), 177–82.

Duflo, E., R. Hanna, and S. P. Ryan (2012). Incentives work: Getting teachers to come to school. *American Economic Review 102*(4), 1241–78.

Filippas, A., J. J. Horton, and J. Golden (2018). Reputation inflation. In *Proceedings of the 2018 ACM Conference on Economics and Computation*. ACM.

Forman, C., A. Ghose, and A. Goldfarb (2009). Competition between local and electronic markets: How the benefit of buying online depends on where you live. *Management science 55*(1), 47–57.

Frechette, G. R., A. Lizzeri, and T. Salz (2016). Frictions in a competitive, regulated market evidence from taxis.

Gans, J. S., A. Goldfarb, and M. Lederman (2017). Exit, tweets and loyalty. Technical report, National Bureau of Economic Research.

Goldfarb, A. and C. Tucker (2017). Digital economics. Technical report, National Bureau of Economic Research.

Greenwood, B. N. and S. Wattal (2017). Show me the way to go home: An empirical investigation of ride-sharing and alcohol related motor vehicle fatalitie. *MIS Quarterly 41*(1).

Haggag, K., B. McManus, and G. Paci (2017). Learning by driving: Productivity improvements by new york city taxi drivers. *American Economic Journal: Applied Economics 9*(1), 70–95.

Hall, J., C. Kendrick, and C. Nosko (2015). The effects of ubers surge pricing: A case study. *The University of Chicago Booth School of Business*.

Hall, J. V., J. J. Horton, and D. T. Knoepfle (2017). Labor market equilibration: Evidence from uber. Technical report, Working Paper, 1–42.

Hall, J. V. and A. B. Krueger (2015). An analysis of the labor market for ubers driver-partners in the united states. *ILR Review*, 0019793917717222.

Hubbard, T. N. (2000). The demand for monitoring technologies: the case of trucking. *The Quarterly Journal of Economics 115*(2), 533–560.

Hui, X., M. Saeedi, Z. Shen, and N. Sundaresan (2016). Reputation and regulations: evidence from ebay. *Management Science 62*(12), 3604–3616.

Klein, T. J., C. Lambertz, and K. O. Stahl (2016). Market transparency, adverse selection, and moral hazard. *Journal of Political Economy 124*(6), 1677–1713.

Lam, C. T. and M. Liu (2017). Demand and consumer surplus in the on-demand economy: the case of ride sharing.

Lee, M. K., D. Kusbit, E. Metsky, and L. Dabbish (2015). Working with machines: The impact of algorithmic and data-driven management on human workers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 1603–1612. ACM.

Liang, C., Y. Hong, and B. Gu (2016). Effects of it-enabled monitoring on labor contracting in online platforms: Evidence from a natural experiment.

Liu, T., E. Vergara-Cobos, and Y. Zhou (2017). Pricing schemes and seller fraud: Evidence from new york city taxi rides.

Nagin, D. S., J. B. Rebitzer, S. Sanders, and L. J. Taylor (2002). Monitoring, motivation, and management: The determinants of opportunistic behavior in a field experiment. *American Economic Review 92*(4), 850–873.

Overby, E. and C. Forman (2014). The effect of electronic commerce on geographic purchasing patterns and price dispersion. *Management Science 61*(2), 431–453.

Pierce, L., D. C. Snow, and A. McAfee (2015). Cleaning house: The impact of information technology monitoring on employee theft and productivity. *Management Science 61*(10), 2299–2319.

Rajgopal, S. and R. White (2015). Cheating when in the hole: The case of new york city taxis.

Reimers, I., B. R. Shiller, et al. (2018). Proprietary data, competition, and consumer effort: An application to telematics in auto insurance. Technical report.

Staats, B. R., H. Dai, D. Hofmann, and K. L. Milkman (2016). Motivating process compliance through individual electronic monitoring: An empirical examination of hand hygiene in healthcare. *Management Science 63*(5), 1563–1585.

Sudhir, K. and D. Talukdar (2015). The peter pan syndrome in emerging markets: The productivity-transparency trade-off in it adoption. *Marketing Science 34*(4), 500–521.

Tabarrok, A. and T. Cowen (2015). The end of asymmetric information. *Cato Unbound*.

Zhang, Y., B. Li, and K. Ramayya (2016). Learning individual behavior using sensor data: The case of gps traces and taxi drivers.

Appendix

A Theory

Recall that the driver chooses the amount of detour (x) and the speed of driving (equivalent to y) to maximize the following expected payoff function:

$$\begin{aligned}\underset{x,y}{\text{Max}}\ \text{E}\ \{&s[p_0 + p_d d_0(a + x + \epsilon) + p_t(\gamma d_0(a + x + \epsilon) + y)] \\ &- f(x; d_0, \Theta) - g(y; d_0, \Theta) - q_e e_t(\gamma d_0 x + y)\},\end{aligned} \tag{8}$$

The first-order conditions are,

$$f_x(x; d_0, \Theta) + q_e e_t \gamma d_0 - s d_0(p_d + p_t\gamma) = 0, \tag{9}$$

$$g_y(y; d_0, \Theta) + q_e e_t - s p_t = 0. \tag{10}$$

Then, the comparative statics lead to the following testable implications:

Hypothesis 1: *Drivers tend to detour more on longer routes than on shorter routes because longer distance increases detour payoffs unless the demand at the drop-off location is sufficiently high, marginal detour penalty increases significantly with trip distance, or both.*

Proof. $\frac{\partial x^*}{\partial d_0} = \frac{s(p_d+p_t\gamma)-q_e\gamma e_t - f_{xd_0}}{f_{xx}} \lesseqgtr 0$, depending on the sign of $s(p_d + p_t\gamma) - q_e\gamma e_t - f_{xd_0}$.

(1) Clearly, when $f_{xd_0} >> 0$, or marginal detour penalty increases significantly with trip distance, $\frac{\partial x^*}{\partial d_0} < 0$, or drivers detour less on longer routes than shorter routes.

(2) When $f_{xd_0} = 0$, $\frac{\partial x^*}{\partial d_0} > 0$ if and only if $q_e < \frac{s(p_d+p_t\gamma)}{\gamma e_t}$. In the case of taxis in normal NYC traffic, $s = 1$, $p_0 = 2.50$, $p_d = 2.50$, and $p_t = 0$. For simplicity, let $e_t = \frac{p_0+p_d D_e+p_t T_e}{T_e}$, where D_e and T_e are the expected length and duration of the forgone trip, respectively. Then, $\gamma e_t = \gamma \times \frac{2.50+2.50D_e}{\gamma D_e} = \frac{2.50+2.50D_e}{D_e}$, by having $T_e = \gamma D_e$. Therefore, $\frac{\partial x^*}{\partial d_0} > 0$ requires q_e to be less than $\frac{D_e}{1+D_e}$. For example, for an average NYC taxi trip of 3 miles, the threshold for q_e is 0.75. Therefore, drivers detour more on longer routes than shorter routes unless the drop-off demand q_e is sufficiently high. Similar conditions apply to Uber. ■

Hypothesis 2: *Drivers detour more when the rider is a non-local passenger, and they detour less when the rider is a local passenger, as non-local passengers are less likely to notice the detour because they lack knowledge of local geography.*

Proof. For a parameter $\theta \in \Theta$ that increases the marginal detour penalty ($f_{x\theta} > 0$), e.g., the passenger is local, $\frac{\partial x^*}{\partial \theta} = -\frac{f_{x\theta}}{f_{xx}} < 0$. ■

Hypothesis 3: *Drivers detour more (respectively, less) during high surge prices if the increase in marginal detour payoff due to high surge dominates (respectively, is dominated by) the increase in marginal detour penalty due to high surge.*

Proof. $\frac{\partial x^*}{\partial s} = \frac{d_0(p_d+p_t\gamma)-f_{xs}}{f_{xx}} \lesseqgtr 0$, depending on the sign of $d_0(p_d + p_t\gamma) - f_{xs}$. ■

Hypothesis 4: *Drivers detour less (respectively, more) when the demand at the drop-off location is higher (respectively, lower).*

Proof. $\frac{\partial x*}{\partial q_e} = -\frac{e_t\gamma d_0}{f_{xx}} < 0$. ■

Hypothesis 5: *Everything else held constant, taxi drivers have greater incentives than Uber drivers to drive faster than other traffic on the road.*

Proof. Let c denote taxis and u denote Uber. According to Equation 10, $g_y(y*) = sp_t - q_e e_t$. Given $g_y < 0$ for $y < 0$, $g_y > 0$ for $y > 0$, and $p_t = 0$ for taxis, it then follows that $y^{c*} < 0$. The condition $p_t > 0$ for Uber makes speeding less profitable for Uber drivers than for taxi drivers, especially when the penalty cost of deviating from the road traffic is greater for Uber than for taxis, i.e., $g_{uy} < g_{cy}$ for $y < 0$ and $g_{uy} > g_{cy}$

for $y > 0$. Therefore, as long as other parameters (q_e and e_t) are similar between taxis and Uber, $y^{c*} < y^{u*}$.
■

B Figures

Figure A1: Dividing NYC into Voronoi Cells Centered at Street Intersections

Figure A2: Pick-up Locations of Uber and Taxi after Matching Step 1 (taxis in purple; Uber in green)

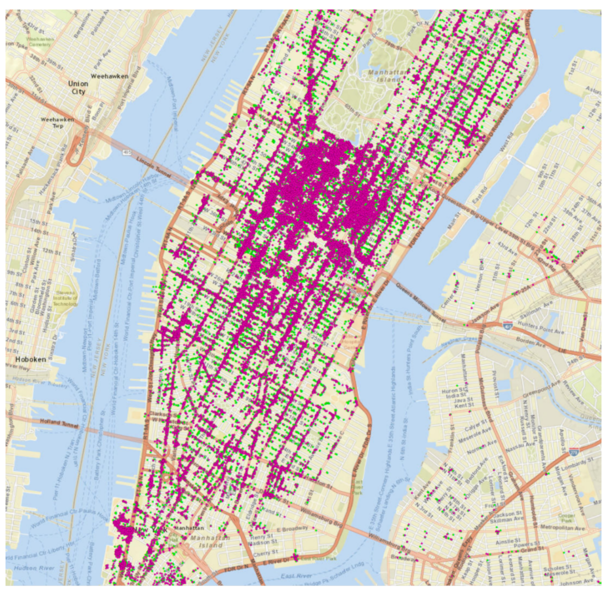
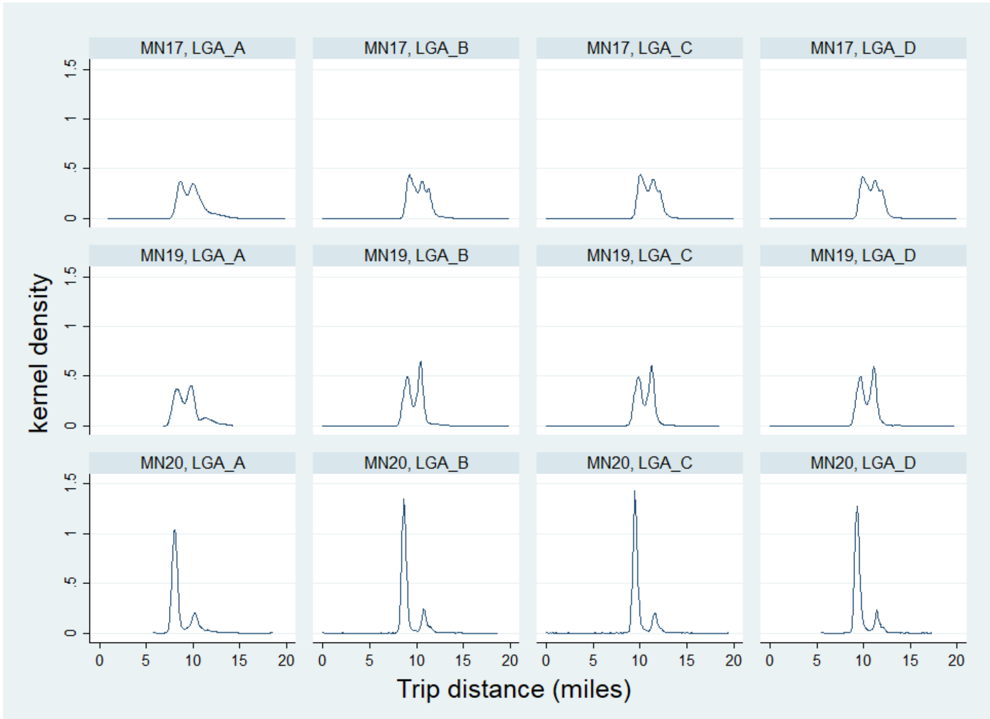
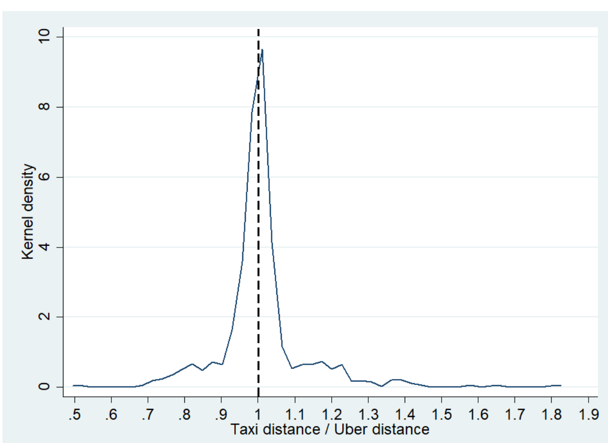


Figure A3: Taxi Trip Distance Exhibits a Bimodal Pattern for Tolled Airport Routes

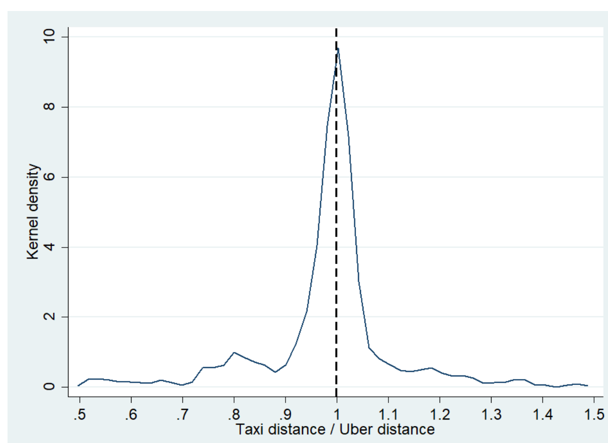


Notes. The plots are based on NYC yellow medallion taxi trips, January to June, 2016. Individual density plots demonstrate bimodal patterns of taxi distance on tolled trips for all routes between three NTAs in Midtown and four LaGuardia terminals. NTA code "MN17" denotes Midtown-Midtown South, "MN19" denotes Turtle Bay-East Midtown, and "MN20" denotes Murray Hill-Kips Bay. LaGuardia terminals are denoted by A, B, C, and D. Similar bimodal patterns are found for UberX trips.

Figure A4: Long-standing Uber Drivers Did Not Become More Honest with Time



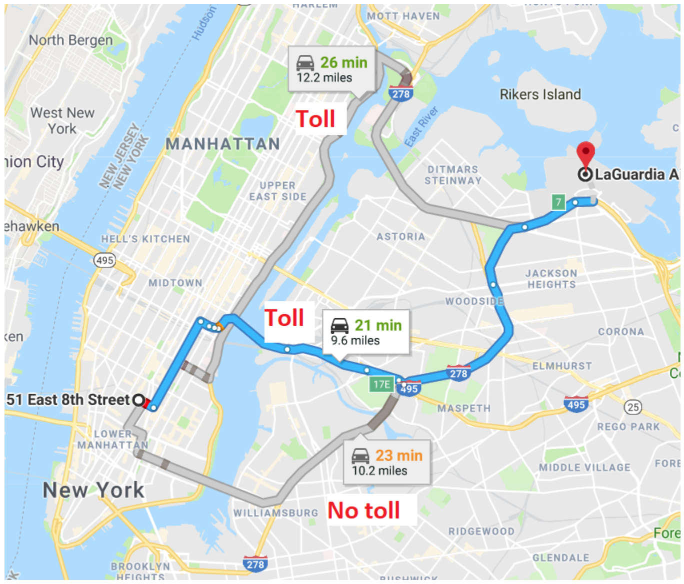
(a) 2013



(b) 2016

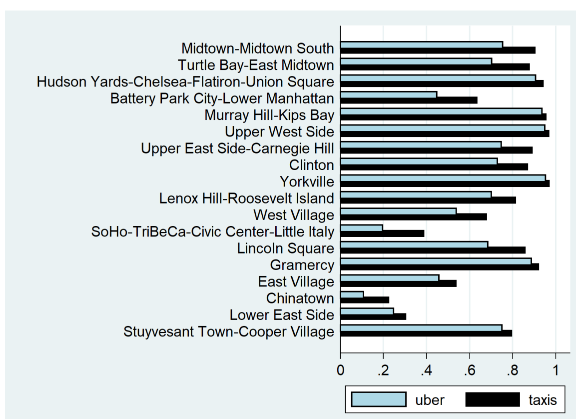
Notes. These plots show that 234 drivers who remained Uber drivers from 2013 to 2016 do not appear to have become more honest in routing, as compared to taxi driver routing in the respective time periods, on JFK taxi flat-fare routes (where taxis are considered the "no detour" benchmark).

Figure A5: An Example Where the Longer, Toll-free Route Can Benefit Passengers



Notes. This is an example where the longer, toll-free route can benefit passengers. The toll-free route of 10.2 miles costs \$1.80 more than the shortest and fastest route of 9.6 miles (0.6 miles $\times$ \$2.50/mile $\times$ 1.20, assuming a tip of 20%), according to the taxi pricing formula. Yet this route saves the passenger a toll of \$5.54. Thus, the longer, toll-free route is preferable by the passenger as long as the cost saving justifies the extra travel time (i.e., if passengers value their time less than \$1.87 per minute ((\$5.54 - \$1.80)/2 minutes).

Figure A6: Taxi and Uber Shares of Tolled Trips between LaGuardia and Manhattan Neighborhoods



Notes. The figure plots the shares of toll roads, separately for taxis and Uber, between major Manhattan neighborhoods and LaGuardia, using 2016 data. These Manhattan neighborhoods are ordered by taxi trip shares.

C Tables

Table A1: Robustness: Various Time Windows

D.V. = Taxi dist. / Uber dist.	**5 min.**	**10 min.**	**15 min.**	**20 min.**	**30 min.**
M_Airport	0.069***	0.067***	0.069***	0.069***	0.069***
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)
JFK	-0.003	0.003	0.006*	0.008***	0.009***
	(0.008)	(0.005)	(0.003)	(0.003)	(0.002)
NonLocal	-0.003	-0.006**	-0.002	-0.003	-0.005***
	(0.005)	(0.003)	(0.002)	(0.002)	(0.002)
M_Airport × NonLocal	0.016***	0.020***	0.016***	0.016***	0.017***
	(0.006)	(0.004)	(0.003)	(0.003)	(0.002)
JFK × NonLocal	0.011	0.008	0.002	0.003	0.002
	(0.009)	(0.005)	(0.004)	(0.003)	(0.003)
Log(Uber_driver_total_trips)	0.002	0.004**	0.004**	0.004***	0.004***
	(0.004)	(0.002)	(0.002)	(0.001)	(0.001)
Log(Uber_rider_total_trips)	-0.000	0.000	0.000	-0.000	-0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Hour of week FE	Yes	Yes	Yes	Yes	Yes
Uber driver FE	Yes	Yes	Yes	Yes	Yes
N	33,830	62,704	90,431	117,220	169,179
R^2	0.551	0.432	0.371	0.332	0.286

Notes. The regression samples consist of matched pairs of taxi and Uber trips on metered airport routes and non-airport routes, for various time windows. For all specifications, standard errors are cluster-robust at the hour-of-week level; *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table A2: Robustness: Taxi Meter Vendor 1

D.V. = Taxi dist. / Uber dist.	**Vendor 2 (Main analysis sample)**	**Vendor 1**	**Vendor 1 and 2**
M_Airport	0.069***	0.091***	0.078***
	(0.002)	(0.003)	(0.002)
JFK	0.006*	0.035***	0.020***
	(0.003)	(0.004)	(0.002)
NonLocal	-0.002	0.002	-0.001
	(0.002)	(0.003)	(0.002)
M_Airport $\times$ NonLocal	0.016***	0.008**	0.013***
	(0.003)	(0.003)	(0.002)
JFK $\times$ NonLocal	0.002	-0.004	-0.001
	(0.004)	(0.005)	(0.003)
Log(Uber_driver_total_trips)	0.004**	0.006***	0.004***
	(0.002)	(0.002)	(0.001)
Log(Uber_rider_total_trips)	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Hour of week FE	Yes	Yes	Yes
Uber driver FE	Yes	Yes	Yes
N	90,431	74,199	164,630
R^2	0.371	0.412	0.298

Notes. The regression samples consist of matched pairs of taxi and Uber trips on metered airport routes and non-airport routes, for Vendor 1 taxis, Vendor 2 taxis, and both taxi vendors, respectively. For all specifications, standard errors are cluster-robust at the hour-of-week level; *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table A3: Uber Driver Ratings by Passengers Are Correlated with Routing Efficiency

	JFK trips with taxi flat fare			**Uber dist. $\geq$ Taxi dist.**	**Uber dist. $<$ Taxi dist.**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Uber_dist/Taxi_dist	-0.152**	-0.150**	-0.132	-0.283**	0.518**
	(0.062)	(0.061)	(0.082)	(0.115)	(0.218)
Uber_dur/Taxi_dur	-0.253***	-0.204***	-0.206***	-0.269***	-0.123
		(0.046)	(0.053)	(0.077)	(0.082)
Uber_driver_rating		0.862***	0.880***	0.904***	0.765***
		(0.096)	(0.104)	(0.149)	(0.154)
Log (Uber_driver_total_trips)		0.013*	0.015*	0.021	0.011
		(0.008)	(0.009)	(0.013)	(0.012)
Surge_multiplier		-0.111**	-0.100*	-0.121	-0.034
		(0.045)	(0.053)	(0.083)	(0.079)
NonLocal		0.019	0.021	0.039	0.007
		(0.017)	(0.023)	(0.031)	(0.030)
Log (Uber_rider_total_trips)		0.028***	0.027***	0.035***	0.018**
		(0.005)	(0.006)	(0.009)	(0.008)
Hour of week FE	No	No	Yes	Yes	Yes
N	5,885	5,885	5,885	3,139	2,746
R^2	0.007	0.026	0.049	0.080	0.078

Notes. For Specifications (3), (4), and (5), standard errors are cluster-robust at the hour-of-week level; *** significant at the 1% level; ** significant at the 5% level. * significant at the 10% level.