Valuing Flexibility: A Model of Discretionary Rest Breaks

Marc-Antoine Schmidt*

November 15, 2018

Job Market Paper
Click here for the latest version

Abstract

As flexible work arrangements become increasingly prevalent in the labor market, more and more workers have discretion over when they take rest breaks—a feature that is likely appealing to many. Yet we do not have a formal economic model of the decision to take breaks, nor do we know how much workers value this 'breaks' flexibility. To fill the gap, I develop and estimate the first dynamic model of daily labor supply that incorporates rest breaks. The model includes several factors that influence the decision to take breaks: fatigue, opportunity costs, preferences across hours of the day, and random utility shocks. I estimate the model using high-frequency data on millions of taxi trips covering over 14,000 drivers in NYC during an entire year. This allows me to characterize heterogeneity across drivers in a flexible and transparent way, estimating the model separately for each driver. Using the estimated parameters, I first evaluate the welfare loss to workers if discretionary breaks were replaced by scheduled breaks. My results show that flexibility is valued highly: the average driver in my sample would require a 22 percent increase in revenue to accept a counterfactual fixed work schedule. Further, I find substantial heterogeneity in this valuation, indicating that for some workers, discretionary breaks bestow a large non-pecuniary benefit. I then use the model to study the effects of a realistic 'mandatory breaks' policy on the frequency of breaks and labor supply. Counterfactual evidence shows that such a policy would substantially increase the frequency of breaks but would reduce labor supply by 6 to 9 percent. This result highlights the need to weigh the benefits of break-oriented policies—including a reduction in accidents—with the negative consequences for labor supply and the welfare of workers. While I use a specific industry to estimate the model, the proposed framework is quite general and can be applied to various other industries to understand how workers make short-term labor supply decisions.

*I am grateful to my advisor, Rob McMillan, as well as Kory Kroft, Philip Oreopoulos, and Aloysius Siow, for their support throughout this project. I also benefited from discussion with Victor Aguirregabiria, Stephen Ayerst, Gustavo Bobonis, Nicolas Gendron-Carrier, Jonathan Hall, Marie-Laurence Lacerte, Jean-William Laliberté, Yao Luo, Mathieu Marcoux, Baxter Robinson, Eduardo Souza-Rodrigues, and seminar participants at the University of Toronto.
Contact: Department of Economics, University of Toronto, 150 St. George Street, Toronto, Ontario, M5S 3G7. Email: ma.schmidt@mail.utoronto.ca.

1 Introduction

With the decline of traditional work arrangements—the nine-to-five workday—and the shift toward greater scheduling flexibility in the labor market,[1] discretionary rest breaks have become an increasingly common feature of many jobs. During their regular workday, freelancers, academics, white-collar managers, and taxi drivers—to name a few—can all choose when and how many rest breaks to take. This appealing flexibility allows workers to tailor the frequency of breaks to their personal preferences, and as such, is likely part of the explanation for the well-documented increase in flexible work arrangements. Yet we do not have a formal economic model of discretionary breaks, nor do we know how much workers value this 'breaks' flexibility.

In this paper, I develop and estimate the first dynamic model of daily labor supply that incorporates rest breaks. The model focuses on forward-looking workers who need to decide each period whether to work, take a break, or end their workday. It accounts for several different factors influencing the labor supply decision: the disutility of fatigue while working, the opportunity cost of a break (forgone earnings), the fixed cost of switching from work to a break, and random utility shocks.

Three types of insight emerge from my analysis. First, the model estimates clarify the relative importance of the factors affecting short-term labor supply decisions, especially breaks. Second, I quantify workers' valuations of 'breaks' flexibility. Third, I am able to simulate the effects of potential policies aimed at reducing work when fatigued or policies seeking to increase the frequency of breaks.

To estimate the model, I focus on the New York City (NYC) taxi industry, using transaction-level data covering the universe of taxi trips. From more than 170 million taxi fares, I construct a dataset which contains the daily labor supply decisions of 14,190 drivers over an entire year.[2] The New York City taxi industry is a suitable setting to

[1] Katz and Krueger (2016) show that between 2005 and 2015, 32% to 52% of net employment growth in the United States was due to 'independent workers,' which include contractors, freelancers, and 'gig-economy' workers.

[2] There are more than 40,000 unique driver identifiers in the data. I restrict my analysis to these 14,090 drivers as they display behaviors indicating that they are rental drivers. The exact restrictions

study the decision to take breaks because drivers have full flexibility over when to take breaks and for how long. It also has important parallels with many recently emerging jobs that feature scheduling flexibility.

The richness of the data allows me to estimate heterogeneity across drivers in a very flexible manner. While I use all drivers to obtain the law of motion for the market wage, I estimate the parameters of the utility function separately for each driver. In doing so, I am able to recover a fully non-parametric distribution of the set of utility parameters. The interpretation of such heterogeneity is transparent: individuals have different costs of effort, different responsiveness to earnings, different preferences for taking breaks at certain hours-of-the-day, etc. Understanding this worker heterogeneity is key if we want to characterize who will be most affected by a given policy, as I discuss below.

Systematic measurement error in the labor supply decision is possible because the dataset does not contain explicit information about whether the driver is taking a break or whether he[3] is searching for a fare. Previous studies have imposed a simple fixed threshold to classify long wait times between fares as a break, an obvious problem with this being the systematic misclassification of long search times as breaks when demand falls. I address this problem using the spatial nature of a taxi driver's working environment: long wait times are categorized as 'breaks' using a threshold that varies with market conditions. I also control for longer driving times when returning from a remote dropoff location by computing the optimal driving time from the previous dropoff to the next pickup.[4]

The model builds in several key reduced-form features of the data. First, it accounts for unobserved heterogeneity, given the differences in the average duration of a shift or the average time on a break observed across drivers. Second, the probability of taking a break increases with the length of continuous work time, suggesting that

are explained in Appendix A.

[3]The taxi industry is still, to this day, male-dominated. According to a report by the NYC Taxi and Limousine commission, fully 98.9% of taxi drivers are male (Taxi and Limousine Commission, 2014).

[4]I use a routing engine with road network data from OpenStreetMap to calculate this distance for the 170 million medallion taxi trips made in 2013.

fatigue plays a role in the decision to take a break. Third, I show that taxi drivers take into account the end of their rental period, indicative of the fact that taxi drivers are forward-looking. Fourth, to show that drivers respond to varying opportunity costs, I replicate a standard methodology in the literature to estimate the labor supply elasticity of earnings. Specifically, I measure labor supply as the time working and find a much higher labor supply elasticity compared to the usual case where labor supply is measured as the time working *or on breaks*. This supports the plausible view that drivers recognize that taking a break when demand is low is less costly than when demand is high.

While the basic formulation is general, I make use of a particular feature of the taxi industry in specifying the model: most taxi drivers have a rental agreement for the vehicle they drive and rent for either the day shift (5 AM to 5 PM) or the night shift (5 PM to 5 AM). The 12-hour rental period makes this a finite-horizon problem, which can be solved by backward induction. The development of the model is also guided by a short survey I conducted in July 2018 with 42 NYC medallion taxi drivers, which indicated that taking a break in the city was a complex decision they faced every day.[5]

To estimate the model, I employ the nested fixed point algorithm of Rust (1987). I find that the fixed cost of taking a break is very important in the NYC taxi industry, with a monetary value of about $25 for the average driver (or about 45 minutes of revenue). Taking a break for a taxi driver requires that they locate a parking place, often a challenge in Manhattan. I also find that the effect of fatigue on utility is small, suggesting that breaks allow drivers to recover only a small portion of their expended energy. Further, random shocks, which could include the luck of finding one of the rare taxi relief stands, are important.[6]

Using the structural estimates of the model, I run several informative counterfactual experiments. First, I look at the difference in workers' surplus between the flexible environment observed in the data and a hypothetical fixed work schedule. I find

[5] Other evidence from the survey indicated that heterogeneity across drivers was important.

[6] There are 68 taxi relief stands in NYC. These street parking places give an opportunity to taxi drivers to leave their vehicles and take care of personal needs.

that workers value scheduling flexibility highly. To accept the hypothetical fixed work schedule, the average NYC taxi driver would require an extra $63 in revenue per day (about 23% of daily revenue), with a standard deviation of $42 indicating a high degree of heterogeneity across drivers.

Second, I study the effects of a 'mandatory breaks' policy, where workers are forced to stop working—by either taking a break or ending their shift—after a certain period of uninterrupted work. Such breaks-oriented regulations are widespread in industries where fatigue can have potentially fatal consequences (e.g. truck driving or air traffic control). In a taxi context, traffic accidents and vehicle insurance create a moral hazard problem whereby the driver might take more risk than what would be socially optimal, perhaps justifying policy interventions. When I compare the current laissez-faire environment (in terms of breaks) to the counterfactual 'mandatory breaks' policy, I find that a regulation of this type would achieve the goal of increasing the frequency of breaks. However, it would also decrease overall labor supply by 6 to 9 percent due to drivers taking more breaks *and* working shorter shifts. This result highlights the need to weigh the benefits of break-oriented policies—including a reduction in accidents—with the negative consequences for labor supply and the welfare of workers.

While I use a specific industry in this study, the model is quite general. Most jobs with discretionary breaks can be thought of as involving hour-by-hour decisions over when to take a break—exactly the model's focus. In this way, it can be applied to various other contexts to understand how workers in a given industry make their short-term labor supply decisions.

In the remainder of this paper, I explain the paper's contribution in the context of prior work in Section 2. Then in Section 3, I describe the data and the methodology used to infer labor supply decisions. Section 4 documents reduced-form patterns in the data that support the modeling assumptions I make. The model is presented formally in Section 5. In Section 6, I describe the estimation strategy and present the parameter estimates. Then I discuss counterfactual experiments and the results from these in Section 7. I conclude in Section 8.

2 Relation to the Literature

This paper builds on several literatures. First, it adds to the recent literature quantifying the value of flexibility in the context of alternative work arrangements. The recent growth in the demand for independent contractors and alternative work arrangements has attracted the attention of many researchers, including economists.[7] It has been suggested that the biggest benefit of independent work is the labor supply flexibility it offers, both in terms of the overall quantity of hours and being able to choose specific times one wishes to work (Oyer, 2016).

Two important studies look at the valuation of scheduling flexibility. In an innovative paper, Chen et al. (2017) seek to estimate the value of flexibility in the labor market using Uber drivers. They estimate the drivers' surplus from the flexibility offered by the platform compared to traditional work arrangements, finding that the drivers' surplus is equivalent to 40 percent of their total pay.[8] My study extends their work in three ways. First, I focus on a smaller time-frame per period (30-minutes periods instead of 1-hour), which allows me to identify the value of short discretionary rest breaks. Second, I use a dynamic discrete choice framework to model the labor supply decision rather than a static model of the reservation wage. Third, I study taxi drivers. Compared to taxi drivers, Uber drivers are not bound by a rental agreement and can work at any time of the day.

In a similar vein, Mas and Pallais (2017) use job postings to recover the valuation of scheduling flexibility based on revealed preference. Their randomized experiment looks at three aspects of flexible work arrangements: working from home, being able to set the number of hours in a week, and being able to choose when to work. Breaks are not the focus of their study. They find a small average willingness to pay for scheduling

[7] There seems to be a consensus around the fact that alternative work arrangements are growing, but the magnitude of the phenomenon is still debated (Jackson et al., 2017; Abraham et al., 2017). Stanford (2017) also argues that this is not a new phenomenon, but a return to previous work organization strategies that were commonplace in previous centuries.

[8] Other researchers have been able to use data from Uber to study labor market outcomes (e.g. Hall et al., 2017; Cook et al., 2018). In 2013, the period for which I have data, Uber accounted for a minuscule share of trips in NYC. Since then, the growth of Uber around the world has been exponential.

flexibility but note that the considerable heterogeneity in valuations suggests that analyses based on the mean could be misleading. Interestingly, they also consider the value of jobs that permit employers to change a worker's schedule on short notice. They find that the average applicant is willing to give up 20 percent of their wage to avoid this alternative arrangement. Because of features of the jobs posted, the authors were not able to offer flexibility within a shift: only the start time and end time of a shift were allowed to vary.

The second literature this paper contributes to focuses on settings with flexible hours to understand short-term labor supply decisions. Since the seminal work of Camerer et al. (1997), data on daily labor supply have been used to test the neoclassical labor supply model against models with reference-dependent preferences. Early papers in this literature, including Camerer et al. (1997), Fehr and Goette (2007), and Crawford and Meng (2011) found evidence supporting reference dependence. However, this conclusion has been challenged (Oettinger, 1999; Farber, 2005, 2008a; Stafford, 2015). In all of these studies, the measure of labor supply employed is always the difference between the end time and the start time of a shift, and so cannot inform of within-shift labor supply decisions such as breaks.

More recently, the availability of the digital records of each taxi trip in New York City—the dataset used in this paper—has generated considerable interest from researchers. Most notably, Farber (2015) replicated the methodology of Camerer et al. (1997) using this newly available dataset. His findings suggest that most taxi drivers behave in a way that supports the neoclassical model. In this paper, I follow this view and model the workers as neoclassical optimizing agents. Recent studies show that reference dependence can explain some specific behaviors of taxi drivers but those are not necessarily of first-order importance in my model.[9,10]

[9]For example, Thakral and Tô (2017) propose a model of adaptive reference points. This can be interpreted as a model midway between the standard reference dependence model and the neoclassical model. In another paper (Schmidt, 2018), I show that taxi drivers respond to idiosyncratic windfall gains in a manner consistent with reference dependence, but respond to variation in the aggregate wage in a way consistent with the neoclassical model.

[10]The NYC taxi dataset has also been used in other contexts that are further away from this paper.

Instead of using a reduced-form approach, two recent studies have used a structural dynamic discrete choice framework to model the daily labor supply decision of taxi drivers. Fréchette et al. (2018) develop a general equilibrium model of the taxi industry in NYC. Within the model, taxi drivers decide endogenously how much labor to supply each hour. Their labor supply model shares some similarity with the one in this paper. First, a 'period' is defined in a unit of time. Many papers have equated a period with a fare (Farber, 2005; Thakral and Tô, 2017; Buchholz et al., 2018), giving rise to periods of different length; and it is unclear how this can be adapted to study within-shift breaks. Second, taxi drivers solve a finite-horizon problem. Third, because the time horizon is short, drivers are assumed not to discount within-day payoffs. For computational simplicity, however, they assume that breaks are exogenous. Buchholz et al. (2018) also model the decision to end a driver's shift as a dynamic discrete choice problem. They find empirical support for both neoclassical and behavioral responses and propose a new estimator that relaxes some assumptions regarding the error term.

None of these previous studies has focused on breaks within the day, although there is an acknowledgment that breaks may be a potential confounding factor and therefore need to be defined. For instance, Farber (2005) and Thakral and Tô (2017) use constant thresholds of 30, 60 and 90 minutes to define breaks, depending on the location. Similarly, breaks are considered exogenous in Fréchette et al. (2018), who use a fixed threshold of 45 minutes between two trips as the definition of a break. In Section 3.3, I explain in more detail the issues arising from previous definitions of breaks, and I also propose solutions to those issues.

A third related literature studies at the consequences of fatigue and breaks for different outcomes. In this paper, I use this literature as a way to understand why workers take breaks. Two important economic outcomes have been studied extensively: productivity and work safety.

For instance, Haggag and Paci (2014) look at tipping behaviors; Mangrum and Molnar (2017) study the marginal congestion of a taxi; and Buchholz (2018) looks at inefficiencies created by regulations in the taxi industry. The dataset has even been used to study informational leakage from the Federal Reserve (Finer, 2018).

The relationship between productivity, working hours, and fatigue has been recognized for a long time (see e.g. Goldmark and Brandeis, 1912; Vernon, 1921). More recently, Brachet et al. (2012) look at the performance of paramedics over the course of their shift, finding a 0.76 percent increase in 30-day mortality of the patient at the end of the shift. Collewet and Sauermann (2017) use data from a call center to estimate the reduction in productivity over a work day. Interesting research by Pencavel (2015, 2016) uses data from WWI munition workers to understand how fatigue and resource recovery affect productivity. Henning et al. (1997) and Pendem et al. (2016) use data entry workers and fruit harvesters, respectively, to document the positive effect of rest breaks on productivity.

An extensive literature studies the effect of breaks on the risk of work accidents and road accidents (Tucker (2003) provides a review). In this literature, the most related study is by Dalziel and Job (1997). They analyze vehicle accidents and fatigue in the context of taxi drivers. Using survey data on 42 Australian drivers, they find a significant negative correlation between break time and the accident rate. Furthermore, they document optimism bias among drivers regarding their capability to avoid accidents and the ability to drive safely while fatigued. This result is essential for understanding why it may be optimal for authorities to implement breaks-related regulations despite the negative effects on the welfare of drivers.

3 The New York City Taxi Data

An obvious barrier to study labor supply microdecisions is the availability of high-frequency data from a setting where these decisions are made in a decentralized way. Furthermore, studying heterogeneity requires many individuals and many observations for each individual. To address these issues, I take advantage of transaction-level data covering the universe of taxi trips in NYC to infer daily and within-day labor supply decisions of drivers. While the official purpose of this dataset is not to study labor supply decisions, it contains an incredible amount of information covering more than 40,000 drivers over an entire year, 2013. I start with a brief description of the institutional environment taxi drivers work in. Then I describe the data and explain how I recover individual labor supply decisions.

3.1 Institutional background

The taxi industry has historically been heavily regulated throughout the world. Very much in line with that, operating a taxicab in New York City requires a medallion (attached to the vehicle) and a NYC Taxi and Limousine Commission (TLC) driver's license. In 2013, the number of medallions was capped at 13,437 (see Taxi and Limousine Commission, 2014). At that time, the market-price of a medallion was at an all-time high, reaching over $1 million. In this environment, most taxi drivers did not own the medallion but rather rented it from the owner or an intermediary (called a taxi garage). The TLC imposes a lease cap whereby the owner of a medallion cannot charge more than a price ceiling. In 2013, the daily lease rate was between $115 to $139, depending on the day of the week and whether it was the morning or the night shift.[11]

The NYC medallion taxi industry has one unique feature compared to the taxi industry in other North American cities: there is no central dispatcher. In other words, it is impossible for a customer to book a taxi trip, as medallion taxis almost exclusively operate by street hailing.[12] Furthermore, medallion taxis have a monopoly over street-hailing customers in Manhattan and at the airports (see Figure A6 for a map of the restricted zone).

[11] On June 20, 2013, the TLC decreased the daily lease cap by $10 but allowed the medallion owner to charge $10 for credit card processing fees per shift, making the new rule revenue-neutral for almost everyone. See `http://home.nyc.gov/html/tlc/downloads/pdf/newly_passed_rules_leasecap_updates.pdf`.

[12] Since the end of 2013, the taxi environment in NYC has changed considerably. In mid-2013, the TLC started a pilot program for an E-Hail smartphone-based application to improve the matching. This pilot was negligible in 2013 and accounted for 0.25 percent of all yellow cab pickups between June and November 2013 (see `http://www.nyc.gov/html/tlc/downloads/pdf/ehail_q2_report_final.pdf`). Another type of taxi, called the Boro taxi, made its debut in 2013. It is possible that taxi drivers changed their strategy following the introduction of Boro taxis. However, Boro taxis are not allowed to compete with yellow medallion taxis in Manhattan and at the airports, where 93.8% of all yellow taxi pickups originate (Taxi and Limousine Commission, 2014).

Fares are regulated by the TLC. There was no fare variation during the sample period, the last change being enacted in September 2012. The regular fare follows a two-part pricing scheme with the fare starting at $3 (including the NY State tax surcharge) and increasing by $0.50 for every 0.2 miles traveled or each minute the cab is stationary.[13] There is also a flat fare of $52 for trips between Manhattan and JFK Airport. Because the fare is fixed, variation in average hourly earnings in the market comes from longer search times.

The airport taxi market is important but should be treated differently from regular street hailing as it is the only place in NYC where taxi stands truly allow a driver to take a break. A long queue of taxis would require the driver to stay in the vehicle to move forward slowly through the queue. In contrast, an airport taxi stand allows the driver to step out of the car and use facilities there while retaining his position in the queue.[14]

The transaction-level data used in this study originate from the Taxicab Passenger Enhancement Project (TPEP), spearheaded by the TLC. I use data from January to December 2013. By 2008, all NYC taxicabs were equipped with a digital system that records detailed information about each trip. This replaced hand-written logs used by the first generation of studies examining the labor supply of taxi drivers.

3.2 The TPEP data

The dataset contains an entry for each of the 173 million taxi trips made during 2013. It contains unique identifiers for the driver and the vehicle, information regarding the date, time, and precise GPS location of both the pickup and dropoff, and detailed information about the regular fare, surcharges, tolls, and tips paid by credit card. All

[13] There are also a night surcharge and a peak hour weekday surcharge of $0.50 and $1, respectively.

[14] The taxi airport holding areas are parking lots comprised of many lanes. The queue starts by filling the first lane. When it is filled, a second lane is opened and taxis start queuing. This process repeats with many more lanes. When customers requests taxis, the first lane advances until the lane is empty. During that time, vehicles in other lanes stay put. This allows the drivers to use the facilities: small restaurant, bathrooms, praying areas, etc.

data are stored electronically and automatically sent to the TLC.

In contrast to many datasets used by labor economists, the TPEP dataset was not created with the goal of conducting research. Most importantly, the dataset does not contain information about what would be called a 'shift.' If the raw data are grouped by drivers, what we see are many transactions, separated by periods of inactivity. It will be helpful to characterize these periods of inactivity into three categories, from shorter to longer: customer search, breaks, and 'time off work.'

I follow Farber (2015) and define a period of time off work as any gap of more than six hours between two trips. This allows me to construct work shifts as all consecutive trips made by a driver between two periods of time off work.

An interesting feature of the New York City taxi industry is the '2-shift rule.' Specifically, the TLC forces the majority of medallions to be operated during two shifts per day. As will be explained in more detail in Section 5, I use the fact that almost all rental agreements between taxi fleets and drivers start or end at 5 AM or 5 PM. The day shift is consequently defined as starting at 5 AM and ending at 5 PM and the night shift starting at 5 PM and ending at 5 AM. In the estimation, I focus on drivers whose behavior during the year indicates that they follow the day-shift or night-shift schedule. The daily labor supply decision then becomes a finite-horizon problem where the driver is forced to end his shift at either 5 PM (day shift) or 5 AM (night shift).[15] In Figure A2, I show the distribution of end time of the shift. Two patterns emerge: first, most drivers in the sample respect the end time, although they sometimes spill in the next driver's rental period; second, the end of the night shift is much more dispersed than the day shift, where most drivers finish their shift at the end of the rental period.

[15] See Fréchette et al. (2018) for a more detailed explanation of the phenomenon. They argue that the reason for the 5 PM transition time is to make day shifts and night shifts equally attractive to drivers in terms of accumulated earnings.

3.3 Identifying breaks

Breaks are the primary focus of this study. The goal is to define breaks in a way that minimizes the misclassification of breaks into customer search time (and vice versa). Previous studies have defined breaks using thresholds that depend only on the location of the pickup and dropoff.[16] That methodology has three important issues: First, market conditions and misclassification errors will be correlated. Second, a fixed threshold does not account for long return trips. Third, waiting for a customer in the airport queue *is* a break.

3.3.1 Market Conditions

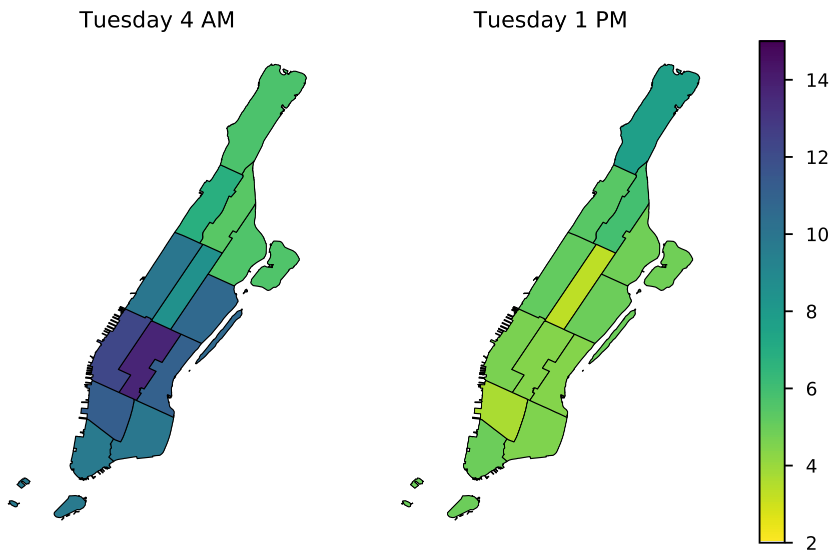
The first reason why using a fixed threshold to define a break is problematic comes from the correlation between market conditions and misclassification errors. To see why, imagine a taxi driver in a busy area in Manhattan's financial district at 4 PM on a weekday. The probability of misclassifying a 30-minute customer search time as a break is very low, given the average search time at 4 PM in that location is well below five minutes. However, at 4 AM in the same location, the average search time is much higher, and the probability of misclassifying a 30-minute customer search time as a break is therefore also much higher.

To address this issue, I compute the average search time for each location during each hour of the week. I compute the mean search time using only observations for which the last dropoff and the next pickup are in the same region. The resulting search time in Manhattan is mapped in Figure 1 for Tuesday at 4 AM and Tuesday at 4 PM. There is a clear pattern, where the search time is generally much longer at the end of the night.

The threshold to classify a break will then depend on this measure of average search

[16] For instance, Farber (2005) defines a break as a 30-minute period between two trips within Manhattan, a 120-minute period between any trip and a trip that started at Laguardia or JFK Airport, and a 60-minute period between all other trips not covered by the first two rules. Thakral and Tô (2017) use similar thresholds.

Figure 1: Average Search Time in Manhattan



Notes: The average search time (in minutes) in each region is computed using only observations with the last dropoff and the next pickup in the same region. Each region is a ‘community board.’ See Section A.3 for further explanations.

time. In areas where the mean search time is higher, a longer period of inactivity will be required to define a break. I multiply the search time by 1.5 so that a period of inactivity slightly longer than the mean search time is not considered to be a break. I also set the minimum time to classify inactivity as a break to be 20 minutes.

3.3.2 Airport Breaks

Farber (2005) and Thakral and Tô (2017) define breaks at airports to have the longest threshold and they consider waiting in the airport queue to not be a break. In July 2018, I conducted informal interviews with taxi drivers. Sixteen of them were located at the Laguardia Airport taxi hold when I interviewed them. All of them indicated that they considered the queuing time as a break. Not considering the queuing time as a break is problematic because, from anecdotal evidence gathered during the survey, some taxi drivers drive to one of the airports when they are in need of a break. Based on this evidence, I consider time waiting in an airport queue as a break. I set the average search time to zero in the area of Laguardia and JFK Airport.[17]

3.3.3 Driving Time

Another issue with the fixed threshold methodology relates to controling for the driving time between the last dropoff and the next pickup. While previous studies used a longer threshold when the wait time originated or ended outside of Manhattan, the TPEP data contains the pickup and dropoff coordinates of each trip, giving us the ability to control for driving time with much more precision.

Specifically, I compute the driving distance and duration between each dropoff and pickup using a shortest path algorithm.[18] While it is possible that the driver did not take the shortest path, this will lead to a much better approximation for the true duration of inactivity.

Using these definitions, I classify a break as a period of inactivity that satisfies the following inequality:

$$(t_p - t_d) - \tilde{d} > \max(20, \bar{s}_{lh}),$$

where t_p and t_d give the time of the next pickup and the last dropoff, respectively; $\tilde{d}$ is the shortest driving time between the two points; and $\bar{s}_{lh}$ is the mean search time in the hour and location of the next pickup.

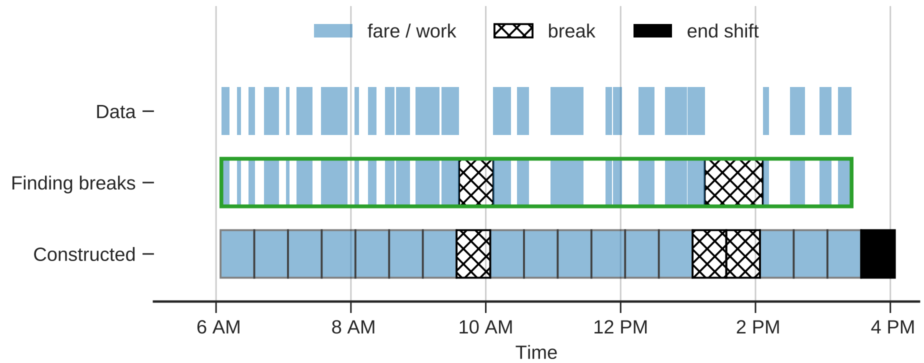
3.4 Constructing the Labor Supply Decisions

From the transaction-level data, I reconstruct the 30-minute labor supply decisions of each driver in my sample. The start of the shift is defined as the pickup time of the first customer. Then, at each subsequent 30-minute mark, I record what the driver's

[17] See Figure A6 for a map of the neighborhoods of NYC. The two airports have their own 'neighborhood.'

[18] I used the Open Source Routing Machine to compute the distance and duration for the 170 million trips. I used data from OpenStreetMap to construct the road network surrounding NYC.

Figure 2: Creation of the Data



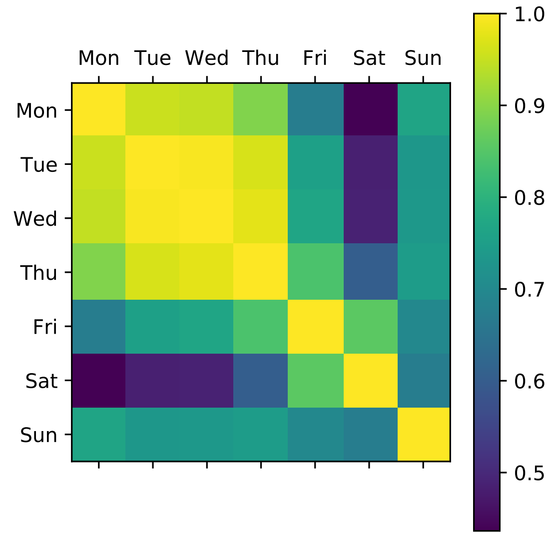
Notes: This figure sketches the transformation made to the raw data to infer the labor supply decisions. The data row represents raw transactions in the data. There are no other trips made by this driver in the 6 hours before or after any trips shown here. The 'finding breaks' row adds the hatched parts where I identified rest breaks. Finally, the constructed row shows the resulting sequence of labor supply decisions for every 30 minutes.

main activity will be during the following 30-minute period.[19] If there is no break and the shift has not ended, I assume the driver has decided to continue working. If a break accounts for more than 15-minutes out of the 30-minute period, I consider that the driver decided to take a break. Finally, if the dropoff time of the last trip of a shift is within the first 10 minutes of a period, I designate the labor supply decision in that period to be 'ending the shift'. Otherwise, the next period will be considered to be the end of the shift.

A simple illustration from a random shift of a driver can be seen in Figure 2. The first row represents what can be directly observed in the data. Each rectangle represents a trip. Then, the second row shows what can be inferred from the data: the hatched sections are breaks, identified using the algorithm proposed in the previous subsection, and the outline, which represents the inferred shift start and end time. The last row presents the resulting sequence of labor supply decisions made each 30 minutes.

[19]If a break accounted for more than 15 minutes of the period, I classify the period as a break.

Figure 3: Correlation Between the Hourly Earnings



Notes: This heatmap shows the Pearson's correlation coefficient between average market hourly earnings for each minute of the day. The methodology to compute the market hourly earnings is detailed in Section A.2. I count the first 6 hours of a day (from midnight to 6 AM) as belonging to the previous day because drivers working during those hours are on the previous day's shift.

3.5 Sample Selection

I restrict the sample in several ways to reduce measurement error. First, I only look at drivers who drove 90 percent of the time within the day shift or the night shift. Furthermore, I restrict the analysis to Monday through Thursday because labor supply patterns across those days are very similar, and I select drivers with more than 75 Monday through Thursday shifts during the year. I also drop irregular shifts with durations below 3 hours or over 12 hours.

To show why I selected shifts from Monday to Thursday, I plot, in Figure 3, the Pearson's correlation coefficient between the time series of market wage per minute during each day of the week. We can clearly see a very high correlation between the first four days of the week. Including Friday, Saturday, or Sunday in the analysis would require that we add another level of fixed effects, which would be computationally expensive.

Summary statistics for the resulting shifts are shown in Table A1. We observe patterns that are consistent with the selection rules. Selected driver are earning less per shift than the remaining drivers. The first reasons is that they work shorter shifts because they are restricted by the 12-hour limit. The second reason is a lower hourly earnings. This can be explained by the fact that owner-drivers are usually much more experienced than rental drivers. Selected drivers also work more shifts per year because I select only the drivers with enough observations to estimate the model consistently.

4 Descriptive Evidence

In this section, I document four patterns related to breaks that can be found in the data. These patterns motivate the major features of the model: heterogeneity across drivers, duration dependence of breaks, forward-looking behaviors, and opportunity costs affecting the decision to take a break.

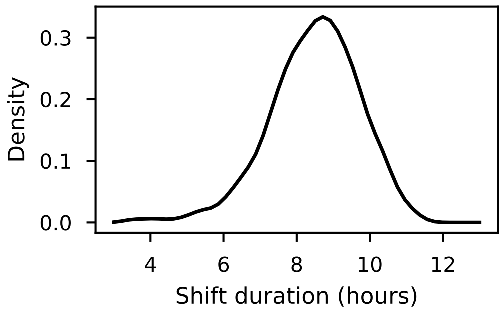
4.1 Heterogeneity across drivers

A central issue with dynamic discrete choice models is the presence of unobserved heterogeneity. As a starting point in accounting for this heterogeneity, I present distributions of summary statistics for each driver.

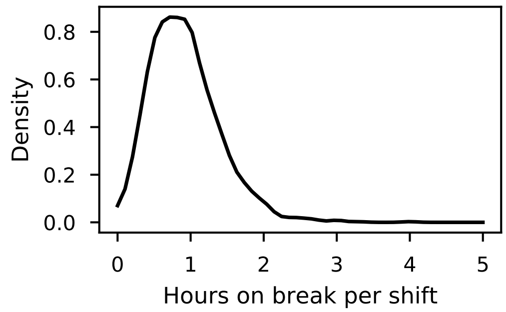
Previous studies have already documented the significant heterogeneity across taxi drivers. Farber (2015), also using the TPEP dataset, estimates a labor supply elasticity for each individual driver. He finds substantial variation across drivers in estimated labor supply elasticities, ranging from well into negative territory to more than 0.75. Mas and Pallais (2017) also find considerable evidence of heterogeneity in valuations, and warn researchers that any analysis that ignores heterogeneity will potentially lead to misleading conclusions.

Figure 4 shows the distribution of key statistics summarizing patterns related to the drivers' labor supply. Each panel shows the distribution of annual averages over

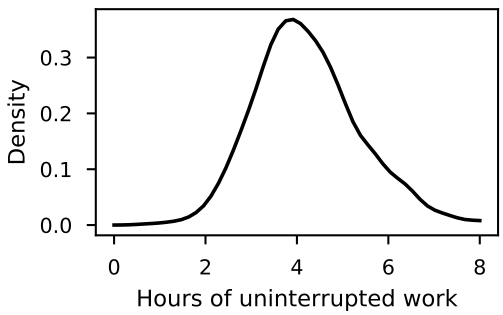
Figure 4: Distributions of Annual Means, by Driver



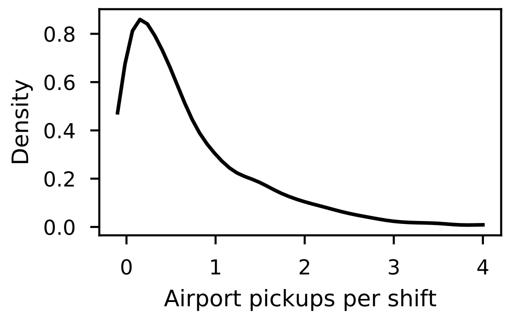
(a) Shift Duration



(b) Hours on Breaks



(c) Uninterrupted Work



(d) Airport Pickups

Notes: To construct these measures, I first obtain the mean of the selected statistic for each driver. Then, I plot the resulting distribution of these averages. The PDFs are estimated using a non-parametric kernel density estimation technique with a Gaussian kernel.

drivers: Figure 4a shows the distribution of average shift duration; Figure 4b shows the distribution of hours spent on breaks during a shift; Figure 4c shows the distribution of average duration of uninterrupted work (i.e. frequency of breaks); and Figure 4d shows the average number of pickups from either Laguardia or JFK Airport per shift. We observe a significant amount of heterogeneity in all dimensions of labor supply. To determine how much of this can be explained by randomness and sample variation, a t-test indicates that the annual average number of hours on break of two drivers needs to differ by about 0.2 to be statistically different from each other. In fact, according to this measure, the mean driver (in terms of time on break) would be statistically different from 66.3 percent of all other drivers.[20]

It is possible that heterogeneity is present in other dimensions. Performing a variance decomposition exercise, I find that differences across weeks of the year (capturing seasonality effects) can only explain 0.2 percent to 0.8 percent of all heterogeneity in the statistics presented in Figure 4. Similarly, differences across weekdays only explain 0.4 percent to 1.8 percent of the heterogeneity. In contrast, between 26.6 percent to 39.3 percent of heterogeneity can be explained by differences across drivers. This supports the choice to focus on heterogeneity across drivers while ignoring seasonality effects.

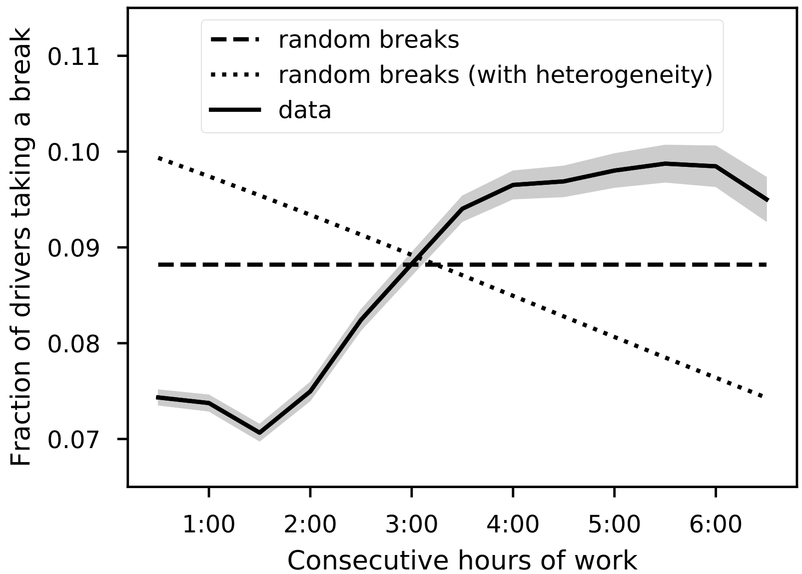
4.2 Duration Dependence of Breaks

It is possible to think of breaks as being generated only by random shocks. For instance, if taxi drivers only take breaks after dropping off a customer at an airport, breaks would be random (i.e. no duration dependence) and would be determined by the arrival rate of an airport-bound customer. For concreteness, suppose the arrival rate is constant throughout the day and equal to λ. It follows that, regardless of the time a driver has been continuously working, the probability of taking a break during the next period is always λ.

Another way to characterize a model where breaks are random is to say they exhibit no duration dependence. The properties of models with duration dependence have been studied extensively in the literature examining unemployment spells. In the present paper, instead of looking at the relationship between the job-finding probability and the duration of the unemployment spell, we look at the probability of taking a break with respect to the duration of a continuous work 'spell.' In this case, the data exhibit positive duration dependence when the probability of the event increases with the spell duration. In contrast, negative duration dependence represents a case where the probability of the event decreases when the spell duration increases.

[20]If hours on break per shift were drawn from the same distribution for all drivers (i.e. no heterogeneity across drivers), then only about 5 percent of all drivers should be statistically different from the mean driver.

Figure 5: Hazard Function of Breaks



Notes: The hazard function of a break under the hypothesis of no duration dependence and heterogeneity assumes that there are two equal-sized groups of drivers. The first group takes a break each period with a probability of 5 percent while the second group does so with a probability of 13 percent.

Duration dependence is graphically represented in Figure 5, where the hazard function of taking a break is plotted for different cases. The dashed line represents a world where breaks have a constant arrival rate and no driver heterogeneity exists. The dashed line is flat because the probability of taking a break does not depend on the duration of continuous work. In contrast, the solid line illustrates the hazard function estimated from the data. Clearly, taking a break is a process that exhibits positive duration dependence.

The above discussion compares the data to a case where breaks arrive at a constant rate throughout the day and the arrival rate is common to all drivers. The previous discussion surrounding the presence of heterogeneity across drivers makes it clear that the data rejects this assumption. When we relax the assumption that the rate is common to all drivers, it is well known that unobserved heterogeneity affects the shape of the hazard function seen in the aggregate data (Baker and Melino, 2000; Van Den Berg, 2001). Intuitively, the reason for this decreasing hazard function is that individuals with a low probability of taking a break represent a larger fraction of drivers that have continuously worked for a longer period of time. Therefore, any form of unobserved heterogeneity would only reinforce the finding that the increasing hazard rate of taking a break is not due to spurious duration dependence.

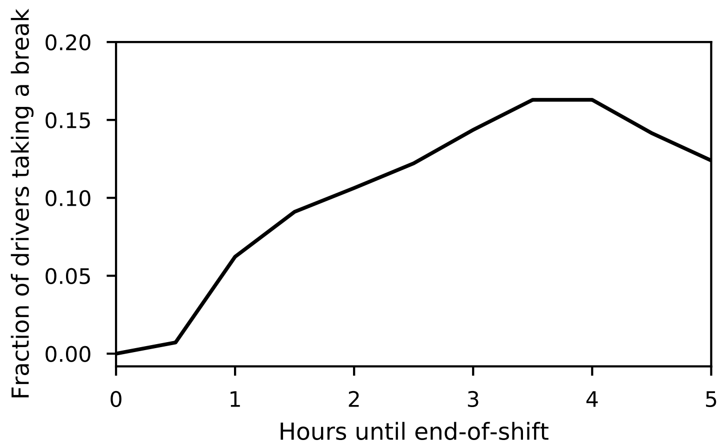
What does positive duration dependence of taking a break suggest? While a random component is still most likely present, positive duration dependence is suggestive that worker fatigue plays a role in the decision to take a break. For example, a driver is more likely to take a break after five hours of continuous work than after one hour because the accumulation of fatigue has decreased the net utility of continuing to work.

4.3 Forward-looking behavior

The hazard function of taking a break is helpful not only to show that fatigue plays a role, but also to demonstrate that drivers are forward-looking within the day. Indeed, if drivers were not forward-looking, then the end of the taxi rental period should not be correlated with a reduction in the probability of taking a break in previous periods. This is because, as a driver approaches the end of a shift, the utility of taking a break decreases, since its fatigue-reducing benefits would have an impact on a short period of time.

Figure 6 shows the hazard function of starting a break but, this time, with respect to time until the end of the taxi rental period. Notice here that the 0 on the horizontal axis represents the end of a shift. We can observe a steadily decreasing probability of taking a break as we get closer to the end of the rental period. Only shifts that end in the last period are included here because the decision to end could lead to a similar pattern without forward-looking preferences. This suggests that drivers are forward-looking because they make their decision incorporating the continuation value, which depends on the terminal period. If they did not expect their shift to end, they would not behave differently.

Figure 6: Probability of Taking a Break at the End of Shift



Notes: Only the drivers that worked until the end of their shift are included in this figure.

While forward-looking behavior can explain the pattern observed in Figure 6, other explanations cannot be ruled out by this simple reduced-form exercise. For instance, because time until the end of the rental period is the same for most drivers, it is impossible to distinguish between the previous explanation and an hour-of-the-day effect. However, these confounding factors will be handled by the structural model presented in the next section.

4.4 Opportunity cost

Simple economic intuition tells us that drivers should respond to pecuniary incentives. If the potential earnings are high, taking a break would be costly. conversely, in times of low demand and low potential earnings, it would be efficient for drivers to take a break. However, the extent to which drivers respond to pecuniary incentives is unclear.

As a simple test, Figure A3 shows the probability of taking a break in $1 increments of a period's deviation from usual potential earnings. As expected, we see a negative correlation, indicating that drivers are more likely to take a break when earnings are low. This suggests that potential earnings during the period are a factor in the decision to take a break.

Table 1: Estimates of Labor Supply Elasticity (2SLS)

	All	Day Shift	Night Shift	Owner
	(1)	(2)	(3)	(4)
Panel A: Gross shift				
log Hourly Earnings	0.256***	0.071***	0.438***	0.463***
	(0.004)	(0.004)	(0.006)	(0.015)
Panel B: Adjusted shift (net of breaks)				
log Hourly Earnings	0.821***	0.624***	1.098***	1.216***
	(0.007)	(0.009)	(0.011)	(0.032)
Driver	Yes	Yes	Yes	Yes
Observations	4,894,002	2,148,223	2,151,782	566,344

Notes: Clustered standard errors in parentheses (driver-level). Controls include weather (temperature and precipitation), location fixed effects (modal pickup neighborhood; 72), holiday fixed effects (9), and fixed effects for the month of the year (11) and the hour of the week (167).

Many studies focus on estimating the daily labor supply elasticity. I replicate their methodology to assess how much the elasticity would change if we account for breaks, not as a control variable, but in the measure of labor supply itself. Define the 'gross' labor supply as the time between the start and end of a shift. Then, we can define 'adjusted' labor supply as the 'gross' labor supply net of all breaks taken during the shift. If breaks happen randomly and are not correlated with potential earnings, we should observe the same labor supply elasticity for both measures of labor supply.

Table 1 shows the estimated labor supply elasticities. The details of the regression, the variables, and the instrument for hourly earnings are explained in Appendix B. In Panel A, the labor supply elasticities are estimated using the gross measure of labor supply. When I use the adjusted measure of labor supply, in Panel B, the elasticities become significantly higher. This is intuitive: some drivers use breaks as a margin of labor supply adjustment. Furthermore, lower potential earnings reduce the cost of taking a break.

5 Labor Supply Model with Discretionary Breaks

In this section, I develop a model of daily labor supply with breaks. The model reflects the flexible environment taxi drivers operate in by letting the drivers make labor supply decisions along two dimensions: taking breaks and ending the shift. This can be set up using the standard dynamic discrete choice framework. At several points over the course of the day, a driver chooses between three options: working, taking a rest break, or ending the shift. His decision is forward-looking and affects the future path of variables in the state space.

Before explaining the details of the model, we need to clarify what constitutes a period. Medallion taxis in NYC are mostly used for very short trips: the median trip duration is 10 minutes and 95 percent of all trips take fewer than 30 minutes. This motivates the use of relatively short periods to model the decision horizon of drivers. In what follows, I define periods as blocks of 30 minutes.[21]

5.1 Fatigue

There is a general consensus in the literature that rest breaks during the workday reduce fatigue and that fatigue is a major factor in the decision to take a break. (Tucker, 2003; Jett and George, 2003; Hideg and Trougakos, 2009).

In the following model, I define fatigue as the net effect of all duration-dependent processes affecting utility during the day. For instance, the need to go to the bathroom or to eat are examples of duration-dependent processes with an associated disutility that grows over time. While this may be an abuse of language, I use 'fatigue' as an umbrealla term to simplify the terminology.

The way I parametrize fatigue captures two crucial features of resource depletion

[21]Many previous papers model a period as being one taxi fare (e.g. Farber, 2005, 2008b; Crawford and Meng, 2011; Thakral and Tô, 2017; Buchholz et al., 2018). In a way, this assumption is natural for taxi drivers as the decision to stop or end a shift can be taken at the end of each fare. However, using this definition, a period is only defined when the taxi driver is working. This becomes problematic when we want to study non-terminal actions (e.g. rest breaks), as in the present paper.

and recovery over the workday. First, breaks can act as a recovery mechanism and lead to a reduction in accumulated fatigue. Second, sleep is crucial for a complete recovery of resources. This also means that we expect the fatigue level of a worker to be higher later compared to earlier in the day, even if the worker is just back from a break. To model those two features, I define fatigue as the sum of two components: recoverable fatigue and non-recoverable fatigue.

In the following section and in the estimation, non-recoverable fatigue enters the utility function as a function of accumulated time since the start of the shift (h_t), where time is denominated in number of 30-minute periods. Recoverable fatigue is defined similarly as a function of the time since the last break (d_t^w), where time is likewise denominated in number of 30-minute periods. In terms of their parametrization, I assume they have a linear and additively separable effect:

$$\text{effect of total fatigue} = \pi_n h_t + \pi_r d_t^w. \tag{1}$$

In the model, breaks are able to reset recoverable fatigue, but have no effect on non-recoverable fatigue. Following the literature on resource recovery, I make the strong assumption that any breaks of at least 30 minutes eliminates recoverable fatigue completely. In contrast, the only way to reduce non-recoverable fatigue is to end one's shift.

As shown in Section 4, the probability of taking a break increases with the time since the last break. It is useful to understand the fatigue process through the lense of Figure 7. Parameters π_n and π_r inform us on the slope of the fatigue process. The larger they are, the more quickly fatigue accumulates.

5.2 Per-Period Utility

Given the above definition of fatigue, let us now define the utility function of the agents. In dynamic discrete choice models, the agent derives utility each period. The utility depends on the value of the state variables as well as the action taken by the agent. While

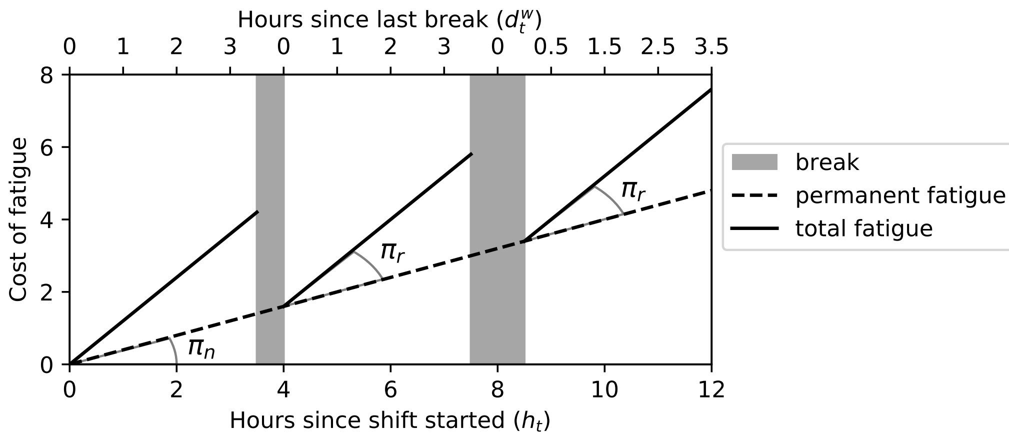


Figure 7: Graphical Representation of the Fatigue Process

agents seek to maximize more than just their utility in the current period, understanding the factors that influences this utility is key to understanding the intuition behind the estimated parameters and how they affect the driver's decision.

Formally, an agent derives per-period (flow) utility, which depends on his action ($a_t \in (w)$ork, (b)reak, (e)nd shift) as well as the state space (x_t). When the agent works, his flow utility is characterized by the expected earnings to be made during that period (I_t), as well as the disutility from fatigue (Equation (1)). The utility when $a_t = w$ is

$$u_t(a_t, x_t | a_t = w) = \gamma I_t - \pi_n h_t - \pi_r d_t^w + \alpha_c^w + \epsilon_t^w,$$

where α_c^w is an intercept that varies with the clock time. It can capture the fact that working is less pleasant at certain times of the day (e.g. rush hour). It can also be viewed as an hour-of-the-day fixed effect. In the estimation, it is treated as a nuisance parameter. ϵ_t^w is an idiosyncratic utility shock. For tractability, I make the assumption that all idiosyncratic utility shocks follow an extreme value type I distribution. This assumption is sometimes called the logit assumption and is standard in the dynamic discrete choice literature.[22]

[22]Using similar data, Buchholz et al. (2018) relax this assumption in a simpler model that only tracks the decision to end a shift. Relaxing the logit assumption in dynamic models with breaks is left

The contribution of earnings to utility is represented by γI_t. In the estimation, I use the average hourly wage in period t across all drivers as a measure of I_t. For computational purposes, I discretize I_t into eight bins, each of which has a width of \$1 per 30-minute period, and is expressed in terms of deviation from the mean. To be more precise about the evolution of potential earnings, it is possible to decompose it into three separate components: the hourly market earnings, the average deviation from the hourly market earnings, a driver-specific productivity, and a driver-specific idiosyncratic shock. In this model, the hourly market earnings and the driver-specific productivity are part of the hour-of-the-day fixed effects. I assume that the driver-specific idiosyncratic shock is i.i.d. and that the drivers are not able to predict this at the start of the hour, making it irrelevant for the decision. Potential earnings (I_t) is normalized and represents the deviation from usual hourly market earnings.

When the agent takes a break, he incurs a fixed cost (τ) at the start of the break. The accumulated time on break is denoted by d_t^b. The term ψd_t^b captures the idea that breaks have decreasing marginal utility. The utility of a two-hour break is not double the utility of a one-hour break. The utility when the agent chooses $a_t = b$ can be written as

$$u_t(a_t, x_t | a_t = b) = -\tau \cdot \mathbb{1}[a_{t-1} \neq b] - \psi d_t^b + \alpha_c^b + \epsilon_t^b.$$

In addition to the terms mentioned above, the utility from rest breaks also contains an intercept that varies with the time of the day and an idiosyncratic utility shock. The hour-of-the-day fixed effect for breaks represents the fact that workers have preferences across hours of the day for taking breaks (e.g. having lunch at the same time every day).

Finally, the value of ending the shift is normalized to 0 plus an idiosyncratic utility

for future research.

shock. Putting everything together, the per-period utility is:

$$u_t(a_t, x_t) = \begin{cases} \gamma I_t - \pi_n h_t - \pi_r d_t^w + \alpha_c^w + \epsilon_t^w & \text{if } a_t = w; \\ -\tau \cdot \mathbb{1}[a_{t-1} \neq b] - \psi d_t^b + \alpha_c^b + \epsilon_t^b & \text{if } a_t = b; \\ \epsilon_t^e & \text{if } a_t = e. \end{cases} \tag{2}$$

x_t, the vector of state variables, is comprised of: I_t, h_t, d_t^w, d_t^b, a_{t-1}, and the hour-of-the-day c. It is also be useful to define $\epsilon_t \equiv (\epsilon_t^w, \epsilon_t^b, \epsilon_t^e)$, as well as the vector of parameters $\theta \equiv (\gamma, \pi_n, \pi_r, \tau, \psi, \alpha^w, \alpha^b, \theta_f)$, where θ_f is the vector of parameters capturing the evolution of the state space.

5.3 Evolution of the State Space

In dynamic discrete choice models, the agents' beliefs about the future are an essential component of the model. Intuitively, because the utility in future periods is part of the optimization problem, a rational agent must keep track of the evolution of the state space. In the problem we have here, only the potential earnings is stochastic and carries some uncertainty over its evolution. The remaining variables evolve deterministically.

Recall that I_t is a discretized measure of the market's deviation from usual period's earnings. I model the law of motion of I_t as a Markov process that can be represented with an 8×8 transition matrix (θ_f). The elements of θ_f are parameters of the model and are assumed to be identical across drivers (i.e. all drivers have the same expectation about the evolution of the potential earnings).

The other laws of motion in this model are deterministic. The hour of the day follows the clock cycle. The accumulated hours on shift (h_t) and the durations of uninterrupted work or break (d_t^w and d_t^b) follow the laws of motion:

$$h_{t+1} = h_t + 1$$

$$d_{t+1}^w = \begin{cases} d_t^w + 1, & \text{if } a_t = w, \\ 0, & \text{otherwise.} \end{cases}$$

$$d_{t+1}^b = \begin{cases} d_t^b + 1, & \text{if } a_t = b, \\ 0, & \text{otherwise.} \end{cases}$$

The fact that these variables evolve in a deterministic fashion is the key to the computational efficiency of the estimation technique and a central reason for why the model can be estimated separately for each driver in a reasonable amount of time.

5.4 Forward-looking value function

As stated earlier, agents in this model are forward-looking during the day. This implies that when deciding their action at the beginning of each period, their objective is to maximize the utility of the current and future periods within the shift, stopping at the end of their rental period.

The agent's problem is then to choose a_t to maximize the discounted sum of future flow utilities:

$$\max_{a_t} \left[u(a_t, x_t, \epsilon_t) + E \left(\sum_{t'=t+1}^{T} \beta^{t'-t} \left[u(x_{t'}, a_{t'}, \epsilon_{t'}) \right] \right) \right].$$

This is a standard dynamic programming problem. It can be written in a recursive form with the ex-ante value function $\bar{V}_t(a_t, x_t)$:

$$\bar{V}_t(x_t) = \max_{a_t} u(a_t, x_t, \epsilon_t) + \beta \int \bar{V}_{t+1}(x_{t+1}) f(x_{t+1}|a_t, x_t) dx_{t+1}.$$

In the above recursive representation of the individual's problem, $f(x_{t+1}|a_t, x_t)$ is the transition function. The value of the state variables depends only on the last period's value and the last period's decision. It is helpful to define the conditional value function $v_t(a_t, x_t)$ (also called the choice-specific value function) as the present discounted value of choosing a_t and behaving optimally in the following periods:

$$v_t(a_t, x_t) = \tilde{u}(a_t, x_t) + \beta \int \bar{V}_{t+1}(x_{t+1}) f(x_{t+1}|a_t, x_t) dx_{t+1}. \tag{3}$$

$\tilde{u}$ is defined as the utility net of the error terms. Once v_t is recovered, the problem becomes very similar to a static conditional logit model. The agent has three options and chooses the option with the highest forward-looking utility. The researcher does not know what the error term is, but makes an assumption on its distribution. Then, for a given set of parameters θ, we can compute the conditional choice probability $P(a_t|x_t, \theta)$—the probability that an action is taken given the current value of the state space. The conditional choice probability is the central object of the estimation strategy.

There is a large literature looking at the identification of dynamic discrete choice models (Rust, 1994; Magnac and Thesmar, 2002). Especially, the discount factor (β) is non-parametrically unidentified. While the discount factor is an important modeling decision when the time horizon of the forward-looking decision is long, it seems less of an issue for within-day applications. Because my periods have a duration of 30 minutes, even a discount rate of 0.99 translates to an implausibly small annual discount rate.[23] I follow Fréchette et al. (2018) and set $\beta = 1$, assuming there is no discounting within a shift.

To summarize, I have set up in this section a dynamic model of daily labor supply with discretionary rest breaks. The model features several components affecting the decision to take a break. First, breaks allow the worker to reduce accumulated fatigue through its recoverable component (d_t^w). Second, through α_c^w and α_c^b, breaks offer higher

[23]The results presented in the following section are not sensitive to changing the discount factor to any value between 0.95 and 1.

utility at certain hours of the day because of differences in taste for breaks across hours (e.g. lunchtime) or differences in average hourly earnings. Third, demand shocks affect the expected earnings and modify the opportunity cost of a break. Other factors influencing the decision to take a break are included in the random utility shocks (error terms).

6 Estimation

The estimation of this dynamic discrete choice model is done by maximum likelihood. The central pieces of the log-likelihood function are the conditional choice probabilities, $p(a_t|x_t)$. In this section, I first describe the log-likelihood. Then, I discuss intuitively how one may think of the identification of each parameter and describe how heterogeneity is handled. Finally, I present the estimates and discuss how well the model fits the data.

6.1 Estimation Strategy

Define the probability of choosing an action given a realization of the state space by $p_t(a_t|x_t)$. According to the model, this will be equal to the probability that action a is the optimal action. Assuming that the idiosyncratic utility shocks follow a type I extreme value distribution, the probability of an arbitrary choice a_t is given by

$$p_t(a_t|x_t) = \frac{\exp\left[v_t(x_t, a_t)\right]}{\sum_{a'_t} \exp\left[v_t(x_t, a'_t)\right]}. \tag{4}$$

The above conditional choice probability is essential to the estimation strategy. For instance, the maximum likelihood function is comprised of each observation's conditional choice probability of the realized action.

The maximum likelihood function is formed by calculating the probability of the observed actions in the data. For each period and each agent, the likelihood can be factored into two pieces: the conditional choice probability and the transition density function. The log-likelihood function is

$$l(\theta) = \sum_{t=0}^{T} \sum_{n=0}^{N} \left(\ln \left[p_t(a_{nt}|x_{nt}, \theta) \right] + \ln \left[f(x_{nt+1}|x_{nt}, a_{nt}, \theta_f) \right] \right).$$

Because θ_f only enters the second part of the likelihood, it can be obtained independently in a first step. Although this is not as efficient as estimating everything jointly, it greatly reduces the computational cost. This first step is carried out nonparametrically using a simple bin estimator (empirical frequencies):

$$f(x_{t+1}|x_t, a_t) = \frac{\sum_{n=1}^{N} \sum_{t'=1}^{T} \mathbb{1}(x_{nt'+1} = x_{t+1}, x_{nt'} = x_t)}{\sum_{n=1}^{N} \sum_{t'=1}^{T} \mathbb{1}(x_{nt'} = x_t)}.$$

Taking the estimate of θ_f as given, I now estimate the model using a finite-horizon version of the nested fixed point algorithm developed in Rust (1987). This technique is known to be computationally intensive because the agent's problem needs to be solved at every iteration of the likelihood optimization algorithm, hence the 'nested' structure of the problem. First, for a given draw of the utility paramaters, the algorithm solves the agent's problem by backward induction. During the last 30 minutes of the rental period, the agent knows that he is forced to end the shift. Thus, regardless of his action choice, the continuation value will be the same. Using this feature, we can compute the value function (v_t) at each preceding period of the agent's problem. Then, using the values of v_t to form the conditional choice probability, the value of the log-likelihood is computed. New draws are obtained and the process is repeated until the log-likelihood function reaches a maximum.

While the model is identified jointly, the identification of each parameter can be understood intuitively from the source of variation. γ is identified from unexpected variation in potential earnings. This requires that the hour-of-the-day fixed effect captures all expected variation in potential earnings. I estimate the model with shifts from Monday to Thursday to avoid the need to include day-of-the-week fixed effects, since this would be too computationally costly. Fréchette et al. (2018) find that shifts on Monday through Thursday are very similar, suggesting that the identification of parameters does not originate from variations across weekdays.

π_n and π_r are identified from variation in the timing of breaks as well as the timing of the end of the shift. Intuitively, I use the difference in the probability of a break throughout the shift. Keeping everything else constant, if the probability of taking a break after working for five consecutive hours is lower than after two hours, it would mean that fatigue counterintuitively *decreases* the marginal disutility of working (π_r would be of the opposite sign). Similarly, ψ is identified from the probability of ending the break at different break durations.

I can estimate hour-of-the-day fixed effects because I observe taxi drivers starting their shift at different hours of the day. This means that the hour-of-the-day fixed effects remove the unconditional choice probability at every hour of the day. The remaining variation then compares, within the same hour, the difference in choice probabilities between days where the driver has accumulated different amounts of fatigue.

The model presented in Section 5 and the estimation strategy described in this section do not distinguish between data from one driver or thousands of them. As stated by Aguirregabiria and Mira (2010), "in microeconomic applications of single-agent models, we typically have that N is large and T_i is small", where N is the number of agents and T_i is the number of periods the researcher observes agent i. This is not the case for the present application. In both dimensions, the number of observations is large. It is possible to see the data as having a completely new dimension: the shift. For each driver, N can be thought of as the number of shifts and T_j as the number of periods for shift j where we see the agent's labor supply decision.

Using the richness of the data, I handle unobserved heterogeneity in possibly the most flexible way: the model is estimated independently for each driver. This is more flexible than using a fixed effects model because heterogeneity is allowed to enter non-linearly in each parameter. In most applications of dynamic discrete choice models, the prefered method to account for unobserved heterogeneity is to use finite mixture distributions (Arcidiacono and Jones, 2003; Arcidiacono and Miller, 2011). One major limitation of this estimation strategy is that increasing the number of "types" is computationally prohibitive. In fact, most applications typically use fewer than 10 types. In contrast, I allow each taxi driver to have his own type. The resulting heterogeneity is therefore non-parametrically identified.

The interpretation of such heterogeneity is also intuitive. Because of genetic reasons or age, for example, some individuals may have better resource recovery capabilities than others. Therefore, their value of π_n and π_r can vary. Similarly, the fixed cost of taking a break (τ) can be different across individuals. I also allow hour-of-the-day preferences to vary from driver to driver. This is necessary if we believe that drivers are heterogeneous in their taste for lunchtime or if some of them have stronger distaste for rush-hour traffic.

This amount of flexibility in heterogeneity demands a lot from the data. In order to balance this flexibility with precision in the estimation, I force the hour-of-the-day fixed effects to be the same per three-hour blocks. This increases the precision of the estimates while still capturing the fact that drivers have different preferences for breaks across the day.

6.2 Results

As a benchmark, I start by showing the estimation results ignoring heterogeneity. The main reason for doing so is to compute standard errors and compare the implied confidence intervals of the parameters to the distributions I obtain after accounting for heterogeneity.

Table 2 presents the results. The estimation is done separatly for day-shift and night-shift drivers. While all parameters have the expected sign, the magnitudes are somewhat unexpected. The estimate of the fixed cost appears to indicate that it is a crucial factor in the decision to take a break. However, recall that medallion taxi drivers

Table 2: Parameter Estimates (homogeneous drivers)

	Day shift	Night shift	Equal (p-value)
Potential earnings (γ)	0.0458 (0.0050)	0.0923 (0.0044)	< 0.001
	[$1]	[$1]	
Non-recov. fatigue (π_p)	0.0155 (0.0060)	0.0132 (0.0032)	0.649
	[$0.34]	[$0.14]	
Recov. fatigue (π_r)	0.0038 (0.0004)	0.0031 (0.0004)	0.434
	[$0.08]	[$0.03]	
Rate of break util. decline (ψ)	0.2295 (0.0194)	0.3150 (0.0184)	< 0.001
	[$5.01]	[$3.41]	
Fixed cost (τ)	2.6827 (0.0704)	2.7368 (0.0666)	0.991
	[$58.57]	[$29.65]	
Number of drivers	1397	1039	
Observations	2,295,383	1,635,429	

Notes: Standard errors in parenthesis are obtained by bootstrap at the shift-driver level. In brackets are the estimates normalized by the utility value of a $1 deviation in market-level earnings (i.e. dividing by γ). The estimates are obtained from a random 15% sample of drivers.

operate mainly in Manhattan, where the fixed cost of taking a break includes the time costs of finding a parking space, a difficult challenge. The estimates of the fatigue costs are low, even when we take into account that they accumulate during over time. For day-shift (night-shift) drivers, even after four hours of uninterupted work, the cost of recoverable fatigue is only $0.66 ($0.27) per period while the cost of non-recoverable fatigue will have accrued to $2.71 ($1.14) per period.

It is important to note that almost all the difference in valuations between day-shift and night-shift drivers is due to the difference in γ. If we normalize γ to be the same across both groups, then the cost of fatigue and the fixed cost will be statistically similar.

Figure A4 presents a graphical representation of the transition matrix of the hourly earnings. The same estimates are used for the model with homogeneous drivers and for the model with heterogeneous drivers because the market conditions apply to everyone. We observe a significant amount of persistance in this measure; the probability of

Table 3: Mean Parameter Estimate (heterogeneous drivers)

	Day Shift		Night Shift	
	Mean	Std. Dev.	Mean	Std. Dev.
Potential earnings (γ)	0.0951 [$0.73]	0.1748	0.1757 [$1.35]	0.0944
Non-recov. fatigue (π_p)	0.0700 [$0.54]	0.0890	0.0624 [$0.48]	0.0517
Recov. fatigue (π_r)	0.0120 [$0.09]	0.0143	0.0116 [$0.09]	0.0114
Rate of break util. decline (ψ)	0.2899 [$2.24]	1.0287	0.5119 [$3.95]	1.8163
Fixed cost (τ)	3.1545 [$24.32]	1.3072	3.2377 [$24.96]	2.0826
Number of drivers	8100		6090	

Notes: The estimates normalized in brackets are obtained by dividing by the utility value of a $1 deviation of market-level earnings averaged across day-shift and night-shift drivers (i.e. dividing by the average γ across all drivers).

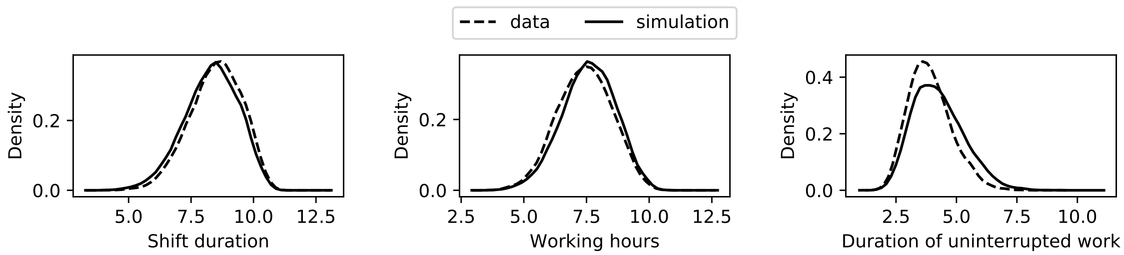
remaining in the same earnings state is always the highest.

The results from the estimation strategy that accounts for heterogeneity are presented in Table 3. The parameter estimates have the same magnitude, but in dollar value, the fixed cost is lower while the parameters governing the fatigue cost are higher. The full distributions are presented in Figure A5. We observe significantly more heterogeneity in every parameter compared to what the standard errors of the estimates from the homogeneous drivers estimation of Table 2 indicated. This heterogeneity will be driving the large range of valuations for labor supply flexibility.

To get a sense of the fit of the model, I simulate shifts from every driver and compare the distribution of some summary statistics at the driver-level. At the start of the simulation, I draw a set of initial values from the observed initial values of this driver. Then, I use the model parameters to infer the driver's decision and the values of the state variables in the next period. I iterate over every period until the driver decides to end the shift or reaches the end of the rental period.

With the simulated shifts in hand, I compare the resulting distributions to their

Figure 8: Distributions of Annual Means, by Driver



(a) Shift Duration (b) Working Hours (c) Time Between Breaks

Notes: The PDF are estimated using a non-parametric kernel density estimation technique with a Gaussian kernel.

empirical counterpart from the data along three dimensions: shift duration, working hours, and duration of uninterupted work. Figure 8 shows the three distributions. The solid line represents the distribution found in the data. The fit is very good along all dimensions. The largest disparity can be found in Figure 8c. However, the difference in means between the data and the simulation is only 21 minutes.

7 Counterfactual Experiments

The estimated structural model can be used to compute the impact of counterfactual experiments. In this section, I review two hypothetical experiments. First, I compare the utility generated by the current environment, where breaks are unconstrained, to that of a fixed schedule. This allows me to compute the compensation that a worker would require in order to accept a fixed working schedule. Second, I study how the frequency of breaks, labor supply, and worker welfare would be affected if we were to impose of a ‘mandatory break’ that seek to limit the number of uninterrupted work hours.

Both counterfactuals can be seen as a modification of the choice set. The fixed-schedule counterfactual is the most extreme, as the driver must select the action dictated by the schedule, completely eliminating any choice over within-day labor supply. The mandatory break policy affects the choice set in a similar fashion, but is more flexible. In this counterfactual, the option to work is removed (for one period) after a predetermined number of periods of uninterrupted work.

7.1 The Value of Discretionary Breaks

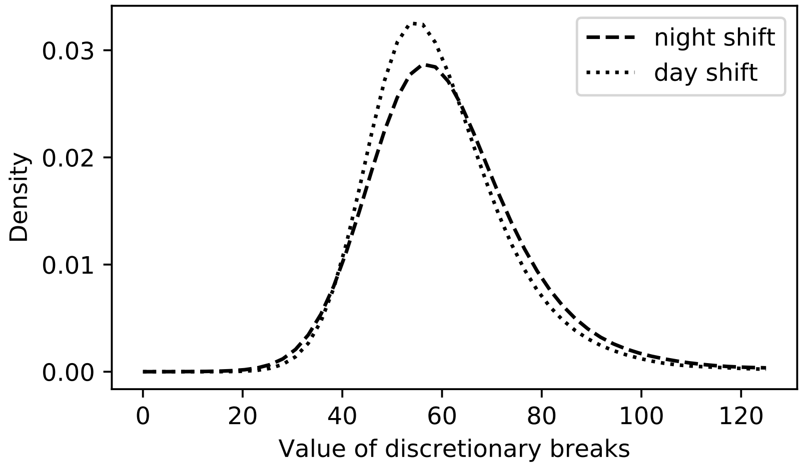
In the first simulation, I quantify the value of discretionary breaks. I compute the reduction in utility per shift of switching from the unconstrained environment to a hypothetical fixed schedule. Because the model is estimated independently, this value is driver-specific.

It should be noted that the value of discretionary breaks depends on how far the fixed schedule deviates from the optimal schedule. For example, an extreme schedule forcing drivers to take a 30-minute break every hour would lead to very low utility due to the high fixed cost. Therefore, the counterfactual schedule I explore follows a realistic work schedule with two breaks: a 30-minute break after 2.5 hours and a 60-minute break after another 2.5 hours of uninterrupted work. After the last break, the driver works for another 2.5 hours. In total, the shift's duration is 9 hours, with 7.5 hours of work and 1.5 hours of break.

The distribution of the value of discretionary rest breaks is presented in figure Figure 9. The average value for day-shift and night-shift drivers is \$61 and \$63, respectively. This indicates that the average driver would require an increase in revenue of about 23 percent to accept the counterfactual fixed schedule. While the distribution of day-shift and night-shift drivers is similar, we see a larger mass at higher values for night-shift drivers. There is a significant amount of heterogeneity over the value of discretionary rest breaks. The standard deviation is \$41 (equivalent to 14.7 percent of average daily earnings).

The high heterogeneity in the value of flexibility has important policy implications. While allowing discretion over the timing and the length of breaks would be highly valuable to some drivers, the welfare implications for other drivers may be much smaller. This heterogeneity implies that the move towards decentralizing labor supply decisions

Figure 9: Value of Discretionary Breaks per Shift



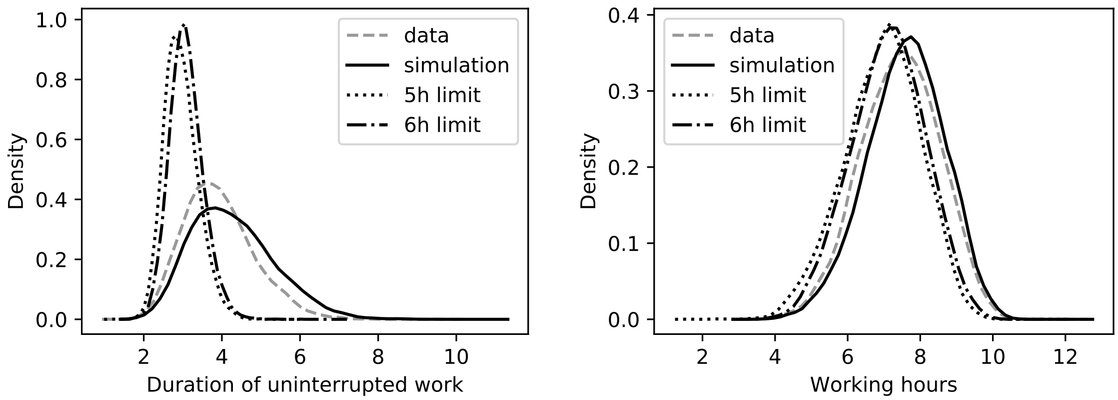
Notes: The PDFs are estimated using a non-parametric kernel density estimation technique with a Gaussian kernel.

from employers to employees will have distributional effects, with some workers gaining more than others. It is also worth noting that the subjects of this study, taxi drivers, have self-selected into this career with highly flexible work hours. Therefore, it is perhaps unsurprising to see a high valuation for discretionary breaks in an industry where discretionary breaks are the status quo.

7.2 The Effect of a Mandatory Break Policy

Policymakers have long recognized the effect of lenghty uninterrupted working hours on worker safety. This is why it is common to see regulations in this area. While the most prevalent type of policy limits the total length of a shift, many jurisdictions impose rules on breaks. For instance, in the United Kingdom, air traffic controllers cannot work for more than two hours without taking a break of at least 30 minutes. In the European Union, commercial truck drivers are required to take a break or breaks totaling at least 45 minutes after no more than four and a half hours of driving. This last rule is enforced with the aid of a digital device called a tachograph, which monitors the speed of the vehicle over a period of time and is mandatory on large vehicles.

Figure 10: Effects of a Mandatory Break Policy



(a) Duration of Time Between Breaks

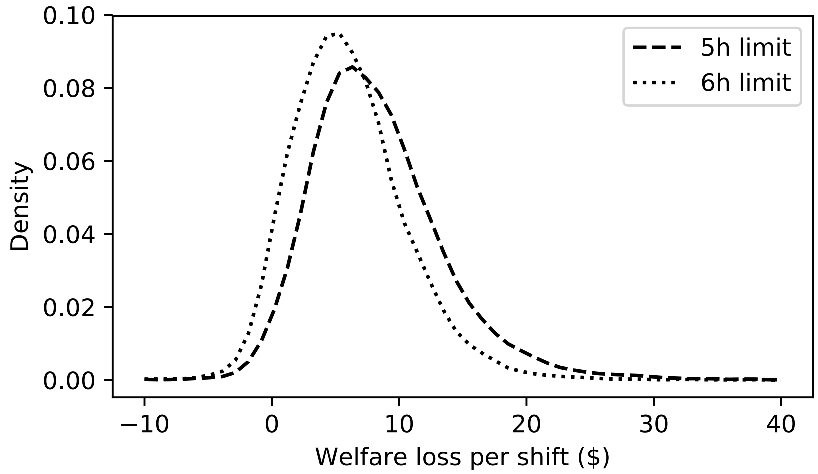
(b) Working Hours

The different thresholds for the policy I consider are 5 and 6 hours. After reaching the threshold, the driver must either take a break or end his shift. In this counterfactual experiment, I use the estimated model to simulate the decisions of each driver. As a starting point, I draw starting values for the state space from the distribution of start times of the same driver. This means that I will be shutting down this potential margin of adjustment.

As we have seen above, taxi drivers value the ability to decide when to take breaks highly. A mandatory break policy, by construction, removes some of the discretion over when to take a break. However, this policy does not imply that drivers will wait until the end of their allowed uninterrupted work time to take a break. In this counterfactual experiment, a driver will take a break before reaching the limit if he receives a large shock to the utility of a break (e.g. dropping off a customer at the airport in previous period).

Following the introduction of the policy, I estimate a significant drop in driver fatigue levels. Figure 10a shows the distribution of average duration of uninterrupted work. There is a clear shift to the left, indicating that the frequency of breaks becomes higher. The difference between the 5-hour and 6-hour policy is rather small. Figure 10b shows the distribution of labor supply. We observe a reduction in labor supply that

Figure 11: Welfare Loss from Mandatory Break Policy, per Shift



is magnified by the overall shift duration which is slightly lower (3.4 percent for the 6-hour limit and 4.4 percent for the 5-hour limit). Overall, labor supply, measured by the number of periods worked, decreases by 6.4 percent for the 6-hour limit and 8.5 percent for the 5-hour limit.

Because this policy places a restriction on taxi drivers, welfare can be affected. I compute welfare with the same method explained in Section 7.1. The results can be found in Figure 11. Even though a mandatory break policy would still leave a lot of flexibility, the average driver would experience a reduction in welfare equivalent to 2.1% to 3.0% of daily revenue.

These results highlight the importance of understanding the incentives at play for taxi drivers. While a 'mandatory breaks' policy will increase the frequency at which drivers take break, they also significantly reduce the drivers' labor supply.

8 Conclusion

The labor market is currently undergoing a profound structural transformation as we move away from traditional employer-employee relationships toward a decentralized marketplace. In this context, this paper seeks to understand how workers decide when to take a break and, more importantly, how much to value this flexibility. Extending the work of Chen et al. (2017) and Fréchette et al. (2018), I develop the first dynamic model of labor supply with discretionary breaks. The model uses a dynamic discrete choice framework and captures worker fatigue, hour-of-the-day effects, fixed costs, and potential earnings. To estimate the model, I use high-frequency data from NYC taxi drivers. The richness of the data allows me to account for unobserved heterogeneity in a very flexible way.

I find that taxi drivers value discretionary breaks highly: they would require a 23 percent increase in daily revenue in order to be induced to accept a fixed break schedule. Furthermore, I explore the effects of a mandatory break policy that limits the number of uninterrupted hours of work. While the policy achieves its goal of increasing the frequency of breaks, we observe a significant reduction in labor supply of 6.4 to 8.5 percent. This result is driven by the taxi drivers taking more breaks *and having shorter shifts.*

These results have broad labor market implications. The high value placed on discretionary breaks may explain the rapid growth of the 'gig' economy in recent years or the move of many employers towards greater employee flexibility. Moreover, quantifying the value of this non-pecuniary benefit is an understudied way to improve the efficiency of labor contracts. The results presented in this paper suggest that some employees are willing to accept reduced wages if they are compensated with flexibility.

The framework developed in this paper is a general model of labor supply with discretionary rest breaks. While the estimation strategy of this paper is performed in a setting with millions of observations, datasets containing information on labor supply decisions are getting larger, more detailed, and span more sectors than ever before, broadening the applicability of my framework. This paper also sets the path for future research focusing on the interaction between discretionary breaks and labor supply decisions over a medium-horizon, such as weekly or monthly hours. Finally, this paper highlights the need for labor economists to account for non-traditional work arrangements as they become more and more prevalent.

References

Abraham, K. G., J. C. Haltiwanger, K. Sandusky, and J. R. Spletzer (2017): "Measuring the Gig Economy: Current Knowledge and Open Issues," Working Paper.

Aguirregabiria, V. and P. Mira (2010): "Dynamic Discrete Choice Structural Models: A Survey," *Journal of Econometrics*, 156, 38–67.

Arcidiacono, P. and J. B. Jones (2003): "Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm," *Econometrica*, 71, 933–946.

Arcidiacono, P. and R. A. Miller (2011): "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity," *Econometrica*, 79, 1823–1867.

Baker, M. and A. Melino (2000): "Duration Dependence and Nonparametric Heterogeneity: A Monte Carlo Study," *Journal of Econometrics*, 96, 357–393.

Borjas, G. J. (1980): "The Relationship between Wages and Weekly Hours of Work : The Role of Division Bias," *Journal of Human Resources*, 15, 409–423.

Brachet, T., G. David, and A. M. Drechsler (2012): "The Effect of Shift Structure on Performance," *American Economic Journal: Applied Economics*, 4, 219–246.

Buchholz, N. (2018): "Spatial Equilibrium, Search Frictions and Dynamic Efficiency in the Taxi Industry," Working Paper.

Buchholz, N., M. Shum, and H. Xu (2018): "Dynamic Labor Supply of Taxicab Drivers: A Semiparametric Optimal Stopping Model," Tech. rep.

Camerer, C., L. Babcock, G. Loewenstein, and R. Thaler (1997): "Labor Supply of New York City Cabdrivers: One Day at a Time," *Quarterly Journal of Economics*, 112, 407–441.

Chen, M. K., J. A. Chevalier, P. E. Rossi, and E. Oehlsen (2017): "The Value of Flexible Work: Evidence from Uber Drivers," Working Paper 23296, National Bureau of Economic Research.

Chou, Y. K. (2002): "Testing Alternative Models of Labour Supply: Evidence from Taxi Drivers in Singapore," *Singapore Economic Review*, 47, 17–47.

Collewet, M. and J. Sauermann (2017): "Working Hours and Productivity," *Labour Economics*, 47, 96–106.

Cook, C., R. Diamond, J. Hall, J. A. List, and P. Oyer (2018): "The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers," Working Paper.

Crawford, V. P. and J. Meng (2011): "New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income," *American Economic Review*, 101, 1912–1932.

Dalziel, J. R. and R. F. S. Job (1997): "Motor Vehicle Accidents, Fatigue and Optimism Bias in Taxi Drivers," *Accident Analysis & Prevention*, 29, 489–494.

Farber, H. S. (2005): "Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers," *Journal of Political Economy*, 113, 46–82.

——— (2008a): "Employment Insecurity: The Decline in Worker-Firm Attachment in the United States," Working Paper 530.

——— (2008b): "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," *Quarterly Journal of Economics*, 130, 1975–2026.

Fehr, E. and L. Goette (2007): "Do Workers Work More If Wages Are High? Evidence from a Randomized Field Experiment," *American Economic Review*, 97, 298–317.

Finer, D. A. (2018): "What Insights Do Taxi Rides Offer into Federal Reserve Leakage?" Working Paper.

Fréchette, G. R., A. Lizzeri, and T. Salz (2018): "Frictions in a Competitive, Regulated Market: Evidence from Taxis," Working Paper 24921, National Bureau of Economic Research.

Goldmark, J. and L. D. Brandeis (1912): *Fatigue and Efficiency: A Study in Industry*, Charities publication committee.

Haggag, K. and G. Paci (2014): "Default Tips," *American Economic Journal: Applied Economics*, 6, 1–19.

Hall, J. V., J. J. Horton, and D. T. Knoepfle (2017): "Labor Market Equilibration: Evidence from Uber," Working Paper.

Henning, R. A., P. Jacques, G. V. Kissel, A. B. Sullivan, and S. Alteras-Webb (1997): "Frequent Short Rest Breaks from Computer Work: Effects on Productivity and Well-Being at Two Field Sites," *Ergonomics*, 40, 78–91.

Hideg, I. and J. P. Trougakos (2009): "Momentary Work Recovery: The Role of within-Day Work Breaks," in *Current Perspectives on Job-Stress Recovery*, Emerald Group Publishing Limited, vol. 7 of *Research in Occupational Stress and Well-being*, 37–84.

Jackson, E., A. Looney, and S. Ramnath (2017): "The Rise of Alternative Work Arrangements: Evidence and Implications for Tax Filing and Benefit Coverage," Office of Tax Analysis Working Paper 114.

Jett, Q. R. and J. M. George (2003): "Work Interrupted: A Closer Look at the Role of Interruptions in Organizational Life," *The Academy of Management Review*, 28, 494–507.

Katz, L. F. and A. B. Krueger (2016): "The Rise and Nature of Alternative Work Arrangements in the United States, 1995-2015," Working Paper 22667, National Bureau of Economic Research.

Magnac, T. and D. Thesmar (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70, 801–816.

Mangrum, D. and A. Molnar (2017): "The Marginal Congestion of a Taxi in New York City," Working Paper, 00000.

Mas, A. and A. Pallais (2017): "Valuing Alternative Work Arrangements," *American Economic Review*, 107, 3722–3759.

Oettinger, G. S. (1999): "An Empirical Analysis of the Daily Labor Supply of Stadium Vendors," *Journal of Political Economy*, 107, 360–392.

Oyer, P. (2016): "The Independent Workforce in America: The Economics of an Increasingly Flexible Labor Market," Technical Report.

Pencavel, J. (2015): "The Productivity of Working Hours," *The Economic Journal*, 125, 2052–2076.

——— (2016): "Recovery from Work and the Productivity of Working Hours," *Economica*, 83, 545–563.

Pendem, P., P. Green, B. R. Staats, and F. Gino (2016): "The Microstructure of Work: How Unexpected Breaks Let You Rest, But Not Lose Focus," SSRN Scholarly Paper ID 2888477, Social Science Research Network, Rochester, NY.

Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55, 999–1033.

——— (1994): "Structural Estimation of Markov Decision Processes," in *Handbook of Econometrics*, Elsevier, vol. 4, 3081–3143.

Schmidt, M.-A. (2018): "The Daily Labor Supply Response to Worker-Specific Earnings Shocks," Working Paper.

Stafford, T. M. (2015): "What Do Fishermen Tell Us That Taxi Drivers Do Not? An Empirical Investigation of Labor Supply," *Journal of Labor Economics*, 33, 683–710.

Stanford, J. (2017): "The Resurgence of Gig Work: Historical and Theoretical Perspectives," *The Economic and Labour Relations Review*, 28, 382–401.

Taxi and Limousine Commission (2014): "2014 Taxicab Factbook," Tech. rep.

Thakral, N. and L. T. Tô (2017): "Daily Labor Supply and Adaptive Reference Points," Working Paper.

Tucker, P. (2003): "The Impact of Rest Breaks upon Accident Risk, Fatigue and Performance: A Review," *Work & Stress*, 17, 123–137.

Van Den Berg, G. J. (2001): "Duration Models: Specification, Identification and Multiple Durations," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. Leamer, Elsevier, vol. 5, 3381–3460.

Vernon, H. M. (1921): *Industrial Fatigue and Efficiency*, G. Routledge.

Appendix A Data Construction and Cleaning

A.1 Data Cleaning Procedures

I conduct several data cleaning procedures to ensure that the results are not driven by measurement errors or outliers. I describe them in this section.

- I flag the trips that have faulty location data or that are not located within NYC 5 boroughs or in New Jersey.
- I flag trips that have negative or zero fare.
- I flag trips that end after they start.
- I flag trips that start before the previous trip has ended.

After aggregating trips into shifts, I remove the entire shift if it contains a flagged trip. I do this because only removing the problematic trip would create a possibility for a false break. Using the resulting dataset, I compute the market wage and the average search time by region without any further restrictions. For the final analysis, I remove the shifts that are outliers following these rules:

- Shifts shorter than 3 hours.
- Shifts longer than 12 hours.[24]
- Drivers with fewer than 75 observed Mon-Thu day shifts or 75 observed Mon-Thu night shifts.

I also further restrict the analysis to drivers exhibiting behavior suggesting they rent the medallion from a taxi garage.

[24] A shift longer than 12 hours suggests the driver is not bound by the regular renting agreement.

A.2 Market Hourly Earnings

I construct the measure of market hourly earnings in a way that is similar to Thakral and Tô (2017).

The measure of hourly earnings depends on how many drivers are working. Drivers that are not working are not included in the measure of average hourly earnings. Each trip is associated with the preceding wait time. As mentioned in the paper, we can categorize wait times into three groups: search time, breaks, and time off work. For the purpose of computing the average hourly earnings, only the search time is relevant. I impute an average search time of 7 minutes to the first trip of the shift and to the first trip after a break.

The measure of revenue per minute from a trip i can be formalized as

$$R_i = \frac{F_i}{S_i + T_i},$$

where F_i is the total fare paid by the customer, S_i is the amount of time the driver searched for the customer, and T_i is the amount of time the trip lasted.

For every minute that the driver was either searching for the customer of trip i or on trip i, R_i will contribute to the measure of average earnings. The average hourly earnings in minute m can then be defined as 60 multiplied by the average of all R_i that are active during this minute.

I construct the measure of hourly earnings for every minutes of the year (each 525,600 of them). When constructing the potential earnings I_t, I compute the average of the next 30 minutes.

A.3 NYC Neighborhoods and Average Search Time

One major solution to correctly identify breaks is to control for market demand. I do this by computing the average search time in an area. In this section, I describe the methodology for doing so.

The simplest way of computing a measure of average search time would be to compute the average search time of all pickup in an area. One issue with this methodology arises because it combines drivers that started looking for a customer directly after a dropoff and drivers that drove to the location. To get a precise measure of search time, I compute search time in a location using only pickups that follow a dropoff in the same neighborhood.

I define a neighborhood in my analysis as a 'community board'[25] plus some large non-residential areas, such as Laguardia Airport, JFK Airport, and Central Park. I also create a zone in New Jersey to capture trips going to Newark Airport or Jersey City. The size and location of each neighborhood can be seen on Figure A6.

If I relied only on pickups during the same hour of the year to estimate the average search time, sample error would be a large issue. It is not rare to observe location-hour pairs with less than 20 observations. To solve this, I instead aggregate the yearly data at the weekly level so that for each location-hour pair, I have about 52 times more observations. This generates a measure of average search time for a maximum of 12,096 location–hour-of-the-week pairs.

The resulting distribution of search time for two specific hours of the week is presented in Figure A7. Not all neighborhood have data. Because more than 93% of all pickups are located in Manhattan or at the airports, some neighborhoods do not have enough observations to consistently estimate an average search time. I drop the shifts for which I do not have data on average search time for a trip.

Appendix B Labor Supply Elasticity

In Section 4.4, I described how the estimates of the labor supply elasticity was affected by the inclusion of breaks in the measure of labor supply. Here, I explain in more detail the underlying model and why an instrumental variable approach is required.

[25] Community boards are comprised of volunteers and only act in an advisory capacity. They can advice the authorities on zoning or services delivery in their community.

One of the strategies used by the daily labor supply literature has been to estimate a regression of hourly earnings on hours worked.[26] The regression equation is:

$$\ln(H_{is}) = \delta \ln\left(E_{is}/H_{is}\right) + \beta X_{is} + \mu_i + \nu_{is}, \tag{5}$$

where H_{is} is the duration of shift s for worker i; (E_{is}/H_{is}) is the hourly earnings; X_{is} are covariates such as date, time, or weather; and μ_i is a driver fixed effect. The labor supply elasticity is measured by δ. There are two issues with this strategy: First, anticipated variation in hourly earnings cannot be used to identify this elasticity because it is possible that workers wanting to work longer shifts only do so when the hourly earnings are higher. Second, the long recognized problem of division bias (Borjas, 1980) is present in this problem because H_{is} appears in both the RHS and the LHS of the equation.

To address the first issue, I follow the literature and use a comprehensive set of controls which leaves only unanticipated hourly earnings variations to identify the labor supply elasticity. The controls include hour-of-the-week dummies, holidays, month of the year, precipitation, temperatures below 0 degree Celsius, and modal neighborhood. The second issue is addressed by instrumenting the hourly earnings of the driver by the average hourly earnings of other drivers with an overlapping shift. To construct the instrument, I use a random sample of 1/3 of the drivers, while the other 2/3 are used for the estimation.

Because elasticities are computed in percentages, reducing labor supply by a constant amount would also increase the elasticity. A 30-minute increase over 7.5-hour shift is proportionately larger than a 30-minute increase over a 9-hour shift. However, we can do a simple back-of-the-envelope calculation to compute how much this would mechanically affect the elasticity. For simplicity, suppose the average shift is 9 hours and drivers take on average 1 hour of break per shift. The elasticity of 0.256 in Table 1 means that a 10% increase in wage will result in a shift 13.8 minutes longer. When breaks are taken

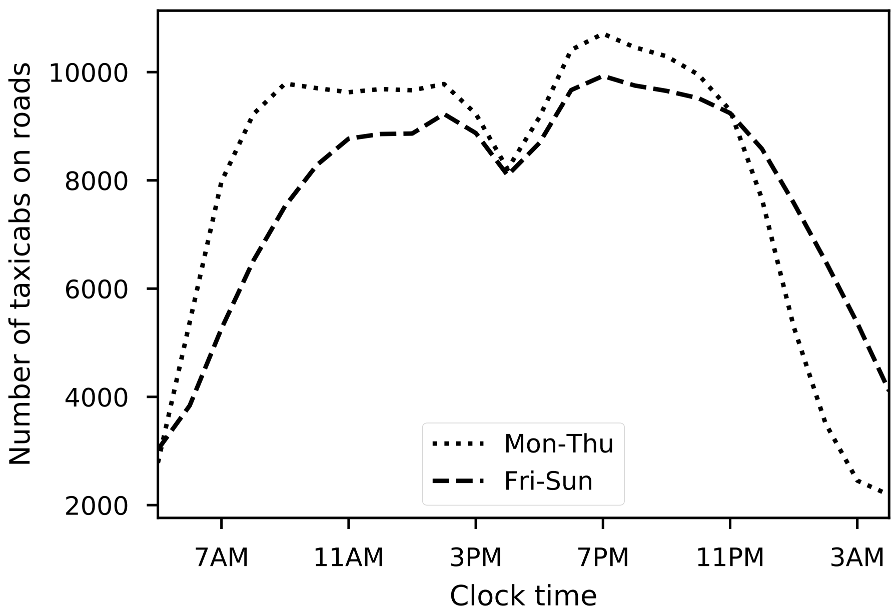
[26] This strategy has been employed by Camerer et al. (1997), Chou (2002), Farber (2015), and Schmidt (2018). Here I replicate this methodology to get a sense of how the parameters would vary.

into account, the average labor supply is now 8 hours. The 13.8 minutes increase is now equivalent to 2.88% (elasticity of 0.288) instead of 2.56%. This mechanical relationship could only be able to explain a difference of 0.032 between the elasticities, representing only 1/20th of observed difference.

As a comparison to the instrumental variable approach, the OLS regression is presented in Table A2. We can see rather than attenuating the negative labor supply elasticity of Panel A, netting out the breaks reduces even more the estimates. This can be explained by the fact that we introduce another variable in both the RHS and the LHS.

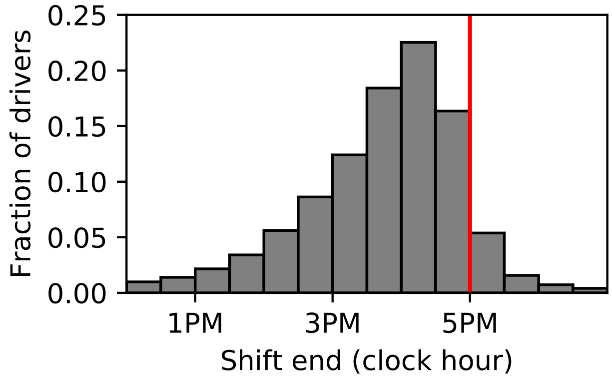
Appendix C Additional Figures and Tables

Figure A1: Number of Taxicabs on the Road, by Clock Time

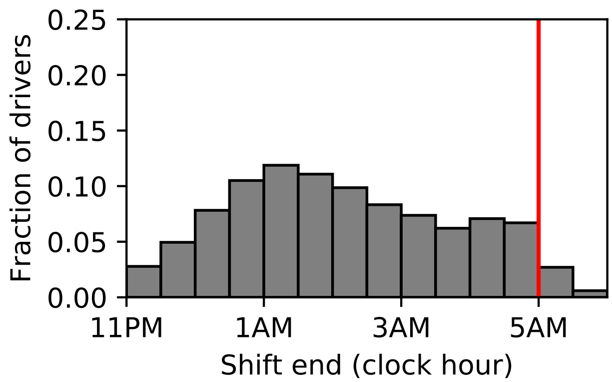


Notes: This plot contains the supply of taxicabs on the road at each hour of the day. This is computed by adding the number of unique medallions that were active (picked up customer) during a particular hour. The fall in supply in the middle of the day is due to the transition time between the day-shift and night-shift drivers. The hours from midnight to 5 AM are associated with the previous day to match the drivers' schedule.

Figure A2: Distribution of End Time of Shifts



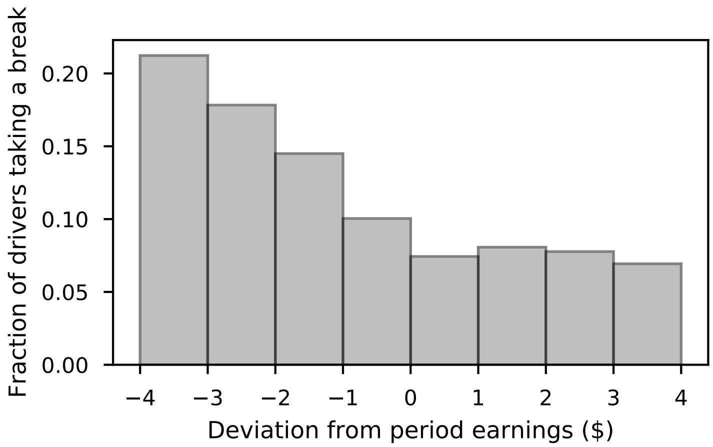
(a) Day-Shift Drivers



(b) Night-Shift Drivers

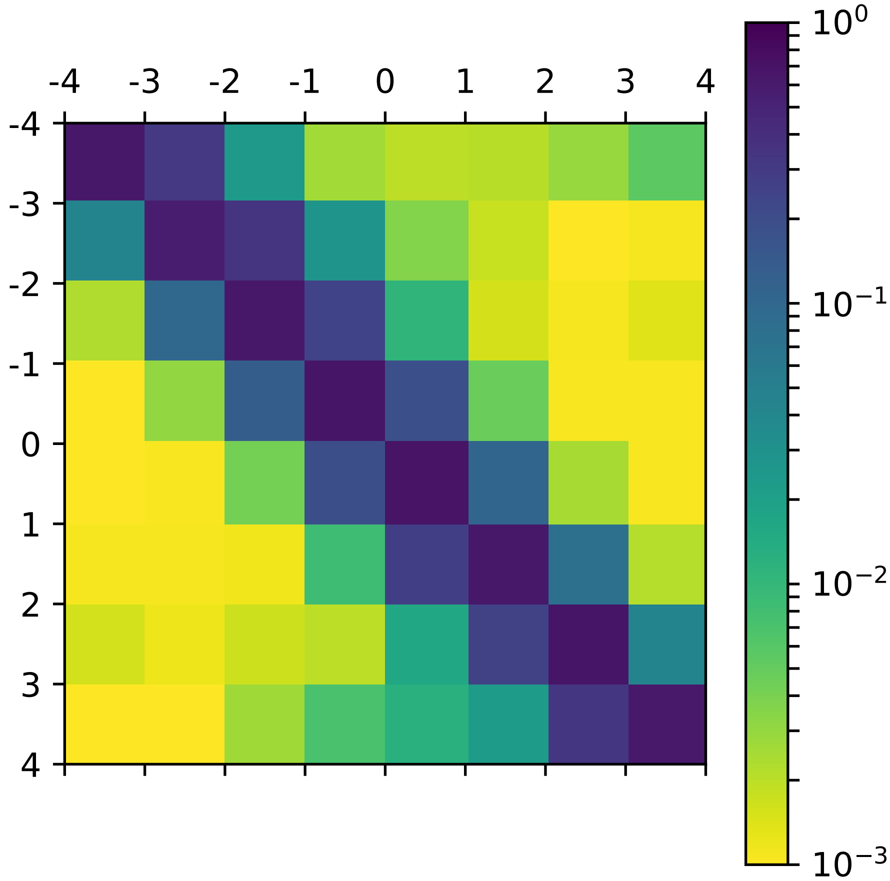
Notes: The end time is the start of the period in which the driver ended his shift. The vertical lines represent the usual transition time (5 AM or 5 PM). Because the drivers need to bring the car back to the garage, typically located in Queens or in Brooklyn, the end of their shift is usually 45 minutes earlier.

Figure A3: Probability of Taking a Break, by Deviation from Usual Earnings



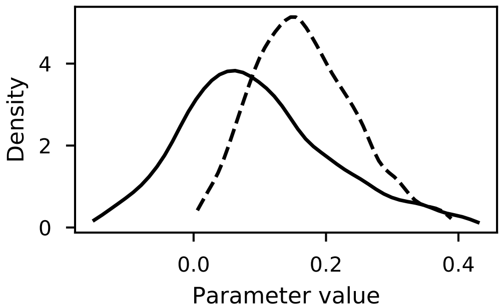
Notes: Each bin represents a $1 increment in the deviation from anticipated earnings. See Section A.2 for a more detailed explanation of the way market hourly earnings are constructed.

Figure A4: Markov Transition Matrix of Hourly Earnings

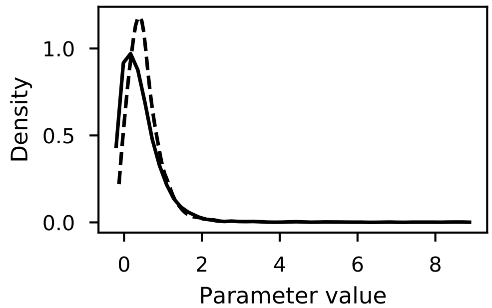


Notes: The transition probabilities between period t (y-axis) and period $t+1$ (x-axis) is represented graphically. Because there is persistance, most of the probability mass is located on the diagonal. Notice that the color scheme was normalized to a log-scale to show the out-of-diagonal values more clearly.

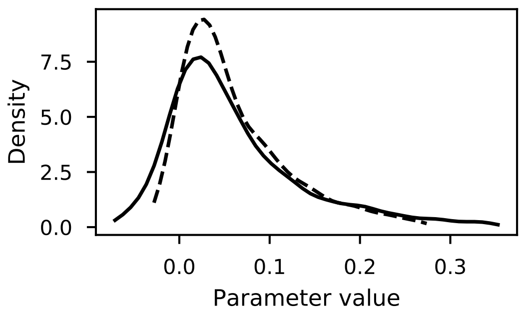
Figure A5: Distribution of Parameter Estimates



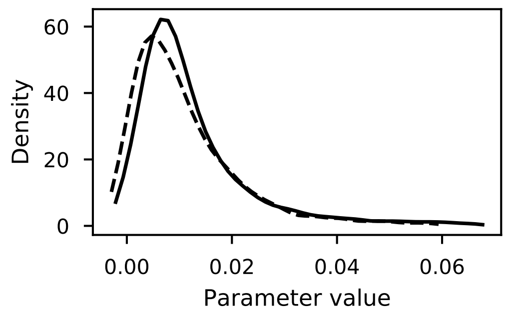
(a) Potential Earnings (γ)



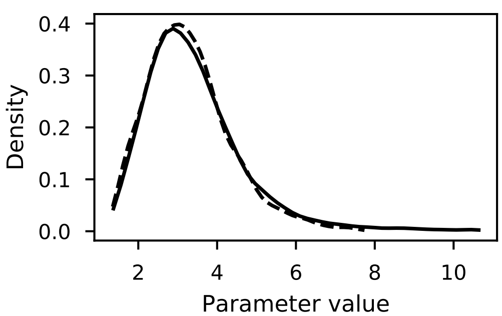
(b) Rate of Break Util. Decline (ψ)



(c) Non-Recov. Fatigue (π_n)



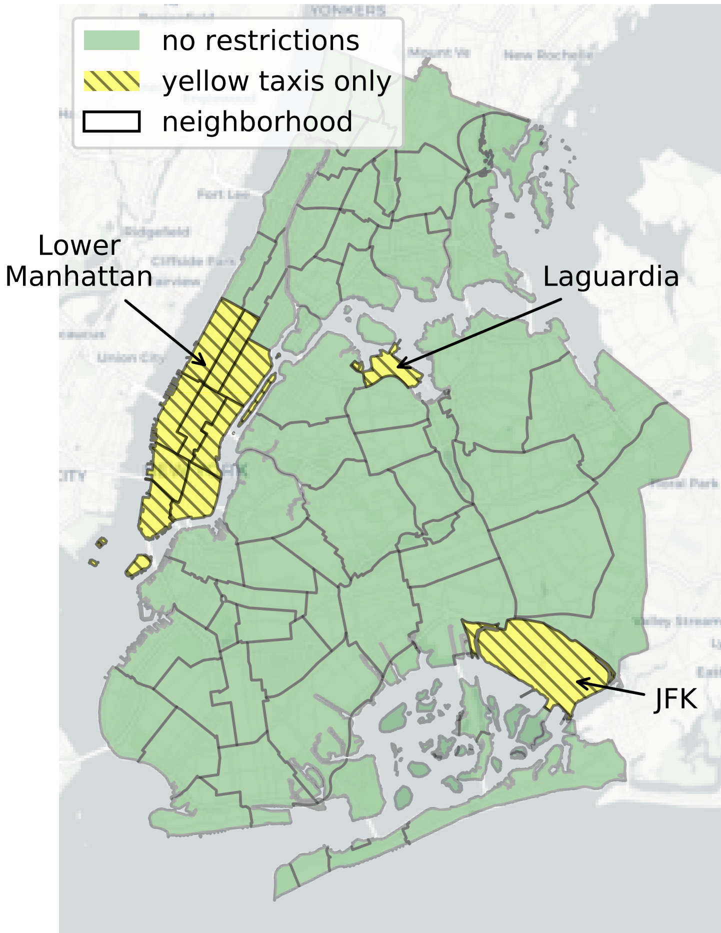
(d) Recov. Fatigue (π_r)



(e) Fixed Cost (τ)

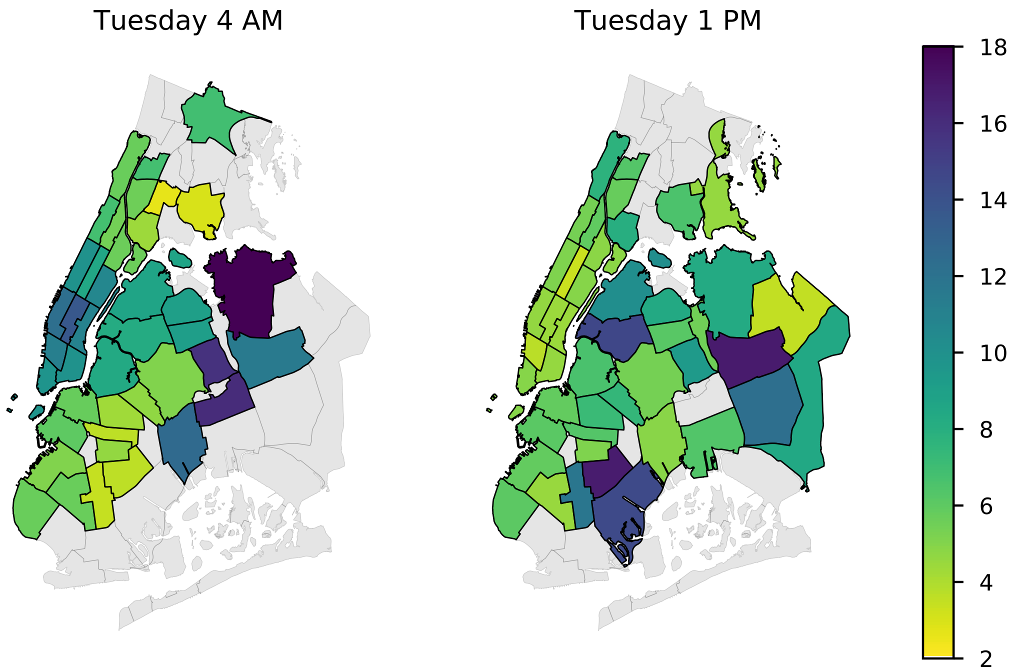
Notes: The PDF are estimated using a non-parametric kernel density estimation technique with a Gaussian kernel. The solid line plot the distribution for day-shift drivers and the dashed line the distribution for night-shift drivers.

Figure A6: Map of NYC with Neighborhoods



Notes: The neighborhoods referred to in this map are called 'community boards' by NYC officials. The agglomeration of 'yellow taxis only' neighborhoods on the left is the southern portion of Manhattan. On the right of Manhattan, the other 'yellow taxis only' zone is Laguardia Airport. The 'yellow taxis only' zone in the bottom right is JFK Airport.

Figure A7: Average Search Time



Notes: The average search time (in minutes) in each region is computed using only observations with the last dropoff and the next pickup in the same region. Each region is a 'community board.' Further explanation is given in Section A.3. The locations of the two airports have been removed because, as explained in Section 3.3.2, I set search time at airports to zero.

Table A1: Summary Statistics

	Selected Drivers			Non-Selected Drivers		
	mean	median	std. dev.	mean	median	std. dev.
Shifts per year	248.2	253	54.8	154.5	151	101.8
Shift length (hours)	8.6	8.8	2.1	9.0	9.0	2.6
Trips per shift	21.8	22	7.8	22.2	22	8.7
Earnings per shift	269.4	266.5	81.0	285.1	280.0	93.0
Hourly Earnings	31.4	31.4	6.0	32.1	32.2	6.5
Number of drivers		14,190			26,079	

Notes: Selected drivers display behaviors that indicates they are renting their medallion with a taxi garage and follow the 5 AM and 5 PM transition times. In general, the selected drivers have a lower level of heterogeneity. Non-selected drivers either do not display behaviors suggesting they rent or they are occasional or irregular drivers and do not appear in enough shifts for a consistent estimation of the model.

Table A2: OLS elasticity estimates

	All	Day Shift	Night Shift	Owner
	(1)	(2)	(3)	(4)
Panel A: Gross shift				
log Hourly Earnings	-0.278^{***}	-0.131^{***}	-0.285^{***}	-0.359^{***}
	(0.003)	(0.003)	(0.003)	(0.008)
Panel B: Adjusted shift (net of breaks)				
log Hourly Earnings	-0.480^{***}	-0.462^{***}	-0.507^{***}	-0.544^{***}
	(0.002)	(0.003)	(0.003)	(0.007)
Driver	Yes	Yes	Yes	Yes
Weather	Yes	Yes	Yes	Yes
Location	Yes	Yes	Yes	Yes
Date/Time	Yes	Yes	Yes	Yes
Observations	4,894,002	2,148,223	2,151,782	566,344

Notes: Clustered standard error in parentheses (driver level). Controls include weather (temperature and precipitation), location fixed effects (modal pickup neighborhood; 72), holiday fixed effects (9), and fixed effects for the month of the year (11) and the hour of the week (167).