Learning by Driving: Productivity Improvements by New York City Taxi Drivers

By KAREEM HAGGAG, BRIAN MCMANUS, AND GIOVANNI PACI*

We study learning by doing (LBD) by New York City taxi drivers, who have substantial discretion over their driving strategies and receive compensation closely tied to their success in finding customers. In addition to documenting learning overall by these entrepreneurial agents, we exploit our data's breadth to investigate the factors that contribute to driver improvement across a variety of situations. New drivers lag further behind experienced drivers when in difficult situations. Drivers benefit from accumulating neighborhood-specific experience, which affects how they search for their next customers.

* Haggag: University of Chicago, Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637 (e-mail: kareem.haggag@chicagobooth.edu); McManus: University of North Carolina, Chapel Hill, Department of Economics CB 3305, Chapel Hill, NC 27599 (e-mail: mcmanusb@unc.edu); Paci: New York University, Institute for the Interdisciplinary Study of Decision Making (e-mail: giovannipaci.work@gmail.com). We thank Ray Fisman, Luca Flabbi, Jonas Hjort, Supreet Kaur, Kiki Pop-Eleches, Bernard Salanie, Chad Syverson, James Roberts, Scott Rockart, and several seminar audiences for helpful comments.

Learning by Doing (LBD) is an economically important phenomenon which can affect several types of market activity. At the level of the individual worker, learning is a source of productivity improvements that can increase wages. At the firm level, LBD can be welfare-improving when it leads to cost reductions that increase output in competitive markets; less beneficial effects can exist in concentrated markets, where LBD may provide incumbents with cost advantages that deter entry. While there exists an extensive literature that documents learning effects in many settings (see Thompson, 2010 and 2012, for surveys), in prior studies it has been difficult to learn which agents in a firm are improving their capabilities, what activities are being improved, and how strongly individual agents are encouraged to find improvements.

We provide new evidence on LBD that addresses some of these gaps in the literature. We use a highly detailed dataset of New York City (NYC) yellow taxi rides to study how drivers make improvements overall, how their performance varies across measurably different situations, and how general and specific experience make different contributions to driver performance and strategies. Taxi service in NYC is characterized by several features that make it an interesting setting for studying LBD.[1] First, drivers have substantial discretion in how they search for new customers, and their pay is closely tied to the earnings they collect through customers' fares. Drivers' agency makes them similar to entrepreneurs, searching for the next fruitful business opportunity. Second, drivers operate in an environment that lends itself to a relatively clean study of learning. While other LBD studies often contend with isolating productivity improvements from confounding factors such as scale economies, improvements in inputs, and shifts in input prices, by contrast all taxi drivers use essentially the same capital equipment, work in the same demand and cost environment, and charge the same prices. Third, we observe drivers visiting a variety of locations a large number of times, which provides opportunities to observe driver performance in different settings. The customer's control over drop-off locations introduces econometrically useful variation in whether a driver arrives in a setting with

[1] Several other authors have studied other aspects of the NYC taxi market. Studies of drivers' labor supply choices have produced results on the role of reference-dependent preferences (Camerer et al. 1997; Crawford and Meng 2011; Doran 2014; Farber 2005, 2008, 2015). Other industry features that have been studied include moral hazard in leasing contracts (Schneider 2010; Jackson and Schneider 2011) and tipping by customers (Haggag and Paci 2014).

strong or weak demand, or within which the driver has little or substantial prior experience.

We perform our analysis using data on all 171 million NYC yellow taxi rides during the full 2009 calendar year. Our main analysis features a sample of about 6,300 drivers, with almost 1,800 identified as new based on the date of the driver's first appearance in the 2009 data. Our main measure of a new driver's stock of experience is his cumulative number of shifts as of a particular day, although it is also possible to use finer units such as the driver's cumulative number of drop-offs. We measure driver output through fare earnings per hour, and in our main analysis we regress new drivers' earnings on measures of their experience. The breadth of our data allows us to control for variation in earning opportunities across individual hours of the sample period, as well as to focus on within-driver learning to avoid attrition problems. In addition to overall measures of learning and performance, we use information on the time and location of every 2009 taxi ride to investigate two additional issues. First, we use experienced drivers' performance across different New York City neighborhoods and different times of the week to characterize whether certain situations represent "easy" or "hard" earnings opportunities. We then analyze how new drivers' performance varies across these settings. Second, we compute each new driver's stock of experience in each neighborhood as of each date in our sample period; we use these data to examine how drivers' location-specific performance is affected by local experience versus overall tenure in the market.

We find that an average new driver's overall productivity increases by 7% between his first and 100th shift, which is worth about \$344 in cumulative fare earnings. Learning is relatively fast, and the overall improvement is fairly small, representing roughly 1.4% of gross fare revenue (or about 2.5% of take-home pay) during the first 100 shifts. We demonstrate that our results are affected substantially by the econometric controls we are able to apply, and in the controls' absence we obtain learning effects that are three times as large.

In addition to estimating the value of learning overall, we describe substantial heterogeneity in learning rates across locations and types of driver experience. We find that new drivers earn considerably less in more difficult situations. Following a drop-off in one of the most difficult locations, a driver in his first shift earns 14% less than an experienced driver in the same situation, but brand new drivers have essentially the same expected earnings as experienced drivers in the easiest settings. Despite the large gap in initial earnings in difficult situations, new drivers' earnings improve fastest there and quickly erase the gap with experienced drivers. Turning to drivers' experience in specific neighborhoods, we find that a driver's amount of local experience has a strong effect on performance after a drop-off in that neighborhood. Accumulated drop-offs in other neighborhoods, which stand-in for general experience, have a less robust effect. Local experience also has a significant role in explaining where new drivers find their next customers. Drivers with a greater share of experience in a drop-off neighborhood are more likely to stay there to find their next fares.

Our results join a large empirical literature on LBD. As we do, Pisano et al. (2001), Rockart and Dutt (2015), and Stith (2013) each consider learning outside of a factory setting. Other research has considered heterogeneity across diverse production settings and offer some controls for this diversity; examples include Thompson (2001), Balasubramanian and Liebman (2010), and Kellogg (2011). An additional branch of the literature analyzes highly detailed data in order to consider a variety of explanations for how improvements occur and how learning is transmitted through the firm; see Hatch and Mowry (1998), Levitt et al. (2013), and Hendel and Spiegel (2014). While the LBD literature largely focuses on output or cost improvements, Levitt et al. (2013) and Mas (2008) consider evidence of quality improvements due to learning. Finally, there exists a literature on learning by entrepreneurs (e.g. Rocha et al. 2013, Lafontaine and Shaw 2014), but this research focuses on learning across entrepreneurial episodes rather than learning about demand within a single business venture.

Our paper also relates to the labor economics literature that studies the returns from worker experience (e.g. Shaw and Lazear, 2008), and additionally the relative benefits of general versus firm- or task-specific experience (e.g. Ost 2014, Gathmann and Schonberg, 2010; and Clement et al., 2007). To our knowledge, however, this literature does not consider how experience's value may vary with the complexity of a job; our study of earnings growth across heterogeneous situations offers novel evidence on this issue. The well-established literature following Mincer (1974) offers additional evidence on the wage-experience relationship, but this literature's life-cycle perspective on experience generally abstracts away from the details of human capital formation within individual jobs.[2]

I. The New York City taxi market

New York City's yellow taxicabs serve customers who hail from the street, plus taxi queues at airports, train stations, and hotels. They are not permitted to accept customers in any other way. Other types of for-hire vehicles (e.g. town cars and limousines) serve the market for pre-arranged and radio-dispatched transportation. In the NYC yellow taxi market, therefore, modern technologies such as GPS do not have a direct impact on the taxi driver's fundamental problem of finding consumers whose demand for transportation has not yet been satisfied by another yellow cab or some other transportation mode.

[2] See Heckman et al. (2003) for a survey. Recent research in this area (e.g. Dustmann and Meghir 2005, Buchinsky et al. 2010, and Bagger et al. 2014) has largely focused on challenges related to endogenous labor force participation, job switching, and wage search.

During our sample period of 2009, there were 13,237 yellow cab taxi licenses ("medallions") and about 40,000 additional for-hire vehicles in New York City. Each medallion corresponds to a single yellow taxi, which may be controlled by an owner-operator, a firm with a fleet of cabs, or an authorized leasing agent. The taxis are operated by drivers licensed by New York's Taxi and Limousine Commission. The TLC reports that the number of licensed drivers was 48,521 at the end of 2009; not all licensed drivers operated a taxi during 2009.

Taxi drivers enter the market through a variety of avenues. Based on conversations we have had with taxi fleet managers, it appears that very few drivers are "seasonal" in the sense that they take multi-month breaks from driving a taxi before returning to the job. Schaller (2006) reports that, in 1991, 20% of TLC license applicants drove professionally in their previous job, with 44% of applicants having ever held a professional driving position (in New York or anywhere in the world). To obtain a TLC license, drivers are required to: hold a valid DMV license; attend either a 24-hour or 80-hour Taxicab School course; and pass tests on New York geography, TLC rules, and English language proficiency. New and experienced drivers are likely to differ on personal characteristics and performance. Schaller (2006) reports that experienced drivers receive fewer complaints for service problems such as refusing passengers, overcharging, treating passengers rudely or abusively, or driving unsafely. Driver experience is also negatively correlated with the number of accidents and traffic violations (Schneider, 2010).

Taxi earnings and costs are structured so that it is in the driver's interest to maximize fare earnings during a shift. Drivers keep all fares and tips. Fares accrue as a function of ride distance and duration, and may include surcharges for nighttime rides, peak weekday rides, and destinations at airports or outside of New York City's five boroughs. Drivers' costs vary depending on their ownership of the taxis they drive. All drivers pay for their own gas ($5,000-$10,000 per year), annual TLC fees ($100), and DMV/TLC fines for driving infractions. Drivers who lease their vehicles will pay a per-day or per-week flat fee; these fees were about $100 per day in 2009. Owner-operators pay for annual maintenance and repair ($4,000-$10,000/year), insurance ($7,000-$13,000/year), and licensing fees ($1,000/year). On average, a driver's take-home pay is 57% of revenue, with the rest divided among expenses paid by the driver and taxi owner (Schaller 2006). Driver earnings vary with the time of day and day of the week, so some shifts are more lucrative than others. Experienced drivers are generally sorted into the more valuable shifts; we document this pattern below. Our conversations with fleet managers reveal that this is largely due to seniority rules for the assignment of drivers to shifts.

II. Data

A. Sample Construction

The data we use in analysis are derived from a fare-by-fare database of yellow taxi activity from the full 2009 calendar year. These data are collected by the TLC as part of an effort to monitor the activity of taxis and their drivers in the NYC area. The database includes all fares for licensed NYC cabs, even if an endpoint of a fare occurs outside of the city's five boroughs.

We begin with a full database of 171 million observed fares received by 41,256 active drivers. Each observation includes a unique driver-specific index, the longitude, latitude, date, and time of the ride's pick-up and drop-off, and payments to the driver. Date and time are recorded to the second, and longitude and latitude are recorded through on-board GPS.[3] The total payment is broken

[3] Longitude and latitude are reported to one-foot precision, but tall buildings and other technical challenges are likely to distort these values somewhat. Small errors in location data will not affect our estimates, which use areas of several city blocks as the finest descriptors of taxi location.

down into the fare, surcharges, tip when the payment is via credit card, and MTA tax. The combination of driver identification codes and specific ride data allows us to construct a complete history of each driver's activity during 2009. We define a shift as a succession of rides with between-ride breaks no longer than 5 hours.[4] We then track each driver's total fare earnings per shift, the shift's duration, and running counts of the driver's number of drop-offs in each neighborhood. We also construct statistics within selected hours of a shift, including: fare earnings in the hour, slack time with no customer in the cab, ride duration, the number of fares collected, and average driving speed. Unfortunately we do not have additional information about the drivers' personal characteristics, employer's identity, contract terms, or costs.

We clean and organize the data in order to conduct analysis of new versus experienced drivers. We remove drivers and shifts that are unlikely to represent the production efforts of a regular driver in the market. In particular, we drop drivers who are associated with fewer than 100 fares, shifts associated with more than one unique car identifier, and shifts that were shorter than 2 hours or longer than 20 hours. This reduces the dataset to 165 million observations. Of the 6 million fare observations dropped in this step, 4 million are due to very long shifts, and 1.4 million are due to the use of multiple cars in a single shift.

Next, we separate drivers into groups by their level of experience. We identify 27,664 drivers who first appeared in the data on or before January 15, 2009. Of these drivers, we retain 22,608 who worked at least 100 shifts between January 1 and December 31, and additionally worked at least 30 shifts before April 1. These drivers are likely to have pre-2009 experience in the market, while also working with enough frequency to maintain their stock of knowledge. To maintain

[4] Farber (2015) uses the slightly different threshold of 6 hours to define a shift transition.

tractability of the data, we select a 20% random sample of 4,522 drivers as our "experienced" drivers in the analysis described below.

From the collection of drivers who fail the criteria for experienced drivers, we identify a subset who are likely to have entered the NYC taxi workforce in 2009 as new and inexperienced, and work with sufficient frequency to indicate an intention to function as a full-time taxi driver in the market. To identify these new drivers, we isolated the sample to 4,033 drivers who were first observed in the data on or after April 1, 2009. Among these drivers, we select the 1,771 who worked at least 50 shifts between their entry date and December 31, 2009, and had no more than five shifts that violated the criteria for inclusion (multiple cars, shift length) mentioned above. It is possible that some of these drivers have prior experience in the NYC taxi market, but we cannot measure the size of this effect in the data. To address the possibility that some drivers are seasonal workers, we perform additional analysis (discussed below) with stricter inclusion criteria; our results are largely unchanged. To complete the selection of drivers and shifts for analysis, we limit both the experienced and new driver samples to shifts that start between April 1 and December 31, 2009.

We use the longitude and latitude information in the fare database to identify the NYC geographic region in which each pick-up and drop-off occurs. We divide the market in two ways. First, we use the boundaries of Public Use Micro Areas (PUMAs) to identify 59 regions within the city's five boroughs. PUMAs represent a fairly coarse division of the city; for example one PUMA is defined by the portion of Manhattan west of Central Park and approximately bounded by the Park's north and south edges. Second, we use the boundaries of 220 Neighborhood Tabulation Areas (NTAs), which are mostly nested within PUMAs and correspond more closely to conventional views of NYC neighborhoods. For example, the Lincoln Square area is a distinct NTA within the Manhattan PUMA described above. We employ both geographic boundaries in order to study the impact of "local" experience in large and small areas.

B. Overall and per-hour driver activity

Our analysis sample includes 1,771 new drivers and 4,522 experienced drivers whose activity we observe between April 1 and December 31, 2009. We present summary statistics on these drivers' productivities in Table 1, separately reporting activity for experienced drivers, new drivers overall, and new drivers in their first 20 shifts. We examine several productivity variables, including average earnings per hour and earnings within a specified hour of the shift. We generally focus on the 60 minutes following a driver's third drop-off of a shift. For convenience we refer to this period as "hour-R3" in our text and tables. Hour-R3 roughly overlaps with a driver's second hour of work.[5] This hour is a microcosm of earning opportunities yet less likely to be affected by considerations about when to stop working or whether to take long mid-shift breaks, which may differ between new and experienced drivers.[6] This also allows us to look directly at the pick-up and drop-off times and locations of the driver's final fare starting immediately before hour-R3, plus the pick-up time and location of the next fare. In computing earnings in the hour following a drop-off, we omit observations in which a driver has no customers for 60+ minutes. We infer that the driver is on a break, and is not attempting to find new customers.[7] When hour-R3's final ride continues beyond the 60 minute cut-off, we calculate the fraction of the ride that took place in hour-R3 and attribute that fraction of the relevant fare to hour-R3.

[5] For the median shift this hour begins at minute 55 of the shift, and shifts at the 25th and 75th percentiles begin at minutes 40 and 81, respectively.

[6] In addition, focus on hour-R3 allows us to partially separate learning from drivers' habituation to driving conditions, such as sitting for many hours, which could be less important early in the shift.

[7] This rule affects 4% of our observations, and has no substantial effect on our results.

[Insert Table 1 Here]

The earnings statistics in Table 1 show that experienced and new drivers collect about the same value in fares per hour (approximately $26), but the median new driver in his first 20 shifts earns $1.39 less per hour (5.3%) than the median experienced driver. Similar differences emerge across drivers when we compare other measures of driver activity. Drivers in their first 20 shifts spend more time without a passenger. Average ride duration is greater for experienced drivers, despite these drivers taking more customers per hour. Average travel speed is about the same across drivers of all experience levels. Finally, new drivers work longer shifts, on average, compared to experienced drivers. Any differences displayed among drivers in this table, however, could be influenced by variation in market conditions when drivers of different experience levels are working.

In addition to differences in driving activity, in Panel B of Table 1 we display some differences in working practices by new and experienced drivers. New and experienced drivers work a similar proportion of days following their first appearance in the analysis sample, although new drivers are observed for fewer shifts because of how they enter the sample. Some notable differences exist in the number of cars associated with each driver. While our data do not provide information on drivers' relationship to cab owners, we observe that experienced drivers are over 200% more likely to be paired with a single taxi during the sample period. This suggests that the proportion of owner-operators and long-term lessees is larger in this population, which could provide experienced drivers with more flexibility in choosing work schedules and earning opportunities. The selection of experienced drivers into more lucrative schedules is evident in the data. Experienced drivers' average fare per hour is negatively correlated (-0.72) with the share of new drivers. Across all 168 day-of-week/hour-of-day combinations, those in the top 10% by new driver share have average experienced driver fares of $21.37 per hour; those in the bottom 10% have average fare earnings of $30.80 per hour.

C. Geographic data and location-specific driver activity

We use the geography data in three ways. First, we use a random sample of experienced drivers to construct statistics on driver performance within each NTA-hour combination within a week (i.e., for each region a separate measure is constructed for each of 7×24=168 unique hours in the week).[8] Using the fare-level data, we calculate each experienced driver's total earnings during the 60 minutes following a drop-off in a specified NTA-hour.[9] We then average these earnings across all drivers in the same NTA-hour, thereby computing a measure of how locations and times may vary in the earning opportunities available for drivers.[10] Some passenger-selected drop-off locations may take the driver to a part of the city where (at a given time of day) it is especially easy to find the next customer, or perhaps find customers who are likely to request rides to high-earning areas. With these measures of average earning within neighborhoods and times, we can characterize some situations as "easy" or "hard" production opportunities, and investigate how new driver performance varies with task difficulty.

In Figure 1 we provide an illustration of how earning opportunities vary across the city, and to what extent they vary across the hours following 8am, 3pm, and

[8] We use data from experienced drivers who are not in the main analysis sample.

[9] When a ride straddles the cut-off time of a particular hour, we calculate the fraction of the ride that occurred in the relevant hour, and then assign this fraction of the ride's total fare to the hour.

[10] The calculation omits rides that are not followed by 4 or more rides within the driver's shift (17.5% of observations) to reduce the influence of endogenous shift-ending behavior. We also omit drivers with earnings of zero in the hour, which coincides with our rule in the regression analysis. Finally, we drop all NTA-hours with fewer than 10 observations over the entire sample period, which corresponds to the bottom 10% of the distribution of NTA-hour observations.

10pm on Tuesdays, and 10pm on Saturday.[11] To construct the figure we first weight the NTA-hour earnings statistics by the number of experienced driver drop-offs within each NTA-hour, and then we split these values into quartiles.[12] The first quartile (lightest shading) represents the lowest average earnings ($20.08 on average) in the hour following a drop-off, and the fourth quartile (darkest shading) contains the highest average earnings ($32.62). Some NTAs (largely on Staten Island) are unshaded because fewer than 10 rides terminated in those neighborhoods during an hour of the week. Lower Manhattan is at its most lucrative around rush hours and on weekend evenings, while the Bronx contains many neighborhoods in the lowest-earning quartile. The eastern portions of Brooklyn and Queens have relatively high average earnings, likely because of their proximity to John F. Kennedy airport, from which the TLC mandated a flat fare of $45 for the most common destinations (i.e. drop-offs in Manhattan).

[Insert Figure 1 Here]

Our second use of geographic data is to construct measures of location specific experience by new drivers. For each new driver and shift we compute as d_{it} the driver's cumulative number of completed drop-offs between the new driver's first shift and ride 3 of shift t. In addition, we maintain running counts of all drop-offs prior to hour-R3 in each NTA and PUMA, which we index by n and p, respectively. Let d_{int} represent driver i's cumulative drop-offs in NTA n through shift t, and let d_{ipt} capture the drop-offs in PUMA p. When working with local drop-offs at the NTA level (e.g. Lincoln Square) we can calculate the number of drop-offs in the same PUMA but outside of the NTA (e.g. Upper West Side

[11] In order to minimize concerns about drivers favoring particular neighborhoods at the ends of their shifts, we select hours that are different from typical shift-end times.

[12] If locations with low mean earnings also have fewer drop-offs, the weighting will partition the NTA-hours so that more than 25% of unique NTA-hours are assigned to the lowest-earning quartile.

excluding Lincoln Square), plus separately the number of drop-offs outside of the NTA's PUMA (e.g. all of NYC outside of Upper West Side). We denote drop-offs in the same PUMA as n but outside of the NTA as $d_{i,-n,p,t}$, and we write $d_{i,-p,t}$ for the count of drop-offs outside of the PUMA in which ride 3's drop-off occurred. The count $d_{i,-p,t}$ (and to a lesser extent $d_{i,-n,p,t}$) stands-in for "general" experience in the sense that location-independent expertise should be developed equally well inside and outside of a drop-off neighborhood. On Table 2 we report some summary statistics on new drivers' cumulative drop-offs. Across all new drivers in the analysis sample, the average value of d_{ipt} is 1711 with 17% (7%) of drop-offs in the same PUMA (NTA) as ride 3's drop-off PUMA (NTA). These drop-off shares hold for new drivers in their first 20 shifts as well, and for these cases represent substantially smaller levels of prior local experience.

[Insert Table 2 Here]

Finally, we use the geography data to connect some individual rides' drop-off locations with the pick-up locations of the next customers. In particular, we create variables to indicate whether a new driver performs his ride 3 drop-off and subsequent pickup (the first new fare of hour-R3) in the same or different NTA or PUMA. While we do not observe search decisions directly, this provides some information on how widely drivers search for their next customer. In Table 2's we show that new drivers, on average, switch NTAs following half of all fares, and new drivers in their first 20 shifts are slightly more likely to switch.

III. Empirical analysis

We estimate a variety of econometric models to satisfy our research objectives. Our models measure how market outcomes (e.g. an hour's fares) vary across drivers with different amounts of experience. We perform our main analysis with a fairly simple econometric framework. Let y_{it} be driver i's productivity during shift t; in most cases this is (a function of) fare earnings during a specified hour of shift t. The new driver's vector of experience at shift t is E_{it}. We specify E_{it} to contain the driver's current shift number (e_{it}), his cumulative number of drop-offs (d_{it}) completed before ride 3 of shift t, and his location-specific numbers of drop-offs (e.g. d_{int}). These counts of experience begin with the new driver's first appearance in the 2009 data; similar variables cannot be constructed for experienced drivers. We measure the impact of E on y with the function $g(E;\theta)$, where θ is our main parameter vector of interest. Across specifications we vary the collection of experience measures included in g as well as g's functional form. In general we regard a new driver's shift number (e) as a better exogenous measure of experience than his count of drop-offs (d), as above-average performance leads directly to a faster accumulation of d. Some location-specific variables are measured through drop-offs, however, so we use d where necessary and after establishing that our main results are robust to either variable.

We introduce several controls for market-, location-, and driver-specific heterogeneity. Let the fixed effect α_{hn} represent the potential production (in terms of y) of an experienced driver during hour h, whose third fare of shift t had drop-off NTA n. In some cases we apply the same fixed effect, $\alpha_h = \alpha_{hn}$, to all locations in the city for the specific hour h. Our models generally include a different α_h value for each distinct hour in the dataset, i.e. one for each of 6,600 hours of our analysis sample between 12AM April 1 through 11PM December 31 2009. Models that include location-specific differences, i.e. α_{hn}, include a full interaction of NTA identifiers with each hour in the dataset. Whether specified as α_h or α_{hn}, the fixed effects account for demand and supply fluctuations over time that may influence all drivers' earnings during t. This may include regular variation in demand (e.g. rush hours), idiosyncratic variation in demand (e.g. weather), and seasonal effects. In addition to the α terms, in many models we include driver-level fixed effects, denoted δ_d, which allow individual drivers to have their own average output levels. In the case of new drivers, the value δ_d represents the individual-specific deviation from α_{hn} when the driver has no experience. In addition to α_{hn} and δ_d, we account for some observable driver and market characteristics in X_{it}. All of our specifications include in X_{it} the log of shift t's full duration in hours; this accounts for potential differences in drivers' effort exertion across long versus short shifts. We implement some additional analysis by adding variables to X_{it}. Finally, we include the error term ε_{it} to account for driver-shift level unobservables in production. Driver production and learning are likely to be correlated within driver over time, so we cluster ε at the driver level during estimation.

We combine the components described above into the econometric model:

$$y_{it} = g(\boldsymbol{E}_{it}, \boldsymbol{\theta}) + \boldsymbol{X}_{it}\beta + \alpha_{hn} + \delta_d + \varepsilon_{it} \quad (1)$$

Across our specifications below we employ a variety of assumptions on y, g, E, X, α, and δ; these variations are explained below as the models are introduced. In our final set of models, in which we describe drivers' choices when searching for their next fares, we use the same collection of explanatory variables and controls.

A. Overall measures of learning

We begin by focusing on driver earnings during hour-R3, and we estimate (1) with y specified as the log of total fare earnings during hour-R3 of a shift. We assess the impact of parametric assumptions in $g(E_{it};\theta)$ with an initial pair of models. In the first model, we assume $g(E_{it};\theta) = \theta\log(e_{it})$. In the second, we specify that e enters g with dummy variables in ten-shift intervals, each with its own coefficient:

$$(2) \qquad g(\boldsymbol{E}_{it};\boldsymbol{\theta}) = \gamma_1 1\{1 \le e_{it} \le 10\} + \gamma_2 1\{11 \le e_{it} \le 20\} + \gamma_3 1\{21 \le e_{it} \le 30\} + \cdots + \gamma_{14} 1\{131 \le e_{it} \le 140\}$$

The second model has new drivers with 141+ shifts as the excluded category. Both models include data from experienced drivers (who have $g = 0$), driver-level fixed effects (δ_i), and a full set of controls (α_{hn}) for each hour-location combination.

In Figure 2 we display results from the initial models, which show rapid productivity improvements among new drivers.[13] The more flexible approach to g shows that new drivers' earnings plateau after about forty shifts. The figure also shows that there is little difference between the two approaches' results, which provides reasonable support for moving forward with more concise log specification in the analysis below.

[Insert Figure 2 Here]

In our next analysis, we investigate the importance of sample selection issues on drivers' overall earnings. We estimate a series of models with experience captured as log(e) or log(d), and we report our results in Table 3. In specification 1 we include data on new drivers only, and without controls for date and time. Without these controls for market conditions, we find a large effect of experience on productivity. Specification 1's results imply a 21% difference in log earnings between the first and 100th shift, or a 23% difference in level earnings.[14]. In

[13] The parameter estimates from the log specification are reported in Table 3, specification 3. The estimates from the dummy variable version are in Online Appendix Table A2, specification 3.

[14] We compute the log earnings difference by multiplying the learning coefficient (0.045) by log(100). To obtain the difference in level earnings we compute exp[0.045×log(100)] – 1.

specification 2 we introduce fixed effects, α_{hn}, that interact the hour of the full dataset with the location of ride 3's drop-off, plus we add data on experienced drivers to provide additional information on demand conditions. The learning rate falls by half in specification 2, and the presence of experienced drivers provides an estimate of the production gap for brand new drivers (9%). In specification 3 we add driver-level fixed effects, δ_d, to measure purely within-driver improvements in y with experience.[15] The learning rate estimate from this specification is roughly 70% of the magnitude in specification 2, but within-driver learning remains statistically significant at $p < 0.01$. The learning coefficient of 0.015 implies that a driver in his 100th shift receives earnings that are 7% greater, on average, than a driver in his first shift. In specification 4 we replace e_{it} with d_{it}, the new driver's stock of drop-offs through ride 3 of shift t; our results are largely unchanged with this alternative experience measure.[16]

[Insert Table 3 Here]

We are able to use the results from Table 3 together with statistics on driver work patterns to calculate the dollar value of learning, i.e. the information cost of driver inexperience in the market. We focus on the first 100 shifts of driver activity, when most learning takes place. If we apply the slope coefficient in specification 3 to predicted earnings under the assumptions of an average shift length (9.19 hours) and per-hour fare earnings ($26.42) for new drivers with 100 shifts, we find that new drivers lose $344 in fare earnings during the first 100 shifts due to inexperience. For the approximately 4,000 new drivers who arrive in

[15] We obtain virtually the same results when including the simpler date-by-hour fixed effect (α_h) in place of α_{hn}. In some analysis below we use α_h instead of α_{hn} because of computational constraints.

[16] Our results are robust to alternative assumptions to identify new drivers. In the Online Appendix we report results from models in which we use a later cut-off date (May 1), and additional models in which we require new drivers to be active during every month after their first appearance in the data. The former approach reduces the chance that an experienced driver on an extended break will be misidentified as new, and the latter reduces the impact of driver attrition.

NYC after March 31, this totals $1.4 million in lost fare earnings due to inexperience in the market. The large learning coefficient of specification 1, by contrast, implies $1016 in lost fares per driver. The difference between the lost earnings values using specifications 1 and 3 suggests that new driver selection into unfavorable shifts reduces earnings by about $672.[17] While our preferred estimate of the overall reduction in fare earnings during the first 100 shifts is small (1.4% of fares), the absolute dollar difference is a larger proportion (about 2.5%) of a driver's take-home pay due to vehicle rental expenses and other fixed costs of operating a taxi.[18] In addition, compensation through tips, which are influenced by both fare earnings and the number of distinct rides, are likely to increase with the driver's experience.

Our estimates imply that learning occurs quickly (i.e. over a few months of shifts) and accounts for a fairly small share of driver earnings. The LBD literature as a whole contains a wide variety of learning rates, only some of which are similar to ours. Jovanovic and Nyarko (1995) estimate learning paths from a variety of industries and firms; some are quite similar to our results while others show greater increases in productivity. Levitt et al. (2013) find that learning occurs over a similar time interval as we do, although the productivity differences they measure are much larger. Hendel and Spiegel (2014) study a case in which firm-level learning occurs continually over many years. The variety of activities observed could be a source of this heterogeneity. Moreover, LBD at the organizational level could offer opportunities for more sustained improvements through changes to communication and coordination among agents.

[17] The results of Table 3's specification 1 are unchanged if we add driver fixed effects.

[18] We calculate the total change using the relevant statistic on take-home pay provided in Section II.

B. Driving activities and outcomes

New drivers' earning improvements are robust in the data, but they leave open the question of what drivers might be doing to generate these improvements. While we do not directly observe all relevant choices by drivers, we can analyze how several outcomes vary with experience. To perform this analysis, we adapt the empirical model in (1) so y represents these additional measures of outcomes and activities.

In specifications 1 and 2 of Table 4 we track improvements in a critical driver activity: reducing the amount of time between customers. The specifications differ in whether driver-level fixed effects are included. While specification 1, which lacks these controls, may be affected by driver attrition, it provides a useful intercept term that describes differences between brand-new drivers and experienced ones. In both specifications we find that driver experience, as captured by cumulative shifts (e), has a significant effect on reducing slack time between passengers. Specification 1's estimated intercept term reveals that brand new drivers spend about 10% more time searching for their next customer. At the experienced driver mean of 13.7 minutes between customers, this implies an additional 1.4 minutes of search time. As in the analysis above, controlling for driver-specific heterogeneity reduces the impact of experience on driver performance, but the coefficient on driver experience remains statistically significant. We interpret this result (and the others in this paper) on improved performance as due to a driver's independent observation of taxi demand in various NYC neighborhoods. An additional mechanism, which we cannot measure with the present data, is that a club of experienced drivers gradually releases information about (or permission to serve) lucrative locations to new drivers who gain tenure in the market. While possible, our results in total suggest that this mechanism is likely to be less important than independent learning.

[Insert Table 4 Here]

Other aspects of drivers' activity change with experience as well. In specification 3 we report that newer drivers have lower-mileage trips, on average, which may be due to the areas in which they search for customers. While our fixed effects in α_{hn} control for the location from which the driver begins searching the next fare, experienced drivers may make different choices about where to search or which customers to pick up following a drop-off. In our final analyses on Table 4 we investigate whether experienced drivers are able to pilot their vehicles more quickly while transporting passengers. (We do not observe miles travelled when the driver has no passenger, so we cannot compute speed at other times.) We find in specification 4 that drivers with more experience are able to travel more quickly. This could be due to the types of fares they take, which may require travel on higher-speed roads, or it could be due to expertise in navigating city streets. We attempt to account for differences in trip composition in specification 5, where we add a control for the average distance of a driver's fare during an hour. Adding this variable, which can account for longer trips on higher-speed roads, only slightly diminishes the impact of experience on speed, with the coefficient remaining positive and statistically significant.

Table 4's analysis leaves open the question of which types of actions or outcomes are most important across drivers, and thus between new and experienced drivers. We address this question in a series of models in which we consider how much hour-R3 earnings variance is explained by the share of the hour spent without a passenger, average ride distance, and average driving speed. (We do not report the models' coefficient estimates because they are unimportant to our analysis.) As a baseline, we estimate a version of equation (1) that contains no experience term (g = 0) but retains the fixed effects α_{hn} and δ_i. As captured through R^2, this model explains 36% of variation in hour-R3 earnings. When we add to the model (in *X*) each driver's share of hour-R3 without a passenger, the new specification explains 74% of earnings variation. The other variables (average distance and speed), by contrast, improve the explained variation by only 6 or 7 percentage points each. When all three variables are included, the model accounts for 84% of the variation in earnings. While the driver's realized time spent without a passenger is affected by a variety of factors (including strategy and luck), these models demonstrate that finding passengers is the critical activity that affects drivers' earnings.

C. Customer-selected destinations as a randomizing device

Some of our empirical analysis below relies on an assumption that drivers are randomized into locations across the city based on the requested destinations of their customers. While drivers can control where they pick up customers, they do not determine the drop-off location.[19] Contrary to our assumption, if drivers gain expertise in selecting customers who are likely to request rides to more lucrative drop-off locations, then some of their measured improvements in fares will be due to the ability to avoid arriving in difficult situations rather than performing well within them. This would be a different sort of learning than interests us here.

We evaluate our randomization assumption by investigating whether new and experienced drivers are different in how lucrative their drop-off locations are. As detailed in Section II.C, we measure the value of an NTA within a particular hour of the week by calculating the mean earnings across experienced drivers for the 60 minutes following a drop-off in that NTA-hour combination; we use the log of this measure to characterize ride-3 destinations. Continuing with our analysis

[19] While drivers may refuse fares occasionally in practice, doing so is against TLC regulations and can result in punishment. 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.

sample of new and experienced drivers, we regress this logged mean earnings variable on measures of experience, captured here with a dummy variable for new drivers and the new drivers' log(e). We report our results on Table 5. Specification 1 includes data from both new and experienced drivers, but no fixed effects at the date-by-hour (α_h) or date-by-hour-by-location (α_{hn}) level. We report that new drivers are transported to significantly lower-earning destinations, and their outcomes improve as they gain experience. This pattern, however, is due to sorting of new drivers into low-earning shifts. In specification 2 we add date-by-hour fixed effects (α_h) to the model, and the estimated coefficients are reduced by over a factor of 20. The coefficients in specification 2 show no economically significant difference in drop-off locations between new and experienced drivers. This result is sustained in specification 3, where we add driver fixed effects to focus on within-driver improvements. Finally, in specification 4 we retain driver fixed effects and replace α_h with α_{hn}, which in this case has l defined as the *starting* NTA of the driver's third ride of shift t. This allows us to ask whether drivers gain expertise in selecting customers from within a given neighborhood; we find no significant effect of driver experience on destination mean earnings. We conclude that drivers do not improve in their ability to identify customers who will take them to high-earning neighborhoods. The improvement, instead, appears to be primarily in locating the next customer following a drop-off. To the extent that drivers tend to refuse customers who may appear to request rides to low-earning areas, this behavior is not correlated with driver experience in our data.

[Insert Table 5 Here]

D. Productivity, learning and difficulty

One strength of our data is that we can examine how new drivers' production improves across a variety of situations. This allows us to draw conclusions about the impact of experience on wages and productivity across different economic conditions. In the broader economy, this would be analogous to estimating different returns to experience across different macroeconomic conditions or circumstances facing a firm. In studies of worker wage dynamics, a similar treatment may be given to estimating different returns to experience across different occupations or tasks, which can differ in their rigor.

For this analysis, we return to the NTA-hour (of week) quartiles that we constructed for Figure 1, including the weighting by number of drop-offs. Difficult (first quartile) NTA-hour combinations will have relatively low earnings in the hour following a drop-off. We estimate a regression model that allows new drivers to have different expected earnings at the start of their career in each difficulty quartile, experience different rates of improvement in each quartile, and allows experienced drivers to vary in their average earnings across difficulty quartiles as well. The base model is:

$$\text{(3)} \qquad \log(\text{hour-R3 fare}_{it}) = \sum_{q=1}^{4}\left(1\{Q_{it} = q\}\times 1\{i \text{ new}\}\times\left[\theta_{0q} + \theta_{1q}\log(e_{it})\right]\right) + \sum_{q=1}^{3} 1\{Q_{it} = q\}\,\mu_q + \boldsymbol{X}_{it}\beta + \alpha_{hn} + \delta_d + \varepsilon_{it}$$

The variable Q_{it} contains the difficulty quartile (q) in which driver i drops-off his third customer, starting hour-R3. $1\{.\}$ is the indicator function. New drivers have a different intercept (θ_{0q}) and learning coefficient (θ_{1q}) for each possible difficulty quartile. These parameters capture, respectively, the productivity lag of brand new drivers across different quartiles and the rate at which new drivers' production improves across quartiles as they accumulate experience (e). Experienced drivers in the bottom three quartiles have earnings that differ by μ_q from the (excluded) earnings of experienced drivers in the easiest quartile. We can estimate the μ_q parameters when we restrict $\alpha_{hn} = \alpha_h$ across all locations, but these coefficients drop out in our more general treatment of α_{hn}. The control variables (X), driver fixed effects (δ), and error term (ε) are all defined as in the models discussed above.

We report the results of this analysis in Table 6. In specification 1 we omit driver fixed effects and restrict $\alpha_{hn} = \alpha_h$. We find that driver productivity, as measured through hour-R3 earnings (log earnings), is 14% (15%) lower for brand new drivers operating in the most difficult quartile than it is for experienced drivers in the same quartile. Experienced drivers in the most difficult quartile, in turn, have earnings (log earnings) that are 31% (38%) below those of experienced drivers in the easiest quartile. New drivers in the middle two quartiles start their driving careers 11% and 8% below the productivity of experienced drivers, while drivers in the easiest quartile are just 2% below experienced drivers. While new drivers perform substantially worse than experienced drivers in difficult situations, their performance improves more quickly in more difficult quartiles.

[Insert Table 6 Here]

We extend our analysis in specification 2, where include fixed effects at the location-hour level, α_{hn}. We also add driver-level fixed effects, which require us to remove one intercept parameter for new drivers, so we normalize new driver performance relative to those who serve the easiest quartile. The qualitative results of specification 1 are preserved in specification 2. New drivers lag behind experienced drivers by substantially more in difficult settings, but new drivers also improve relatively quickly in these settings. If we assume a 10.5% earnings gap between brand new and experienced drivers in the most difficult quartile, the learning rate θ_{01} in specification 2 implies that new drivers require 67 shifts to eliminate the difference between themselves and the incumbents.[20] Drivers in the second-easiest quartile require 55 shifts to eliminate a gap of 6%.

E. General and specific experience

Our next set of analyses employs separate measures of driver experience across locations. In the NYC taxi market, we are able to track how a new driver's performance in a specific neighborhood is related to the driver's prior experience in the same neighborhood and his overall experience. This exercise relates to a large labor literature (e.g. Becker, 1964) that examines the degree to which human capital acquired on the job is general or more specific, and therefore less transferrable across jobs. While the bulk of this literature focuses on general vs. occupation-, industry-, or firm-specific human capital, there is also interest in the degree to which human capital gains are *task*-specific (Gibbons & Waldman 2004).[21] Our data provides a unique opportunity to add to this literature using a large dataset, precise measures of *task*-specific (i.e. *location*-specific) experience, and a simple identification argument. Specifically, for each driver we observe a large number of randomizing events in which customers ask for drop-offs in a variety of neighborhoods, which themselves vary in their earning opportunities and (perhaps) optimal strategies for finding customers.

We perform this analysis using functions of new drivers' drop-off tallies, i.e. d_{int}, d_{ipt}, $d_{i,-n,p,t}$, and $d_{i,-p,t}$. We vary whether PUMA or NTA is the finest

[20] We assume a 10.5% earnings gap because this is the sum of the first-to-fourth quartile gap for brand new drivers in specification 2 (-0.083), plus the fourth quartile gap between new and experienced drivers in specification 1 (−0.022). To calculate the number of shifts (e) before new and experienced drivers converge in hardest-quartile earnings, we solve -0.105 + 0.025×log(e) = 0.

[21] For example, Ost (2013) studies teacher grade switches to evaluate whether productivity gains are specific to a *task* (e.g. years teaching the 5th vs. 7th grade) or more general (years teaching overall).

geographic area for which we include a local experience measure. Throughout this analysis we allow each experience measure, d, to enter $g(E; \theta)$ as $\theta_j \log(1 + d)$, where j indexes an individual entry in θ. In all models we include the fixed effects α_{hn} and δ_i. In these models the location-specific fixed effect helps control for the possibility that drivers accumulate more experience in neighborhoods that have greater demand, which could create positive correlation between d_{int} and earnings due to forces unconnected to learning.

We report in Table 7 our results on local and general experience. Our first specification adapts Table 3's specification 4 by including d_{ipt} and $d_{i,-p,t}$ to account for PUMA-level experience. The results indicate that both types of experience are significantly correlated with fare earnings and have similar impacts. A driver in his 100th shift averages about 382 drop-offs in a PUMA, and the coefficient estimate of 0.007 suggests that these drop-offs lead to earnings that are 4.2% greater than those of a driver with no drop-offs in a PUMA. In specification 2 we narrow our focus by including the new driver's (logged) number of previous drop-offs in ride 3's destination NTA (d_{int}), the number in the same PUMA but in different NTAs ($d_{i,-n,p,t}$), and the number outside of the NTA's PUMA ($d_{i,-p,t}$). We find that NTA-level experience has a significant impact on fare earnings, as does experience outside of the drop-off PUMA. The number of drop-offs in the same PUMA but different NTAs has an insignificant impact on earnings after ride 3.

[Insert Table 7 Here]

Table 7's specifications 1-2 imply that non-local experience has a significant impact on earnings, which might be due to the benefits of overall experience. An additional explanation, however, is that earnings for the full hour-R3 include outcomes following a driver's fourth (and later) drop-offs, and these are often in neighborhoods other than ride 3's destination, and therefore the driver could be benefiting from knowledge about different specific locations. We address this possibility by considering in specification 3 the driver's time without a passenger after ride 3, using the same logged duration variable as in Table 4. This analysis places a tighter focus on drivers' outcomes immediately after ride 3, where performance following later rides does not affect the dependent variable. In specification 3 we find that the driver's slack time falls as prior experience in the drop-off PUMA increases, but experience outside of the PUMA has no effect on slack time. Specification 4 repeats the analysis of specification 2, and we find that drivers locate passengers faster when they have more experience in the same NTA. Experience outside of the drop-off NTA, however, has no significant effect on the duration until the driver finds his next passenger. While these results suggest that local experience may be all that matters for taxi drivers, we note that drivers' objectives are likely to be more closely tied to maximizing fare earnings than to reducing time between a pair of rides. In total, Table 7's results consistently indicate the importance of local experience in new drivers' productivity growth.

F. Search decisions and neighborhood-switching

While we cannot study the full details of drivers' decisions with the present data, we are able to analyze whether a driver has switched neighborhoods to find his next fare. We take this outcome as a reasonable proxy for the driver's choices about where to search. For this analysis we continue in the same vein described above, replacing y in (1) with an indicator variable for whether a driver's ride-4 pick-up is in a different NTA than his ride-3 drop-off. We use linear probability models to estimate this outcome. To tie switching decisions to drivers' experience, we continue with the location-specific measures of experience that we employ in Table 7: d_{int}, $d_{i,-n,p,t}$, and $d_{i,-p,t}$. We include driver-level (δ) and location-hour (α_{hn}) fixed effects in each model. The data on experienced drivers serves to provide average switching frequencies for individual location-hour combinations from which new drivers might deviate.

Our results, which we report in Table 8, show that drivers switch locations more frequently as they gain overall experience in the market, and they are less likely to leave individual neighborhoods in which they have more experience. In specification 1 of Table 8 we report results from the full analysis sample, and we find that overall experience (outside of the drop-off's PUMA), within-PUMA experience, and NTA-level experience each have a significant effect on whether the driver changes neighborhoods following a drop-off. The largest effect is due to local experience, which suggests that drivers are more inclined to search for their next customers in areas where they have previously spent more time observing the market. The positive coefficients on $d_{i,-n,p,t}$ and $d_{i,-p,t}$ suggest that drivers may be more willing to leave a neighborhood, holding local experience constant, if they have more experience elsewhere. These results are contrary to what we would expect if strategic experimentation incentives dominated; if so, new drivers would be more likely to switch NTAs when their local experience is relatively large.

[Insert Table 8 Here]

One additional concern is that our results are influenced by drivers having personal favorite neighborhoods that they favor, and the drivers actively seek opportunities to stay within these neighborhoods, thereby creating correlation between switching activity and experience. While we cannot address the question of driver preferences directly, we show that the same relationship between local experience and switching appears across different NYC locations. In specification 2 we focus on activity following drop-off activity in lower Manhattan (i.e., PUMAs below Central Park's northern border, where most rides terminate), while in specification 3 we limit the analysis to all PUMAs except those in lower Manhattan. Finally, in specification 4 we extend specification 3 by excluding observations from the NTA (outside of lower Manhattan) where a new driver has his greatest number of drop-offs over the full sample period. This removes the possibility of a personal "favorite" NTA in which a driver prefers to focus his effort.[22] Across the final three subsamples we find very similar effects of NTA-level experience on switching outcomes. While the positive relationship between overall experience and switching persists in specifications 2-4, experience outside of an NTA but within the same PUMA is negatively correlated with switching behavior. The difference between specifications 1 and 2, which largely use the same sample, may be due to collinearity among the drop-off measures, while the significantly negative coefficients on $d_{i,-p,t}$ in specifications 3 and 4 may be due to relationships among nearby NTAs outside of Manhattan, where between-NTA differences in demand may be substantially different than they are in higher-traffic areas. We leave the issue of drivers' full search strategies to future work.

IV. Discussion and conclusions

We have described learning patterns by New York City yellow taxi drivers while controlling for a wide variety of potential factors that can confound empirical studies of learning by doing. In our preferred model, we find that an individual new driver's total earnings would be about $344 greater, on average, if he were able to skip directly to the productivity of a driver with 100 shifts of experience. The difference between this figure and the one we obtain ($1016)

[22] The results are unchanged if we omit the driver's top 5 NTAs outside of lower Manhattan or his most frequent one or two PUMAs. We focus on areas outside of lower Manhattan because when the full city is included it appears that all drivers' favorite neighborhoods are near Midtown.

with a simpler approach suggests caution in inferring learning's value based on direct comparisons of new versus experienced workers' wages. Policies to train workers may fall short of expectations if wage gaps are driven by selection or seniority-favoring practices.

In addition to estimating the (relatively small) impact of overall learning, we provide evidence on driver performance across situations that differ in difficulty for experienced drivers. The difference between new and experienced drivers' performance is greatest in the most difficult situations, although in these settings the new drivers' performance improves relatively quickly. While other occupations are likely to vary in the details of where performance differences are greatest between new and experienced workers, our analysis suggests a novel type of heterogeneity to investigate when targeting training resources toward to the situations where the payoffs are greatest.

Finally, we use location-specific experience measures to obtain two findings. We find that neighborhood-specific experience is important in improvements to driver productivity, and that local experience has a significant impact on drivers' search strategies following drop-offs. The importance of local human capital suggests that, in this market, efforts to create general (i.e., market-wide) human capital could have very limited benefits. While a new driver's time between passengers, a critical determinant of earnings, is smaller when the driver's history includes more drop-offs in the same neighborhood, this measure of driver performance is unaffected by the driver's total city-wide experience.

We present our results with two main contributions relative to the prior literature. First, the taxi drivers in our study are able to choose their own work strategies, and through fare-based compensation they are rewarded directly for strategies that are more successful. This is novel relative to the manufacturing settings that are commonly studied in the LBD literature, where workers' actions may be relatively constrained by production line conventions and rigid pay practices. Second, the TLC's fare-level data allow us the opportunity to examine a large number of economic agents moving though a precisely documented environment. The large population allows us to benchmark new drivers' production relative to experienced drivers' contemporaneous efforts, which is often impossible in studies that examine single firms or one-time production processes. The location- and time-specific data on each fare allow us to quantify both task difficulty and keep track of individual drivers' neighborhood-specific and total experience.

Our data and analysis leave open several important questions about learning in general and performance by NYC taxi drivers specifically. First, we do not observe whether drivers selectively refuse fares to certain neighborhoods. While this is prohibited by the TLC and can result in the loss of a hack license, we cannot gauge the frequency of this activity and its correlation with driver experience. Second, the absence of driver-characteristic data prevents us from describing which types of drivers learn most quickly, and whether learning is affected by the driver's social circle or the organizational arrangements of the medallion owner. Third, our analysis prevents us from characterizing fully the precise mechanisms of driver learning or their welfare benefits to both drivers and consumers. Finally, additional analysis is needed to assess how strongly our empirical arguments (on selection, etc.) and results apply in settings outside of the NYC taxi market.

REFERENCES

Bagger, Jesper, Francois Fontaine, Fabien Postel-Vinay, and Jean-Marc Robin. 2014. "Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics." *American Economic Review* 104: 1551-96.

Balasubramanian, Natarajan and Marvin B. Lieberman. 2010. "Industry Leaning Environments and the Heterogeneity of Firm Performance," *Strategic Management Journal* 31: 390-412.

Becker, Gary. 1964. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. University of Chicago Press.

Buchinsky, Moshe, Denis Fougere, Francis Kramarz, and Rusty Tchernis. 2010. "Interfirm Mobility, Wages, and the Returns to Seniority and Experience in the United States." *Review of Economic Studies* 77: 972-1001.

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler. 1997 "Labor Supply of New York City Cabdrivers: One Day at a Time." *Quarterly Journal of Economics* 112: 407–441.

Crawford, Vincent P. and Juanjuan Meng. 2011. "New York City Cab Drivers Labor Supply Revisited: Reference-Dependent Preference with Rational Expectations Targets for Hours and Income." *American Economic Review* 101: 1912-1932.

Clement, Michael B., Lisa Koonce, and Thomas J. Lopez. 2007. "The Roles of Task-specific Forecasting Experience and Innate Ability in Understanding Analyst Forecasting Performance." *Journal of Accounting and Economics* 44 (3): 378–398.

Doran, Kirk. 2014. "Are Long-Term Wage Elasticities of Labor Supply More Negative than Short-Term Ones?" *Economic Letters* 122: 208-210.

Dustmann, Christian and Costas Meghir. 2005. "Wages, Experience, and Seniority." *Review of Economic Studies* 72: 77-108.

Farber, Henry S. 2005. "Is Tomorrow Another Day? The Labor Supply of New York City Cab Drivers." *Journal of Political Economy* 113: 46-82.

______. 2008. “Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers.” *American Economic Review* 98: 1069-1082.

______. 2015. “Why You Can’t Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers.” *The Quarterly Journal of Economics* 130 (4): 1975-2026.

Gathman, Christina and Uta Schonberg. 2010. “How General is Human Capital? A Task-Based Approach.” *Journal of Labor Economics* 28 (1): 1-49.

Gibbons, Robert and Michael Waldman. 2004. “Task-Specific Human Capital.” *American Economic Review Papers and Proceedings* 94: 203-207

Haggag, Kareem and Giovanni Paci. 2014. “Default Tips.” *American Economic Journal: Applied Economics* 6(3): 1-19.

Hatch, Nile and David Mowry. 1998. “Process Innovation and Learning by Doing in Semiconductor Manufacturing.” *Management Science* 44 (11): 1461-1477.

Heckman, James J., Lance J. Lochner, and Petra E. Todd. 2003. “Fifty Years of Mincer Earnings Regressions.” NBER working paper 9732.

Hendel, Igal and Yossi Spiegel. 2014. “Small Steps for Workers, a Giant Leap for Productivity.” *American Economic Journal: Applied Economics* 6 (1): 73-90.

Jackson, C. Kirabo and Henry S. Schneider. 2011. "Do Social Connections Reduce Moral Hazard? Evidence from the New York City Taxi Industry." *American Economic Journal: Applied Economics* 3(3): 244-267.

Jovanovic, Boyan, and Yaw Nyarko. 1995. “A Bayesian Learning Model Fitted to a Variety of Empirical Learning Curves.” *Brookings Papers on Economic Activity. Microeconomics* 1995: 247-305.

Kellogg, Ryan. 2011. “Learning by Drilling: Interfirm Learning and Relationship Persistence in the Texas Oilpatch.” *The Quarterly Journal of Economics* 126 (4): 1961–2004.

Lafontaine, Francine and Kathryn Shaw. 2013. “Serial Entrepreneurship: Learning by Doing.” NBER working paper 20312.

Levitt, Steven D., John List, and Chad Syverson. 2013. "Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant." *Journal of Political Economy* 121 (4): 643–681.

Mas, Alexandre. 2008. "Labour Unrest and the Quality of Production: Evidence from the Construction Equipment Resale Market." *Review of Economic Studies* 75 (1): 229-258.

Mincer, Jacob. 1974. *Schooling, Experience, and Earnings*, New York: NBER Press.

Ost, Ben. 2014. "How Do Teachers Improve? The Relative Importance of Specific and General Human Capital." *American Economic Journal: Applied Economics* 6 (2): 127-51.

Pisano, Gary P, Richard M.J. Bohmer, and Amy C. Edmondson. 2001. "Organizational Differences in Rates of Learning: Evidence from the Adoption of Minimally Invasive Cardiac Surgery." *Management Science* 47 (6): 752–768

Schaller, Bruce. 2006. "The New York City Taxicab Fact Book." Brooklyn, NY: Schaller Consulting.

Schneider, Henry S. 2010. "Moral Hazard in Leasing Contracts: Evidence from the New York City Taxi Industry." *Journal of Law and Economics* 53 (4): 783-805.

Shaw, Kathryn and Edward Lazear. 2008. "Tenure and Output." *Labour Economics* 15: 710-724.

Stith, Sarah S. 2013. "Organizational Learning-by-Doing: Evidence from Liver Transplantation." University of New Mexico working paper.

Rocha, Vera, Anabela Carneiro, and Celeste Amorim Varum. 2013. "Serial Entrepreneurship, Learning by Doing, and Self-Selection." CEF.UP working paper.

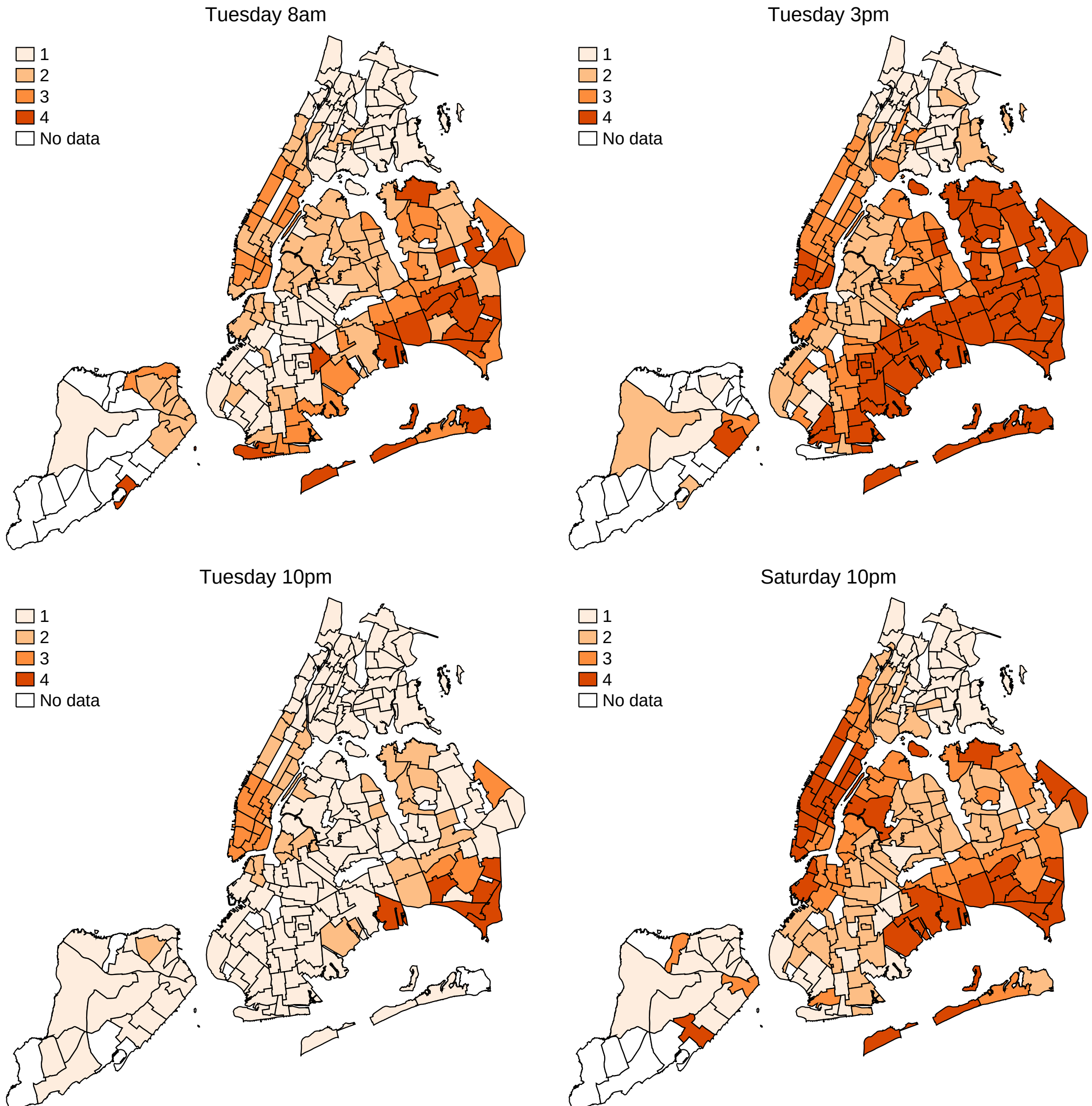
Rockart, Scott and Nilanjana Dutt. 2015. "The Rate and Potential of Capability Development Trajectories." *Strategic Management Journal* 36 (1): 53-75.

Thompson, Peter. 2001. "How Much Did the Liberty Shipbuilders Learn? New Evidence for an Old Case Study." *Journal of Political Economy* 109 (1): 103–137.

———. 2010. "Learning by Doing." Chapter 11 in the *Handbook of Economics of Technical Change*, edited by Bronwyn H. Hall and Nathan Rosenberg, North-Holland/Elsevier.

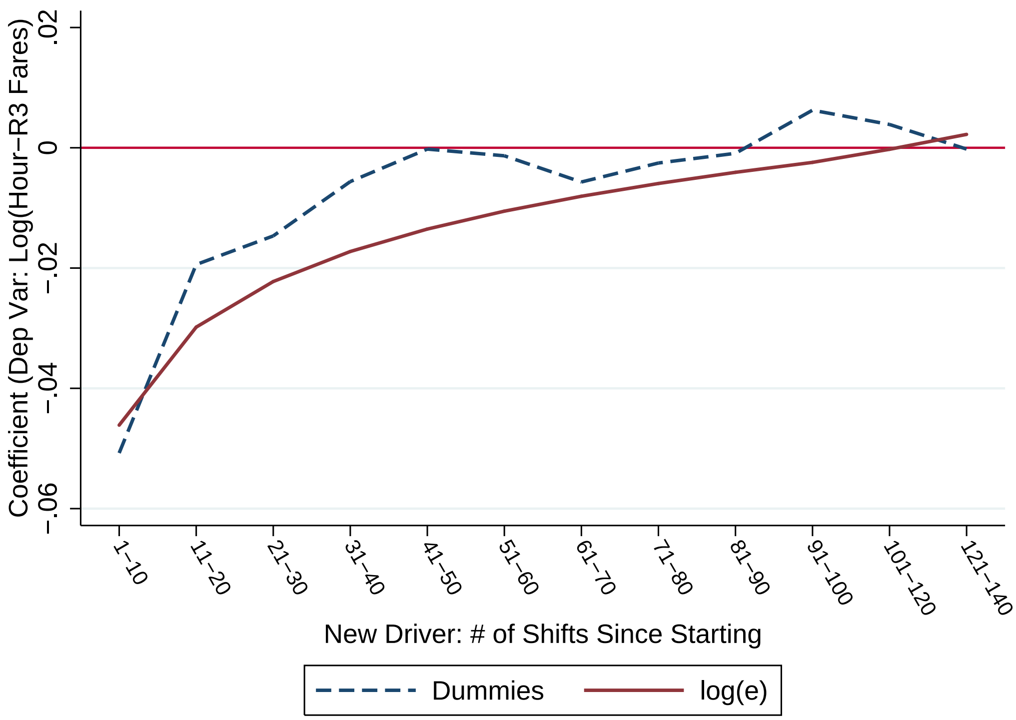
———. 2012. "The Relationship between Unit Cost and Cumulative Quantity and the Evidence for Organizational Learning-by-Doing." *The Journal of Economic Perspectives* 26 (3): 203–224.

Figure 1: NTA/DOW/HH-Level Difficulty Measures by Location of Drop-Off



Notes: Sample corresponds to 80% of Experienced Drivers (January 1 - December 31, 2009). Difficulty is defined by an NTA/Day-of-the-Week/Hour-of-the-Day-specific measure of the total fare earnings in the 60 minutes following a drop-off. This variable is split into quartiles, weighted by the number of rides. The fourth quartile corresponds to the "easiest" quartile (i.e. highest average fare). 8am corresponds to the hour falling between 8am and 9am (and similarly for the other reported hours). The maps exclude five NTAs that straddle geographic boundaries and are difficult to render in the figure.

Figure 2: Evaluation of parameteric (log) vs non-parametric (dummy) approaches



Notes: The solid line corresponds to Table 3, Specification 3 and uses a natural log specification for g(E). We impose an intercept value of -.07 on the solid line in order to match the intercept of the dashed line; the intercept value is not identified due to the inclusion of driver fixed effects. The dashed line corresponds to Appendix Table 2, Specification 4 and uses a collection of dummy variables for g(E).

Table 1: Summary Statistics

	Experienced Drivers			New Drivers			New Drivers, Shift≤20		
	N=839,888			N=213,051			N=34,420		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Panel A: Shift-Level Summary Statistics									
Avg. Hourly Fare Earnings in Shift	25.87	26.00	5.63	25.84	25.86	5.05	24.56	24.61	5.13
Fare Earnings in Hour-R3	24.94	26.13	11.41	25.15	26.13	11.14	23.66	24.30	11.49
Percent of Hour-R3 Empty	48.85	46.53	22.99	48.87	46.67	22.78	52.10	50.00	22.90
Empty Time (Min.) after Ride 3	13.74	5.43	25.80	12.94	6.00	21.84	14.30	7.00	22.51
Fares (Dropoffs) in Hour-R3	2.79	3.00	1.40	2.83	3.00	1.36	2.68	2.83	1.34
Avg. Ride Duration (Min.) in Hour-R3	12.55	10.71	8.38	12.34	10.62	7.90	12.20	10.45	8.47
Speed Per Ride (MPH) in Hour-R3	12.75	11.32	17.89	12.60	11.25	9.43	12.64	11.32	8.15
Shift Duration (Hrs.)	8.99	9.00	2.75	9.32	9.48	2.20	9.49	9.68	2.15
Panel B: Driver-Level Summary Statistics									
	N=4,522			N=1,771					
Driver Shifts	185.73	194.00	52.04	120.30	111.00	51.58			
Share of Days Active	0.71	0.74	0.17	0.70	0.72	0.20			
Number of Cars Per Driver	9.62	2.00	17.86	18.52	10.00	19.71			
Driver in Single Car? (Y=1)	0.49	0.00	0.50	0.15	0.00	0.36			

Notes: Sample contains shifts between April 1, 2009 and December 31, 2009 (Full Sample of New Drivers, 20% Sample of Experienced Drivers). N in Panel A corresponds to the main fare variable (Fare earnings in hour-R3); some counts are smaller due to missing data.

Table 2: Summary Statistics (Continued)

	New Drivers				New Drivers, Shift≤20			
	N	Mean	Median	SD	N	Mean	Median	SD
Drop-offs prior to ride 3 of shift t	212,918	1711.08	1420.00	1304.93	34,391	220.03	210.00	143.96
Prior drop-offs in same PUMA as ride 3	210,444	289.65	191.00	317.39	34,037	37.83	27.00	38.18
Prior drop-offs in same NTA as ride 3	208,953	127.75	69.00	171.33	33,819	16.65	9.00	20.78
Switch NTA after ride 3? (Y=1)	208,428	0.50	0.00	0.50	33,739	0.53	1.00	0.50

Notes: Sample contains new drivers' shifts between April 1, 2009 and December 31, 2009.

Table 3: Driver Experience and Productivity Improvements

Specification:	1	2	3	4
Dependent Variable:	log(Hour-R3 Fares)	log(Hour-R3 Fares)	log(Hour-R3 Fares)	log(Hour-R3 Fares)
New Driver		-0.085		
		(0.007)		
New× log(Shift)	0.045	0.021	0.015	
	(0.002)	(0.002)	(0.001)	
New× log(Drop-offs)				0.014
				(0.001)
Constant	2.968	3.201		
	(0.022)	(0.007)		
Includes Experienced Drivers?	N	Y	Y	Y
Date X Hour X Drop-Off NTA Fixed Effects?	N	Y	Y	Y
Driver Fixed Effects?	N	N	Y	Y
N	204,406	977,947	908,042	908,042
R^2	0.008	0.397	0.357	0.357

Notes: Robust standard errors, clustered at the driver level, in parentheses. "Hour-R3 Fares" is the sum of all fares within the first 60 minutes following the 3rd drop-off of a shift (including fractional fares for those rides that straddle the end of the hour). The variable log(Shifts) corresponds to the logarithm of the shift count (where shift=1 for the first shift observed for the driver in 2009); the variable log(Drop-offs) corresponds to the total number of drop-offs observed for the driver in 2009, prior to the 3rd drop-off of the current shift. All models control for the logarithm of the current shift's total length (in hours).

Table 4: Changes to Activities and Outcomes

Specification:	1	2	3	4	5
Dependent Variable:	log(Empty Time Between Fares)	log(Empty Time Between Fares)	log(Hour-R3 Avg Miles Per Ride)	log(Hour-R3 Avg Speed)	log(Hour-R3 Avg Speed)
New Driver	0.101				
	(0.012)				
New$\times$ log(Shift)	-0.025	-0.016	0.007	0.013	0.011
	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)
log(Hour-R3 Avg Miles Per Ride)					0.390
					(0.002)
Date X Hour X Drop-Off NTA Fixed Effects?	Y	Y	Y	Y	Y
Driver Fixed Effects?	N	Y	Y	Y	Y
N	1,031,633	953,870	911,793	911,296	911,296
R^2	0.435	0.378	0.219	0.412	0.691

Notes: Robust standard errors, clustered at the driver level, in parentheses. "Empty Time Between Fares" is the number of minutes between the 3rd drop-off and the 4th pick-up. "Hour-R3 Avg Miles Per Ride" is the number of miles travelled with passengers divided by the number of fares within the hour. "Hour-R3 Avg Speed" is total miles with passengers divided by the number of minutes with passengers in the taxi. The variable log(Shifts) corresponds to the logarithm of the shift count (where shift=1 for the first shift observed for the driver in 2009). All models control for the logarithm of the current shift's total length (in hours).

Table 5: Tests of Randomization Assumption

Specification:	1	2	3	4
Dependent Variable:	log(Mean Earnings at Ride 3 Drop-Off)			
New Driver	-0.0415	-0.0015		
	(0.0045)	(0.0006)		
New× log(Shift)	0.0088	0.0004	0.0004	-0.0000
	(0.0011)	(0.0002)	(0.0002)	(0.0002)
Constant	3.2793	3.2779		
	(0.0014)	(0.0001)		
Date X Hour Fixed Effects?	N	Y	Y	N
Date X Hour X Pick-Up NTA Fixed Effects?	N	N	N	Y
Driver Fixed Effects?	N	N	Y	Y
N	1,031,352	1,031,352	1,031,346	979,833
R^2	0.001	0.840	0.843	0.879

Notes: Robust standard errors, clustered at the driver level, in parentheses. The dependent variable's mean earnings refers to the average across experienced drivers in the same NTA, day-of-week, and hour-of-day. The variable log(Shifts) corresponds to the logarithm of the shift count (where shift=1 for the first shift observed for the driver in 2009). All models control for the logarithm of the current shift's total length (in hours).

Table 6: Productivity across Different Difficulties

Specification:	1	2
Dependent Variable:	log(Hour-R3 Fares)	log(Hour-R3 Fares)
New× Q1 Location	-0.154	-0.083
	(0.015)	(0.017)
New× Q2 Location	-0.113	-0.068
	(0.010)	(0.013)
New× Q3 Location	-0.076	-0.038
	(0.010)	(0.010)
New× Q4 Location	-0.022	
	(0.009)	
New× Q1 × log(Shifts)	0.034	0.025
	(0.004)	(0.004)
New× Q2 × log(Shifts)	0.026	0.020
	(0.002)	(0.002)
New× Q3 × log(Shifts)	0.021	0.015
	(0.002)	(0.002)
New× Q4 × log(Shifts)	0.009	0.004
	(0.002)	(0.002)
Q1 Location	-0.376	
	(0.005)	
Q2 Location	-0.179	
	(0.003)	
Q3 Location	-0.097	
	(0.003)	
Date X Hour Fixed Effects?	Y	N
Date X Hour X Drop-Off NTA Fixed Effects?	N	Y
Driver Fixed Effects?	N	Y
N	977,744	908,042
R^2	0.181	0.357

Notes: Robust standard errors, clustered at the driver level, in parentheses. Difficulty is defined by an NTA/Day-of-the-Week/Hour-of-the-Day-specific measure of the total fare earnings in the 60 minutes following a drop-off. This variables is split into quartiles, weighted by the number of rides. The fourth quartile corresponds to the "easiest" quartile (i.e. highest average fare). The variable log(Shifts) corresponds to the logarithm of the shift count (where shift=1 for the first shift observed for the driver in 2009). All models control for the log of the shift length (in hours).

Table 7: Location-Specific Experience and Productivity

Specification:	1	2	3	4
Dependent Variable:	log(Hour-R3 Fares)	log(Hour-R3 Fares)	log(Empty Time Between Fares)	log(Empty Time Between Fares)
New× Log(Drop-Offs Outside PUMA)	0.008	0.008	-0.001	-0.002
	(0.002)	(0.002)	(0.003)	(0.003)
New× Log(Drop-Offs Inside PUMA)	0.007		-0.015	
	(0.002)		(0.003)	
New× Log(Drop-Offs Inside PUMA, Outside NTA)		0.003		0.001
		(0.002)		(0.003)
New× Log(Drop-Offs Inside NTA)		0.004		-0.016
		(0.002)		(0.003)
Date X Hour X Drop-Off NTA Fixed Effects?	Y	Y	Y	Y
Driver Fixed Effects?	Y	Y	Y	Y
N	908,042	908,042	953,870	953,870
R^2	0.357	0.357	0.378	0.378

Notes: Robust standard errors, clustered at the driver level, in parentheses. All models control for the logarithm of the current shift's total length (in hours).

Table 8: Switching Locations

Specification:	1	2	3	4
Dependent Variable:	Switch NTA	Switch NTA	Switch NTA	Switch NTA
New× Log(Drop-Offs Outside PUMA)	0.013	0.023	0.042	0.030
	(0.002)	(0.003)	(0.006)	(0.006)
New× Log(Drop-Offs Inside PUMA, Outside NTA)	0.008	-0.004	-0.022	-0.026
	(0.002)	(0.003)	(0.006)	(0.006)
New× Log(Drop-Offs Inside NTA)	-0.028	-0.025	-0.049	-0.034
	(0.002)	(0.003)	(0.006)	(0.006)
Locations:	All	Lower Manhattan	Excluding Lower Manhattan	Excluding Lower Manhattan and Most Common NTA
Date X Hour X Drop-Off NTA Fixed Effects?	Y	Y	Y	Y
Driver Fixed Effects?	Y	Y	Y	Y
N	949,315	875,130	74,075	62,090
R^2	0.244	0.223	0.519	0.559

Notes: Robust standard errors, clustered at the driver level, in parentheses. "Switch NTA" is an indicator for whether the 3rd drop-off is in a different NTA than the 4th pick-up of the shift. Column 4 excludes each driver's most common drop-off NTA (outside of Lower Manhattan). The variable log(Drop-offs) corresponds to the total number of drop-offs observed for the driver in 2009, prior to the 3rd drop-off of the current shift. All models control for the logarithm of the current shift's total length (in hours).