Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

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

Meng Liu , Erik Brynjolfsson , Jason Dowlatabadi

MANAGEMENT SCIENCE
Vol. 67, No. 8, August 2021, pp. 4665–4685
ISSN 0025-1909 (print), ISSN 1526-5501 (online)
http://pubsonline.informs.org/journal/mnsc

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

Meng Liu,[a] Erik Brynjolfsson,[b,c] Jason Dowlatabadi[d]

[a] Washington University in St. Louis, St. Louis, Missouri 63130; [b] Stanford University, Stanford, California 94305; [c] National Bureau of Economic Research, Cambridge, Massachusetts 02138; [d] Uber Technologies, San Francisco, California 94103

Contact: mengl@wustl.edu, https://orcid.org/0000-0002-5512-7952 (ML); erik@brynjolfsson.com, https://orcid.org/0000-0001-5285-4436 (EB); jasond@uber.com (JD)

Received: April 12, 2019
Revised: November 1, 2019; March 12, 2020
Accepted: May 17, 2020
Published Online in Articles in Advance: February 26, 2021

https://doi.org/10.1287/mnsc.2020.3721

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 nonlocal 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 flat fare; and (3) long routing in general leads to longer travel time, instead of saving passengers time. These findings are consistent with digital platform designs reducing driver moral hazard, but not with competing explanations such as driver selection or differences in driver navigation technologies. We also find evidence of Uber drivers' long routing on airport trips in times of surge pricing, suggesting that the tech-enabled market designs may not be binding in our setting.

History: Accepted by Chris Forman, information systems.
Supplemental Material: The e-companion is available at https://doi.org/10.1287/mnsc.2020.3721.

Keywords: moral hazard • information asymmetry • Uber • digital platforms • market design

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. Although markets in general tend to suffer from information asymmetry, many digital platforms appear to have design features than can enhance market transparency. This is often in the form of new technologies and incentive systems, such as ratings of buyers and sellers, real-time monitoring, and low-cost complaint channels. For example, using our data, we find that 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 6,300 trips.

One of the biggest barriers to market efficiency is moral hazard, as a result of asymmetric information. Do digital platforms reduce moral hazard and improve service quality, compared with traditional settings? In this paper, we study this question by comparing Uber with traditional taxis. Our findings will be of broad interest to economists and policy makers 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, Liu et al. 2019). 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. 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, such that the matched drivers are subject to the same real-time optimal routes, even though 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 highly efficient when compared with 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 NYC 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 moral hazard, where the driver decides on detour and travel speed to maximize payoff. The key tension is a trade-off between the cost and benefit of detour—detour increases driver earning while also increasing the penalty cost (in terms of expected monetary and reputation cost of cheating), as well as the opportunity cost in terms of expected forgone earnings (detouring usually prolongs travel time, reducing opportunities for additional trips). Drivers may lack detour incentives when driving short trips in dense markets (e.g., Manhattan trips), because of low return due to short distance and high opportunity cost due to high subsequent demand. 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 *nonlocal 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 with taxi drivers. However, the GPS effect[1] should be at least as salient for JFK trips, because 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 route longer but save passengers time. We focus on a popular route between Midtown Manhattan and LaGuardia airport where we can infer 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 with shorter routes taken by taxi drivers completing essentially the same trips at the same time, and this long-routing strategy is more seen in taxi drivers with more route-specific experience. Therefore, this hypothesis of distance-time trade-off is not broadly supported by the data.

Lastly, we investigate driver selection, which may also explain the data. First, driver types may select differently across routes. We rule this out by controlling for driver fixed effects and finding no material changes in our results. Second, taxi and Uber may represent different distributions of driver types. We 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 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, digital payment, and passenger monitoring. Also, it is important to re-evaluate the current pricing scheme that rewards driver 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.

Digital platforms are not perfect. We find that Uber drivers tend to take a longer route in times of higher surge pricing, compared with their matched taxi drivers and particularly on airport trips. This suggests that although dynamic pricing is often praised to be an effective device at balancing supply and demand (Cohen et al. 2016, Lam et al. 2017), the scope it creates for driver strategic behavior may have been neglected. In addition, we document a puzzling empirical pattern that when Uber drivers take short routes, passengers tend to give low ratings likely because off-GPS routing may come across as suspicious behavior. As a result, tech-enhanced monitoring may create constraint for driver initiative and discretion. These issues and pitfalls that emerge on digital platforms should be taken seriously by platform designers.

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 one location 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 closely relates to the literature 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 (Bresnahan et al. 2002, Aral et al. 2012), 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 that 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. (2019) 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.

Although 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 with 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.

This study also contributes to the literature on digital disruption and online-offline competition (Bakos 1997, Brynjolfsson and Smith 2000, Brown and Goolsbee 2002, Brynjolfsson et al. 2003, Forman et al. 2009, Overby and Forman 2014, among many others. See Goldfarb and Tucker 2017 for a review). In the context of taxis and ride-hailing platforms, studies have shown that emerging ride-hailing platforms improve market efficiency through real-time technologies and dynamic pricing (Hall et al. 2015, Castillo et al. 2017), which leads to greater capacity utilization (Cramer and Krueger 2016, Hall et al. 2019), greater service quality (Athey et al. 2019), and consumer surplus (Cohen et al. 2016, Lam et al. 2017). Besides, welfare impacts are also found on drivers due to the flexible work arrangements (Hall and Krueger 2018, Chen et al. 2017) and on the general public for reasons such as reduced drunk driving (Greenwood and Wattal 2017) and traffic congestion (Erhardt et al. 2019). 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 from the agency perspective.

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. (2019) have identified nonlocal passengers from local passengers based on the destinations of trips originating at NYC's airports and have found that taxi drivers defraud nonlocals 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 build on and contribute to this literature by examining how moral hazard can be mitigated by tech-aided market design innovations.

2. NYC Taxis vs. Uber: Market Design and Pricing

The market design for the Uber platform differs significantly from that of taxis. First, GPS navigation is integrated into the Uber app and extensively used by Uber drivers, whereas taxi drivers mainly navigate without GPS. When an Uber 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). As such, Uber passengers can readily monitor drivers' routing in real time, by either checking their own smartphone app, or looking at the driver's app, because the driver's phone is usually mounted in a way that it is visible to passengers. Therefore, passenger monitoring is improved on Uber.

Second, 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 concentrated with a mean of 4.74 out of 5 (see online appendix Figure A1a). Drivers with low ratings are constantly warned by Uber. Uber starts to consider deactivating a driver when the driver rating falls below a threshold.[2] Drivers appear to be very concerned about their ratings,[3] and perhaps as a result, the actual deactivation risk is relatively low (about 3%).

Thanks to electronic trip records, verification and complaints can be made with little friction on Uber. 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 through the app, and Uber handles the conflict resolution by evaluating the trip records. Among various reasons of fare adjustments on Uber, the number one reason is "inefficient route" (see online appendix Figure A1b). 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. Based on our calculation, in 2016, taxi complaints were 1 in every 6,300 trips, whereas Uber fare adjustments were 1 in every 170 trips.

Pricing also differs between taxis and Uber. NYC taxi fares are set by the TLC.[4] 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).[5,6] A flat rate of \$52 applies to routes between Manhattan and JFK Airport. Some taxi drivers are medallion-owners who essentially run the business as an entrepreneur; others lease the medallions on a daily, weekly, or monthly basis, and they collect all revenues minus gasoline and some vehicle maintenance costs. In both cases, drivers are residual claimants who are incentivized to maximize earnings.

Unlike taxi pricing, 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 $2 \times (\$2.55 + \$0.35 \times 10 + \$1.75 \times 2) = \$19.10$. Almost all Uber routes in NYC are metered according to the same pricing formula, except for extremely short-length trips that are subject to a minimum fare.[7] 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 choose to rent a vehicle from fleet owners due to TLC requirements such as commercial insurance.

3. Illustration of Detour Incentives

In this section, we explain the essential trade-offs drivers face when choosing among alternative routes and deciding on whether and how much to detour the passengers. A more detailed model is presented in the online appendix. For both taxi and Uber drivers, detour benefit is a direct outcome of the pricing formula—the more the driver drives, the more she gets paid for the trip. However, the driver has to balance the benefit with the cost of detour, which consists of two parts: the penalty cost and the opportunity cost.

3.1. Detour Penalty Cost and Monitoring

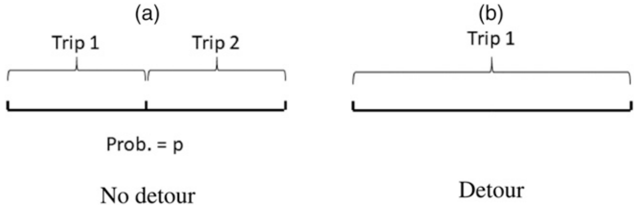
The penalty cost is the penalty charged multiplied by the probability that the cheating behavior is detected. When detour is detected by passengers, taxi drivers can lose tips, or even be reported by passengers to TLC, which may lead to monetary punishment (Haggag et al. 2017). For Uber drivers, the penalty may be money back to passengers and/or low ratings. Therefore, factors that enhance passengers' ability to monitor and detect detour will increase the penalty cost to drivers. To the extent that Uber market designs enhance passenger monitoring, it is possible that the same detour results in a greater penalty cost for Uber drivers than for taxi drivers, even when the actual penalty conditional on being caught is higher for taxi drivers.[8]

3.2. Detour Opportunity Cost and Drop-off Demand

We use a simple example to illustrate the opportunity cost of detour (our modeling of opportunity cost essentially follows Liu et al. 2019). For simplicity, suppose the driver can complete two trips per unit of time, and each trip takes one unit of distance and half unit of time to finish. Assume that each trip gives the driver a payment of 5, equal to the fixed component of 2.5 plus the per-unit-distance rate of 2.5 times one distance unit. Lastly, assume for now that the driver picks up the second passenger with certainty, immediately after the first trip. If the driver completes both trips as shown in Figure 1(a), the driver gets a payment of 5+5×1=10. If, however, the driver chooses to detour by doubling the distance of Trip 1, the driver gets a payment of 2.5+2.5×2 = 7.5, as shown in Figure 1(b). Then the opportunity cost of detour in this case is 10, which is the forgone earnings from the no-detour alternative.

The opportunity cost is a decreasing function of the probability of getting Trip 2 at the drop-off time and location of Trip 1 (when there is no detour on Trip 1). Let this probability be denoted as p and we present four cases with varying values of p in Table 1. The benchmark example above is the case when $p = 1$. When $p = 0.8$, the driver's payment without detour decreases from 10 to 9, because Trip 2 is a less likely event. When $p = 0.5$, the detour payment is equal to the opportunity cost. Lastly, when $p = 0.2$, the detour payment is greater than the opportunity cost by 1.5,

Figure 1. A Graphical Example of Detour Opportunity Cost



thus creating the strongest detour incentive among all four cases. Therefore, the more likely the driver is able to pick up subsequent fares, the less profitable detour is.

3.3. Trip Length May Affect Detour Benefit and Opportunity Cost

As shown above, p is the sole factor that determines whether detour is profitable. Thus, for situations where p is high, drivers have no detour incentive. However, when p is low enough, there is incentive to detour on all trips—short and long—but the detour incentive is greater for long-length trips. We use a simple example to illustrate this intuition by making the following assumptions: (1) trips are homogeneous, 3-mile long, and takes 10 minutes to complete; (2) drivers only *very rarely* get long-length trips of 9 miles that they can finish in 30 minutes;[9] (3) at any time and any location, the probability of getting a 3-mile trip is 0.7 and the probability of getting a 9-mile trip is negligible. For simplicity, the penalty cost of detouring is assumed to be 0 for both short- and long-length trips.[10]

How does a 100% detour (i.e., doubling the trip length) affect the hourly earning when the focal trip is 3 miles, compared with when the focal trip is 9 miles? For easy illustration, we assume that drivers do not expect any more 9-mile trips in the future, although the calculation can easily adapt to a nonzero probability of 9-mile trips. The hourly earning when the driver detours by 100% on a 3-mile trip is 45.5, which is equal to the earning from the initial 6-mile trip (i.e., 2.5+2.5×6), plus the expected earning of four subsequent regular 3-mile trips at the probability of 0.7 (i.e., 0.7×4×(2.5+2.5×3)). However, the opportunity cost is 45, equal to the earning of the first 3-mile trip, plus 0.7 × earning of five subsequent 3-mile trips. Therefore, the net benefit is 0.5. Detouring by 100% on a 9-mile trip leads to a benefit of 47.5, equal to the base fare of 2.5 plus 18 miles × 2.5. If the driver does not detour, she can make an earning of 46 from the first 9-mile trip and then with a probability of 0.7, three subsequent regular trips (i.e., 2.5+2.5×9+0.7×3×(2.5+2.5×3)). Then the net benefit is 1.5, greater than the net benefit of detouring when the focal trip is 3 miles. Therefore, detour can be more profitable when the focal trip is a long-length trip *and* when the drop-off demand is low or mild, essentially because there is more "room" to detour in these cases.

Taken together, whether drivers detour crucially depends on several key factors—the penalty cost and passenger's ability to monitor, the demand at the drop off location and time, as well as trip length. These observations will motivate our empirical set-up to test moral hazard, which will be made more clear after we discuss the data and the matching.

Table 1. A Numerical Example of Detour Opportunity Cost

p	Detour payment	No-detour payment (opportunity cost)	Net benefit
1	7.5	5+5×1 = 10	–2.5
0.8	7.5	5+5×0.8 = 9	–1.5
0.5	7.5	5+5×0.5 = 7.5	0
0.2	7.5	5+5×0.2 = 6	1.5

4. Empirical Model and Results

4.1. Data and Matching

Our data combine the universe of NYC yellow medallion taxi trip records and the universe of Uber's proprietary UberX trip records for January to June 2016, and July to December 2013. The taxi trip records are published by the TLC.[11] The UberX trip records are proprietary data of Uber. Throughout the paper, we use the 2016 data for the main analysis for the large sample size, given that Uber was significantly larger in 2016 than in 2013. As we explain later, we also leverage the 2013 data because anonymized taxi driver ID's are available in the 2013 data but not in the 2016 data, where driver ID enables several key analyses of this paper, such as controlling for taxi driver fixed effects.

Taxi trip records contain detailed information such as pick-up and drop-off time and longitude/latitude, trip distance and duration, and various fares and fees. UberX trip records contain similar information as taxi trips, plus extra information such as the surge multiplier at the time of trip request, anonymized driver ID, driver total number of trips on Uber prior to a given trip, passenger total number of trips on Uber prior to a given trip, driver lifetime rating, and driver and passenger rating for a given 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.

Our discussions in Section 3 imply that the "treatment effect" of Uber may be reflected in the difference between taxi and Uber driver routing decisions, when they complete identical trips. Therefore, 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., minutes apart on the same day). 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[12] (see online appendix Figure A2) 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.

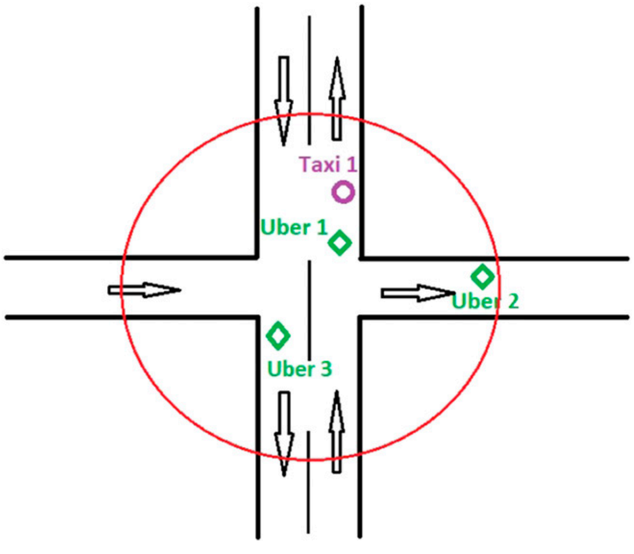
Step 2 (same street): We then restrict matched pick-ups to be on the same street,[13] because pick-ups on different streets can be subject to different optimal routes even when they are going to the same destination. In Figure 2, this means that Taxi 1 will be matched with Uber 1 and Uber 3, but not 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.[14] 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.

Figure 2. (Color online) Pickup-Dropoff-Time Matching of Taxi and Uber Trips



Using the 2016 taxi and Uber data, we obtain 173,770 pairs of matched trips. Seventy percent of the matched pairs are airport trips, which is not surprising because the strict matching criteria make nonairport trips more difficult to match than airport trips. The vast majority (95%) of matched nonairport 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 nonairport trips (e.g., trips between Brooklyn and Queens) from the sample, which leads to a 1.6% reduction of sample size (173,770–170,942). 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 of the range (0.5–1.5), in order to prevent extreme cases from affecting our results. This leads to a 3.7% drop in sample size (170,942–164,631). 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,[15] which we provide evidence in the robustness section. Given this, we drop all matched pairs that involve Vendor 1 taxi trips (with a 45.1% drop in sample size, i.e., 164,631–90,431).

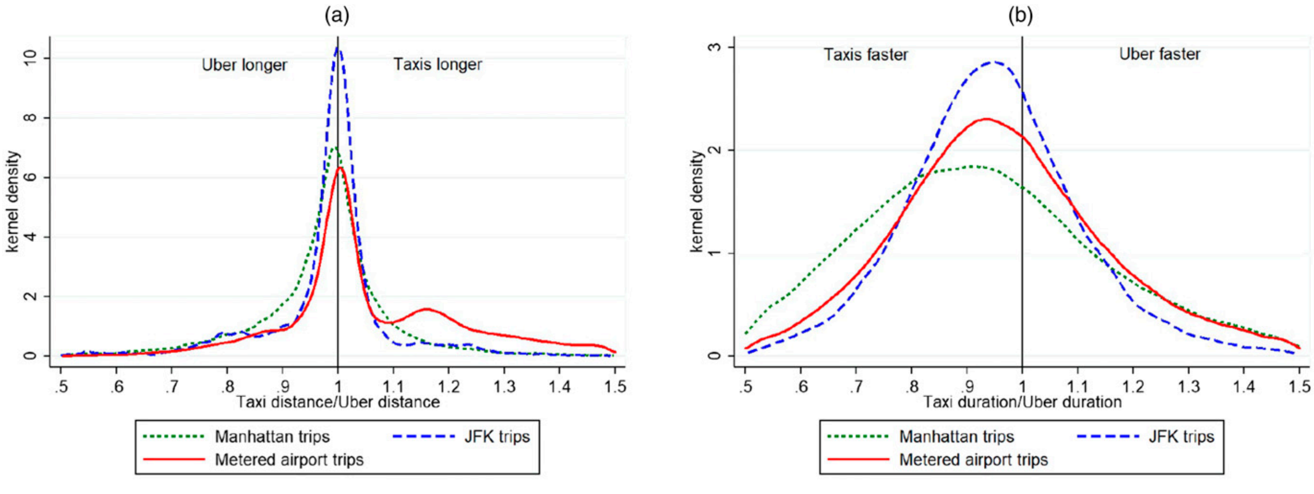
After above-mentioned sample restrictions, our sample contains 90,431 matched pairs, with 23,484 unique Uber drivers (about 54% of all UberX drivers in the same period). Table 2 summarizes the sample, where the unit of observation is a matched taxi-Uber pair. Nonlocal passenger takes the value of 1 if the billing zip code of a given Uber passenger is outside of NYC, or the passenger's city of most Uber trips is not NYC in the case of missing billing zip code. 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 overrepresented by airport trips, especially metered airport trips, compared with the population of taxi trips and Uber trips, due to the matching. Three route types emerge: (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 3(a) shows how matched taxi and Uber trips compare in distance. For Manhattan trips, taxi-Uber distance ratios are on average 0.98, with a standard deviation of 0.12 and a median of 0.99. The distance ratio is below 1 with statistical significance, which is plausibly due to the difference in the way taxi and Uber drivers pick up and drop off passengers. We show in online appendix Figure A3 that after Matching Step 1, taxi pick-ups (purple) are more concentrated on major

Table 2. Summary Statistics (Unit: A Matched Taxi-Uber Pair)

Variable	Mean	Standard Deviation	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 flat fare; Uber metered fare)	0.10	0.29	0	0	0
Nonlocal 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	2,491.12	2,010.00	358	2,021	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				

Figure 3. (Color online) Distance and Duration Ratios of Matched Taxi and Uber Trips



Distance ratios for matched trips

Duration ratios for matched trips

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. Even after Matching Step 2 and 3, it is still common that taxi drivers pick up and drop off passengers at convenient locations such as street corners, instead of picking up and dropping off passengers at their doorsteps as Uber drivers usually need to do. Therefore, these differences in practice may have contributed to the slight favorable bias toward taxi distance in Manhattan.

When focusing on metered airport trips, we find that taxi trips are significantly longer in distance than matched Uber trips, where the distance ratio has a mean of 1.06 and a standard deviation of 0.16, and it is different from 1 with statistical significance. In addition, the distribution exhibits a second mode around 1.15, as well as a fatter right tail. As we will show in later sections, this second mode is primarily caused by taxi drivers' tendency to take a different bridge/tunnel than the optimal one, which more often than not leads to unnecessary travel. However, on JFK trips with flat taxi fares, the distance ratio appears to have a similar distribution as that of Manhattan trips. Specifically, the distance ratio of JFK trips has a mean of 0.99 and a standard deviation of 0.11, and it is different from 1 with statistical significance.

Figure 3(b) shows that across route types, taxis tend to travel at a greater speed and arrive faster than Uber. This is consistent with the difference in pricing formulas—taxi drivers have a large incentive to drive fast because their time is not directly paid for, but rather the more distance they cover in a time unit, the more they get paid.[16] However, it is also possible that taxi drivers choose longer-distance routes that can save passenger time. We investigate this alternative explanation in a later analysis.

Although these differences in driver routing behavior can map to a variety of explanations such as driver ability and navigation technologies, our primary goal of this paper is to tease out the factors that affect detour incentive from others, by leveraging these three route types. We discuss the empirical strategy and identification in the following section.

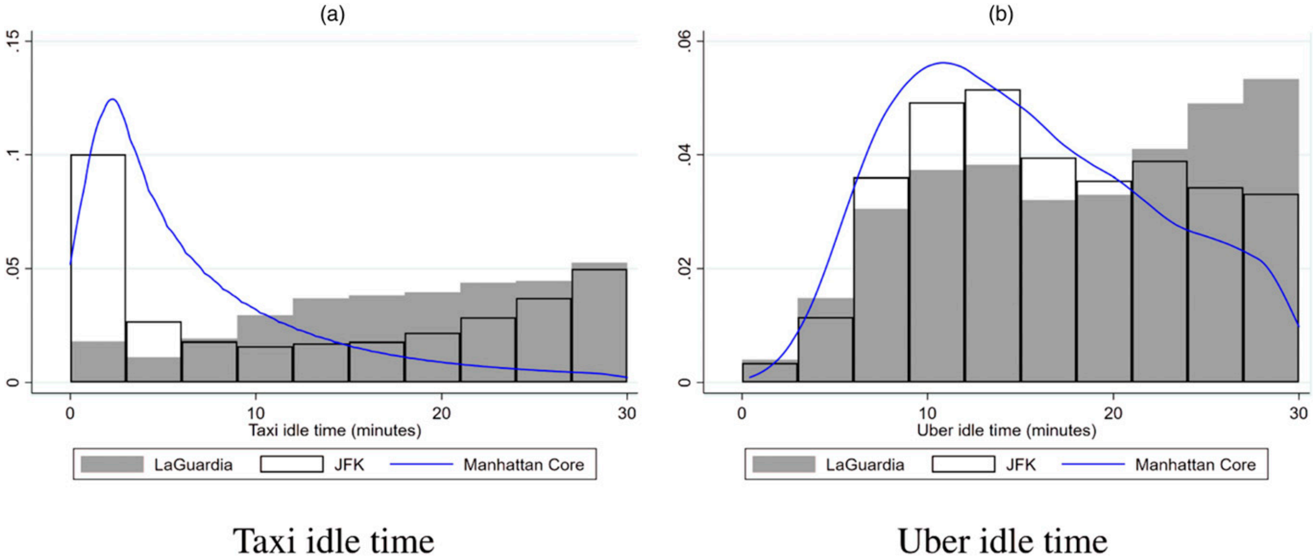
4.2. Empirical Strategy

4.2.1. Empirical Strategy and Baseline Specification.

Our "treatment" of interest is the Uber market design that increases the penalty cost of driver detour. In a situation where both the matched taxi and Uber drivers have high detour benefit, this treatment effect should be reflected in the difference in routing behavior. In this section, we leverage the rich empirical setting that consists of the three major route types to establish identification.

For Manhattan trips, both taxi and Uber drivers should have little or no incentive to detour, because of the short trip length and the high opportunity cost of detouring when finding another ride at drop-off is easy. In Figure 4, we show the distributions of driver idle minutes (i.e., the minutes between trips) at a given location, separately for LaGuardia, JFK, and Manhattan drop-offs. Note that we use 2013 taxi data to construct the idle time because we only observe anonymized taxi driver IDs in the 2013 data. For comparability, we use 2013 Uber data to construct Uber driver idle time. It is clear that for taxi drivers, Manhattan drop-offs mean a significantly lower average idle time, compared with airport drop-offs. This is also true for Uber drivers in 2013 (pairwise t statistics indicate statistical significance at the 1% level), although the difference is not as pronounced as for taxi drivers.[17] In addition, we show high routing

Figure 4. (Color online) Taxi and Uber Idle Time by Destination



Notes. Computing idle times requires following the trip history of the same driver. Because of driver ID availability, we use July 2013 taxi trip records to compute and plot taxi driver idle times in Figure (a). Figure (b) is plotted using July 2013 Uber trip records. In all plots, we drop idle times that are longer than 30 minutes due to the possibility that such long idle times can be breaks or lunch times. The LaGuardia and JFK distributions are plotted in histograms with a bin size of three minutes, whereas the Manhattan distribution is plotted in kernel density with a one-minute half width.

efficiency of taxi and Uber drivers in Manhattan by comparing their routing with Uber's internal routing engine called GURAFU. In a nutshell, GURAFU returns the long-run optimal routing for a given pickup and drop-off location pair. We feed the 2016 taxi and Uber Manhattan trips in GURAFU and then examine how the actual distance compares with the GURAFU-returned distance. We show in online appendix Figure A4 that in Manhattan, both taxi and Uber trip distances closely concentrate on the GURAFU benchmark. Backed by these empirical evidence, we consider Manhattan trips as the no detour benchmark throughout the paper.

For metered airport trips, both taxi and Uber drivers' detour benefit increases because potentially low drop-off demand *and* long length creates scope for detouring. As we discuss in Section 3, drivers have incentive to detour if they believe the drop-off demand is not sufficiently high, and the long length of airport trips creates more room to detour. However, the detour penalty cost is higher for Uber drivers than for taxis, which may reduce Uber drivers' detour incentives. If this taxi-Uber difference in penalty cost is nontrivial, it leads to a testable difference in routing choices—taxi drivers may route longer than Uber drivers, producing a positive taxi-Uber distance ratio, when compared with 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, it will be less than the no detour benchmark. Let c denote taxis and u denote Uber. Let r denote a given matched taxi-Uber pair. We specify and estimate the following empirical model:

$$\frac{d_r^c}{d_r^u} = \alpha_0 + \alpha_1 M\_Airport_r + \alpha_2 JFK_r + \phi_{t(r)} + \epsilon_r. \quad (1)$$

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 at the time of 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, distance ratio is expected 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., $\alpha_2 = 0$).

4.2.2. 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 inefficient drivers (e.g., due to low ability) 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 with passengers of certain destinations is close to randomly assigned, because (1) passengers do not select taxis because taxi cabs are ex ante homogeneous to passengers; (2) taxi refusal of passengers is heavily penalized by the TLC refusal law.[18] However, taxi drivers can indeed form expectations of passenger destinations and route profitability and develop their own geographical search strategies (Zhang et al. 2015, Haggag et al. 2017, Zhang et al. 2018), 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 bias. In the absence of taxi driver IDs in 2016, we demonstrate in Section 6 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.

Drivers may be more likely to detour when driving nonlocal passengers due to information asymmetry.[19] To the extent that Uber market designs reduce information asymmetry, the strategic routing inefficiency for nonlocal 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 "nonlocal" should be smaller on airport routes than on nonairport routes, because it is more likely that the taxi rider and the Uber rider are either both locals or both nonlocals when they head to/from the airport from/to the same specific place at the same time (e.g., a hotel). For metered airport trips, taxi-Uber distance ratios are 1.5% greater when passengers are nonlocal than when they are local, and the difference is statistically significant at 1% level. This difference is 0.4% for Manhattan trips (statistically significant at the 5% level) and not statistically different from 0 for JFK trips. Leveraging the abovementioned sources of variation, our enhanced regression equation is the following,

$$\begin{aligned}\frac{d_r^c}{d_r^u} = {} & \alpha_0 + \alpha_1 M\_Airport_r + \alpha_2 JFK_r + \alpha_3 NonLocal_r \\ & + \alpha_4 M\_Airport_r \times NonLocal_r + \alpha_5 JFK_r \\ & \times NonLocal_r + X_r\Omega + \eta_{i(r)} + \phi_{t(r)} + \epsilon_r. \end{aligned} \tag{2}$$

where $NonLocal_r$ is equal to 1 if the passenger is nonlocal and 0 otherwise; X_r is a set of Uber driver and Uber rider characteristics, including Uber driver's total trips prior to the focal trip, Uber driver rating, and Uber rider' total trips completed prior to the focal trip; $\eta_{i(r)}$ is the Uber driver fixed effects. We use α_3, α_4, and α_5 to detect additional taxi-Uber routing difference when passengers are nonlocal, for Manhattan trips, metered airport trips, and JFK trips, respectively. Uber driver experience, measured by the driver's total trips driven prior to the current trip, is expected to positively correlate with Uber driver routing efficiency if there is a learning-by-doing effect. Uber driver routing efficiency is expected to positively correlate with Uber driver rating, as routing efficiency is an important metric in overall driver quality. More experienced Uber riders (measured by the total number of trips completed) 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.

4.3. Results Are Consistent with Moral Hazard

Table 3 reports regression results of Equation (1) and Equation (2). 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 may be the positive bias favoring taxis because of the way taxis and Uber handle pick-ups and drop-offs as we discussed earlier. Nonetheless, it becomes smaller in size and loses statistical significance as more controls and fixed effects are included in the regression. In

Table 3. Taxi-Uber Routing Difference

D.V. = Taxi distance/Uber distance	(1)	(2)	(3)	(4)	(5)
	Baseline	Nonlocal	More controls	Driver FE	Surge
M_Airport	0.080***	0.070***	0.070***	0.069***	0.072***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
JFK	0.009***	0.007***	0.008***	0.006*	0.007**
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
NonLocal		−0.005**	−0.005**	−0.002	−0.002
		(0.002)	(0.002)	(0.002)	(0.002)
M_Airport × NonLocal		0.019***	0.019***	0.016***	0.015***
		(0.003)	(0.003)	(0.003)	(0.003)
JFK × NonLocal		0.005	0.004	0.002	0.002
		(0.003)	(0.003)	(0.004)	(0.004)
Log (Uber_driver_total_trips)			0.000	0.004**	0.004***
			(0.000)	(0.002)	(0.002)
Uber_driver_rating			0.037***		
			(0.006)		
Log (Uber_rider_total_trips)			0.000	0.000	0.000
			(0.000)	(0.000)	(0.000)
Surge					0.004
					(0.002)
Surge × M_Airport					−0.018***
					(0.003)
Surge × JFK					−0.004
					(0.006)
Hour of week FE	Yes	Yes	Yes	Yes	Yes
Uber driver FE	No	No	No	Yes	Yes
N	90,431	90,431	90,431	90,431	90,431
R^2	0.069	0.070	0.071	0.371	0.371

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

addition, we find that the coefficient of *M_Airport* and the coefficient of *JFK* are statistically different from each other at the 1% level, and this significant difference holds across specifications.

When we add nonlocal and its interaction with metered airport trips and JFK airport trips, we find that taxi drivers route additionally longer when the passenger is nonlocal than when the passenger is local (1.9%) when driving metered airport trips (Table 3; 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 nonlocal 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 nonlocal passengers) in situations with large detour benefit (i.e., metered airport trips), and Uber drivers do not exhibit such strategic responses compared with 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. Particularly, we observe that Uber driver rating is positively correlated with the taxi-Uber distance ratio, likely because better-rated Uber drivers are more efficient, thereby increasing the taxi-Uber distance ratio by reducing the denominator. We provide a supplementary analysis in online appendix (Table A1) to show that Uber drivers' rating and experience are indeed positively correlated with their routing efficiency, when compared with other Uber drivers completing essentially the same trip. We also show in online appendix (Table A2) that the estimated detour is more pronounced when we use alternative efficiency benchmarks, for example, only Uber drivers with top decile ratings.

The estimated effects hardly change when we control for Uber driver fixed effects, as shown in Specification (4).[20] 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). Therefore, 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.

4.3.1. Surge Pricing and Moral Hazard. Surge pricing on Uber is praised by many to be effective at clearing the market. However, it may create room for driver moral hazard as high surge pricing makes detouring more profitable. To explore whether this is compatible with the data, we report in Specification (5) of Table 3 the regression results with the surge pricing dummy (i.e., surge multiplier > 1) and its interaction with *M_Airport* and *JFK*.[21] 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 is negative and statistically significant, suggesting that Uber drivers route less efficiently on metered airport trips, and that they are more likely to engage in moral hazard when surge pricing is in effect. The surge effect for JFK trips, however, is small in size and insignificant. Overall, the evidence presented in Table 3 Specification (5) reveals that moral hazard can be mitigated by Uber's market designs, but not completely eliminated as surge pricing may have created a new margin for moral hazard.

5. Navigation-Related Alternative Explanations

5.1. GPS Effect or Moral Hazard?

One threat to our identification of moral hazard is that the increased GPS usage among Uber drivers may have improved their routing behavior compared with taxis drivers. Although NYC taxi drivers are well-known for their driving experience and sophistication, GPS can still be effective, especially in situations where real-time traffic information is valuable. For example, taxi drivers may 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 they often involve the same set of bridges or tunnels to cross the river. Therefore, 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.

5.2. 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. This is done by investigating taxi driver routing decisions via a case study, which focuses on routes between LaGuardia Airport and a Midtown Manhattan area[22] (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 "Short"), whereas the other two routes have tolls of the same amount,[23] with the top route via FDR Drive generally longer in distance (we call this route "Long") than the bottom route via the Midtown Tunnel (we call this route "Medium"). Taxi and Uber trip distance indeed exhibits a clear bimodal pattern for these two tolled routes (online appendix Figure A5), 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 matched trips. If, say, a taxi driver who takes Long finishes the trip later than an Uber driver who takes Medium, then that suggests inefficient routing and moral hazard. However, as we discussed before, taxi drivers generally drive faster than Uber drivers in all cases. Thus, comparing taxi-Uber duration cannot cleanly infer moral hazard. Therefore, we need to *match taxi drivers with taxi drivers*, and investigate whether taxi drivers save passengers time by choosing longer routes. We follow the same matching approach as we did for taxi and Uber trips. Because we focus on airport trips 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.

Figure 5. (Color online) Alternative Routes Between Midtown Manhattan and LaGuardia

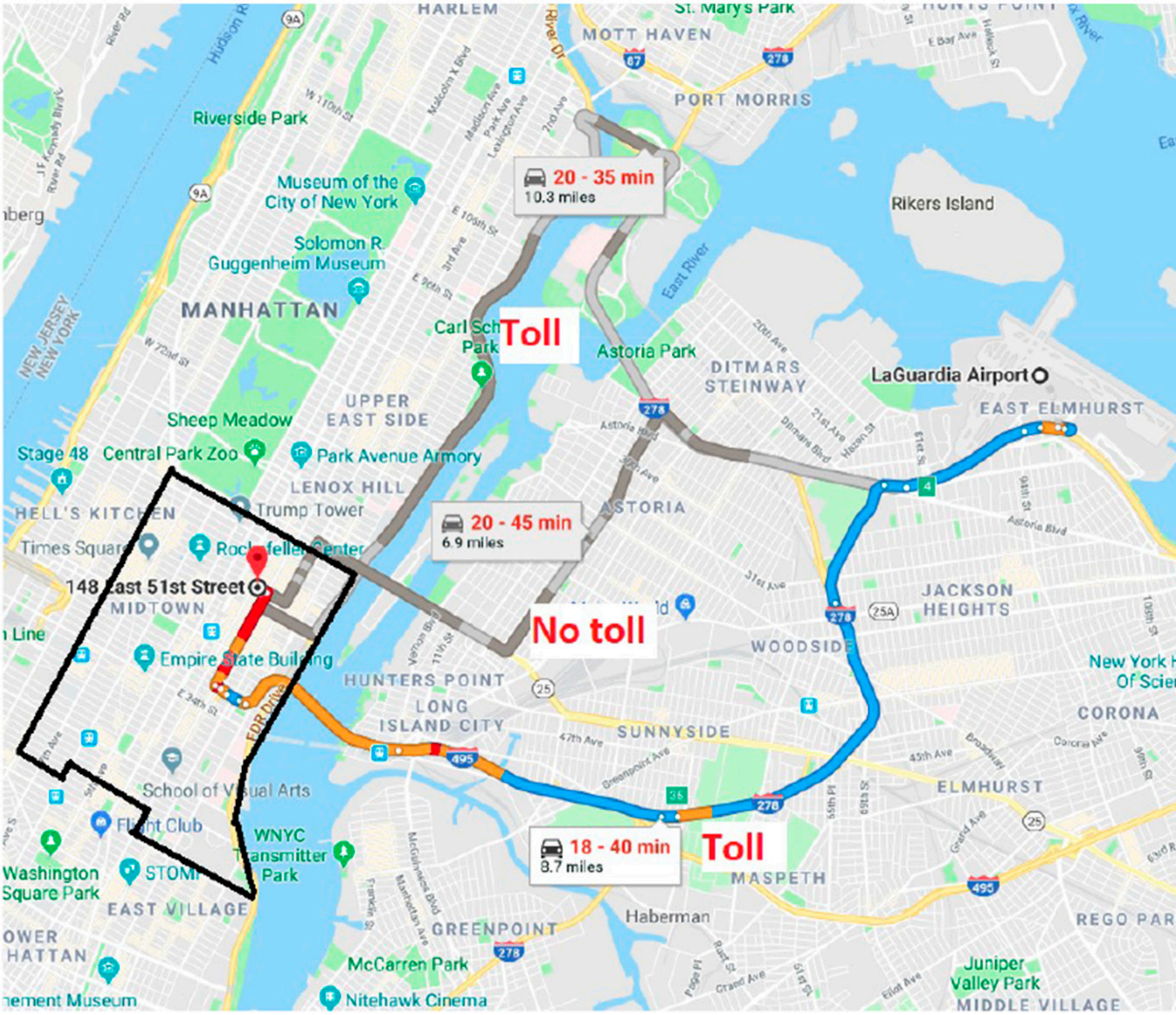


Table 4 shows how focal taxi drivers' travel time compares with that of the matched taxi drivers across situations when they take the same or different routes. For clean illustration, the cases Short-Medium (focal is Short and matched is Medium), Short-Long, and Medium-Long are not reported because they are symmetric to Medium-Short, Long-Short, and Long-Medium, 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 deviation. Numbers in braces are numbers of observations in each cell. 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 (i.e., Short-Short, Medium-Medium, and Long-Long), they spend about the same amount of time driving. When the focal driver takes Medium while the matched driver takes Short (i.e., Medium-Short), the focal driver indeed performs better in travel time (90%). In the case of Long-Medium, the focal driver on average needs 5% more time to complete the trip. Although Long saves passenger time compared with Short (0.94), it is still inferior to Medium because it is both longer in distance and longer in driving time. Therefore, it appears that Medium is often the time-efficient route, which is in line with our casual checks with Google Maps.

Then the question becomes: are taxi drivers more likely to take Long, which is most financially rewarding, when the optimal route is usually Medium? 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 indeed significantly more likely to choose Long than Uber drivers do, and they appear

Table 4. Do Longer Routes Save Passengers Time? A Comparison of Matched Taxi Drivers

		Focal taxi		
		Short	Medium	Long
Matched taxi	Short	1.00 (0.19) [9,325]	0.90*** (0.20) [20,036]	0.94*** (0.20) [31,605]
	Medium		1.01*** (0.17) [162,782]	1.05*** (0.20) [121,487]
	Long			1.00***(0.17) [229,248]

Notes. This table shows how focal taxi drivers' travel time compares with that of the matched taxi drivers 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 deviation. Numbers in brackets are numbers of observations in each cell. In the case of Medium-Short, the mean trip duration of focal drivers is 31.94 minutes with the standard deviation of 12.11 minutes. The mean trip duration of matched drivers is 36.15 minutes with the standard deviation of 12.51 minutes. The focal-matched time difference is statically significant at the 1% level. In the case of Long-Short, the mean trip duration of focal drivers is 33.60 minutes with the standard deviation of 11.18 minutes. The mean trip duration of matched drivers is 36.55 minutes with the standard deviation of 12.20 minutes. The focal-matched time difference is statically significant at the 1% level. In the case of Long-Medium, the mean trip duration of focal drivers is 34.79 minutes with the standard deviation of 11.36 minutes. The mean trip duration of matched drivers is 34.31 minutes with the standard deviation of 12.39 minutes. The focal-matched time difference is statically significant at the 1% level.

***Stands for 1% confidence level of the *t*-test against 1.

slightly less likely to take Medium routes than Uber drivers do. Therefore, it appears that taxi drivers on average cost passengers more time, instead of saving passengers time, by choosing longer routes.[24]

Finally, we find 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 7(a) 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., Long-Medium and Medium-Short), *and* the focal taxi driver arrives later than the matched driver. Figure 7(b) shows a more pronounced pattern with the definition *Detour 2*, where the focal driver both logs more distance and more time, no matter what routes they choose.[25] 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.

6. Moral Hazard vs. Adverse Selection

The observed difference in efficiency can be driven by selection, instead of behavior. Two types of selection may be present in our context—on the intensive margin, it may be that driver types select differently across route types; on the extensive margin, drivers may select differently into platforms (i.e., taxis or Uber). In this section, we demonstrate that our empirical patterns are not entirely accounted for by these selections.

6.1. Driver Selection on the Intensive Margin

As discussed earlier, the unobserved driver types in the error term may correlate with route types. It is reassuring 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

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

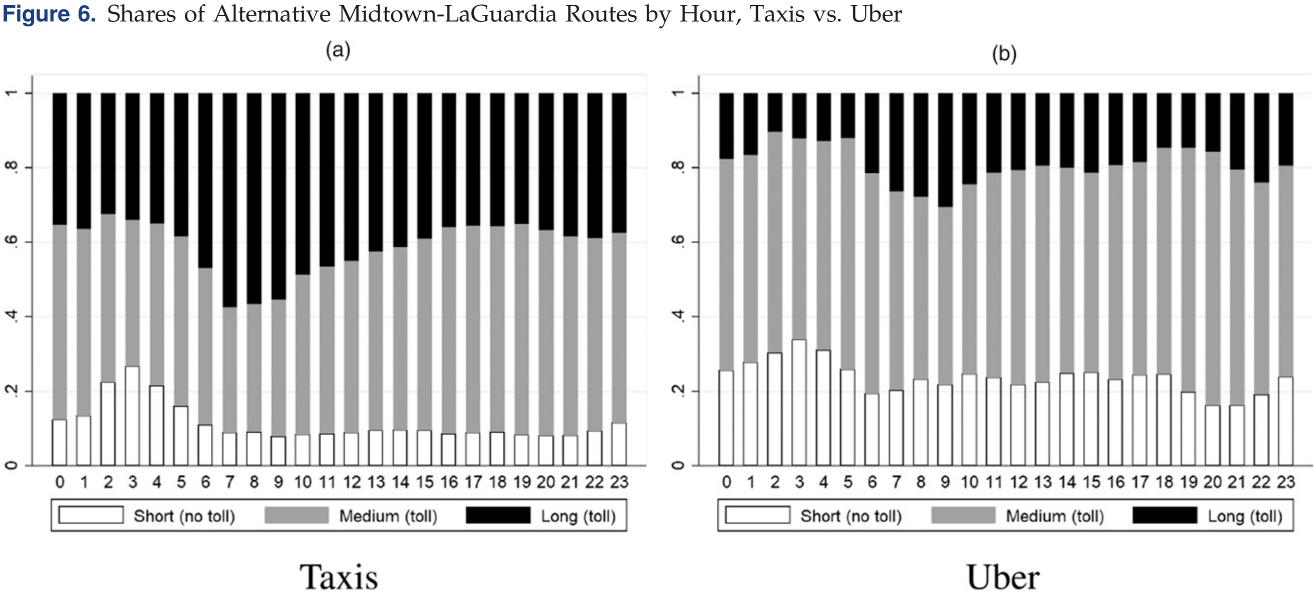
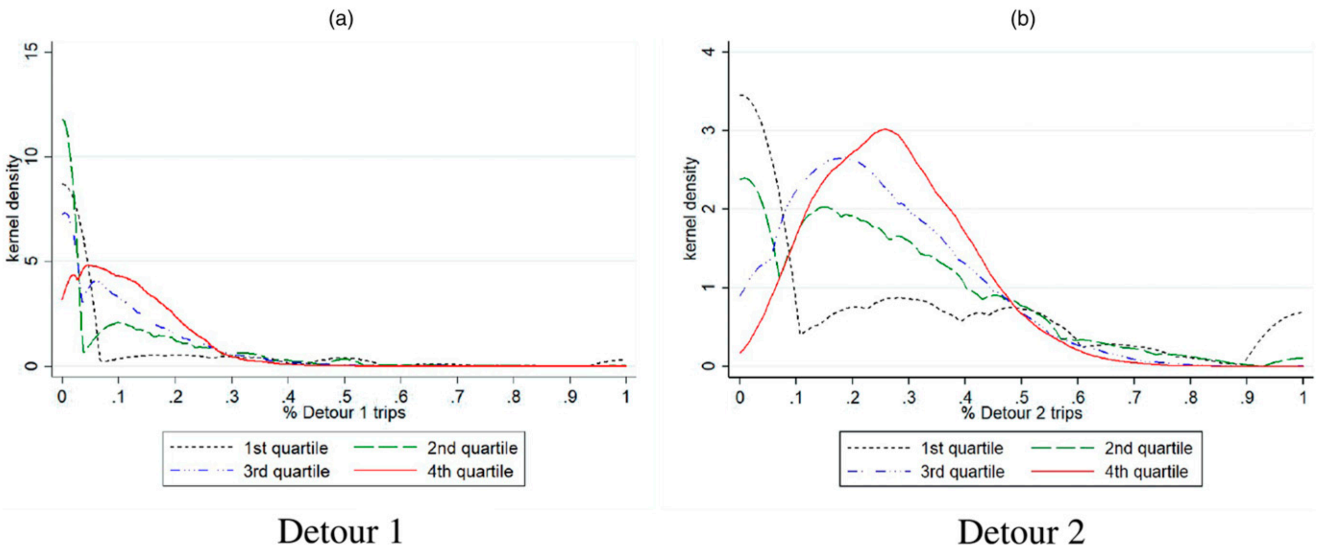


Figure 7. (Color online) Drivers with More Same-route Experience Tend to Detour More



Notes. For both detour definitions, we conduct pairwise comparisons across driver quartiles and find that the differences are overall statistically significant. Detailed test statistics will be shared upon request.

taxi-driver fixed effects are accounted for. 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 (< 1 at the 1% level), because the taxi distance has a downward measurement bias due to the taxi-Uber difference in pick-ups and drop-offs (refer to online appendix Figure A3). Thus, regressions that follow Equation (1) and Equation (2) should lead to an upward bias in coefficient estimates of *M_Airport* and *JFK*. The distance ratio is on average 1.05 (>1 at the 1% level) for metered airport trips and 1.01 (>1 at the 1% level) for JFK trips.

Table 5 Specifications (1) and (2) report 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. Across specifications, the coefficient of *M_Airport* and the coefficient of *JFK* are statistically different from each other at the 1% level. 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.

6.2. Driver Selection on the Extensive Margin

Drivers of varying types may select differently into taxi and Uber. If this is the case, then the observed moral hazard would be driven by 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.

We track the status of a driver across time by linking their TLC IDs. TLC driver IDs are a license system that applies to all types of taxi-like services in NYC. We observe TLC driver IDs for all Uber drivers[26] and only anonymized TLC IDs for taxi drivers in the 2013 data. However, TLC used a simple encryption method called MD5, which is a one-way hash, where the same input always produces the same MD5 hash. Given this property, we applied MD5 hashing to Uber drivers' TLC IDs, which then allowed us to uniquely identify drivers who switched. In the entire matching process, we only used anonymized IDs without characteristics such as actual IDs or driver names. 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 taxis to Uber at some point between 2013 and 2016. We refer to these drivers as switchers. We focus on 1,999 switchers who appear both in the 2013 matched sample and the 2016 matched sample.

Table 5. Moral Hazard vs. Adverse Selection

	2013 taxis vs. 2013 Uber		Switchers vs. 2013 Uber		Switchers vs. 2016 taxis	
	D.V. = Taxi distance/Uber distance		D.V. = Switcher distance/Uber distance		D.V. = Taxi distance/Switcher distance	
	(1)	(2)	(3)	(4)	(5)	(6)
M_Airport	0.124*** (0.002)	0.114*** (0.007)	0.120*** (0.005)	0.105*** (0.019)	0.054*** (0.002)	0.054*** (0.004)
JFK	0.077*** (0.004)	0.068*** (0.013)	0.079*** (0.011)	0.078*** (0.029)	0.005 (0.004)	0.010* (0.006)
NonLocal		–0.013*** (0.004)		–0.016 (0.011)		0.001 (0.004)
M_Airport × NonLocal		0.025*** (0.008)		0.025 (0.022)		0.003 (0.005)
JFK × NonLocal		0.022 (0.015)		0.012 (0.035)		–0.005 (0.008)
Log (Uber_driver_total_trips)		–0.002 (0.001)		–0.001 (0.004)		0.003 (0.003)
Log (Uber_rider_total_trips)		0.001 (0.001)		0.000 (0.003)		–0.001 (0.001)
Hour of week FE	Yes	Yes	Yes	Yes	Yes	Yes
Taxi driver FE	No	Yes	—	—	—	—
Switcher FE	—	—	No	Yes	No	Yes
N	23,774	23,774	4,030	4,030	16,363	16,363
R^2	0.123	0.591	0.112	0.616	0.036	0.239

Notes. We use the term "Switchers" to denote drivers who drove taxis in 2013 but switched to Uber at some point between 2013 and 2016. For Specifications (1) and (2), because of small sample size, we do not include Uber driver FEs in the regression analyses. For all specifications, standard errors are cluster-robust at the hour-of-week level. FE, fixed effect.

***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

We first show that the routing behavior of switchers, who were taxi drivers in 2013, exhibits detour patterns when compared with that of Uber drivers in 2013. *T*-tests performed on the sample yield a mean of distance ratio at 0.92 (< 1 at the 1% level) for Manhattan trips, 1.04 (> 1 at the 1% level) for metered airport trips, and 1.00 (not statistically different from 1) for JFK trips. Table 5 Specifications (3) and (4) report the regression results using the 2013 matched sample of these 1,999 switchers and Uber drivers. Because 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 5 Specifications (5) and (6), we perform the same regression analyses on the matched sample consisting of taxi drivers in 2016 and switchers, who were Uber drivers in 2016.[27] 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. In addition, the *M_Airport* coefficient and the *JFK* coefficient are statistically different at the 1% level for both specifications. Therefore, evidence in Table 5 Specifications (3)–(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 also match trips of switchers with trips of regular Uber drivers (i.e., drivers who likely never drove taxis before) in 2016 to see if they are similar.[28] We find a positive, small effect of *M_Airport* (1.1%–2.1%, depending on specifications) and essentially zero effect of *JFK* (detailed analysis in online appendix Table A3), indicating that the switchers still exhibit some degree of detouring on metered airport trips, and no detouring is observed on JFK flat fare trips. This suggests that switchers may still cheat a little, when compared with regular Uber drivers who likely never drove taxis before, and this is perhaps due to old habits.

In addition, we identify 178 drivers who were driving for both taxis and Uber in 2013, which we denote as multihomers. We examine how multihomers' routing compares with single-homing drivers, using the matched 2013 sample, separately for their Uber and taxi roles. We find that multihomers exhibit strategic long routing behavior when they drove taxis, but long routing is much less pronounced when they drove Uber. Detailed discussions are provided in online appendix Table A4.

7. Other Issues and Robustness Checks

In this section, we discuss other issues that may bias our identification and we report several robustness checks that alleviate these concerns.

7.1. Passenger Selection

Recall that our main specifications control for the experience of Uber passengers, which can mitigate the endogeneity threat that Uber passenger types select differently across routes. However, another type of passenger selection may exist—passenger types may select differently into taxis and Uber. For example, passengers who are more sophisticated and demanding may select into Uber, which can lead to reduced inefficiency, independently from the treatment effect of Uber's incentive mechanisms. The ideal solution to this problem is to control for both taxi and Uber passenger fixed effects into the model. Unfortunately, we do not have information on taxi passenger IDs to perform such practice. Nonetheless, we find this alternative explanation inconsistent with taxi-Uber efficiency difference across airport trip types. For example, if Uber passengers are overall more sophisticated than taxi passengers, we would expect a similar taxi driver inefficiency on JFK airport trips as on metered airport trips.

7.2. Tolls

There are cases when drivers can make passengers better off by choosing longer routes that minimize tolls (see online appendix Figure A6 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 A7 in the online appendix 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.

7.3. Various Time Windows in Matching

In the main analysis of driver detour, we constrained the matched taxi and Uber trips to be 15 minutes apart. When we use alternative time windows (namely 5, 10, 20, and 30 minutes), the estimated effects are stable and consistent across time window lengths (see online appendix Table A5). 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.

7.4. Taxi Meter Measurement Error

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. Estimating the main regressions using only Vendor 1 leads to an upward bias (see online appendix Table A6). We find this upward bias, instead of a downward one, consistent with our main findings.

8. Mechanisms

Our empirical results imply that market designs implemented by Uber (monitoring, rating, conflict resolution) enhance market transparency and effectively deter opportunistic behavior in most cases, as Uber drivers do not appear to detour on JFK airport routes, where the potential gain from detouring is large. However, to the extent that our treatment is the combination of market designs implemented by Uber, it is difficult to quantify the contribution of specific mechanisms to the observed patterns. An ideal empirical setting would be one where one mechanism can be varied while the others are held the same. Nonetheless, we shed light on the importance of enhanced passenger monitoring by leveraging the direction of metered airport trips.

Specifically, we run the baseline specifications by splitting metered airport trips into "To_M_Airport" and "From_M_Airport" trips and splitting JFK trips into "To_JFK" and "From_JFK" trips, based on airport trip direction. Reported in online appendix Table A7, we find a smaller tendency of taxi long routing when taking passengers to airports, where the coefficients for "To" and "From" are statistically different from each other at 1% level for both metered airport trips and JFK trips. At first glance, this is not quite consistent with what the demand at drop-off would dictate—drivers should have a greater detour incentive when driving to the airport because idle time at airports are substantially longer than that of Manhattan (Figure 4). However, passengers are plausibly more time sensitive when they go to the airport than leave the airport for the city. Because of this, passengers should be more motivated to monitor the routing efficiency on to-airport routes.

In a separate analysis, we examine taxi passenger tips[29] and find that tips respond to driver routing efficiency on LaGuardia trips, more so on to-LaGuardia trips than on from-LaGuardia trips (see online appendix Table A8). This suggests a stronger time pressure of passengers who go to LaGuardia that may cause them to monitor the routing more closely. Interestingly, we find no statistically significant correlation between tips and routing efficiency for JFK trips, or JFK trips split by direction. This may be because passengers do not perceive routing inefficiency on flat-fare routes as intentional long routing, given that the interests of both parties are fully aligned. Taken together, these empirical patterns point to an important role of enhanced passenger monitoring.

One necessary condition for a working rating system to penalize strategic behavior is the negative correlation between passengers' ratings to the drivers and driver routing inefficiency. We use the JFK flat-fare subsample (recall that taxis are the efficiency benchmark on these trips) to show that this correlation is robust (online appendix Table A9). Specifically, passenger rating of a given trip is regressed on the routing efficiency of the Uber driver, measured by the Uber-taxi distance and duration ratios. The estimates suggest that when the Uber trip costs more distance and time as compared with taxis, passenger rating tends to be lower. 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.

9. Concluding Remarks

In this paper, we study whether digital platforms affect moral hazard and service quality, when compared with 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 nonlocal. This long routing is not found for short, within-Manhattan trips or airport trips with a flat 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.

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.[30] In general, once the smart phone infrastructure is in place, these features can be rolled out at very low marginal cost.

That said, the incentives for creating and using driver expertise and discretion may be reduced by the Uber platform, relative to reliance on technologies such as GPS.[31] This is confirmed by an interview with Loai Yousef,[32] 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 the rider to just cross the street. Taxi drivers often drop passengers off a short distance from exact address." As a result, Uber drivers might not be as motivated as taxi drivers to use their discretion, because off-GPS routing might come across as suspicious behavior to the riders, which can result in bad ratings and complaints. 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.

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.

Acknowledgments

We thank Keith Chen, Andrey Fradkin, John Horton, Xiang Hui, and Erina Ytsma, as well as seminar participants at MIT, Uber, Washington University in St. Louis, Boston Federal

Reserve, Federal Communications Commission, NBER Summer Institute, and other conferences 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.

Endnotes

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

[2] See https://www.businessinsider.com/leaked-charts-show-how-ubers-driver-rating-system-works-2015-2. Last accessed on 07-30-2020.

[3] 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" pp. 1608.

[4] Refer to the official language on the pricing rule: http://www.nyc.gov/html/tlc/html/passenger/taxicab_rate.shtml. Last accessed on 07-30-2020.

[5] In slow traffic, taxi rides are often "stop-and-go," so usually part of the trip is metered by duration and the other by distance.

[6] Using the June 2016 taxi trip records, we compute the average taxi speed, defined as the trip distance (miles) divided by the trip duration (hours). After removing the top and bottom 1% as extreme values, the average taxi speed in NYC is 11.16 mph and the median is 10.17 mph.

[7] It appears that the minimum fare was \$7 in our 2016 sample and it may have increased to \$8 in the last month of our sample period.

[8] Studies have shown that factors such as social desirability can also affect self-dealing (Covey et al. 1989, Filiz-Ozbay and Ozbay 2014).

[9] This assumption on the rare occurrence of long-length trips is consistent with the small share of airport trips in reality.

[10] Although we assume no penalty cost here for simplicity, in reality it can be true that penalty cost is a function of trip length. In the theory part in the appendix, we account for this possibility and discuss the implications.

[11] Taxi trip records can be downloaded either directly from the TLC's website or from NYC Open Data. We obtained the 2016 taxi data by direct download. We obtained the 2013 taxi data, which contained anonymized (by the TLC) taxi driver IDs, from Chris Wong's original FOIL (Freedom of Information Law) request of NYC taxi data (https://chriswhong.com/open-data/foil_nyc_taxi/). Last accessed on 07-30-2020. When Wong requested these data, there was no such data publicly available on the TLC website or NYC Open Data. At a later time, the TLC decided to put the taxi data online, but the published data did not include the anonymized driver IDs or medallion IDs. Therefore, we chose to use the data shared by Wong.

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

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

[14] We use the GIS shapefile of NYC streets published by NYC Department of City Planning (see https://data.cityofnewyork.us/City-Government/LION/2v4z-66xt/data) Last accessed on 07-30-2020 to associate the pick-up and drop-off coordinates to the closest street. A nice feature of the shapefile is that it records all unique directions of the same street. Because a majority of streets in Manhattan are single-directional, we automatically know the direction of the traffic the trip had to follow. For streets with multiple directions, we were able to associate the pick-up point with the direction of the street that it is closest to.

[15] 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.

[16] It should be noted that the difference in speed can also be an outcome of experience, given that taxi drivers on average tend to be more experienced than Uber drivers.

[17] This is plausibly due to the lack of network effect in Manhattan given the size of Uber in 2013. When we use the 2016 Uber data to compute idle time, we show that Manhattan drop-offs indeed represent a significantly lower average idle time than airport drop-offs. The plot is available upon request.

[18] 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" pp. 85. Although TLC strictly enforces the refusal law, anecdotal evidence exists that taxi drivers sometimes refuse passengers in spite of penalties.

[19] We use the Uber passenger's billing zip code to identify local passengers from nonlocal passengers. Although Uber drivers do not have such information and thus may not have the exact knowledge of the localness of passengers, there are ways for them to tell. For example, if a passenger goes from the airport to a hotel in Manhattan, then that passenger is most likely nonlocal. Similarly, languages and/or accents of the passengers can be used to tell if they are local. Therefore, on average drivers can distinguish with a positive probability. Although drivers' partial knowledge does not affect the rationale of using local versus nonlocal for our analysis, the size of effect could be biased downward, if the measurement error of drivers' ability to distinguish is a classical one that adds random noise to the localness measure. An extreme case would be where drivers have no way to tell (i.e., knowledge is purely noise), then there should be zero difference between how they drive locals and how they drive nonlocals.

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

[21] Two features of Uber surge pricing 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 earnings from the base fare. Second, surge multipliers are extremely volatile and difficult for individual drivers to predict. As shown in Lam et al. (2017), Uber surge multipliers are volatile and difficult to predict even after accounting for highly granular location-time fixed effects. Although drivers may have the incentive to "chase" the surge, market designs on Uber make this not worthwhile. Uber drivers commonly agree on this view, based on our conversations with Uber drivers in NYC and Boston, Massachusetts. For example, Uber drivers can check the across-location distribution of real-time surge pricing from the app

and move to any location as they desire. However, this "surge hotspots" feature is shut down once the driver accepts a pick-up request. This limits driver strategic cancellation of rides, because without such an automatic shutdown, drivers may cancel on the current passenger when they discover higher surge on their way over. Therefore, these features alleviate the endogeneity concern of surge pricing in our empirical setting, especially when hour-of-week fixed effects are accounted for.

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

[23] In 2016, the cash rate of the tolls was $8.00 for taxi-like vehicles. Taxi and Uber drivers normally pay the discounted rate of $5.54 by using an E-Z Pass.

[24] One caveat, as shown by Figure 6, is that taxi drivers are significantly less likely to take Short routes than Uber drivers do. Short routes, as we show by the duration ratios (Table 4), are usually longest in time but shortest in distance. Plus, it is toll-free. When we focus on the Medium-Short cell, Short routes on average take 36.15 minutes and cost $32.71 (the time-and-distance fare calculated by the meter), whereas Medium routes on average take 31.94 minutes and cost $37.48 (the time-and-distance fare by the meter plus tolls). That is, on average the 4.21 minutes saved come at the cost of $4.77. Given this, which option is better depends on how passengers trade off low fares with lost minutes. If passengers value time sufficiently (i.e., about $1.00 per minute), then taxi drivers' avoiding Short routes may benefit passengers. On the contrary, when the time value is not as high, passengers can be made worse off by not having the Short option. Therefore, we do not draw conclusions based on Short and instead we focus on Medium versus Long where one option appears to dominate the other.

[25] Detour 2 is a more relaxed measure in that Detour 2 also captures cases where both drivers take the same bridge/tunnel, but one is more efficient in both distance and time.

[26] All drivers who drive with Uber in NYC share their TLC license (ID) with Uber. Given that TLC license is a regulatory requirement imposed by the city and enforced by the TLC, Uber is not permitted to list drivers on its platform who do not have this license.

[27] *T*-tests performed on the sample yield a mean of distance ratio at 0.98 (< 1 at the 1% level) for Manhattan trips, 1.04 (> 1 at the 1% level) for metered airport trips, and 0.99 (< 1 at the 1% level) for JFK trips.

[28] With our current data, the identification of Uber drivers who were never taxi drivers is imperfect because we cannot link taxi license IDs except for 2013. For example, it is possible that a driver who shows up in our Uber driver pool and not in the 2013 taxi driver pool may be an ex-taxi driver who stopped driving taxi as early as 2012. That said, we define regular Uber drivers as Uber drivers who are not switchers.

[29] Although passengers can freely choose how much to tip, tips can be influenced by the default options set up by taxi companies (Haggag and Paci 2014).

[30] Detouring leads to welfare loss in the form of lost passenger time. To compute how much time would be lost when taxi drivers take longer routes, we use the within-taxi matched sample and regress the duration ratio on the distance ratio. The coefficient estimate of 0.20 suggests that when a taxi driver increases distance by 100%, trip duration is expected to increase by 20%. This means that for the average 8% detour by taxi drivers on metered airport trips, as estimated in our main specification, the extra travel time is roughly 1.6%. 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. Passengers also pay more when drivers detour, although the payment is an income transfer to drivers rather than an efficiency loss. The detour on an average airport trip is about 0.7 miles, which translates into 12,600 added miles per day. This is equal to an extra passenger payment of $37,800 per day (12,600 miles × $2.50 per mile × (1+20%), assuming a 20% tip).

[31] 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.

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

References

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

Athey S, Castillo JC, Chandar B (2019) Service quality in the gig economy: Empirical evidence about driving quality at Uber. https://www.tse-fr.eu/sites/default/files/TSE/documents/ChaireJJL/wp/athey_ubernudges.pdf.

Bakos JY (1997) Reducing buyer search costs: Implications for electronic marketplaces. *Management Sci.* 43(12):1676–1692.

Balafoutas L, Beck A, Kerschbamer R, Sutter M (2013) What drives taxi drivers? A field experiment on fraud in a market for credence goods. *Rev. Econom. Stud.* 80(3):876–891.

Balafoutas L, Kerschbamer R, Sutter M (2017) Second-degree moral hazard in a real-world credence goods market. *Econom. J. (Lond.).* 127(599):1–18.

Bresnahan TF, Brynjolfsson E, Hitt LM (2002) Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *Quart. J. Econom.* 117(1):339–376.

Brown JR, Goolsbee A (2002) Does the Internet make markets more competitive? Evidence from the life insurance industry. *J. Political Econom.* 110(3):481–507.

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

Brynjolfsson E, Smith MD (2000) Frictionless commerce? A comparison of Internet and conventional retailers. *Management Sci.* 46(4):563–585.

Castillo JC, Knoepfle D, Weyl G (2017) Surge pricing solves the wild goose chase. *Proc. 2017 ACM Conf. Econom. Comput.* (ACM, Cambridge, MA), 241–242.

Chen MK, Chevalier JA, Rossi PE, Oehlsen E (2017) The Value of flexible work: Evidence from Uber drivers. NBER Working Paper No. 23296, National Bureau of Economic Research, Cambridge, MA.

Cohen P, Hahn R, Hall J, Levitt S, Metcalfe R (2016) Using big data to estimate consumer surplus: The case of Uber. NBER Working Paper No. 22627, National Bureau of Economic Research, Cambridge, MA.

Cook C, Diamond R, Hall J, List JA, Oyer P (2018) The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. Accessed June 25, 2020, https://web.stanford.edu/~diamondr/UberPayGap.pdf.

Covey MK, Saladin S, Killen PJ (1989) Self-monitoring, surveillance, and incentive effects on cheating. *J. Soc. Psychol.* 129(5):673–679.

Cramer J, Krueger AB (2016) Disruptive change in the taxi business: The case of Uber. *Amer. Econom. Rev.* 106(5):177–182.

Duflo E, Hanna R, Ryan SP (2012) Incentives work: Getting teachers to come to school. *Amer. Econom. Rev.* 102(4):1241–1278.

Erhardt, GD, Roy S, Cooper D, Sana B, Chen M, Castiglione J (2019). Do transportation network companies decrease or increase congestion? *Sci. Adv.* 5(5), eaau2670.

Filippas A, Horton JJ, Golden J (2018) Reputation inflation. *Proc. 2018 ACM Conf. Econom. Comput.* (ACM, New York), 483–484.

Filiz-Ozbay E, Ozbay EY (2014) Effect of an audience in public goods provision. *Exp. Econ.* 17(2):200–214.

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

Gans JS, Goldfarb A, Lederman M (2017) Exit, weets, and oyalty. NBER Working Paper No. 23046, National Bureau of Economic Research, Cambridge, MA.

Goldfarb A, Tucker C (2017) Digital conomics. NBER Working Paper No. 23684, National Bureau of Economic Research, Cambridge, MA.

Greenwood BN, Wattal S (2017) Show me the way to go home: An empirical investigation of ride-sharing and alcohol related motor vehicle fatalities. *Management Inf. Syst. Quart.* 41(1):163–187.

Haggag K, McManus B, Paci G (2017) Learning by driving: Productivity improvements by New York City taxi drivers. *Amer. Econom. J. Appl. Econom.* 9(1):70–95.

Haggag K, Paci G (2014) Default tips. *Amer. Econom. J. Appl. Econom.* 6(3):1–19.

Hall J, Kendrick C, Nosko C (2015) *The Effects of Uber's Surge Pricing: A Case Study* (The University of Chicago Booth School of Business, Chicago, IL).

Hall JV, Horton JJ, Knoepfle DT (2019) Pricing efficiently in designed markets: The case of ride-sharing. john-joseph-horton.com. Accessed January 18, 2021, https://www.semanticscholar.org/paper/Pricing-Efficiently-in-Designed-Markets%3A-The-Case-Hall-Horton/c96ad4e51587c3d7ed900f7a90da1d2dd86337bb?p2df.

Hall JV, Krueger AB (2018) An analysis of the labor market for Uber's driver-partners in the United States. *Ind. Labor Relat. Rev.* 71(3):705–732.

Hubbard TN (2000) The demand for monitoring technologies: The case of trucking. *Quart. J. Econom.* 115(2):533–560.

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

Klein TJ, Lambertz C, Stahl KO (2016) Market transparency, adverse selection, and moral hazard. *J. Political Econom.* 124(6):1677–1713.

Lam CT, Liu M, Hui X (2020) The geography of ridesharing: A case study of New York City. Preprint, submitted May 18, https://dx.doi.org/10.2139/ssrn.2997190.

Lee MK, Kusbit D, Metsky E, Dabbish L (2015) Working with machines: The impact of algorithmic and data-driven management on human workers. *Proc. 33rd Annual ACM Conf. Human Factors Comput. Systems* (ACM, New York), 1603–1612.

Liang C, Hong Y, Gu B (2016) Effects of IT-enabled monitoring on labor contracting in online platforms: Evidence from a natural experiment. Preprint, submitted September 30, https://dx.doi.org/10.2139/ssrn.2844920.

Liu T, Vergara-Cobos E, Zhou Y (2019) Pricing schemes and seller fraud: Evidence from New York City taxi rides. *J. of Industrial Econom.* 67(1):56–90.

Nagin DS, Rebitzer JB, Sanders S, Taylor LJ (2002) Monitoring, motivation, and management: The determinants of opportunistic behavior in a field experiment. *Amer. Econom. Rev.* 92(4): 850–873.

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

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

Rajgopal S, White R (2015) Cheating when in the hole: The case of New York City taxis. *Accounting, Organ. and Soc.* 79:101070.

Reimers I, Shiller BR. (2019) The impacts of telematics on competition and consumer behavior in insurance. *J. of Law and Econom.* 62(4):613–632.

Staats BR, Dai H, Hofmann D, Milkman KL (2016) Motivating process compliance through individual electronic monitoring: An empirical examination of hand hygiene in healthcare. *Management Sci.* 63(5):1563–1585.

Sudhir K, Talukdar D (2015) The "Peter Pan syndrome" in emerging markets: The productivity-transparency trade-off in it adoption. *Marketing Sci.* 34(4):500–521.

Tabarrok A, Cowen T (2015) The end of asymmetric information. Cato Unbound. Accessed January 18, 2021, https://www.cato-unbound.org/print-issue/1915.

Zhang Y, Li B, Krishnan R, Liu S (2015) Learning from the offline trace: A case study of the taxi industry. Working paper, Carnegie Mellon University, Pittsburgh.

Zhang Y, Li B, Ramayya K (2018) Learning individual behavior using sensor data: The case of GPS traces and taxi drivers. *Inform. System Res*. 31(4):1304–1321.