RETHINKING REFERENCE DEPENDENCE: WAGE DYNAMICS AND OPTIMAL TAXI LABOR SUPPLY

NICHOLAS BUCHHOLZ*, MATTHEW SHUM†, AND HAIQING XU‡

ABSTRACT. Workers with variable earnings and flexible hours offer unique opportunities to evaluate intertemporal labor supply elasticities. Existing static analyses, however, have generated well-known puzzles, suggesting evidence of downward sloping labor supply curves. Using a large sample of shifts of New York City taxicab drivers, we estimate a dynamic optimal stopping model of drivers' work times and quitting decisions. Our analysis demonstrates that patterns previously interpreted as behavioral biases arise from rational, forward-looking optimization. We use our model to provide new estimates of individual earnings elasticities and show that taxi drivers have similar elasticities to workers in markets where experimental evidence has been obtained. Finally, we demonstrate that market-level labor supply responses to fare changes are much smaller than individual-level responses due to equilibrium effects. This finding suggests that recent estimates of the benefits to recent earnings legislation in the taxi and ride-hail industries are overstated.

Keywords: Labor supply elasticity, Optimal stopping, Dynamic discrete choice, Taxi industry

JEL: J20, D91, C41, L91

Date: Monday 10th February, 2025.

The authors are grateful to Laura Doval, Hank Farber, Kate Ho, Adam Kapor, Jakub Kastl, Henrik Kleven, Ricardo Lagos, Alessandro Lizzeri, Aniko Öry, Ariel Pakes, Rob Porter, Eric Richert, Richard Rogerson, Nicola Rosaia, Jesse Shapiro, Neil Thakral, and Kevin Williams for helpful comments and suggestions.

*Department of Economics, Princeton University, nbuchholz@princeton.edu.

†Division of Humanities and Social Sciences, California Institute of Technology, mshum@caltech.edu.

‡Department of Economics, University of Texas at Austin, h.xu@austin.utexas.edu.

1. Introduction

Workers with variable earnings and flexible hours offer unique opportunities to evaluate intertemporal labor supply elasticities. Using detailed trip-level data from New York City taxicab drivers, we document two novel empirical patterns that resolve longstanding puzzles in the labor supply literature. First, drivers take more frequent breaks as their shifts progress, leading to mechanical declines in average hourly earnings and a spurious correlation between hours worked and wages. Second, high-paying trips to outer boroughs are systematically followed by periods of low earnings due to reduced passenger density in those areas, leading to negative serial correlation in hourly wages. These patterns explain why previous static analyses have generated well-known puzzles, suggesting evidence of downward sloping labor supply curves.

We formalize these insights through a dynamic optimal stopping model of drivers' work times and quitting decisions. Our data-driven modeling approach allows us to avoid imposing assumptions about the equilibrium search and matching process while still capturing the rich patterns of serial correlation in earnings. The estimated model demonstrates that once we account for forward-looking driver behavior, the empirical patterns previously attributed to behavioral biases can be fully rationalized within a neoclassical framework. We further use our model to provide new estimates of individual earnings elasticities and show that taxi drivers exhibit similar elasticities to workers in markets where experimental evidence has been obtained.

Studies using observational data aggregated across multiple industries and spanning several years have encountered challenges in measuring labor supply elasticities because observed wage changes typically do not hold all else equal. As access to rich micro-data has become more available, economists have turned to flexible-work settings, in which workers have the ability to immediately adjust their work time in response to changes in earnings opportunities. Such settings offer a clearer path to estimating substitution effects. Yet, the literature has raised several new questions of methodology and interpretation and has therefore not settled on how to obtain the elasticities of interest. A seminal paper is Camerer, Babcock, Loewenstein and Thaler (1997), which finds evidence for negative income elasticities of labor supply among taxi drivers – that is, drivers appear to work fewer hours when faced with higher wage rates. Such a finding is inconsistent with textbook neoclassical labor supply models and instead congruent with alternative "behavioral" labor supply models, such as income targeting, in which agents with flexible hours set an income target and work until the target is reached. This work continues to be highly influential and widely cited in labor economics and public finance, generating a large follow-up literature that has corroborated, expanded or challenged these findings (Ashenfelter, Doran and Schaller (2010), Duong, Chu and Yao (2022), Farber (2005, 2008, 2015), Crawford and Meng (2011), Thakral and Tô (2021)).

Our first contribution is that our approach explains multiple puzzles in the literature: both the negative wage elasticities, as recovered in Camerer *et al.* (1997) as well as the fact that drivers appear more likely to quit after earnings shocks late into the shift, a "recency" bias documented in Thakral and Tô (2021). In the case of Camerer *et al.* (1997), in which authors analyze end-of-shift data and regress hours worked on average daily wages, we find that the declining relative productivity of drivers on longer shifts mechanically leads to negative wage elasticities; a driver with a longer shift will, on average, end the day at a lower productivity state than a similar driver who ends earlier, generating a correlation between low average earnings per hour and longer work hours. In the case of Thakral and Tô (2021), the authors regress drivers' binary decisions to quit on cumulative earnings interacted with the time of day in which the earnings accrued. Their estimates show that drivers exhibit a higher probability of quitting associated with late-in-shift earnings shocks. We find that the particularly large negative autocorrelation of earnings on long trips, for example trips from the New York boroughs of Manhattan to Queens or Bronx, produces these effects. Long trips appear in the data as earnings shocks and lead to subsequently low expected future earnings. Viewed through the lens of our dynamic stopping model, a positive current earnings shock imposes a less attractive future path of earnings relative to the time cost of continuing to work, leading to a higher probability of quitting for the day. We validate our model by showing that it is capable of generating the patterns in the data that have been attributed to behavioral explanations. Our results therefore demonstrate that the apparent non-standard behavior can be rationalized as optimal decision making.

A second contribution is to provide methodology to estimate the short-run inter-temporal substitutability of labor and leisure, or the Frisch elasticity of labor supply, from observational data of dynamically optimizing agents. Our approach treats drivers' labor supply decisions as emerging from an optimal stopping problem, where drivers weigh expected future earnings against the opportunity costs of continuing to work. While reduced-form approaches could attempt to control for forward-looking behavior through various proxy controls, this is inherently difficult since forward-looking behavior involves a value function that recursively depends on agents' optimal decisions. Our dynamic model addresses this difficulty by providing a systematic framework that accounts for how drivers incorporate the forward-looking continuation value into their decisions. While our model is consistent with equilibrium models of search, matching and spatial sorting in the taxi industry, it offers a simpler and more implementable approach that avoids many additional modeling assumptions. Our specific research questions do not require treating most aspects of these equilibrium outcomes as endogenous. Instead, we are able to specify the model in a tractable and parsimonious way for our counterfactuals of interest, allowing us to estimate driver preferences as a single-agent dynamic problem.

We use our estimated model to explore a key policy question: how do regulated fare increases affect driver earnings? This analysis serves two purposes. First, it demonstrates that labor supply elasticities with respect to market-wide price changes are distinct from elasticities with respect to individual earnings variation - a distinction overlooked in some of the literature. Second, it shows how our framework can inform policy decisions, such as the New York Taxi and Limousine Commission's recent fare increases designed to raise driver earnings. The Commission estimates that a January 2023 fare change, raising trip prices by 23%, will lead to a 33% increase in driver earnings. This estimate necessarily implies that labor supply elasticities with respect to fares must be greater than 0.40. We evaluate this elasticity by using data spanning a comparable change, a fare hike that raised average prices by 18% in September 2012. Our methodology allows us to analyze these effects without explicitly modeling the interaction of supply and demand, instead exploiting changes in earnings transition densities before and after the policy. We find that the elasticity of labor supply with respect to market prices is approximately 0.12, around one sixth of our overall Frisch elasticity estimates. This implies that the benefits to individual drivers of a large price increase will be substantially less than predicted by regulators.

Literature. A growing literature has arisen aiming to estimate labor supply elasticities in markets where labor supply is continuously adjustable. Several of these papers have studied the taxi industry, because taxi drivers are typically able to choose their own hours. Moreover, as automated data collection has been implemented to meet regulatory standards, detailed trip data has become available in some of the largest U.S. taxi markets.[1] Camerer *et al.* (1997) analyzes hand-collected data from New York City and considers the hours worked during individual driver shifts. The authors conduct a series of regressions of log hours worked on the log of average daily wages and find evidence for negative wage elasticities. The authors argue that negative elasticities are consistent with income-targeting on the part of drivers: for example, a labor supply policy of the form "I will work today until I earn $200." Farber (2005, 2008, 2015) consider static optimal stopping models of labor supply. The first paper develops a stopping rule model which explores similar forces to our model, showing that drivers' stopping is most reliably predicted by hours instead of income. The latter two papers integrate reference-dependent utility, which is the notion that agents' utility is not only a function of income but also reference-points or targets, where the marginal utility of income increases more quickly before the target is met than after it is met. While Farber (2008) finds mixed evidence for the existence of reference-dependence, Farber (2015) uses more comprehensive data and finds stronger evidence that drivers have, on average, upward sloping earnings elasticities. Nevertheless, Farber finds that just under a third of drivers exhibit behavior consistent with negative elasticities. Using data on taxi inspections, Ashenfelter *et al.* (2010) finds that drivers who worked before and after fare hikes tended to work slightly less on average afterwards, suggesting a small negative earnings elasticity.[2] Our paper highlights new data evidence that drivers' relative earnings productivity tends to decline with hours worked. This fact generates a downward bias in the elasticity estimate in the wage regressions due to a selection effect, one which may reconcile these

[1]There is also an older empirical literature that recovers the elasticities of intertemporal substitution as a part of estimating lifecycle models of labor supply, for example Heckman and MaCurdy (1980), Browning *et al.* (1985), MaCurdy (1981), and Altonji (1986). The data used are highly aggregated and often cross-sectional and cross-industry in nature. These studies generally predict small positive labor supply adjustments as a result of increased wage rates. Within the literature, authors regularly highlight significant data limitations and modeling assumptions: for example one must assume that workers are free to choose their own hours and that wage variation is exogenous, which is unlikely to hold in the analyzed settings. It is also a challenge to separate wealth effects from substitution effects, even when long-run panels are used. Nevertheless, our work corroborates the findings of positive substitution elasticities but at a much finer, intra-daily scale instead of workers' lifecycle.

[2]There are additional studies outside of the context of taxis which consider related questions: Fehr and Goette (2007) demonstrate positive labor supply elasticities in an experiment providing higher payments to bicycle messengers. Andersen, Brandon, Gneezy and List (2014) also show positive labor supply elasticities in an experiment among market vendors in India, explicitly testing for and rejecting reference-dependence. Oettinger (1999) documents an equilibrium increase in labor effort on high demand days among stadium vendors.

disparate findings. We further show through our model that the key puzzles are reproducible when simulating data directly from the model.

A newer thread in the literature presents evidence that drivers' labor supply behavior is dependent on the time of day in which they earn revenue. Crawford and Meng (2011) specifies and estimates a dynamic model of labor supply incorporating reference-dependence in both income and hours-worked during a shift, finding that drivers' types of reference-dependence depends on whether their earnings are high or low early in the shift relative to their long-run average. Thakral and Tô (2021) also takes up the question of whether there is a timing dimension to behavioral biases in drivers stopping decisions, showing that more recent income is a stronger determinant of quitting than income earned earlier in a shift. Our paper proposes an explanation for this behavior by showing that earnings shocks are associated with subsequently negative earnings opportunities. Incorporating this fact into our model, we show that we can reproduce the apparent time-inconsistent behavior.

We also contribute to a literature on structural models of labor supply and market equilibrium in taxi and ride-hail settings. Our model is closely related to the taxi labor supply model of Frechette, Lizzeri and Salz (2019), in which taxi drivers decide how long to work by weighing the utility of earning revenue against the disutility of working longer. One important difference is that we do not model endogenous search frictions or strategic entry, and instead we rely on non-parametric estimates of the dynamic path of earnings. This data-driven approach captures the intra-daily dynamics of earnings at more granular level, but at the same time abstracts from modeling the underlying mechanisms of market clearing. In essence we substitute with extra data part of the analysis that would otherwise require an additional equilibrium model of search and matching as well as several assumptions to make such a search and matching model tractable. Chen, Rossi, Chevalier and Oehlsen (2019) estimate labor supply and the value of flexibility in the setting of Uber drivers. Because Uber drivers are able to supply labor in irregular schedules, often as secondary jobs, the labor supply problem is fundamentally different from that of professional taxi drivers (Hall and Krueger, 2018).[3] The authors focus on quantifying driver preferences for the flexibility offered by the platform and the opportunity costs of working at different times of day. Buchholz (2022) considers a model of endogenous spatial equilibrium among taxi drivers. We do not model drivers' location choices directly, however these decisions are embedded in our model's transition densities, allowing drivers to account for divergent continuation values associated with distant locations and use these to condition labor supply decisions. This suggests a useful approach to accommodate spatial models built from the framework of Lagos (2000), including many additional settings and applications (e.g. Brancaccio *et al.* 2020; Castillo 2022; Rosaia 2023), when counterfactuals address aggregate moments. Petterson (2022) considers a dynamic labor supply model among taxi drivers to estimate a model of reference dependence. In contrast, our model assumes no reference dependence and derives predictions that are consistent with static evidence of reference dependence.

Finally, we contribute to a broad literature that demonstrates how a variety of static empirical puzzles can be rationalized through dynamic models of firm behavior. Examples of these include the puzzle of procyclical labor productivity, rationalized in part by models of labor hoarding among forward-looking firms (Rotemberg and Summers 1990; Burnside *et al.* 1993; Lagos 2006), the puzzle of firms pricing below cost, rationalized through models of learning by doing or predatory pricing (Benkard 2004; Besanko, Doraszelski and Kryukov 2014), and the puzzle that incurring costly regulations would benefit firms, rationalized through a model of rising rival entry costs (Ryan, 2012).

In Section 2 we present the data used, document important stylized facts and review the main findings of past literature on taxi driver labor supply. In Section 3 we present a dynamic model of drivers' labor supply decision. In Section 4 we discuss the estimation and identification of our model. Section 5 provides estimation results and revisits the literature in light of them. Section 6

[3] A parallel literature examines labor supply in ride-hail settings using experiments, leading to results that echo our estimates and findings: Angrist *et al.* (2021) find large and positive labor supply elasticities, while Hall *et al.* (2023) study how market-level elasticities differ from individual responses to fare increases, with Caldwell and Oehlsen (2018) and Christensen and Osman (2023) providing additional evidence on heterogeneous responses across driver demographics and broader equilibrium effects. Chen, Ding, List and Mogstad (2020) also leverage Uber's flexible work arrangements to estimate labor supply parameters.

uses the estimated model to conduct counterfactuals that measure labor supply elasticity in the context of both individual and market-wide wage fluctuations. Section 7 concludes.

2. Data and Evidence

We start by introducing the dataset and presenting some descriptive results. In 2009, the Taxi and Limousine Commission of New York City (TLC) initiated the Taxi Passenger Enhancement Project, which mandated the use of upgraded metering and information technology in all New York medallion cabs. The technology includes the automated data collection of taxi trip and fare information. This data set represents a complete record of all trips operated by licensed New York medallion taxis. We primarily use TLC data on all medallion cab rides given from July 1, 2012 to September 3, 2012, the last day before a fare change. The sample analyzed here consists of 27,830,861 trips. Data include the exact time and date of pickup and drop-offs, trip distance, trip time, fare information and car and driver identifiers. Table 1 provides summary statistics on trips in Panel I and driver shifts in Panel II. In Section 6 we complement this data with two additional months of trips that occur after the fare change.

Recent work has made broad use of this data set (See, e.g., Haggag *et al.* (2017), Frechette *et al.* (2019), Thakral and Tô (2021), Buchholz (2022)). Earlier research, including work devoted to explicitly measuring labor supply elasticities, employs smaller samples and less reliable taxi trip data. While there is continued debate about model specification and the presence of behavioral biases, prior studies argue that the TLC data obviates most lingering worries about sample size and measurement error (Farber, 2015).

There are several regulatory statutes governing TLC licensed taxis that are relevant for analyzing the labor supply of drivers. The TLC divides licenses into several categories, including yellow taxis, liveries, para-transit, and special charter vehicles. In 2012, yellow taxis represented, by far, the highest volume of the license classes, providing around 175 million rides per year among roughly 50,000 drivers and 13,437 cars. In 2013 the TLC began licensing "green cabs" or "boro cabs", which grants authorization to pick up passengers outside of Manhattan. Later, they permitted high-volume for-hire vehicles that include platform-based ride-hail companies such as Uber and Lyft.

Yellow taxi medallions are sub-divided into operational types. The most common is the mini-fleet type, representing about 60% of all yellow taxis, in which companies own multiple cars with attached medallions and drivers lease the taxi on a daily or weekly basis, paying a fixed leasing fee subject to regulated caps. Daily leases impose strict shift limitations, where day-shift drivers are required to return cars to bases, typically located in Queens, by 4-5pm for the evening shift, and night-shift drivers are required to return the car by 4-5am. The remaining types involve a driver-owned car and a leased medallion, or a driver-owned car and driver-owned medallion, neither of which are subject to daily shift restrictions.

Yellow taxis are required to locate passengers through street hail and cannot schedule rides in advance.[4] During the sample period the fare was fixed by the TLC at $2.50 fixed fee plus $2.00 per mile.

Table 1 contains summary statistics for the full sample of trips in Panel I and driver shifts in Panel II, as well as subsamples by weekday vs. weekend, morning vs. evening shift, and fleet status. Morning (AM) and evening (PM) shifts are defined similarly to Farber (2015): AM shifts start between 4am – 10am and PM shifts start between 2pm – 8pm. Fleet status is not directly observed but can be partially inferred from data. We're specifically interested in capturing whether an AM-shift driver is obligated to return the taxi by 5pm, or similarly for a PM-shift driver at 5am.[5] Fleet drivers here are interpreted as daily lease drivers, whereas non fleet drivers include owner-operators with and without leased medallions as well as fleet drivers who operate on longer-term leases.[6]

Panel I shows that nearly all driving yields similar distributions of trips in terms of fares and duration. However weekdays, afternoon shifts, and non-fleet drivers face slightly higher fares and longer durations. These differences reflect a slightly lower concentration of central Manhattan trips. Panel II shows that evening shifts earn drivers about 6% more revenue across

[4]Permission to engage in app-based e-hailing of yellow taxes, a limited type of scheduling and search aid, was approved by the TLC in mid-December 2012 and took effect in February 2013.

[5]Though exact shift times are not recorded, we take shifts to be drivers' total span of work without breaks longer than five hours. This definition is adopted by some of the literature using this data set (e.g., Haggag *et al.* (2017), Frechette *et al.* (2019)) and close to the six hour definition used in the rest (e.g., Farber (2015), Thakral and Tô (2021))

[6]While our data do not directly record the type of license used by each driver, they do identify the medallion ID separately from the driver ID. We categorize leased vs. owner-operator medallions based on the number of unique drivers observed using each medallion and the probability that a medallion is turned over during the "witching hour" between 4pm – 5pm. See Section A.1 for more detail.

TABLE 1. Trip-level Summary Statistics

Variable	Data Sample	Obs.	10%ile	Mean	90%ile	S.D.
	I. Trip Summary Statistics					
Trip Revenue ($)	Overall	27.9M	5.60	12.35	21.50	9.23
	Mon-Fri	18.6M	5.70	12.47	21.82	9.39
	Sat-Sun	9.3M	5.50	12.11	20.90	8.90
	AM Shift	17.5M	5.40	11.80	20.30	8.98
	PM Shift	7.1M	6.40	13.39	23.29	9.33
	Fleet	14.9M	5.52	12.06	20.78	8.75
	Non-Fleet	13.0M	5.70	12.69	22.68	9.74
Trip Minutes	Overall	27.8M	4.00	12.15	22.70	8.61
	Mon-Fri	18.6M	4.00	12.50	23.00	8.64
	Sat-Sun	9.3M	4.00	11.45	21.17	7.89
	AM Shift	17.5M	4.00	11.96	22.23	8.47
	PM Shift	7.1M	4.52	12.56	23.00	8.36
	Fleet	14.9M	4.00	11.90	22.00	8.20
	Non-Fleet	13.0M	4.00	12.45	23.00	8.76
	II. Shift Summary Statistics					
Shift Revenue ($)	Overall	1,287K	150.26	267.25	378.59	89.94
	Mon-Fri	880K	153.32	262.18	358.20	81.90
	Sat-Sun	406K	144.77	278.27	419.33	104.46
	AM Shift	766K	174.97	270.10	359.52	77.07
	PM Shift	331K	166.57	285.60	410.00	92.59
	Fleet	647K	168.06	277.24	380.76	85.33
	Non-Fleet	640K	138.35	257.21	375.64	93.28
Shift Minutes	Overall	1,287K	287.92	497.91	659.58	149.60
	Mon-Fri	880K	304.12	497.69	641.27	140.11
	Sat-Sun	406K	262.85	498.37	689.87	168.35
	AM Shift	766K	380.07	525.71	649.57	125.30
	PM Shift	331K	303.24	485.97	662.10	139.74
	Fleet	647K	331.28	516.79	658.79	137.13
	Non-Fleet	640K	260.59	478.81	661.03	158.97

This table uses TLC Data from August–September, 2012. Panel I summarizes revenues and total trip times over all trips in the sample, with subsampling on different types of shifts as indicated. Panel II summarizes total cumulative revenues and cumulative work time across each driver-shift. Shift minutes includes both the time spent vacant and the time spent occupied.

shifts with about 7% less time. Though owner-operators have essentially free entry across shifts, the fact that their shifts are similarly distributed as lease drivers suggests that evening work may impose higher opportunity cost on drivers despite being more valuable overall. A similar albeit weaker pattern is true for weekend shifts compared to weekdays.

Drivers are highly variable in their daily working hours. The 10th to 90th percentiles among various driver types falls between about 4.5–11 hours. A simple variance decomposition reveals that about 34% of the total variation in driver hours is attributable to differences between drivers. In contrast, the within-driver variance accounts for about 66% of the total variation, suggesting a large degree of idiosyncratic variation in quitting behavior among drivers. In the next section we provide a histogram of driver quit times and compare this distribution against patterns of earnings.

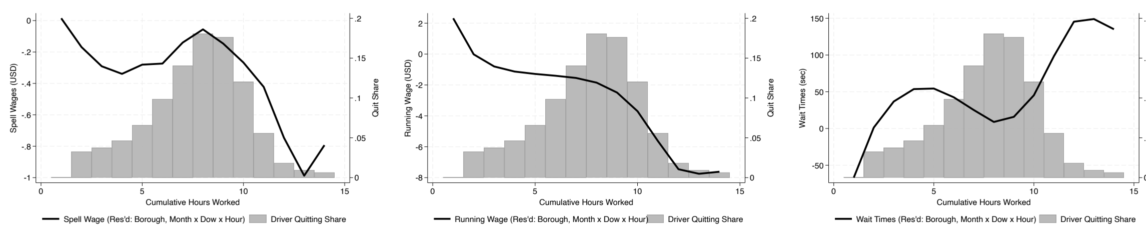
2.1. **Earnings Rates Across the Day.** In this section we highlight patterns in how drivers accrue earnings over time. While trip prices are determined by the TLC's fare schedule, drivers face uncertainty over hourly earnings because they need to first search for passengers in order to earn fare revenue. The amount of search time required to find a passenger is highly uncertain, generating variability in the realized productivity of a driver's time. To capture this in a simple way, we define a *spell* as the length of time between passenger drop-offs. A spell is therefore the sum of time spent searching for a passenger and the time spent traveling with a passenger, and every shift can be characterized as a sequence of spells from the time the driver begins working until the end of the day. We now define a driver's *spell wage* as the total revenue earned over a spell divided by length of the spell. The driver's spell wage is thus a trip-by-trip effective wage. We also define a *running wage* as the total earnings divided by total hours worked at any given point in time. Note that at the end of a driver's shift, the weighted average spell wage is equal to the average realized wage over a given driver's shift, a common moment used in the literature. We express spell wage and running wage in units of dollars per hour.

We document two key stylized facts about drivers' earnings. First, spell wages and running wages tend to decline with time spent on a shift. Figure 1 shows how spell wages and running wages evolve with time spent on the shift. Both wage series depict hourly means residualized over location fixed effects and month by day-of-week by hour fixed effects.[7] To see where in the day drivers tend to quit relative to their wage dynamics, both panels depict a histogram

[7]Our empirical strategy will pool these effects across drivers similarly to this figure. However, in Figure 4 we show that these effects persist net of driver-level fixed effects as well.

indicating the share of quitting relative to hours worked. The change in running wages across a shift implies that drivers earn, on average, around $2 less per hour relative to other drivers from the start to the end of a typical eight- to nine-hour shift. Importantly, these effects are residual to the traditional hourly and weekly shifters of earnings as well as location. Thus, even two drivers who begin work at different hours on the same day will, at any point in the day and at the same location of the city, have different expected earnings rates.

FIGURE 1. Rate of Earnings by Cumulative Hours Worked



(a) Spell Wage (b) Running Wage (c) Search Times

TLC from January to August 2012. This figure shows how earnings evolve with cumulative hours worked. Panel (a) shows the relation between spell wage (residual to location and time fixed effects) and cumulative hours worked. Panel (b) shows the relation between the residualized running wage and cumulative hours worked. Both panels depict the share of hours in which drivers quit, with units on the right y-axis.

A natural follow up question is to ask why driver earnings are falling as they work more hours. Panel (c) shows that, while trip revenues are modestly increasing with hours worked, indicating a small average increase in longer and higher-fare trips, the predominant reason for declining earnings rates is that time spent *between* trips grows as drivers work longer hours. To explain the puzzles in the labor supply literature, we can largely remain agnostic in understanding why search times are increasing. However, it is instructive to decompose why this pattern occurs.

Drivers' increasing search times could arise from increases in active working time or increases in break time. Active working time may grow if drivers become less efficient at searching for passengers, perhaps through less intensive or less productive search. Alternatively, drivers may take more or longer breaks between trips, for example to eat a meal or rest. Either explanation might qualify as an effect of workplace fatigue, which has been documented in myriad studies of other workplace settings.[8]

The evidence points to increased break time being the predominant force driving these patterns. If drivers became less efficient at search, we would expect that they would drive longer distances on average to find passengers. In Figure 5 in the Appendix we show that the search radius, defined as the straight line distance between a driver's pickup position and subsequent drop-off position, is flat in hours worked, until around the 9th hour, after which the radius grows at a slow rate for each additional hour worked. This effect is small, however, as few drivers work beyond 9 hours. The increase in radius by a driver's 11th hour into the shift is about 0.1 miles, or 1/2 the length of a standard New York City block. By 14 hours, the effect is around one city block in length, yet less than 1% of drivers are observed to work this many hours. Because drivers' search distances are not materially changing, we believe drivers are most likely taking breaks.[9]

2.2. **Breaks and Bias in Static Wage Regressions.** A more plausible explanation for the pattern of declining earnings is that drivers take more frequent or longer breaks the longer they work. The possibility of driver break periods has been discussed to some extent in the literature. Though results are not reported directly Camerer *et al.* (1997) claims to recover similar results when removing inter-trip break times beyond 30 minutes when computing labor hours. Farber (2005) and Thakral and Tô (2021) similarly characterize breaks as waiting times in excess of 30 minutes for most trips and 60–120 minutes for trips outside of Manhattan. Each paper finds that conditioning on such breaks do not alter the respective conclusions. Schmidt (2019) approximates break time as the minimum between any waiting time 1.5 times greater than the

[8]Economists have long discussed the possibility that workplace productivity declines with hours worked (Leveson, 1967; Barzel, 1973). Evidence for these effects are documented in both worker productivity (Pencavel, 2015) and in the context of physician errors (Ricci *et al.*, 2007; West *et al.*, 2009).

[9]While explicitly modeling drivers' intradaily break decisions could provide additional insights, we find that break patterns do not vary systematically with intraday earnings (see Section A.3), suggesting breaks can be reasonably treated as exogenous to the quitting decision. Moreover, since breaks are not directly observed in the TLC data, any model of endogenous break-taking would require strong assumptions to infer break periods from waiting times between trips.

average search times and twenty minutes, and finds that breaks defined this way increase with hours worked. [10]

We next investigate how sensitive the definition of break taking is to the results in the literature. We first consider a break as any waiting time over 30 minutes, consistent with the above papers. Because lower percentiles of residual waiting times also increase with hours worked, we next define a break as waiting time over 10 minutes. Removing these from work hours will further control the effect of increased future waiting times (and therefore, future declining spell wages) on hours worked. Finally, we regress waiting times on a rich set of time, location and driver covariates to generate *expected* waiting times (averaging over all hours-worked variation). Using these, we recompute work hours as time on trips plus expected waiting times, essentially purging all residual factors that induce drivers to wait more or less.

With our measures of work hours net of break-taking, we use data from Jan. 1, 2012 – Sep. 3, 2012 to reproduce the Camerer *et al.* (1997) (henceforth CLBT) regression specification, using identical instruments and controls. We left out post-fare hike data after September 4, 2012 to only include variation from a constant-fare period as in CLBT. Next, we repeat the exercise with the full 2012 data and reproduce Farber (2015) (henceforth F2015), with the F2015 instruments and controls.

Following CLBT, we instrument for wages using the 25th, 50th and 75th percentiles of average shift wages across all other drivers on a given day, with additional controls for shift type and weekday. Our F2015 replication uses mean wages of other drivers as the instrument, with richer controls including day-of-week and weekly indicators, more shift types, a pre/post fare increase dummy, and driver fixed effects. Both regressions are conducted at the shift level. For both specifications, the dependent variable is log hours worked on a shift and the key independent variable is log average wage over a shift.

Table 2 show the results of this exercise. Baseline specifications are reproduced in columns 1 and 2. Our baseline CLBT-like estimate is -0.680 compared to CLBT's published estimates of -0.319 and -0.975 (across different samples). By removing 30 minute breaks, our estimate shrinks in magnitude to -0.365. With 10 minute breaks removed, our estimate turns positive to

[10] The remainder of the literature mostly ignores breaks. Farber (2008), Farber (2015), and Crawford and Meng (2011) do not discuss breaks or handle them in any distinct way.

TABLE 2. Wage Regressions Replication with Excess Breaks Removed

	Baseline		Removing Breaks					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	CLBT 1997	Farber 2015	CLBT wait<30	CLBT wait<10	CBLT mean wait	Farber wait<30	Farber wait<10	Farber mean wait
log daily wage	-0.680***	0.516***	-0.365***	0.047***	0.416***	0.970***	1.507***	1.872***
	(0.014)	(0.010)	(0.016)	(0.017)	(0.017)	(0.013)	(0.015)	(0.017)
Instrument + Controls	CLBT	F2015	CLBT	CLBT	CLBT	F2015	F2015	F2015
Observations	786,808	786,808	786,808	786,808	786,808	786,808	786,808	786,808

This table shows wage regression estimates. The dependent variable is log(work hours). Specification (1) replicates the final specification in Camerer *et al.* (1997) Table III including the same set of controls and a a 25% sample of data from Jan 1, 2012 – Sep 3, 2012. Specification (2) replicates Farber (2015) Table 5 including the same (richer) set of controls and data from Jan 1, 2012 – Dec 31, 2012. The remaining columns repeat these replications except with alternative definitions of work hours. In columns (3) and (6) we define work hours as all cumulative work time minus all waiting times beyond 30 minutes. In columns (4) and (7) we instead subtract waiting times beyond 10 minutes. In columns (5) and (8) we reconstruct total shift time using average waiting times by hour of day, day of week, location and driver. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

0.047. Finally, by constructing total hours with expected wait times only, our estimate becomes 0.416. We repeat the analysis for F2015, obtaining a baseline estimate of 0.516. Although our data is the same as with our CLBT analysis, we are now controlling for a richer set of covariates, including day-of-week and weekly indicators, more shift types, a pre/post fare increase dummy, and driver fixed effects. Our baseline closely compares to the published F2015 estimate of 0.589. By removing breaks, again our estimate increases to 0.970 (removing 30 minute breaks), 1.507 (removing 10 minute breaks), and 1.872 (using average wait times). These results suggest that our replication can closely match estimates in the literature, and that controlling for drivers break times, which are increasing with hours worked, can flip the sign of the original behavioral results.

Collectively, we interpret these results as evidence that drivers' increased break-taking with longer work hours is driving a decline in earnings and introducing a negative correlation between average hourly earnings and hours worked. Importantly, this correlation introduces a bias in the static wage regression framework. We refer to this bias as the *technology bias*, referring to the technology of human workers and associated requirements of rest with long hours. Without wage instruments, Camerer *et al.* (1997) highlight the problem of division bias, where hours worked appears on both side of the OLS equation. The technology bias, however, is a problem that even existing wage instruments cannot fully mitigate. This is because an instrument that impacts the daily earnings distribution does not impact the relation between hours worked and break taking. As a result, even if, all else equal, drivers work longer in response to a high-wage day compared to a low-wage day, the longer-hours drivers will also face longer break times in expectation, leading to a downward bias in prevailing average wages. The measured effect in the IV regression is a composition of (positive) labor supply elasticity and a (negative) technology bias.

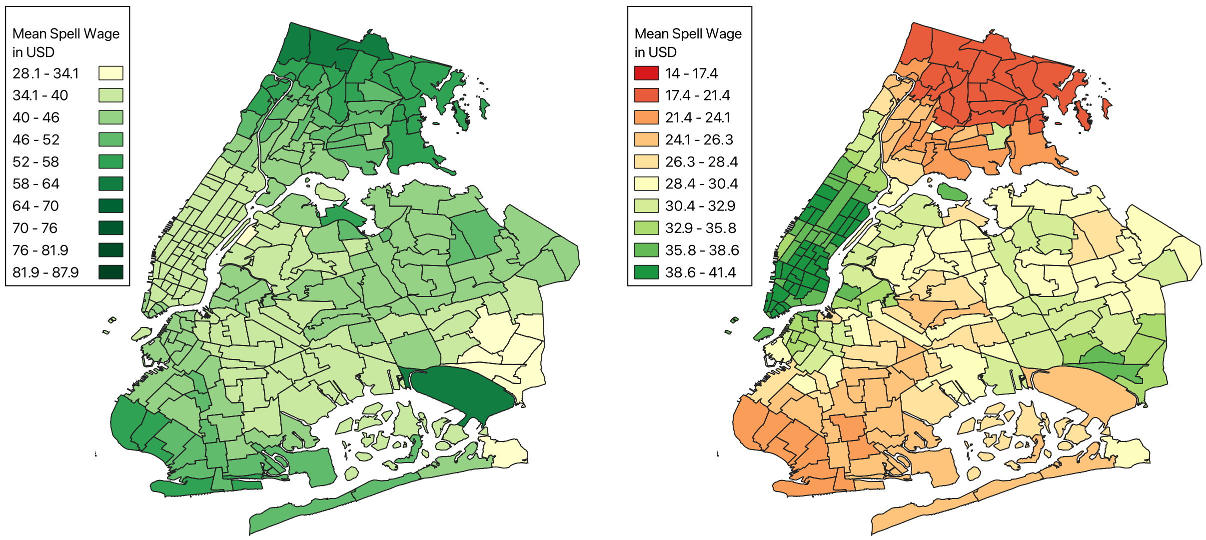
2.2.1. *Is break-taking endogenous?* In Section A.3 we provide evidence that drivers do not modulate effort in response to transient earnings shocks within a day, which suggests that the decrease in the earnings rate to be exogenous with respect to cumulative earnings. These patterns reflect a declining within-driver productivity: earnings continue to slow down even when market conditions become more favorable for drivers. One important implication of this declining productivity is that, all else equal, driver shifts with longer cumulative work hours will be associated with lower average shift wages.

2.3. **Negative Serial Correlation and Location Effects in Earnings.** The second key fact about driver earnings is that positive shocks to earnings (i.e. earnings exceeding that of a typical weekday and time of day) exhibit a strong negative serial correlation. For example, a driver who earns an extra $20 per hour compared with other drivers on a Monday 2pm will face, in expectation, a rate of earnings well below average in the subsequent hour. This pattern arises because many long and high-earning trips will leave taxi drivers in less desirable locations that require additional search time to find the next passenger.

Below we display map of the TLC taxi zones and mean spell wages by zones, first, as destinations (Panel I) and then, as origins (Panel II). Panel I shows that positive income shocks, the darker shaded regions, are associated with drop-offs farther from Manhattan. This is because most rides originate in Manhattan, and travel times to these destinations will be substantially longer than inter-Manhattan trips, leading to a high degree of hourly occupancy among drivers serving these destinations. However, Panel II shows that subsequent spell wages faced by drivers in the positive-shock destinations are quite low. This is not surprising, as drivers are generally unlikely to find passengers in these more distant locations and frequently travel back to Manhattan without passengers, leading to very low next-spell earnings rates.

FIGURE 2. Taxi Zones vs. Mean Spell Wage

I. Spell Wages (During Trip) II. Spell Wages (After Trip)



This figure uses TLC data from Jan. 1, 2012 to Sep 3, 2012. Panel I shows the average spell wage for trips which lead to a drop-off in each of the 260 TLC-defined taxi zones. Panel II shows the average, forward-looking spell wage for each taxi zone conditional on having dropped-off passengers in that zone. These quantities represent the expected earnings per hour for any spell which ends (Panel I) and begins (Panel II) in a given taxi zone. Taxi Zone shape files are obtained from https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page.

Table 3 provides a more direct quantification of these effects. Across non-Manhattan destinations, average earnings increase from $5 to $55 per hour (or 12% to 140%) over the baseline of about $39.40 per hour. However, those same high-earnings destinations become almost equivalently bad as places to subsequently search for passengers, with average declines in hourly earnings of $7 to $25 (or -17% to -61%).

The negative auto-correlation shown here plays an important role in the dynamic model because it helps explain why drivers often exhibit an increased likelihood of quitting after positive earnings shocks near the end of their shift. To test whether location-based negative serial correlation explains the apparent recency bias, we can replicate the baseline results in

TABLE 3. Spell Wage Effects During and After Borough Trips, Relative to Manhattan

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bronx (trip destination)	14.14***	13.72***	14.69***				16.74***
	(0.0672)	(0.0656)	(0.0669)				(0.0657)
Brooklyn (trip destination)	4.610***	3.294***	3.445***				6.419***
	(0.0222)	(0.0222)	(0.0221)				(0.0222)
EWR Airport (trip destination)	53.28***	54.43***	54.95***				54.73***
	(0.126)	(0.123)	(0.121)				(0.119)
Queens (trip destination)	10.92***	11.15***	11.63***				13.61***
	(0.0211)	(0.0208)	(0.0208)				(0.0206)
Bronx (start search)				-16.48***	-18.42***	-18.42***	-19.47***
				(0.0696)	(0.0693)	(0.0693)	(0.0678)
Brooklyn (start search)				-6.649***	-8.668***	-8.668***	-10.31***
				(0.0229)	(0.0226)	(0.0226)	(0.0227)
EWR Airport (start search)				-25.26***	-24.87***	-24.87***	-25.31***
				(0.129)	(0.124)	(0.124)	(0.121)
Queens (start search)				-10.40***	-10.75***	-10.75***	-13.10***
				(0.0219)	(0.0215)	(0.0215)	(0.0213)
Constant	39.41***	39.46***	39.42***	41.11***	41.22***	41.22***	40.32***
	(0.00474)	(0.00464)	(0.00459)	(0.00475)	(0.00459)	(0.00459)	(0.00463)
Obs.	15,422,836	15,422,836	15,419,838	15,420,001	15,419,999	15,419,999	15,406,011
Dow x Time FE	NO	YES	YES	NO	YES	YES	YES
Driver FE	NO	NO	YES	NO	NO	YES	YES

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

This table shows the impact of trips to different boroughs on spell wage relative to trips to/from Manhattan. Columns (1)-(3) shows the effect on spell wage from taking a trip to the stated borough on the *current* spell. Columns (3)-(6) shows the effect on spell wage from ending the *previous* trip in each borough. Column (7) combines both sets of controls.

Thakral and Tô (2021) (hereafter, TT) while splitting our sample based on trip destinations. If recency bias is driven by negative serial correlation in earnings after long trips to outer areas, then the effect should be stronger when restricting to earnings from trips to low continuation value neighborhoods (those with below-median future expected earnings) and weaker or absent when restricting to high continuation value neighborhoods (those with above-median future expected earnings). This provides a direct test of whether the apparent greater sensitivity to recent earnings is actually capturing rational responses to temporarily high earnings that predict lower future opportunities.

Table 4 shows our results. The coefficients represent the elasticity of the probability of stopping with respect to earnings in each hour, where a positive coefficient indicates that higher earnings in that hour increase the probability of stopping after hour 8. Columns (1) and (2) replicate TT's elasticity estimates in our sample. Column (3) makes a technical adjustment to the sample window to address potential mechanical effects (see Section A.2.3 for details). Columns (4)-(5)

TABLE 4. Elasticity of Stopping at 8.0-8.5 Hours with Respect to Income

	(1) Full Set	(2) Combined Incomes	(3) Adjusted	(4) High Cont'n Value Income	(5) Low Cont'n Value Income
Cumulative Income	-0.952***				
	(0.228)				
Income in Hour 1	1.166***	0.214	0.112	0.384***	-0.285**
	(0.247)	(0.247)	(0.121)	(0.146)	(0.117)
Income in Hour 2	0.900***	-0.047	0.005	0.067	-0.261
	(0.250)	(0.250)	(0.133)	(0.187)	(0.172)
Income in Hour 3	0.673**	-0.279	0.015	-0.313	0.219
	(0.291)	(0.291)	(0.155)	(0.208)	(0.198)
Income in Hour 4	0.917***	-0.035	-0.251	-0.102	0.129
	(0.293)	(0.293)	(0.157)	(0.216)	(0.211)
Income in Hour 5	0.912***	-0.040	0.460***	0.159	0.287
	(0.292)	(0.292)	(0.157)	(0.218)	(0.219)
Income in Hour 6	1.497***	0.545***	0.253	0.092	0.390*
	(0.291)	(0.291)	(0.156)	(0.213)	(0.222)
Income in Hour 7	2.009***	1.057***	1.135***	0.640***	0.122
	(0.287)	(0.287)	(0.157)	(0.217)	(0.241)
Income in Hour 8	2.015***	1.063***	0.108	-1.619***	4.480***
	(0.323)	(0.323)	(0.167)	(0.259)	(0.403)
Observations	234,349	234,349	261,389	129,153	129,153
R-squared	0.339	0.339	0.320	0.367	0.369

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This table shows the effect of cumulative income in each hour of the day on the probability of stopping after the eighth hour of cumulative work time. All estimates are presented as elasticities relative to the baseline probability of stopping. The sample consists of a 25% sample of New York TLC data from 1/1/2012-9/3/2012, further limited to trips ending at 8h0m to 8h40m. All specifications include controls for driver, day-of-week x hour, week-of-year, drop-off neighborhood, cumulative income and cumulative work time. Columns (1) and (2) replicate Table 2 of Thakral and Tô (2021) in our sample. Column (3) adjusts the sample window to 8h0m to 8h10m. Column (4)-(5) repeat the specification of column (3) separately for earnings coming only from high- or low-continuation value neighborhoods based on the top and bottom 50% of New York neighborhoods with respect to one-period-ahead spell-wages conditional on drop-off in each neighborhood. Neighborhoods are defined as each of the 260 Taxi Zones defined by the NY TLC here: `https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page`

repeat this specification separately for earnings coming from high and low continuation value neighborhoods.

The results strongly support our hypothesis that recency bias reflects rational responses to location-based earnings dynamics rather than behavioral effects. For high continuation value locations (column 4), most hourly coefficients are small and insignificant, with a negative coefficient in hour 8. In contrast, for low continuation value locations (column 5), we see a large positive coefficient (4.480) in hour 8, indicating drivers are much more likely to quit after high-earning trips to areas with poor future earnings prospects. These results provide strong evidence for our claim that earnings shocks induce higher quitting probabilities late in the day because the associated destinations offer limited earnings opportunities. In Appendix Table 10 we show these results are robust to alternatively defining high/low continuation value areas using a simpler Manhattan/Brooklyn versus outer borough split.

2.4. **Motivating a dynamic model.** Below we present a dynamic optimal stopping model to reconcile the disparate results and puzzles in one simple framework. We believe the model makes two contributions. First, from a negative perspective we show that it can explain key behavioral puzzles in the literature, which posit negative wage elasticities. Concretely, we demonstrate that a simple "neoclassical" model of dynamic optimization will generate the static behavioral results given the earnings patterns that we document in the descriptive analysis.

Second, from a positive perspective, we show how to use this model to estimate the intensive-margin labor supply elasticity, which is the key object of interest in this literature. While our analysis of the early literature illuminates the key econometric problems with the static analyses, our approach using the dynamic model allows for systematically controlling for drivers' forward-looking expectations in a way that reduced-form specifications with additional controls cannot fully achieve. Rather than adding various proxy controls for future earnings opportunities in an ad hoc way, the model provides a coherent framework that fully accounts for how drivers incorporate the stochastic evolution of wages into their decisions. This comprehensive treatment of dynamics enables us to disentangle the technology of earnings (i.e., declining productivity) from preferences in order to obtain the elasticity of interest as a counterfactual. Our approach is straightforward and may closely apply to studying other flexible labor supply settings such as other gig-economy jobs like food delivery or owner-operated trucking.

3. Model

Taxi drivers drive around the city searching for customers. They earn fare revenue by providing rides and work until deciding to quit for the day. We model the quitting decisions of individual drivers, indexed by i, engaged in daily shifts indexed by j.

We group drivers and shifts into a set of discrete types $k(i, j)$ which characterize the combinations of all possible driver types and shift types. For example, a driver type could be whether the driver owns or leases a taxi. A shift type could be whether the shift takes place during a weekday morning or weekend evening. Market conditions on a given shift are described by x_j, a discrete index of daily earnings quintiles that captures profitability for specific shifts. x_j summarizes market-wide variation in expected average earnings-per-hour, arising from demand factors like weather conditions and events. This variation provides a natural source of exogenous shifts in drivers' expected earnings, analogous to the wage instruments used in the reduced-form analysis of Section 2.

Drivers earn a payoff from cumulative earnings and cumulative time spent working, denoted as $u_k(r, h)$. In the empirical analysis below, we make the assumption that money is fungible throughout the day and further normalize the scale of utility to the dollar, so that $\frac{\partial u}{\partial r} = 1$ for all r and h. The function $u_k(r, h)$ is the main primitive object of interest, as it describes how drivers value their time at various levels of work time.

We assume the mean utility of drivers has the following structure, with r equal to cumulative earnings and h equal to cumulative number of hours worked:

$$u_k(r, h) = r + \theta_{1k}h + \theta_{2k}h^2 \tag{1}$$

where $\boldsymbol{\theta}_k^c = \{\theta_{1k}, \theta_{2k}\}$ is a vector of unknown cost function parameters to be estimated.[11]

Each decision period is discrete and occurs when a passenger is dropped off. After a trip is completed, the driver observes the current state $\{r, h, \ell\}$ and faces a decision. He may quit for the day, in which all fares and costs earned up to that point are collected and all possible future fares for that day are foregone. Alternatively, the driver may choose to keep working for one more spell, in which case he draws from a distribution of new fares dr, new work times dh (consisting of both search time and then time-on-trip) and new locations ℓ', such that

$$(dr, dh, \ell') \sim \mathcal{F}_{k,x}(r, h, \ell). \tag{2}$$

[11] In practice, we will treat drivers as homogenous within eight discrete *types*: AM/PM shifts $\times$ Weekday/Weekend shifts $\times$ owner-operator/fleet licenses.

Every new spell drawn adds to the stock of cumulative earnings, i.e. $r' = r + dr$ and similarly for work time $h' = h + dh$. In other words, the additional payoff associated with a new spell is the difference between payoffs at the new state from the previous one: $u_k(r + dr, h + dh) - u_k(r, h)$. At the end of this trip, the driver once again faces the decision to quit or not.

$\mathcal{F}$ plays a critical role in the model. Fixing a shift-type k and market conditions index x, draws from this distribution capture the transition path of the state variable, and, by extension, the path of spell wages, as determined each period as $\frac{dr}{dh}$. The descriptive patterns documented in Section 2.1 are directly encoded in $\mathcal{F}$ through its dependency on current state variables r, h and ℓ. Moreover, $\mathcal{F}$ represents the equilibrium earnings process for the entire shift, summarizing the search process and the efficiency of driver search as a consequence of the thin/thick externalities on both sides of the market. Our counterfactuals of interest operate on $\mathcal{F}$ in a natural and straightforward way, which allows us to circumvent the need to model the underlying search and matching process and avoid the need to take a stance on its form.[12]

Before making a decision, drivers draw an unobserved opportunity cost ϵ_{kty} associated with each driver-shift k, period t, and each quit decision, $y \in \{0, 1\}$. We assume each ϵ_{kty} is i.i.d and distributed as Type I Extreme Value with mean zero and scale parameter σ_k.

Driver i's decision problem in period t can then be characterized by a value of quitting or continuing depending on the state $\{r_t, h_t, \ell_t\}$, market conditions x, and a draw of $\epsilon = \{\epsilon_{kt0}, \epsilon_{kt1}\}$. Omitting the subscript t on state variables for simplicity, we can write the value function of each driver-shift type k as follows:

$$V_{kt}(r, h, \ell, \epsilon) = \max\left\{u_k(r, h) + \epsilon_{k0}, \int_{\mathcal{F}_{k,x}(dr,dh,\ell'|r,h,\ell),G(\epsilon')} V_{k,t+1}(r + dr, h + dh, \ell', \epsilon') + \epsilon_{k1}\right\} \quad (3)$$

Equation 3 summarizes the timing of the decision problem: after dropping a passenger, drivers observe their private signal ϵ about the relative value of continuing work over quitting. If they decide to continue, they then draw an updated work time and income. We therefore assume they either quit or commit to working until finding another passenger and revisiting

[12] Our approach relates to Huang and Smith (2014), which models the non-stationary evolution of fishery stocks, in part a function of a complex biological process, through a flexible specification of transition densities.

the decision after that passenger is dropped off.[13] Since the entire shift occurs within a single day, we assume there is no discounting. However, the draws dh do impact model timing. First, the wage process $\mathcal{F}$ depends on h, so a long spell dh impacts future draws. Second, the driver's cumulative work time also increments by dh, which increases total costs according to the shape of $u_k(\cdot,\cdot)$. Motivated by data, we make the assumption that drivers always quit by 15 hours, so that $V_{kt}(r, h = 15, \ell, \epsilon) = u_k(r, h = 15) + \epsilon_{k0}$.

Existence and Uniqueness of the Optimal Stopping Rule.

Theorem 1. *Denote* $\mathbf{r}_t \equiv (r_1, \cdots, r_t)$ *and* $\mathbf{h}_t \equiv (h_1, \cdots, h_t)$*, and* $\boldsymbol{\ell}_t \equiv (\ell_1, \cdots, \ell_t)$*. Assume:*

(i) *For each* t*,* ϵ_t *is independent of* $(\mathbf{r}_t, \mathbf{h}_t, \boldsymbol{\ell}_t)$*, i.e.,* $\epsilon_t \perp (\mathbf{r}_t, \mathbf{h}_t, \boldsymbol{\ell}_t)$*; Conditional on* $(\mathbf{r}_t, \mathbf{h}_t, \boldsymbol{\ell}_t)$*, the errors* ϵ_s *are i.i.d. across* $1 \leq s \leq t$*; and* ϵ_{ty} *are i.i.d. across* $y = 0, 1$*, with a distribution that is absolutely continuous with respect to the Lebesgue measure on* $\mathbb{R}$ *and have a mean of zero.*

(ii) *The support of* (r_t, h_t, ℓ_t) *and* $\bigcup_{t=1}^{\infty} Supp(r_t, h_t, \ell_t)$ *are compact.*

(iii) *The mean per-period utility function* $u(r, h)$ *is continuously differentiable in* (r, h)*.*

Given these assumptions, a unique solution exists to the optimal stopping rule.

We provide a proof of the above Theorem in Section A.8.[14] Note that Assumptions (ii) and (iii) are not essential. They can be relaxed, though doing so would lengthen the proof and increase the notational complexity of the analysis.

Entry. We abstract from modeling daily entry costs and instead assume drivers' entry timing decisions are fixed. While this is in part supported by regulatory constraints, as daily lease drivers are bound to morning and evening shift timing windows, owner-operators are generally free to begin work at any hour. Because drivers' actual starting time decisions and constraints are not observable, identifying entry costs would require a model of strategic interaction in daily entry timing. We nevertheless believe the most natural way to estimate the substitution elasticity of labor supply, as we do in Section 6.2, is to fix starting times and predict work hours

[13] This assumption is motivated by limitations of the TLC dataset; we cannot observe drivers engaged in search who give up half-way through and quit their shift. Therefore we assume drivers' decision to quit occurs at the point of passenger drop-off.

[14] Note that Assumptions (ii) and (iii) are not essential. They can be relaxed, though doing so would lengthen the proof and increase the notational complexity of the analysis. Additionally, it should be noted that our model assumes that the per-period utility is additive in ϵ_t, which can also be relaxed.

as a function of changes to earnings. This approach offers a direct comparability between our estimates and those obtained in other labor markets. We also want to emphasize that even a simple dynamic model is capable of reconciling multiple puzzles that arise in the labor supply literature. Nevertheless, this assumption limits the types of counterfactuals one can consider using our framework.

Competition. Drivers compete for fares with other drivers. This interaction is encoded in the distribution $\mathcal{F}$ of work time, earnings and location draws. All else equal, more drivers in the market tends to shift the mass of this distribution towards longer realizations of h for a given r, as a thicker supply-side leads to increased driver search times.

Modeling the underlying mechanisms that give rise to the equilibrium embedded in $\mathcal{F}$ poses unique challenges, as it requires a model that maps strategic entry decisions to hourly earnings (Frechette *et al.*, 2019) and a model that maps drivers' endogenous location search to location-specific earnings (Buchholz, 2022). Each of these challenges, as addressed by the literature, entails a number of substantial assumptions for computational tractability.

For our questions of interest, we can circumvent many of these difficulties. In estimation, we leverage the fact that individual drivers are small relative to the market, and treat $\mathcal{F}$ as a data object, and holding it fixed conditional on a broad set of observables. In our counterfactuals, we operate directly on $\mathcal{F}$ in different ways. We further argue that, conditional on an equilibrium of interest, in both estimation and computing counterfactuals, $\mathcal{F}$ can be regarded as common and exogenous across drivers. This approach enables us to conduct our analysis through the lens of a single-agent problem.[15]

4. Empirical Strategy

In this section, we discuss the computation and estimation of the model presented in Section 3. We first describe the forms of heterogeneity we account for. Then we describe how the descriptive facts of declining earnings and location effects are incorporated into the model. Finally, we turn to details on computing value functions and estimating model parameters.

[15] We discuss this assumption further in Section 2.1 and Section A.3.

4.1. **Driver Heterogeneity:** To capture the sources of preference heterogeneity consistent with the literature, we focus our analysis of heterogeneity on eight discrete driver-shift types, indexed by k, according to whether a driver's shift is classified as *AM* or *PM* and *Weekday* or *Weekend*, and whether the driver is classified as *Owner-operator* or *Fleet*. Drivers within each group (e.g., AM-Weekday-Fleet) are assumed to have common cost function parameters and common scale parameters on unobserved shocks. Denote the full vector of parameters for type k drivers as $\boldsymbol{\theta}_k = \{\theta_{k1}, \theta_{k2}, \sigma_k\}$ and denote the cost-function-specific parameters as $\boldsymbol{\theta}_k^c = \{\theta_{k1}, \theta_{k2}\}$.

4.2. **Market-level Heterogeneity:** Some days are more profitable for drivers than other days, for example weekdays versus weekends, or simply days in which demand is very high or very low. This variation has persistence within a particular shift and will therefore enter into drivers' earnings expectations, represented by $\mathcal{F}_k$ on any given day. We use this type of daily market-level variation for two reasons. First, it allows us to finely construct driver expectations with respect to market observables. Second, by using quintiles of average daily earnings, we exploit the same type of exogenous earnings variation as the instrumental variables strategy employed in Section 2, where wage instruments were constructed from percentiles of daily earnings. This parallel identification strategy allows us to isolate changes in expectations that are plausibly exogenous to any individual driver's choices. We incorporate this heterogeneity into our model along two dimensions of k: *AM/PM* shifts, *Weekday/Weekend* shifts, and we further separate shifts into five types of days, denoted as x_j or the daily earnings quintile x associated with shift j of type $k(\cdot, j)$. We compute x_j by first computing the average spell wage of all drivers across each driver-shift type k and categorizing these into five quintiles, which represent how productive driving is on average. The three above shift characteristics combine to create $2 \times 2 \times 5$ types that make up market-level heterogeneity entering the model as $\mathcal{F}_k$.

4.3. **Serial Correlation in Earnings:** Our model accounts for the two forms of within-driver-shift serial correlation discussed in Section 2.1. We describe each of these in turn.

(i) *Declining spell wages*, or the phenomenon in which drivers tend to become less productive as their shift grows longer, are patterns that enter our model through the state transition probability matrix. State transitions determine the relative probabilities of advancing in cumulative earnings and cumulative time conditional on a location ℓ and daily shift-type

x_j. Transition probabilities incorporate declining spell wages documented in Section 2.1 because, as the cumulative time state grows, a driver's probability of reaching higher earnings states declines relative to his probability of reaching higher time states.

(ii) *Location effects*, or the chance that future spell wages fall following trips to outer boroughs, also enter drivers' expectations and impact their quitting decisions. Due to dimensionality concerns with adding more locations we treat the space of locations coarsely, dividing ℓ into six categories as detailed in Section 4.4.1. Each time drivers in location ℓ draw a new location ℓ' they earn fare, cumulate work time, and then face a new decision at location ℓ'. Drivers' labor supply choices depend on their location through the expectations of future earnings, which depends on their current location.

4.4. **Estimation.** Model estimation is split into two separate parts: estimating the transition process $\mathcal{F}_{kx}$ and estimating payoff parameters θ_k for each driver-shift type k. Below we describe each part of the estimation.

4.4.1. *Estimating the transition process.* $\mathcal{F}_{kx}$ describes the probability distribution over new state variables (hours, earnings and locations) as a function of current state variables on a shift of type k. Our large sample size enables us to estimate $\mathcal{F}_{kx}$ non-parametrically for most trip types by finely discretizing the state (r, h, ℓ) and computing empirical transition probabilities between cells from spell to spell. Specifically, we create twenty uniformly divided bins between the lowest and highest observed values of earnings and time worked within a shift. This leads to cumulative earnings (r) bins from \$2.50 to \$753.33 in twenty intervals of about \$36 and cumulative time (h) bins from 0 minutes to 1,008 minutes in twenty intervals of about 47 minutes.

We first divide ℓ into six categories: Manhattan, Brooklyn, Bronx, Queens, Staten Island, and Newark Airport.[16] However, three of these regions, Staten Island, Newark Airport and Bronx are geographically isolated and represent less than 1% of pickups or drop-offs. As a result, we do not have sufficient data to estimate $\mathcal{F}_{kx}$ for these regions and therefore drop shifts with such trips from our estimation sample. We estimate non-parametric transition probabilities between

[16] Note that both New York City Airports, LaGuardia and JFK, are contained within Queens.

(r, h, ℓ) across Manhattan, Brooklyn and Queens. Together these three grids (20 x 20 x 3) form a state space of 1,200 discrete bins over which drivers face value functions and policy functions.

4.4.2. *Solving value functions.* We use a nested fixed-point procedure in which, for each guess of the parameter vector, we first solve value functions in Equation 3. All drivers quit by the 15th hour and receive a payoff equal to their current state according to Equation 7. Our data cleaning procedure throws out shifts beyond this length, because they are likely due to computer error or electronic testing (see Section A.1). We compute value functions in each state by backwards induction, beginning at the terminal time state, updating values over the discretized grid of remaining state variables, and then incrementing backwards through each time period. Because values for each location state are calculated sequentially, there is some residual inconsistency in value functions after a single round of backward induction. Therefore, we iterate on this calculation until finding a fixed point.

4.4.3. *Estimating payoff functions.* Our model parameters reflect the tradeoff between earnings and time as revealed by drivers' quitting decisions conditional on observed states. To estimate the model parameters governing this tradeoff we specify a likelihood function. The function represents the likelihood that drivers are observed to react as they do to a sequence of state variables as they are observed in the data. In other words, a driver who is observed to quit $(y_{it} = 1)$ at time t has, by definition, chosen to not quit $(y_{it} = 0)$ for all states $(r_{i,s}, h_{i,s}, \ell_{i,s})$ where $s < t$. Given all unique driver-shifts $\iota \in \mathcal{I}$ with driver-shift-type k and a trip index t, the log-likelihood function is as follows:

$$
\begin{aligned}
LL(\boldsymbol{\theta}_k) = &\sum_{\iota \in \mathcal{I}} \{ y_{\iota t} \cdot [\ln P(y_{\iota t} = 1 | r_{\iota t}, h_{\iota t}, \ell_{\iota t}, x_j; \boldsymbol{\theta}_k) \\
&+ \sum_{s=1}^{t-1} (1 - y_{\iota s}) \ln P(y_{\iota s} = 0 | r_{\iota s}, h_{\iota s}, \ell_{\iota s}, x_j; \boldsymbol{\theta}_k) \Big] \Big\}
\end{aligned} \tag{4}
$$

The quitting probability $P(y_{\iota t} = 1 | r_{\iota t}, h_{\iota t}, \ell_{\iota t}, x_\iota; \boldsymbol{\theta}_k)$ is obtained via the equilibrium value functions (Equation 3), which we solve by applying a backwards induction solution conditional on our estimate of $\mathcal{F}_{kx}$ and conditional on each candidate parameter vector $\boldsymbol{\theta}_k$.[17]

$$P(y_{\iota t} = 1 | r_{\iota t}, h_{\iota t}, \ell_{\iota t}, x_i; \boldsymbol{\theta}_k) = \frac{exp(u(r_{\iota t}, h_{\iota t} | \boldsymbol{\theta}_k^c)/\sigma_k)}{exp(u(r_{\iota t}, h_{\iota t} | \boldsymbol{\theta}_k^c)/\sigma_k)) + exp(V_k(r_{\iota t}, , h_{\iota t}, \ell, x_j | \boldsymbol{\theta}_k^c)/\sigma_k)} \quad (5)$$

Our estimator maximizes Equation 4 separately for each k, representing eight observable driver-shift types.[18]

5. Results

In this section we present and discuss the empirical results of the dynamic labor supply model presented in Section 3. Estimates of the driver cost parameters are reported in Table 5, Panel I. We produce estimates on eight samples, dividing drivers into owner-operator or fleet and their shifts into day or evening and weekday or weekend. As discussed in Section 2, these are natural divisions across which opportunity costs should differ. Because we normalize the scale of utility to dollars, we can also interpret cost functions in dollar terms. The raw parameter values show that time costs are decreasing and convex, with steeper costs on weekends compared to weekdays and evening shifts compared to day shifts.

In Table 5, Panel II, we use the mean cost parameters θ_{1k} and θ_{2k} to compute drivers' marginal cost of time at the typical hour of quitting. This value is computed as the mean of the derivative in total time cost with respect to hours worked (i.e., $mc_k(h) = \theta_{1k} + 2\theta_{2k}h$) where h is the final hour of each driver's shift. For example, the first column shows that owner-operators during daytime weekday shifts have a cost of time at the average hour of quitting of \$40.73 per hour. We contrast with Panel III, which shows the average shift earnings in that hour to be \$36.94. The

[17] Note that the full likelihood function normally also includes the contribution of the evolving state variables. But due to the conditional independence assumption, we can separate the likelihood function into two additive terms, one involving state transitions and one involving choices. Since we assume drivers are small in this market, they do not impact state transitions so that additive part drops out of the likelihood function.

[18] In practice, we impose a small penalty term on high values of σ_k.This is because the solver otherwise overweights matching the many "continue" choices, or $P(y_{\iota,s} = 0)$ compared with the relatively few"quit" choices, or $P(y_{\iota,s} = 1)$, and converges to degenerate distributions of θ and very high σ_ϵ to explain quitting. We find that by penalizing σ_k we require the estimator to match the quitting probabilities instead using the cost curves. This method provides cost curves that fit the data well. We offer more details in Section A.5.1

TABLE 5. Model Estimates

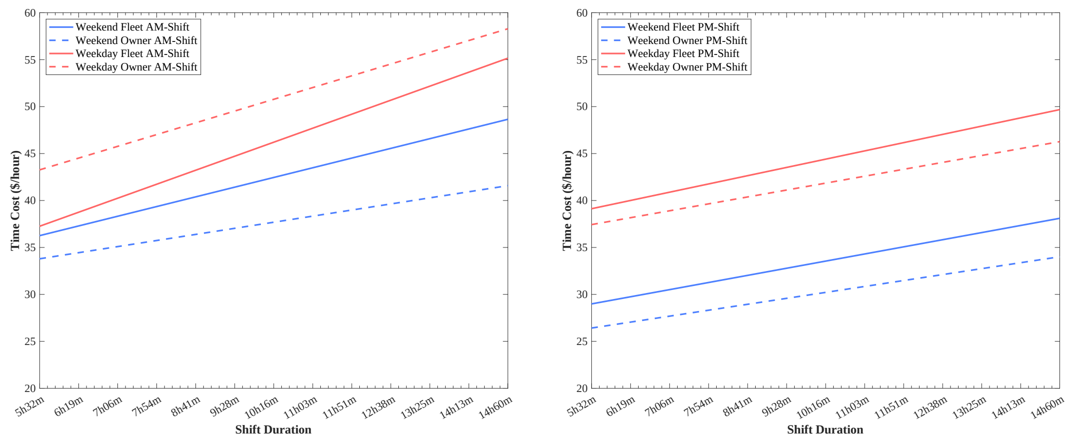
Driver-Type:	Owner-Operated		Fleet		Owner-Operated		Fleet	
Shift-Type:	AM		AM		PM		PM	
	Mon-Fri	Sat-Sun	Mon-Fri	Sat-Sun	Mon-Fri	Sat-Sun	Mon-Fri	Sat-Sun
I. Estimates								
σ_ϵ	14.65	13.78	13.43	17.42	17.04	12.54	19.14	17.94
	(5.35)	(4.96)	(5.11)	(6.42)	(5.59)	(5.12)	(7.08)	(6.69)
θ_1	208.70	136.91	240.04	210.15	155.75	128.38	185.25	175.05
	(7.34)	(4.76)	(8.62)	(7.43)	(5.47)	(4.89)	(7.30)	(6.44)
θ_2	-7.93	-4.11	-9.44	-6.54	-4.66	-4.01	-5.57	-4.81
	(0.27)	(0.14)	(0.33)	(0.23)	(0.18)	(0.15)	(0.23)	(0.18)
II. Implied by Estimates								
Last Hour Time Cost ($/hr.)	40.73	54.02	40.59	59.46	52.30	43.51	50.38	40.59
	(18.81)	(16.16)	(17.39)	(19.20)	(17.35)	(14.62)	(20.10)	(17.19)
III. Data Comparison								
Last Hour Earning ($/hr.)	36.94	39.68	37.57	40.07	38.55	40.89	37.80	41.26
Avg. Shift Minutes	472	399	509	456	577	568	550	567
Avg. Trips Per Hour	3.8	3.9	3.8	3.8	3.7	4.0	3.6	3.9

This table shows model estimates by shift, owner-status and weekday/weekend. Panel I shows parameter estimates as well as standard errors. Panel II shows the average marginal time cost of drivers at the time of quitting. This cost is computed as the mean (across drivers in each group) of model estimates of drivers' time costs at the time that driver quits. Panel III displays average cumulative earnings and work time for driver shifts within each group. Standard errors are obtained by resampling entire driver shifts, with replacement, and re-estimating state transitions, parameter estimates (panel I) and associated moments (panel II) within each driver- and shift-type. We conduct the estimation across 200 samples for each group and report standard deviations in parentheses.

discrepancy between the two rows is due to the role of unobservables. The bottom row of Panel III shows the expected number of trips per hour across shifts (computed as the average of the inverse spell length, in hours, across trips of each shift type). This value informs us how many draws of ϵ_{kty} are obtained by drivers in each hour. Thus, the model predicts that driver facing systematic hourly costs of $40.73 are likely to quit in an hour when average earnings are $36.94, because total costs are likely to surpass these earnings. Moreover, since time costs are rising rapidly by 7–8 hours due to the quadratic term, these estimates appear to broadly rationalize drivers' quitting behavior around the observed times. Finally, Table 5 Panel III also displays average shift durations by group. Average shift durations across groups are between 7h 40m and 9h 40m.

FIGURE 3. NYC Taxi Driver Marginal Cost of Time Estimates

I. AM-Shifts II. PM-Shifts



This figure shows estimates of drivers' marginal cost of time by day/evening and weekday/weekend shifts.

While the empirical specifications in Table 5 are simple – payoff functions have only three parameters – the behavioral implications of the dynamic model are quite rich. In Figure 3 we present drivers' marginal cost of time, or $\delta u(r,h)/\delta h$. These marginal cost functions are linear as $u(\cdot,\cdot)$ is quadratic in cumulative time-worked. An immediate implication of the positive slopes is that individual driver labor supply elasticities are positive. As earnings grow, all else equal, a driver's optimal amount of cumulative work time will increase on average. Day shift preferences are more alike than evening shift preferences; the higher marginal costs in the day shifts are almost uniformly higher. The largest slope, among AM-fleet drivers, appears to demonstrate the influence of the 5pm shift change period, in which fleet drivers have to return to bases to hand over leased medallion vehicles. Day shift cost differences may also reflect a preference for traditional work hours. We also see that marginal costs are slightly less steep on weekend shifts compared with weekday shifts, implying lower driver opportunity costs on weekends.

Nevertheless, a point of emphasis here is that once taxi drivers' quitting decision are modeled in a dynamic optimal stopping framework, the existing static wage or hazard models are hard to interpret as they largely require drivers to react to past outcomes instead of forward-looking tradeoffs. However, in the dynamic context we can also see why past outcomes may matter in more subtle ways: drivers react to past earnings shocks (i.e., observed "high" draws from $\mathcal{F}_k$) because there is autocorrelation in these shocks, at times positive and other times negative. Drivers react to cumulative time because it moves them up their cost curves as illustrated in Figure 3. Both types of serial correlation lead static models to identify real effects of variation in contemporaneous time and income, but absent the dynamic model the interpretation of these effects is difficult.

5.1. **Generating the Behavioral Results.** In this section we show that data generated from our estimated model produces driver behavior that appears as non-standard or "behavioral" when analyzed in a static model. Our first step is to use our model to simulate data that takes the same form as our actual data set. To do this we simulate a large set of driver shifts conditional on driver types (owner-operator or fleet drivers), shift-types (morning or evening shift periods), and day-types (one of five levels of average daily earnings), where drivers begin at an initial state and take draws from the joint distribution of earnings and time when they choose to keep working. We design this exercise to generate a new data set such that driver-, shift-, and day-types are represented in proportions identical to those found in the original data set. We then conduct a series of regressions analogous to those used in the literature and show that the results indicate apparent downward-sloping labor supply curves. In Section A.7 we provide a detailed description of the simulation exercise.

In Table 6 we report estimated wage regressions next to analogous regressions from the original data set. Both data and estimates underlying the simulation come from a two-month period of July 1, 2012 to August 31, 2012. The first two columns display instrumented wage regressions comparable to those in Camerer *et al.* (1997).[19] We report results using two separate instruments representing the distribution of wages among drivers on a given day. *IV 1* denotes the Camerer *et al.* (1997) wage instrument, or the 25th, 50th and 75th percentiles of the average shift wages of other drivers on the same day. *IV 2* denotes an alternative instrument as used in Farber (2015) equal to the average of the daily shift wage rates across all drivers. Under all specifications we find significant negative coefficients on log wage, indicating that shifts in which drivers who earn higher wages are correlated with shifts in which drivers work less

[19]The exception is that our estimates are not conditioned on weather and temperature variables so we omit these covariates across the table.

TABLE 6. Simulated Wage Regressions Comparison

	(1) Simulated (IV 1)	(2) Simulated (IV 2)	(3) Data (IV 1)	(4) Data (IV 2)
Log Wage	-0.602**	-0.602**	-0.190**	-0.158**
	(0.049)	(0.049)	(0.017)	(0.018)
Weekday	0.026**	0.027**	0.047**	0.046**
	(0.003)	(0.003)	(0.003)	(0.003)
PM Shift	0.110**	0.112**	0.013**	0.011**
	(0.004)	(0.003)	(0.003)	(0.003)
Owner Operator	0.090**	0.090*	-0.055**	-0.056**
	(0.002)	(0.003)	(0.002)	(0.002)
N	69,258	69,258	382,241	382,241

TLC Data from July-August, 2012. Panels (1)-(2) use simulated driver shift data. Data record the final cumulative hours and average wage earned as of the last trip of each simulated driver-shift. IV 1 denotes wage instruments are the 25th, 50th and 75th percentile across all driver wages in each day type, weekday/weekend and am/pm shift. IV 2 denotes wage instruments is the mean of average hourly driver wages across all drivers in each day type, weekday/weekend and am/pm shift. Panels (3)-(4) use TLC data and report the same regressions, where IV 1 denotes the wage instruments are the quartiles of hourly driver wage each day and IV 2 denotes wage instruments that are the average hourly driver wages across all drivers. Standard Errors clustered at the driver-shift level. Asterisks (**) indicate significance at or above the 1% level.

time.[20] Columns (3)-(4) replicate identical specifications by directly using the data. While we replicated the specifications of Camerer *et al.* (1997) in Table 2, here we demonstrate that the negative coefficients still obtain when we compare them to our simulated regressions. While our model inherently abstracts from the richness of the decisions taken by actual drivers on the street, our elasticity estimates are slightly *more* negative than those produced using the actual data. This suggests that, despite its simplicity, our model fully captures the "behavioral" aspects of driver behavior as documented in the prior literature.

Despite the apparent negative wage elasticities, we know that our model is, by construction, fully consistent with standard or neoclassical preferences for earnings at all states. To see why we obtain negative coefficients on wage in hours worked, and positive coefficients on earnings in the probability of quitting, we turn back to Section 2.1. There we document how spell wages of a given driver tend to decline relative to other drivers the longer his shift grows. Thus, by

[20]In columns (1)-(2), which use data simulated from our estimated model, a "day" is simply the combination of *day types* d and weekday or weekend. In columns (3)-(4), which use actual data, a day is defined as the calendar day in which the shift began. Within a day, average shift wages are separated between am and pm shift workers.

evaluating the average wage and the total time worked, we find that longer shifts are associated with a lower average earnings-per-hour compared to drivers who worked shorter shifts as a result of the pattern of declining spell wages.

This pattern holds despite an instrument for wages that generates exogenous variation in the average earnings per hour. This result indicates that a selection bias is present: drivers who work for longer hours are more likely to be drivers with lower average wages, given the negative correlation between hours and average wage. In the IV regressions, this effect appears to dominate the standard effect in which longer hours are worked due to higher earnings. We show in Section 6.2 that the dynamic model can be used to disentangle these two channels in order to estimate the actual impact of a persistent increase in earnings on total work hours.

Next, we turn to the case of time-inconsistent preferences highlighted in Thakral and Tô (2021). This result again suggests a non-standard, downward-sloping labor supply curve in certain periods of time close to the end of the work day.

By incorporating negative serial correlation into the dynamic model, our simulations also produce data that align with the time-inconsistent preferences. To show this, we estimate the following equation:

$$Pr(y_{int} = 1) = \sum_{\ell} \beta^{\ell}(h_{int}) r^{\ell}_{int} + X_{int}\gamma + \epsilon_{int} \tag{6}$$

Here y is the binary decision for driver i to quit or not on specific shift n at time t. As in Thakral and Tô (2021) we allow drivers' decisions to depend on cumulative earnings r which are earned in hour ℓ of the shift. X includes controls for hour and shift.[21]

We collect only the observations where drivers have a pickup or drop-off in the middle 20 minutes of the eighth cumulative hour (i.e., drivers that are observed between 8:20-8:40 minutes into their shift), and we control for income earned in the previous hours starting at the fourth cumulative hour, not including the final hour.[22]

[21] Note we do not include the full set of weather controls and driver fixed effects because these are not estimated separately in our model. However, these controls are not pivotal to the outcomes documented in Thakral and Tô (2021).

[22] By limiting the length of the window we mitigate concerns about selection within the window, where higher earnings mechanically correlate with higher hours. A similar approach appears in Thakral and Tô (2021).

In Table 7, we use this sample to replicate the analysis and specifications of Table 4 in Section 2.3, instead using our simulated data. While our sample is derived from a model, which is inherently coarser compared to actual data, we nevertheless reproduce statistically significant positive coefficients in the seventh hour, a result that is robust to the different specifications. As discussed in Section 2.3, we only anticipate late-in-shift effects in this instance because this when the continuation value differences between locations are more likely to induce differential quitting decisions. We also note that Table 4 also finds strongest effects in the seventh hour.

Although our simulated sample is much smaller and, by construction, less rich than the TLC data itself, these results suggest we can replicate patterns consistent with adaptive reference dependence or late-in-day income targeting without any explicit modeling of these phenomena, solely through incorporating forward looking decisions into an otherwise simple preference specification.[23]

6. Analysis of Labor Supply Elasticity and the Effect of Rising Fares

In this section we examine intertemporal labor supply elasticity from both the individual and equilibrium perspective. The existing literature primarily evaluates this elasticity from the individual perspective; in most markets, increased labor supply of some workers does not directly impact wages of other workers. However, in gig-economy settings and particularly in taxi and ride-hail markets, increases in labor supply can directly dampen the returns to all drivers through increased waiting times. With this dynamic in mind, we evaluate the equilibrium labor supply effects of taxi fare increases, such as recent policy designed to increase driver wages among NYC taxi drivers. In this case, fare increases would be expected to impact both supply, through the aforementioned channel, as well as demand.

[23]As we assume driver breaks are fully captures by the state variable, one might ask if our model embeds some behavioral aspects — such as effort targeting — into the estimated transition functions $\mathcal{F}$. To accommodate this possibility, we consider an effort targeting hypothesis whereby drivers whose income surpasses a certain level exert less effort to find additional rides, instead of quitting outright; this results in "soft-quitting" whereby rising incomes causes drivers to soft-quit by exerting progressively lower effort (It is debatable whether such a model would be either rational or boundedly rational). Nevertheless, as long as the effort policy function can be characterized by the driver's state variable of hours worked and earnings accrued, we can identify these effects non-parametrically and control for them by conditioning drivers' decisions on their state. We have performed additional analysis in Section A.3.1 to control for drivers' exact states and show that cumulative earnings at a fixed state is not predictive of total shift work time. This casts doubt on the hypothesis that higher earnings (ceteris paribus) triggers "soft quitting", which otherwise would be manifested in shorter predicted shift work times going forwards. As a result, we do not believe that $\mathcal{F}$ embeds the effort-targeting hypothesis.

TABLE 7. Simulated Elasticity of Stopping at 8.0-8.5 Hours with Respect to Income

	(1) Full Set	(2) Combined Incomes	(3) Adjusted
Total Income	-0.498*	-	-
	(0.286)	-	-
Income in Hour 1	0.244	-0.255	-0.793
	(0.360)	(0.359)	(0.828)
Income in Hour 2	0.241	-0.258	0.498
	(0.400)	(0.364)	(0.654)
Income in Hour 3	0.242	-0.257	0.795
	(0.450)	(0.349)	(0.645)
Income in Hour 4	0.993*	0.495	0.329
	(0.460)	(0.361)	(0.652)
Income in Hour 5	1.003*	0.504	-0.496
	(0.464)	(0.364)	(0.658)
Income in Hour 6	0.051	-0.447	-0.071
	(0.469)	(0.365)	(0.671)
Income in Hour 7	1.197***	0.698*	1.334**
	(0.461)	(0.365)	(0.668)
Income in Hour 8	0.261	-0.238	0.144
	(0.493)	(0.351)	(0.633)
Observations	18,345	18,345	18,345

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This table replicates Table 4 with simulated data, and is based on Table 2 in Thakral and Tô (2021). Data are simulated from the dynamic choice model. Asterisks (*) and (**) indicate significance at or above the 5% and 1% level.

6.1. **Incorporating Latent Heterogeneity.** Because this section surrounds quantification of an important and oft-debated parameter, we adapt our model to be more realistic by incorporating unobserved heterogeneity among drivers. We allow for each observed driver-shift type k to further have two latent shift types $\alpha \in \{1, 2\}$. For example, drivers may be more or less tired at the start of the day and operate according to different costs of time. In this extension, we allow cost functions to vary flexibly by each observable type k and latent-type α, i.e.,

$$u_{k,\alpha}(r, h) = r + \theta_{1,k,\alpha} h + \theta_{2,k,\alpha} h^2. \tag{7}$$

By definition, we do not observe latent type classifications α. To estimate the model, we use an EM-algorithm that allows for jointly estimating parameters and the latent type distributions via an iterative maximum likelihood routine as developed in Arcidiacono and Jones (2003). In this method, a candidate type distribution is assigned (the E-step) and, given this, cost parameter estimation is conducted (the M-step) identically to the approach outlined in Section 4.4. Likelihood values for the type distribution are then updated via Bayes' rule and the process is iterated until type distributions converge. In Section A.5.2 we provide additional results without latent heterogeneity, which reveal elasticities that are very similar.

6.2. **Estimating Individual Labor Supply Elasticity.** Our estimated model allows us to recover an estimate of the labor supply elasticity of taxi drivers with respect to wage rates. Given the day-to-day nature of earnings variation and driver labor supply choices, it is natural to interpret these as short-run, intensive-margin inter-temporal substitution (Frisch) elasticities. While these estimates are the target of the literature reviewed in Section 2, we have shown that existing approaches are flawed when it comes to identifying them. We estimate these elasticities by constructing counterfactual wage rate increases and simulating driver shifts and expected work hours with and without the increases. For example, we consider a 10% increase in all earnings available to a *single driver* across the joint distribution of earnings and time draws. This would represent a 10% increase in the measurement of wages preserving the stochasticity of wages as well as the negative auto-correlation discussed in Section 2.1. This counterfactual assumes that demand as well as the supply of all rivals is held fixed. In other words, we make a large markets assumption: when a single driver faces the earnings increase, this driver's impact on demand and any spillover effects to other drivers are negligible.

Table 8 displays the estimated elasticities of hours worked with respect to earnings rates. They show that for a range of earnings increases from 5–25%, the estimated mean hours worked, across all driver types, increases by up to 105 minutes, implying individual elasticities between 0.64–0.93. Table 5 in Farber (2015) estimates aggregate elasticities to be 0.589. That our estimates are slightly higher is not surprising: there is a downward bias in the reduced form approach due to drivers' within-shift declining productivity of earnings.

We refer to these estimates as *individual* labor supply elasticities because they hold search times fixed, implying that the labor supply of other drivers and consumer demand are both held

TABLE 8. Individual Labor Supply Elasticity

Earnings Change	Hours Worked			Implied Elasticity	
	P25	Mean	P75	Mean	Std. Err.
Baseline	5.72	7.47	9.38	.	.
5% increase	5.95	7.71	9.63	0.638	0.284
10% increase	6.22	8.02	10.00	0.730	0.281
18% increase	6.59	8.58	10.77	0.826	0.289
25% increase	7.09	9.21	11.68	0.933	0.294

This table reports the distribution of estimated work-hours resulting from simulating shifts at the baseline as well as across a series of increases to the earnings from each trip, assuming that demand and all other taxis behavior remains fixed. Implied elasticities are computed at the mean of hours worked. Standard errors are obtained by resampling entire driver shifts, with replacement, and re-estimating model parameters for each driver-shift type and re-simulating data in equal proportion to how each driver-shift is distributed. We conduct the exercise across 200 samples, compute the weighted average elasticity across shift types, and report standard deviations in parentheses.

fixed. If a wage increase were implemented on the entire market, other drivers would change their behavior and consequently lead to endogenous changes in search times. In addition, in this exercise we are holding fixed passenger prices: if passenger prices increased, this would decrease demand for taxi rides and again lead to longer search times. Therefore our estimates can be interpreted as the effect of an exogenous subsidy to an individual driver holding all else equal. Finally, driver preferences are identified from transient earnings variation, so our elasticity estimates can be interpreted as changes due to transient earnings shocks; drivers may exhibit different tradeoffs when earnings changes are permanent.

Looking to other comparable settings, our individual elasticity estimates are quite close to the intertemporal elasticity estimate of 0.704 in response to a 10% earnings shock as reported in Pistaferri (2003), which studies labor supply responses to earnings shocks and expectations among households in Italy. Angrist *et al.* (2021) conduct a leasing experiment among Uber drivers and estimate intertemporal substitution elasticities around 1.2 – 1.8. There, the authors remark that these elasticities are likely to be larger than those for total hours worked, since many Uber drivers have other jobs (Hall and Krueger (2018)). [24]

[24] See also Table 1 of Chetty *et al.* (2011), which among other statistics provides an extensive meta-analysis of quasi-experimental evidence on intensive-margin labor supply elasticities, highlighting an average inter-temporal substitution elasticity of 0.54.

We implicitly assume that in a given day the earnings process $\mathcal{F}_{kx}$ is exogenous. One threat to our strategy would be the case that drivers endogenously choose their levels of search effort, perhaps at higher cost, when the earnings profile changes. In Section A.3 we conduct a simple test for this by analyzing drivers who achieve unexpectedly high earnings early in the shift. We find that such high earnings do not predict subsequently higher earnings, implying that drivers who perceive the day to be more profitable than it is do not earn higher profits in later periods. This suggests that endogenous effort is not confounding our analysis of earnings elasticities.

Although individual elasticities are an important quantity of interest, we can also use our data and empirical approach to learn more. The reason for including the specific case of 18% is that in September, 2012 the NY TLC hiked fares by 18%, which affected search times through both passenger demand and equilibrium effects of labor supply. In the next section we use this change in fares to evaluate and compare individual labor supply elasticities with aggregate or equilibrium labor supply elasticities and consider these effects in the context of recent minimum wage legislation applied to taxi and ride-hail workers in New York City.

6.3. **Equilibrium Elasticities and Wage Policy.** On December 19, 2022 the New York Taxi and Limousine Commission implemented the first fare hike in ten years under a proposal known as "Raise For All".[25] Fares increased through a mix of increased base fares and increased surcharges during rush hour trips and trips to airports. The net effect is estimated by the TLC to increase average passenger fares by 23%.[26]

We investigate how much work hours will be impacted by the rising fares and whether the data generated by such a change can be informative about drivers' earnings elasticities. Data are not available to study the 2022 event as the TLC no longer provides access to driver and medallion identifiers. We instead study this question by considering the last time in which fares were changed. On September 4, 2012 the NYC TLC raised base taxi fares across for all drivers and customers. The distance fee increased from $\$2.00$ per-mile to $\$2.50$ per-mile, the JFK airport flat-fee increased from $\$45$ to $\$52$ and the Newark Liberty Airport surcharge increased from $\$15$ to $\$17.50$. Collectively these changes amounted to an 18% increase in the expected cost of a trip.

[25] See details in Hartwell *et al.* (2022).

[26] This estimate seems to compare the existing distribution of trips and routes and multiplies these trips by the new fares. In reality, even if we ignore potential supply-side effects, the demand response alone on more affected routes would likely dampen this aggregate estimate.

At the same time, drivers also became more competitive: the number of daily active medallions grew by 24.5% between a two-month period before and after the fare hike.[27]

Fare increases of this form impact the earnings process through three channels. First, drivers directly earn more from each trip according to the new fare schedule. Second, rival drivers change their labor supply behavior and the level of competition and therefore the expected searching times are impacted. Third, there is a demand response: higher trip prices depress demand and increase the expected search times for taxi drivers. In determining individual labor supply elasticities we only considered the impact of the first channel. Market-wide price changes will give rise to different work-hours elasticities because they also impact the other two channels.

An important caveat is that our parameter identification strategy allows us to estimate drivers' short-run Frisch elasticities in the absence of wealth effects or other long-run considerations. The 2012 fare change was permanent and therefore may have induced such effects. Nevertheless, our primary aim is twofold: first, we want to contrast drivers' elasticities from direct changes to individual earnings changes from those induced by a comparable change in product prices. Second, we want to demonstrate that the New York Taxi and Limousine Commission overestimates driver benefits from fare hikes. For this goal, we argue that our Frisch estimates provide good upper-bounds on the long-run uncompensated (Marshallian) elasticities. This is because the latter also captures wealth effects, which counteracts the substitution effect. Moreover, the evidence suggests that even utility-compensated long-run (Hicksian) elasticities are less than Frisch elasticities (see, e.g., Table 1 of the survey Chetty *et al.* (2011)).

To measure the short-run elasticity of hours worked with respect to fares, we use our model and estimated parameters to simulate data under the state transitions from (1) the two months before and (2) the two months after the September 2012 fare increase. Critically, the state transitions embed all relevant information drivers need to formulate new stopping rules. Therefore, by simply observing the earnings process encountered by drivers before and after the change took effect, we can avoid modeling the search and matching process and thereby also avoid the

[27]While we do not explicitly model entry, this change implies an extensive margin elasticity around 1.36. Paired with our intensive margin elasticity estimates, our findings are broadly consistent with the aggregate elasticity estimate of 1.8 in Frechette, Lizzeri and Salz (2019). The extensive-margin effect seems to be driven mostly by increased activity in the overnight hours, when the regulatory medallion caps rarely bind.

additional restrictive assumptions that are necessary to estimate such a model. By simulating driver shifts and hours worked across the *pre-* and *post*-fare hike periods, we can compute market-wide labor supply elasticities.

TABLE 9. Individual vs. Market Labor Supply Elasticity

	Baseline	**I. 18% Individual Wage Increase**			**II. 18% Market Fare Increase**		
	Hours	Hours	Elasticity	Δ Welfare	Hours	Elasticity	Δ Welfare
Overall	7.47	8.58	0.826	+13.10%	8.06	0.122	+7.79%
Weekday, AM	7.30	8.28	0.660	+11.89%	7.58	0.217	+8.04%
Weekday, PM	7.92	9.08	0.707	+12.09%	8.08	0.110	+7.99%
Weekend, AM	7.48	8.67	0.762	+12.75%	7.11	-0.280	-3.82%
Weekend PM	7.30	8.66	0.868	+18.06%	7.55	0.188	+15.45%

This table shows the labor hours and welfare effects of an 18% individual earnings increase (panel I) and an 18% market fare increase (panel II). The baseline and Panel I are estimated on a sample from Aug 1, 2012 – September 3, 2012. Panel II is estimated on a sample from September 4, 2012 – October 31, 2012, just after the fare change. Welfare is computed as $\sigma_\epsilon^{-1}\sum_i log(exp((r_i + C(h_i|\theta_{\tau(i)})/\sigma_\epsilon) + 1/\sigma_\epsilon))$ for each driver-shift i and shift type $\tau(i)$.

Table 9 reports the mean work hours across all simulated shifts before and after the fare change along with implied work-hours elasticities and changes to driver welfare. Averaging across shifts and drivers, the overall elasticity with respect to fares is 0.122, or about 85% less than the elasticity with respect to individual earnings. This implies smaller benefits to a single driver once we account for the market adjustment to the wage increase. This comparison is varied across different types of shifts, likely reflecting divergence in both demand and driver preferences across groups in the pre- and post-periods. Our findings align with Hall *et al.* (2023), who study Uber fare increases and find that market re-equilibration due to increased labor supply and reduced demand lead to very limited effects of driver earnings. Day-shift weekend simulations suggest that drivers actually work less after the fare hike; there, drivers on average work 30 minutes *less* when the fare increases. This result captures the fact that earnings opportunities after the September 4, 2012 fare increase decline in hours worked for this shift type. At high levels of cumulative work time, the post-fare change average spell wages are below those of the pre-fare change period.[28] We also derive welfare estimates to evaluate drivers'

[28]We detail these patterns in Section A.4, and note that the patterns observed in the data will likely also embody longer-run elasticities due to the permanent nature of the fare hike.

overall benefits net of costs. An average driver is about 13% better off from an individual earnings increase, compared to 8% better off from a fare increase.

The main goal of the exercise in Table 9 is to demonstrate the importance of distinguishing market-wide price and earnings variation from individual earnings variation in the assessment of labor supply elasticity. Since prices are regulated and therefore set exogenously, price changes induce both a demand and supply response as well as an equilibrium adjustment to expected search times. More generally, earnings variation may be induced by shifts in demand or supply. In all of these cases, drivers' forward-looking tradeoffs and therefore labor supply decisions are impacted in unpredictable ways: demand elasticities may vary spatially, shifts in demand may be local depending on what events are taking place, etc. The mission to measure labor supply elasticity from observational data is inherently complicated by these factors. This is where our counterfactual approach can offer a clean and clear alternative; by simulating driver shifts subject to a uniform earnings increase, we are able to replicate a wage experiment without the confounding effects of equilibrium adjustments.

7. Conclusion

We use a comprehensive dataset of trips and work hours among New York City taxi drivers to take a new approach to a long-running question of drivers' wage elasticities by modeling taxi drivers' labor supply decisions as emerging from a dynamic *optimal stopping* problem. Our model explicitly assumes that drivers have standard preferences for labor and leisure, implying standard behavior stemming from upward sloping labor supply curves.

We estimate our model and use it to simulate a panel of driver shifts. We then conduct a static analysis of drivers' labor supply behavior that is analogous to specifications used in previous literature. We demonstrate that we can replicate the same patterns in the literature, in which labor supply elasticity may appear to have a negative sign. We show that these patterns arise because of previously unexplored intra-daily dynamics in earnings per hour. In particular, drivers become less productive as they work longer, and there is also negative autocorrelation in long trips that generate apparent earnings shocks in a static framework.

Our results reconcile a twenty-five-year debate in this literature. More broadly, these findings suggest that once we account for the dynamic incentives in taxicab drivers' labor supply decisions, there is no need to add behavioral parameters to the model to explain their quitting behavior. However, we note that behavioral patterns may serve as useful heuristics for drivers that happen to coincide with the more complicated dynamic optimization problem.

Finally, our model is also capable of answering the question of what is the intertemporal labor supply elasticity of New York City taxi drivers. We find individual elasticities of 0.64–0.93. These values are strikingly close to estimates obtained in other settings. However, to evaluate wage policy among drivers, such as the recent fare hike among New York taxi drivers, we also have to acknowledge the equilibrium impact of wage changes as it transmits from demand elasticities as well as the spillovers from all drivers re-optimizing their search behavior. We find that the average market-level elasticity to a 18% wage hike is 0.12, or about a sixth of the elasticity obtained under a direct 18% earnings increase. This result implies that earnings variation due to price changes cannot be directly used to measure labor supply elasticities, particularly in settings where there are network effects such as platform-based labor markets.

References

ALTONJI, J. G. (1986). Intertemporal substitution in labor supply: Evidence from micro data. *Journal of Political Economy*, **94** (3, Part 2), S176–S215.

ANDERSEN, S., BRANDON, A., GNEEZY, U. and LIST, J. A. (2014). *Toward an understanding of reference-dependent labor supply: Theory and evidence from a field experiment*. Tech. rep., National Bureau of Economic Research.

ANGRIST, J. D., CALDWELL, S. and HALL, J. V. (2021). Uber versus taxi: A driver's eye view. *American Economic Journal: Applied Economics*, **13** (3), 272–308.

ARCIDIACONO, P. and JONES, J. B. (2003). Finite mixture distributions, sequential likelihood and the em algorithm. *Econometrica*, **71** (3), 933–946.

ASHENFELTER, O., DORAN, K. and SCHALLER, B. (2010). A shred of credible evidence on the long-run elasticity of labour supply. *Economica*, **77** (308), 637–650.

BARZEL, Y. (1973). The determination of daily hours and wages. *The Quarterly Journal of Economics*, **87** (2), 220–238.

BENKARD, C. L. (2004). A dynamic analysis of the market for wide-bodied commercial aircraft. *The Review of Economic Studies*, **71** (3), 581–611.

BESANKO, D., DORASZELSKI, U. and KRYUKOV, Y. (2014). The economics of predation: What drives pricing when there is learning-by-doing? *American Economic Review*, **104** (3), 868–897.

BRANCACCIO, G., KALOUPTSIDI, M. and PAPAGEORGIOU, T. (2020). Geography, transportation, and endogenous trade costs. *Econometrica*, **88** (2), 657–691.

BROWNING, M., DEATON, A. and IRISH, M. (1985). A profitable approach to labor supply and commodity demands over the life-cycle. *Econometrica: journal of the econometric society*, pp. 503–543.

BUCHHOLZ, N. (2022). Spatial equilibrium, search frictions, and dynamic efficiency in the taxi industry. *The Review of Economic Studies*, **89** (2), 556–591.

BURNSIDE, C., EICHENBAUM, M. and REBELO, S. (1993). Labor hoarding and the business cycle. *Journal of Political Economy*, **101** (2), 245–273.

CALDWELL, S. and OEHLSEN, E. (2018). Monopsony and the gender wage gap: Experimental evidence from the gig economy. *Massachusetts Institute of Technology Working Paper*.

CAMERER, C., BABCOCK, L., LOEWENSTEIN, G. and THALER, R. (1997). Labor supply of new york city cabdrivers: One day at a time. *The Quarterly Journal of Economics*, **112** (2), 407–441.

CASTILLO, J. C. (2022). Who benefits from surge pricing? *Available at SSRN 3245533*.

CHEN, K.-M., DING, C., LIST, J. A. and MOGSTAD, M. (2020). *Reservation wages and workers' valuation of job flexibility: Evidence from a natural field experiment*. Tech. rep., National Bureau of Economic Research.

CHEN, M. K., ROSSI, P. E., CHEVALIER, J. A. and OEHLSEN, E. (2019). The value of flexible work: Evidence from uber drivers. *Journal of political economy*, **127** (6), 2735–2794.

CHETTY, R., GUREN, A., MANOLI, D. and WEBER, A. (2011). Are micro and macro labor supply elasticities consistent? a review of evidence on the intensive and extensive margins. *American Economic Review*, **101** (3), 471–75.

CHRISTENSEN, P. and OSMAN, A. (2023). *The demand for mobility: Evidence from an experiment with uber riders*. Tech. rep., National Bureau of Economic Research.

CRAWFORD, V. P. and MENG, J. (2011). New york city cab drivers' labor supply revisited: Reference-dependent preferences with rational expectations targets for hours and income. *The American Economic Review*, **101** (5), 1912–1932.

DUONG, H. L., CHU, J. and YAO, D. (2022). Taxi drivers' response to cancellations and no-shows: New evidence for reference-dependent preferences. *Management Science*.

FARBER, H. S. (2005). Is tomorrow another day? the labor supply of new york city cabdrivers. *The Journal of Political Economy*, **113** (1), 46–82.

— (2008). Reference-dependent preferences and labor supply: The case of new york city taxi drivers. *The American Economic Review*, **98** (3), 1069–1082.

— (2015). Why you can't find a taxi in the rain and other labor supply lessons from cab drivers. *The Quarterly Journal of Economics*, **130** (4), 1975–2026.

FEHR, E. and GOETTE, L. (2007). Do workers work more if wages are high? evidence from a randomized field experiment. *American Economic Review*, **97** (1), 298–317.

FRECHETTE, G. R., LIZZERI, A. and SALZ, T. (2019). Frictions in a competitive, regulated market: Evidence from taxis. *American Economic Review*, **109** (8), 2954–92.

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

HALL, J. V., HORTON, J. J. and KNOEPFLE, D. T. (2023). *Ride-sharing markets re-equilibrate*. Tech. rep., National Bureau of Economic Research.

— and KRUEGER, A. B. (2018). An analysis of the labor market for uber's driver-partners in the united states. *Ilr Review*, **71** (3), 705–732.

HARTWELL, A., METZ, T. and DIGIOVANNI, J. (2022). 11/15/22 commission meeting.

HECKMAN, J. and MACURDY, T. (1980). A dynamic model of female labor supply. *Review of Economic Studies*, **47** (1), 47–74.

HUANG, L. and SMITH, M. D. (2014). The dynamic efficiency costs of common-pool resource exploitation. *American Economic Review*, **104** (12), 4071–4103.

LAGOS, R. (2000). An alternative approach to search frictions. *Journal of Political Economy*, **108** (5), 851–873.

— (2006). A model of tfp. *The Review of Economic Studies*, **73** (4), 983–1007.

LEVESON, I. F. (1967). Reductions in hours of work as a source of productivity growth. *Journal of Political Economy*, **75** (2), 199–204.

MACURDY, T. E. (1981). An empirical model of labor supply in a life-cycle setting. *Journal of political Economy*, **89** (6), 1059–1085.

OETTINGER, G. S. (1999). An empirical analysis of the daily labor supply of stadium vendors. *Journal of political Economy*, **107** (2), 360–392.

PENCAVEL, J. (2015). The productivity of working hours. *The Economic Journal*, **125** (589), 2052–2076.

PETTERSON, M. S. (2022). Estimation of a latent reference point: Method and application to nyc taxi drivers, working Paper.

PISTAFERRI, L. (2003). Anticipated and unanticipated wage changes, wage risk, and intertemporal labor supply. *Journal of Labor Economics*, **21** (3), 729–754.

RICCI, J. A., CHEE, E., LORANDEAU, A. L. and BERGER, J. (2007). Fatigue in the us workforce: prevalence and implications for lost productive work time. *Journal of occupational and environmental medicine*, **49** (1), 1–10.

Rosaia, N. (2023). *Competing Platforms and Transport Equilibrium: Evidence from New York City*. Tech. rep., mimeo, Harvard University.

Rotemberg, J. J. and Summers, L. H. (1990). Inflexible prices and procyclical productivity. *The Quarterly Journal of Economics*, **105** (4), 851–874.

Ryan, S. P. (2012). The costs of environmental regulation in a concentrated industry. *Econometrica*, **80** (3), 1019–1061.

Schmidt, M.-A. (2019). Valuing flexibility: A model of discretionary rest breaks, working Paper.

Thakral, N. and Tô, L. T. (2021). Daily labor supply and adaptive reference points. *American Economic Review*, **111** (8), 2417–43.

Wan, Y. and Xu, H. (2014). Semiparametric identification of binary decision games of incomplete information with correlated private signals. *Journal of Econometrics*, **182** (2), 235–246.

West, C. P., Tan, A. D., Habermann, T. M., Sloan, J. A. and Shanafelt, T. D. (2009). Association of Resident Fatigue and Distress With Perceived Medical Errors. *JAMA*, **302** (12), 1294–1300.

Online Appendix

Appendix A. Data: Additional Details

A.1. **Data Cleaning and Preparation.**

A.1.1. *Data Cleaning.* We begin with raw data obtained from the New York City Taxi and Limousine commission consisting of all yellow taxi trip and fare data from July 1 to September 3, 2012. The raw files consist of 29,939,090 observations. To obtain a structure suitable for estimating driver labor supply behavior, we throw out any data that appears contaminated by severe measurement error, key missing information, or highly unusual patterns.

We closely follow previous work using the same TLC dataset to prepare our data for the analysis of shifts (e.g., Haggag *et al.* (2017), Thakral and Tô (2021)). We start by cleaning the raw TLC data of obvious measurement errors or apparent extraneous data produced by duplicate or false entries or those data produced from electronic testing. The data are organized trip-by-trip, where we observe medallion identifiers and exact pickup and drop-off date-times. Our first step is to establish criteria for the change of shifts. To do this, for every driver we measure the time between trips and, for trip times with gaps beyond five hours, we define a change of shift (See discussion in Section 2). Next, we classify every driver-shift with a unique identifier. There are 1,408,646 driver-shifts in our full sample. When we identify a potentially erroneous trip or data problem, we flag the entire shift as having a problem and drop it from our analysis sample.

To begin, we flag the following problematic observations:

(i) Duplicates on medallion and pickup date-time
(ii) Missing or zero entry for trip distance or trip duration
(iii) Trip duration far less than feasible for trip distance
(iv) Trip distance far longer than feasible for trip duration
(v) Trip fare less than minimum TLC prices for normal fares (indicated as fare code=1)
(vi) Trip times less than 10 seconds
(vii) Trip times less than 60 seconds with fare greater than $10
(viii) Trips from Manhattan to JFK Airport with trip duration less than ten minutes
(ix) Trips from Manhattan to JFK Airport with trip distance less than 10 miles
(x) Trips with pickup time occurring before the previous trip's drop-off time
(xi) Latitude or longitude of trip could not be mapped to a destination within New York City or Newark Airport

Next, we flag the following problematic shifts:

(i) Shifts with more than one car per driver within a shift
(ii) Shifts with total duration longer than 18 hours or shorter than 2 hours
(iii) Shifts with 3 or fewer trips in total

We drop shifts with the above errors, leaving us with a data set of 8,220,299 observations, 30,231 drivers, and 444,317 unique shifts across July 1 to September 3, 2012.[29]

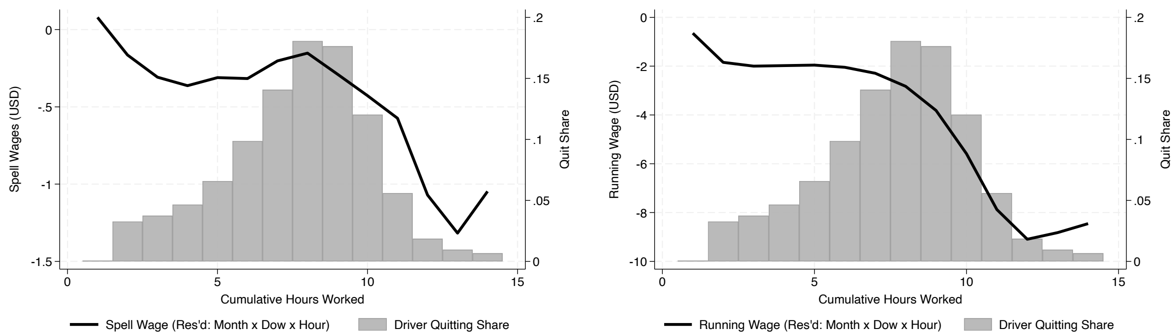
A.1.2. *Predicting Medallion Types.* The TLC issues different types of medallions with different restrictions that may impact driver incentives. Although we do not have data on medallion types, we screen medallions for patterns that indicate a higher likelihood of being *fleet* medallions (subject to higher levels of minimum usage and often stringent turnover hours) vs. owner-operator medallions (which have less onerous requirements). Our screen is constructed as follows:

(i) Number of drivers per medallion less than 4
(ii) Number of trips per medallion greater than 200

The first criterion checks that the taxi medallion is only utilized by a small number of individuals, for example licensed individuals within a family. The second criterion ensures that the small number

[29]Note we conduct an analogous data cleaning routine for the post-fare change period, September 4, 2012 to October 31, 2012, in constructing our sample for analyzing the fare hike counterfactual in Section 6.

FIGURE 4. Rate of Earnings by Cumulative Hours Worked



(a) Spell Wage (b) Running Wage

TLC from January to August 2012. This figure shows how earnings evolve with cumulative hours worked. Panel (a) shows the relation between spell wage (residual to location and time fixed effects) and cumulative hours worked. Panel (b) shows the relation between the residualized running wage and cumulative hours worked. Both panels depict the share of hours in which drivers quit, with units on the right y-axis.

of individuals is not a consequence of scant usage of the medallion. Although this screen is simple and coarse, it predicts well the probability that the medallion will be used during the witching hour, between 4–5pm, in which fleet medallions turn over to the evening shift. The benefit of using this screen without incorporating the witching hour directly is that some owner-operators are only active in the evening shift, for which there is no equivalent of the witching hour.

A.2. **Additional Descriptive Evidence.**

A.2.1. *Declining Earnings Conditional on Driver and Day-of-week by Hour.* Figure 4 shows that the declining earnings pattern persists after controlling for driver-level fixed effects.

A.2.2. *Effect of Hours Worked on Search Radius.* Figure 5 shows how the distance of passenger drop-off to subsequent passenger pickup changes as drivers work longer hours. The figure overlays a histogram of quitting probability by cumulative hours worked. It shows that average search distance does not change until about nine hours of work, after which it slowly increases. For reference, an average (long) city block in New York City is about 900 ft., or about .17 miles.

A.2.3. *Thakral and Tô (2021) Specification Discussion.* The estimates of recency bias in Table 2 of Thakral and Tô (2021) (hereafter, TT) include a control variable *total cumulative income earned*. Since this specification also controls for cumulative income earned during each hour up to hour 8:00, and because the data

FIGURE 5. Search Distance by Cumulative Hours Worked

This figure shows the increase in average straight-line distance of passenger drop-off to subsequent passenger pickup, by cumulative hour worked.

sample consists of trips and quitting decisions that occur from 8:20–8:40, the *cumulative income* variable includes any income earned during work hours hours 8:00–8:20.

In order to evaluate the elasticity of quitting with respect to earnings in a specific hour, TT first sums the coefficients for each preceding hour with the coefficient on cumulative income, and then scales this sum by a factor to transform it from a level effect into an elasticity. However, we found the summed coefficient is likely confounded with the effect of additional earnings in the 8:00-8:20 period.

To illustrate this, consider a simplified specification in which there are just two hours, 1 and 2: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3(x_1 + x_2 + x_3) + \epsilon$. Where $y \in \{0, 1\}$ is the quit decision, x_k is income earned in hour k and x_3 is total cumulative income. Thus β_3 is the coefficient on total cumulative income. To measure the effect of hour 2 income, the TT approach would report the sum $\beta_2 + \beta_3$. However, since x_3 is correlated with x_2, the effect of interest is $\frac{\delta y}{\delta x_2} = \beta_2 + \beta_3 + \beta_3 \frac{\delta x_3}{\delta x_2}$.

The sign of this bias in a given hour is somewhat ambiguous for reasons we explore in this paper. For example, we showed that some income shocks induce large negative serial correlation in earnings in subsequent hours, implying a negative bias. On the other hand, a day with a high demand shock will induce positive correlation in earnings across all hours, implying a positive bias. However, these biases tend to be more positive when we condition earnings on specific locations (in effect, purging a key source of the negative bias), resulting in $\frac{\delta x_3}{\delta x_2} > 0$ and β_3<0 in estimation, $\frac{\delta y}{\delta x_2} < \beta_2 + \beta_3$. Thus, there is likely an upwards bias in the reported results for income once we condition on earnings in a specific location.

By implementing our preferred adjustment of the time window, we nevertheless obtain the TT result to an extent by showing there remains evidence of recency bias, albeit weaker. We then show this remaining recency bias largely disappears when conditioning on high continuation-value regions of earnings, and yet amplifies when conditioning on low continuation-value regions of earnings.

A.2.4. *Elasticity of Stopping By Earnings Hour: Alternative Location Definition.* In Table 10 we display the Manhattan/Brooklyn vs. Outer Borough alternative of Table 4. Outer borough locations, analogous to the low-continuation value locations in Table 4, show positive probability of quitting in the final hour. Manhattan/Brooklyn locations are similarly captured by high continuation value locations.

A.3. **Prior Earnings and Endogenous Effort.** In this section we investigate whether taxi drivers engage in higher (or lower) "effort" when they perceive earnings opportunities to be high (or low). We specifically characterize effort as the extent to which drivers can choose their earnings per hour across some support by paying an extra cost. Endogenous effort, or effort that varies with demand or supply shocks, would imply that quantities such as earnings per hour, wage spells, and average wages are all equilibrium objects that cannot be imposed as policy counterfactuals.

To test for endogenous effort, we evaluate whether drivers who, through lucky draws of trips, experience consistently higher or lower earnings relative to other drivers within the first four hours of the same shift (and who therefore perceive higher or lower overall demand) achieve different expected payoffs in the subsequent four hours. The question is whether drivers who perceive high demand but in reality were lucky will adjust their effort levels to be more productive later in the day. We evaluate this via a regression that compares, among those drivers who work at least eight hours, the expected earnings of a driver in hours four through eight conditional on their first four hours. To account for the fact that there are predictable hourly patterns of earnings, we control for the hour in which the drivers' shifts begin.

Table 11 reports the results. We see that, conditional on four hour earnings, each dollar earned in the first four hours is associated with \$0.031 additional earnings in the second four hours. In other words, take a one standard deviation increase in hours 0-4 income (net of date, driver and starting hour controls) of \$31.87. A 1 SD increase in income over this period would predict a subsequent gain of \$0.99 over the next four hours. To verify that our selection of hours 4 and 8 are not driving these results, we report alternative versions of the test in each cell and see similar results.

We interpret these results to imply that drivers' day to day success with earnings, insofar as it shapes their beliefs about future earnings within the same day, does not lead to meaningful changes in their ability to earn income compared to other drivers in the same market. Effort is difficult to measure, but these results give us some confidence that it is not driving the earnings schedules of drivers.

TABLE 10. Elasticity of Stopping at 8.0-8.5 Hours with Respect to Income (Alt. Definitions)

VARIABLES	(1) Manhattan/Brooklyn	(2) Outer-Boroughs
Manh./Bkln. Income in Hour 1	0.407***	
	(0.145)	
Manh./Bkln. Income in Hour 2	0.070	
	(0.187)	
Manh./Bkln. Income in Hour 3	-0.256	
	(0.206)	
Manh./Bkln. Income in Hour 4	-0.034	
	(0.215)	
Manh./Bkln. Income in Hour 5	0.133	
	(0.215)	
Manh./Bkln. Income in Hour 6	0.267	
	(0.210)	
Manh./Bkln. Income in Hour 7	0.741***	
	(0.214)	
Manh./Bkln. Income in Hour 8	-0.244	
	(0.260)	
Outer-Boro Income in Hour 1		-0.318***
		(0.117)
Outer-Boro Income in Hour 2		-0.277
		(0.171)
Outer-Boro Income in Hour 3		0.162
		(0.196)
Outer-Boro Income in Hour 4		0.054
		(0.209)
Outer-Boro Income in Hour 5		0.278
		(0.217)
Outer-Boro Income in Hour 6		0.220
		(0.218)
Outer-Boro Income in Hour 7		-0.114
		(0.239)
Outer-Boro Income in Hour 8		3.447***
		(0.477)
Observations	129,153	129,153
R-squared	0.367	0.368

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This table shows the effect of cumulative income in each hour of the day on the probability of stopping after the eighth hour of cumulative work time. All estimates are presented as elasticities relative to the baseline probability of stopping. The sample consists of a 25% sample of New York TLC data from Jan 1, 2012 – Sep 3, 2012. Specifications include controls for driver, day-of-week x hour, week-of-year, drop-off neighborhood, cumulative income and cumulative work time. Column (1) replicates Table 2 of Thakral and Tô (2021) but limits hourly cumulative earnings to trips with origins in Manhattan and Brooklyn, includes a withheld category of hour 1 earnings, and uses observations from hour 8:00 to 8:20. Column (2) repeats the exercise with earnings from Bronx, Queens and EWR Airport.

A.3.1. *A non-parametric test.* In addition to the above, we present non-parametric evidence that drivers' labor supply decisions are not responsive to past earnings. We isolate 1.14M instances in which two taxi

TABLE 11. Evidence for Exogenous Effort

	Effect of 1 SD increase in earnings by:		
	Fourth Hour	**Fifth Hour**	**Sixth Hour**
E[Addt'l Earning] by Hour 7	$0.83	$0.22	$-0.02
E[Addt'l Earning] by Hour 8	$0.99	$0.35	$-0.01
E[Addt'l Earning] by Hour 9	$1.32	$0.62	$-0.00

This table shows the predicted change in earnings by the indicated hour in each row as a function of a one standard deviation growth in expected earnings by the hour indicated in each column. These values are obtained by regressing each row variable on shift income at the hour on each column along with date, driver and starting hour fixed effects.

drivers who have each worked the same number of minutes on the same day dropped off passengers less than 200 meters apart within a minute of each other. This precise matching allows us to control for nearly identical labor conditions and, crucially, isolate the effect of differences in cumulative earnings on subsequent work duration. We then demonstrate that differences in cumulative earnings at any point are, in fact, not predictive of different shift durations.

Table 12 summarizes this evidence. It presents a regression of total shift duration on driver earnings, where the sample is to limited to matched instances of two drivers with identical time spent on shift and coinciding passenger drop-offs as described above, and where we implement fixed effects for each matched pair. By comparing drivers in nearly identical situations, we neutralize the influence of differing future earnings expectations, focusing solely on the impact of earnings up to that precise moment. Our findings indicate that variation in past earnings, up to the point of each controlled drop-off, do not predict different shift lengths. This suggests that any influence of past earnings on labor supply decisions is likely through their correlation with anticipated future payoffs, rather than through direct effects.

We now turn to an analysis of dynamic forces that explain the past history of labor supply puzzles among taxi drivers. The next two subsections detail how these dynamics—specifically related to earnings expectations in light of drivers' intra-daily productivity evolution and spatial search effects —shape drivers' labor supply decisions in ways that purely static models fail to capture.

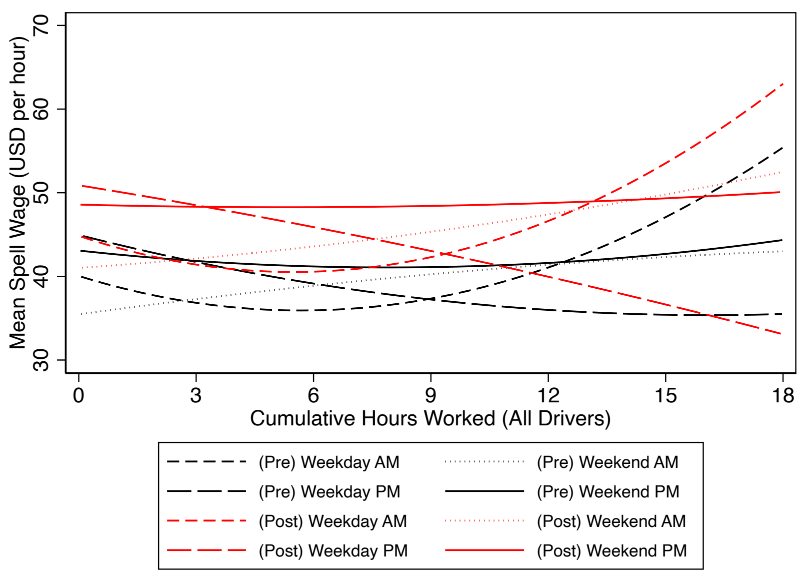
A.4. **The Evolution of Average Spell Wage by Cumulative Hours and Shift Types.** Figure 6 shows how average spell wages evolve with cumulative work hours by shift type. This figure does not directly describe hourly patterns in spell wage, because shift start times are distributed across several hours in the day and evening shifts. Its purpose, however, is to reconcile Table 9 which shows that, following an

TABLE 12. Prior Earnings Effects on Labor Supply

Variable	Dependent Variable: *Log(Shift minutes)* (1)	*Shift minutes* (2)
Log(Cumulative Earnings)	0.0012	.
	(0.0015)	.
Cumulative Earnings	.	-0.0023
	.	(0.0047)
Constant	6.239	530.2
	(0.007)	(0.645)
Match Pair Fixed Effects	✓	✓
N. Obs.	1,135,323	1,135,323

Note: Regression controls include fixed effects for each unique matched pair of driver drop-offs. Standard errors are clustered at the matched-pair level.

FIGURE 6. The Time Path of Earnings by Cumulative Hours Worked



This figure depicts, for each stated category of shifts, the expected spell wage earned by drivers with each level of cumulative hours worked. "Pre" denotes the sample period September 4 – October 31, 2011. "Post" denotes the sample period September 4 – October 31, 2012. The expectation is fitted via regression on cumulative hours and cumulative hours squared.

18% fare increase, work hours decreased for Weekday PM drivers. In Figure 6 we see that this category is the only one in which earnings opportunities decline relative to the pre-fare change period.

A.5. **Estimation: Additional Detail.**

A.5.1. *Likelihood Details.* In this section we discuss our maximum likelihood procedure in more detail. Our estimation is configured to maximize the log-likelihood function Equation 4. We found that, given a high degree of within-driver idiosyncratic variation in drivers' quitting decisions, for a few driver-type subsets the solver converges on solutions with constant marginal costs and large idiosyncratic variation. However, these parmeterizations do not fit quitting patterns well (drivers quitting probabilities are uniform over time); in essence they over-weight continuation decisions. As a means of re-weighting the likelihood to capture drivers' quitting choices, we impose a penalty therm γ and estimate a likelihood function via maximizing Equation 4 minus $\gamma * \sigma_\epsilon^2$ for each driver-shift type k. To choose γ we estimated the model over a gridded range of γ and found the smallest value such that no parameter solutions across driver types led to constant marginal costs, obtaining $\gamma = 0.25$.

A.5.2. *Heterogeneous Model Estimates.* This section includes extra results for both the heterogeneous model and basic model.

Table 13 shows the parameter estimates for the heterogeneous model.

Table 14 shows the elasticity estimates for the simple model.

A.6. **Model Fit.** In this section we show evidence on model fit with respect to quitting probabilities at different levels of hours worked. Figure 7 Panel A shows the fit of the dynamic logit model without latent heterogeneity. Panel B shows the fit of the model with latent heterogeneity. The main difference is that the latter model is able to better fit the cases of quitting after 12 hours. Note that only about 5% of drivers quit after 12 hours of work. As such, these observations are not weighted heavily within the maximum likelihood estimator. The model fits the data well in the vast majority of driver states that appear in the data.

A.7. **Simulation Details.** In Section 5.1 we use the estimated model to simulate driver shifts and then use those simulated data to reproduce key behavioral puzzles in the literature. For each simulated driver of type d, we draw a sequence of trips as (dr_t, dh_t) from the empirical distribution $F(dr, dh|r, h, x)$, draw a sequence of drivers' net shocks to the outside option ϵ_t, and for each decision point t together with parameter estimates, we compute whether drivers' value of quitting exceeds the value of continuing search. As with estimation, there are eight persistent driver types i and five categories of demand heterogeneity $k \in \{1, ..., 5\}$, defined as quintiles over average daily spell wages.

TABLE 13. Heterogeneous Model Estimates

Driver-Type:	Owner-Operated		Fleet		Owner-Operated		Fleet	
Shift-Type:	AM		AM		PM		PM	
	Mon-Fri	Sat-Sun	Mon-Fri	Sat-Sun	Mon-Fri	Sat-Sun	Mon-Fri	Sat-Sun
Estimates: Type I								
σ_ϵ	9.97	10.35	9.50	10.72	10.79	8.50	9.91	8.84
	(0.69)	(1.11)	(0.87)	(1.15)	(0.56)	(0.76)	(0.86)	(1.18)
θ_1	131.28	106.41	170.43	145.56	101.54	77.68	102.31	82.65
	(7.55)	(5.62)	(6.96)	(6.19)	(8.06)	(7.30)	(7.23)	(6.47)
θ_2	-5.08	-3.58	-6.96	-5.33	-3.59	-2.55	-3.68	-2.63
	(0.25)	(0.10)	(0.20)	(0.16)	(0.30)	(0.20)	(0.21)	(0.14)
Estimates: Type II								
σ_ϵ	9.31	9.69	8.88	10.03	10.11	7.90	9.29	8.25
	(0.65)	(0.47)	(0.79)	(0.86)	(0.55)	(0.71)	(0.78)	(0.84)
θ_1	122.90	99.54	159.87	136.33	95.14	72.06	95.93	76.96
	(7.05)	(3.37)	(6.51)	(6.22)	(8.11)	(6.82)	(6.76)	(6.06)
θ_2	-4.79	-3.37	-6.56	-5.02	-3.39	-2.40	-3.47	-2.47
	(0.23)	(0.13)	(0.18)	(0.12)	(0.37)	(0.19)	(0.19)	(0.11)

This table shows heterogeneous model estimates by latent type, shift, owner-status and weekday/weekend. Standard errors are obtained by resampling entire driver shifts, with replacement, and re-estimating state transitions, parameter estimates (panel I) and associated moments (panel II) within each driver- and shift-type. We conduct the estimation across 200 samples for each group and report standard deviations in parentheses.

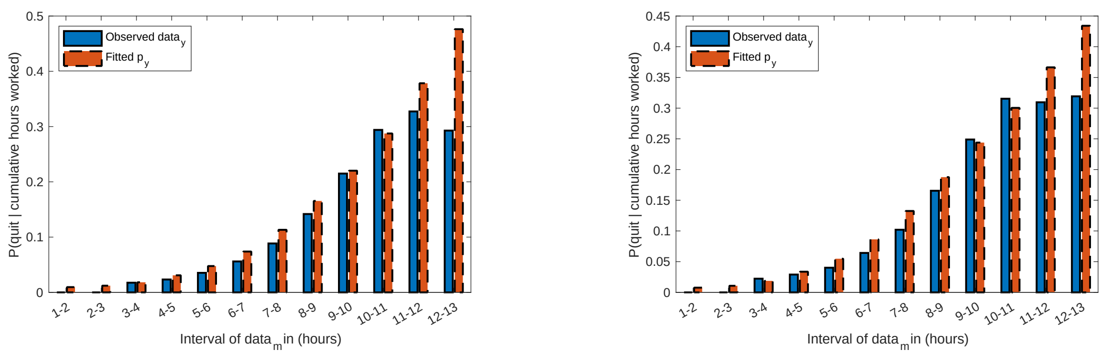
TABLE 14. Individual Labor Supply Elasticity

Earnings Change	**Hours Worked**			**Implied Elasticity**	
	P25	Mean	P75	Mean	Std. Err.
Baseline	5.70	7.39	9.27	.	.
5% increase	5.83	7.59	9.54	0.548	0.284
10% increase	6.10	7.87	9.80	0.657	0.281
18% increase	6.44	8.34	10.45	0.718	0.289
25% increase	11.10	8.80	11.10	0.764	0.294

This table reports the distribution of estimated work-hours resulting from simulating shifts at the baseline as well as across a series of increases to the earnings from each trip, assuming that demand and all other taxis behavior remains fixed. Implied elasticities are computed at the mean of hours worked. Standard errors are obtained by resampling entire driver shifts, with replacement, and re-estimating model parameters for each driver-shift type and re-simulating data in equal proportion to how each driver-shift is distributed. We conduct the exercise across 200 samples, compute the weighted average elasticity across shift types, and report standard deviations in parentheses.

FIGURE 7. Model Fit Comparison

(A) Simple Model (B) Latent Heterogeneity



Simulation Steps. To simulate data we adopt the following procedure:

(i) Uniformly draw a demand quintile k. We assume drivers observe the day's demand type and have expectations that are consistent with the empirical state transition matrix for days of type k.

(ii) Each shift begins at the initial state $(0, 0)$, with zero cumulative income and time worked. For each of 25,000 simulated shifts, we draw from the empirical distribution of *realized* sequences of trips. This begins by selecting a random shift s from a driver-day of type i and demand quintile k. We take the first trip of s and simulate spells as they occurred in the data. Denote the simulated shift by $\hat{s}$.[30]

(iii) At the end of each spell, drivers weigh the opportunity to quit and receive an immediate payoff against the option value of continuing to work longer and accrue additional earnings and time costs.

(iv) If our simulated shift $\hat{s}$ exceeds the set of trips observed in s, we append draws from another randomly selected shift s' given i and k by additionally matching the origin of the first draw in s' with the destination of our final observed trip in s.

(v) Finally, we combine each driver-day-type simulation together in proportion to their appearance rates in the data to assemble the simulated data set.

A.8. **Existence and Uniqueness of the Optimal Stopping Rule.**

[30]Drawing from the realized distribution of trips ensures our simulated spells will follow the wage process outlined in Section 2.1.

Proof. Fix i and j, and therefore we omit the subscripts i and j for notation simplicity. We begin by writing out the Bellman equation representing the driver's optimal stopping problem:

$$V_t^*(r_t, h_t, \ell_t, \epsilon_t) = \max\left\{u(r_t, h_t) + \epsilon_{t0}, E\left(V_{t+1}^*(r_{t+1}, h_{t+1}, \ell_{t+1}, \epsilon_{t+1})\middle| r_t, h_t, \ell_t, \epsilon_t\right) + \epsilon_{t1}\right\}$$

By Assumptions (i),

$$E\left(V_{t+1}^*(r_{t+1}, h_{t+1}, \ell_t, \epsilon_{t+1})\middle| r_t, h_t, \ell_t, \epsilon_t\right) = E\left(V_{t+1}^*(r_{t+1}, h_{t+1}, \ell_{t+1}, \epsilon_{t+1})\middle| r_t, h_t, \ell_t\right).$$

Then $V_t^*(r_t, h_t, \ell_t, \cdot)$ is a monotone increasing function, and the optimal stopping rule can be described as follows: Let $y_t = 1$ denotes the "quit" decision at time t, and $\delta_t = \epsilon_{t1} - \epsilon_{t0}$ as the difference of choice-specific errors. Then the driver's optimal stopping rule is given by the following threshold-crossing strategy:

$$y_t = 1\left\{\delta_t \geq \delta_t^*(r_t, h_t, \ell_t)\right\},$$

where $\delta_t^*(r_t, h_t, \ell_t)$ is obtained from

$$u(r_t, h_t) + \delta_t^* = E\left(V_{t+1}(r_{t+1}, h_{t+1}, \ell_{t+1}, \epsilon_{t+1})\middle| r_t, h_t, \ell_t\right).$$

In the above derivation, the independence assumption between ϵ_{t+1} and ϵ_t can be relaxed to a high-level condition allowing for correlation, as in e.g. Wan and Xu (2014).

Consider the following numerical solution by iteration: let $\delta_t^{(0)}(\cdot) = -\infty$ be the threshold of the stopping rule. In this case, the driver always chooses to quit at each decision period. It follows that

$$V_t^{(0)}(r_t, h_t, \ell_t, \epsilon_t) = u(r_t, h_t) + \epsilon_{t0}.$$

Now update the optimal decision and value function as follows:

$$\delta_t^{(1)}(r_t, h_t, \ell_t) = -u(r_t, h_t) + E\left[u(r_{t+1}, h_{t+1})\middle| r_t, h_t, \ell_t\right],$$

and

$$V_t^{(1)}(r_t, h_t, \ell_t, \epsilon_t) = \max\left\{u(r_t, h_t) + \epsilon_{t0}, E\left[u(r_{t+1}, h_{t+1})\middle| r_t, h_t, \ell_t\right] + \epsilon_{t1}\right\}.$$

Clearly, $\delta_t^{(1)} \geq \delta_t^{(0)}$ and $V_t^{(1)} \geq V_t^{(0)}$. Moreover, $\delta_t^{(2)}$ is obtained by solving $\bar{\delta}_t$ from the following equation

$$u(r_t, h_t) + \bar{\delta}_t = E\left[V_{t+1}^{(1)}(r_{t+1}, h_{t+1}, \ell_{t+1}, \epsilon_{t+1})\middle| r_t, h_t, \ell_t\right]$$

and then

$$V_t^{(2)}(r_t, h_t, \ell_t, \epsilon_t) = \max\left\{u(r_t, h_t) + \epsilon_{t0}, E\left[V_{t+1}^{(1)}(r_{t+1}, h_{t+1}, \ell_{t+1}, \epsilon_{t+1})\middle| r_t, h_t, \ell_t\right] + \epsilon_{t1}\right\}.$$

We now show by induction that $\delta_t^{(s+1)}(\cdot) \geq \delta_t^{(s)}(\cdot)$ and $V_t^{(s+1)}(\cdot) \geq V_t^{(s)}(\cdot)$: Because

$$V_t^{(s)}(\cdot) \geq V_t^{(s-1)}(\cdot),$$

it follows that

$$V_t^{(s+1)}(\cdot) \geq V_t^{(s)}(\cdot).$$

Thus,

$$E\left(V_{t+1}^{(s+1)}(r_{t+1}, h_{t+1}, \ell_{t+1}, \epsilon_{t+1}) \middle| r_t, h_t, \ell_t, \epsilon_t\right) \geq E\left(V_{t+1}^{(s)}(r_{t+1}, h_{t+1}, \ell_{t+1}, \epsilon_{t+1}) \middle| r_t, h_t, \ell_t, \epsilon_t\right)$$

holds almost surely. Therefore, $\delta_t^{(s+1)}(\cdot) \geq \delta_t^{(s)}(\cdot)$. Moreover, by Assumption (iii), $\delta_t^{(s)}$ is bounded above, i.e.,

$$\delta_t^{(s)} \leq -u(r_t, h_t) + \sup_{r,h \in Supp(r,h)} u(r, h).$$

Then Monotone Convergence Theorem implies the existence of an optimal stopping rule δ_t^* and its corresponding value function V_t^*.

Next, we show the uniqueness of this solution by contradiction. Let $\tilde{V}_t$ be the value function of another equilibrium. By induction, $V_t^{(s)}(\cdot) \leq \tilde{V}_t(\cdot)$ and $\delta_t^{(s)}(\cdot) \leq \tilde{\delta}_t(\cdot)$ for all $s \geq 1$. Therefore, $V_t^{\infty}(\cdot) \leq \tilde{V}_t(\cdot)$. Let $\Delta \equiv \sup_{r,h,e} \tilde{V}_t(r, h, \ell, e) - V_t^{\infty}(r, h, \ell, e) > 0$. Note that for any (r, h, ℓ) in the support, $\tilde{V}_t(r, h, \ell, e) - V_t^{\infty}(r, h, \ell, e) = 0$ for all $e : e_0 - e_1 \geq \tilde{\delta}_t(r, h, \ell)$. Then for an arbitrary $(r_t, h_t, \ell_t, \epsilon_t)$:

$$\begin{aligned}
& \tilde{V}_t(r_t, h_t, \epsilon_t) - V_t^{\infty}(r_t, h_t, \epsilon_t) \\
= \quad & \max\left\{u(r_t, h_t) + \epsilon_{t0}, E\left(\tilde{V}_{t+1}(r_{t+1}, h_{t+1}, \ell_{t+1}, \epsilon_{t+1}) \middle| r_t, h_t, \ell_t, \epsilon_t\right) + \epsilon_{t1}\right\} \\
- \quad & \max\left\{u(r_t, h_t) + \epsilon_{t0}, E\left(V_{t+1}^{\infty}(r_{t+1}, h_{t+1}, \ell_{t+1}, \epsilon_{t+1}) \middle| r_t, h_t, \ell_t, \epsilon_t\right) + \epsilon_{t1}\right\} \\
\leq \quad & \max\left\{u(r_t, h_t) + \epsilon_{t0}, \Delta + E\left(V_{t+1}^{\infty}(r_{t+1}, h_{t+1}, \ell_{t+1}, \epsilon_{t+1}) \middle| r_t, h_t, \ell_t, \epsilon_t\right) + \epsilon_{t1}\right\} \\
- \quad & \max\left\{u(r_t, h_t) + \epsilon_{t0}, E\left(V_{t+1}^{\infty}(r_{t+1}, h_{t+1}, \ell_{t+1}, \epsilon_{t+1}) \middle| r_t, h_t, \ell_t, \epsilon_t\right) + \epsilon_{t1}\right\} \\
< \quad & \Delta.
\end{aligned}$$

Therefore,

$$\sup_{\epsilon_t} \tilde{V}_t(r_t, h_t, \ell_t, \epsilon_t) - V_t^{\infty}(r_t, h_t, \ell_t, \epsilon_t) < \Delta.$$

By Assumptions (ii) and (iii),

$$\sup_{r_t, h_t, \ell_t, \epsilon_t} \tilde{V}_t(r_t, h_t, \ell_t, \epsilon_t) - V_t^{\infty}(r_t, h_t, \ell_t, \epsilon_t) < \Delta,$$

which is a contradiction. □