Daily Labor Supply and Adaptive Reference Points

Neil Thakral
Brown University

Linh T. Tô
Harvard University

October 2018*

Abstract

This paper provides field evidence on how reference points adjust, a degree of freedom in reference-dependence models. To examine this in the context of cabdrivers' daily labor-supply behavior, we ask how the within-day timing of earnings affects decisions. Drivers work less in response to higher accumulated income, with a strong effect for recent earnings that gradually diminishes for earlier earnings. We estimate a structural model in which drivers work towards a reference point that adjusts to deviations from expected earnings with a lag. This dynamic view of reference dependence reconciles the "neoclassical" and "behavioral" theories of daily labor supply.

*Thakral: Department of Economics, Brown University, Box B, Providence, RI 02912 (email: neil_thakral@brown.edu). Tô: Department of Economics, Harvard University, Littauer Center, 1805 Cambridge Street, Cambridge, MA 02138 (email: linhto@fas.harvard.edu). We are grateful to Isaiah Andrews, Pedro Bordalo, Stefano DellaVigna, Henry Farber, Josh Feng, Benjamin Golub, Jerry Green, Maximilian Kasy, Lawrence Katz, Marc Kaufmann, David Laibson, Ted O'Donoghue, Matthew Rabin, Joshua Schwartzstein, Jesse Shapiro, Andrei Shleifer, and Elie Tamer for advice and suggestions.

In the classical economic model of labor supply, individuals choose hours of work to trade off the utility of additional income against the disutility of additional effort. Based on an analysis of daily decisions about work hours among New York City (NYC) taxi drivers, Camerer et al. (1997) propose the alternative hypothesis that drivers quit working upon reaching a target level of earnings. An ongoing debate since then focuses on the question of whether workers exhibit such reference-dependent behavior with respect to daily earnings.

The broader question of what determines the reference point poses a challenge for evaluating the importance of reference dependence in any given setting. The highly influential work on prospect theory by Kahneman and Tversky (1979) describes the implications of the existence of a reference point but leaves the reference point itself largely unspecified. In an attempt to discipline the theory, Kőszegi and Rabin (2006) endogenize the reference point through assuming that it coincides with expectations. Even under this perspective, there remains an implicit degree of freedom in the theory—the speed of adjustment of the reference point—which can substantially affect its empirical predictions.

A convincing answer to the first question, about the relevance of reference-dependent preferences for daily labor supply, requires a thorough investigation of the second question, namely what determines the reference point. We propose to address both of these questions by using comprehensive trip-level data on all NYC cab fares in 2013 to identify the timing of reference-point effects. To interpret our results, we offer a conceptual framework which emphasizes the role of reference-point adaptation. By characterizing the dynamics of reference points in the context of daily labor-supply decisions, this paper provides field evidence on a degree of freedom in one of the central models of behavioral economics (DellaVigna, 2009). The framework also helps to organize, explain, and reconcile conflicting interpretations of previous evidence.

The earliest work on reference dependence and labor supply focuses on estimating daily wage elasticities on the intensive margin for cabdrivers, uncovering a negative relationship between average wages and the number of hours worked each day (Camerer et al., 1997; Chou, 2002). As Camerer et al. (1997) acknowledge, estimating a daily wage elasticity requires that wages vary across days but remain relatively constant within days, which they do not find evidence against. Since the data consist of many drivers on a small set of days (about 1,200 drivers across all three of their datasets, with an average of 1.5 shifts for each driver), the authors instrument for a driver's average wage on a given day with summary statistics of the distribution of other drivers' wages on the same day, in response to potential concerns about division bias.[1] To explain their puzzling finding of a backward-bending labor-supply curve, Camerer et al. (1997) argue that a cabdriver's marginal utility of income must drop sharply around the level of average daily income due to loss aversion, resulting in a probability of quitting for the day that rises substantially when a driver gets near their target.[2] Farber (2005) reiterates the division-bias issue and expresses two additional concerns with their analysis. First, he disputes the premise of relatively constant within-day wages. Second, he points out that the instrumental-variables approach does not purge the elasticity estimates of day-specific factors that affect both wages and aggregate labor supply.

Motivated by econometric problems with estimating daily wage elasticities, subsequent work uses a hazard specification to examine directly whether stopping decisions respond to accumulated daily earnings, as a test of reference dependence. Farber (2005) introduces a different dataset, consisting of only 21 drivers but over many days (with an average of 28 shifts for each driver), which allows for the estimation of a stopping model with driver fixed effects. With these data, Farber cannot reject the null hypothesis of the neoclassical model despite finding a positive association between accumulated daily earnings and the probability of ending a shift conditional on hours worked, qualitatively consistent with reference dependence.[3]

---

[1]That is, average hourly wages obtain from dividing daily income by hours, so measurement error in hours can lead to a spurious negative relationship between wages and hours.

[2]Chou (2002) lacks data on hourly wages but points out that with such data, "income targeting may be tested more rigorously. . . utilizing a hazard specification" in which "the probability that a driver quits for the day at any point in time may be parameterized as a function of the cumulated income and the expected marginal wage. . . Short-horizon targeting predicts that quitting is related to cumulative same-day income. . . "

[3]Farber (2008) uses the same data to estimate a structural version of the stopping model, which consists a latent underlying distribution of daily income targets and accommodates a threshold effect of exceeding the income target, and concludes that the variation in driver-day targets leaves

The availability of large-scale administrative data since then provides an opportunity to settle these unresolved issues. Farber (2015) uses a sample of 13 percent of all NYC cabdrivers between 2009 and 2013 to revisit the earlier studies. He presents a model of reference dependence with a fixed daily income target but does not find any support in the data for its prediction of a wage elasticity of -1. In addition, he finds an important role for hours worked in the decision to end a shift, as the neoclassical model suggests.

However, applying the previous approaches to the comprehensive new dataset leaves a puzzle, as some traces of income-targeting behavior emerge. In particular, Farber (2015) finds that negative wage elasticities appear for one-third of day-shift drivers and one-seventh of night-shift drivers, and accumulated daily income has a small but statistically significant influence on the decision to quit working during day shifts.[4] Existing work implicitly tends to offer a binary characterization of behavior, with a negative elasticity corresponding to daily income targeting and a positive elasticity corresponding to the neoclassical model, or a positive marginal effect of earnings on stopping corresponding to daily income targeting and a null effect corresponding to the neoclassical model. The mixed evidence for the presence of reference dependence, albeit modest, suggests the need to go beyond a dichotomous view of behavior and suggests a possible misspecification of the reference point. Developing an understanding of what constitutes the reference point requires a new approach.

We pose a new question which aims to shed light on reference-point formation by uncovering when reference-point effects emerge: Does the effect of additional accumulated earnings depend on recency? Using data from all 40,000 NYC cabdrivers

the model with little predictive value despite finding a threshold effect. Crawford and Meng (2011) use these data to estimate a structural stopping model that allows for reference dependence in both daily income and hours, where drivers' expectations of daily income and hours based on previous shifts determine the driver-day targets following the ideas from Kőszegi and Rabin (2006), and conclude that the data support this model.

[4]Morgul and Ozbay (2014) use the full set of over 30,000 drivers in four separate months of 2013 to revisit the earlier studies as well, finding a negative wage elasticity for the month of January as well as a positive relationship between daily earnings and stopping conditional on hours during all four months (though not disaggregated by day and night shifts).

in 2013, we find strong income effects for recent earnings—despite the fact that higher recent earnings predict better opportunities from continuing to work—and behavior that appears neoclassical in response to earnings accumulated earlier in the day. We also document using simulations that previous methods fail to deliver consistent estimates of the marginal effect of earnings on the decision to stop working for the day, and we propose an alternative. Overall, for a driver who finishes a trip after 8.5 hours of work, a 10 percent increase in accumulated daily earnings corresponds to a 3 percent increase in the probability that he stops working for the day. The effect size changes to 10 percent if the additional earnings come in the most recent hour and gradually declines for earnings accumulated earlier. Earnings from the first four hours have little or no effect on the decision of whether to end a shift at 8.5 hours.

Neither the neoclassical view nor the static income-targeting view can account for our findings. The neoclassical model does not predict a daily income effect, and a daily-level income target does not make any distinction based on the recency of income. The Kőszegi and Rabin (2006, 2009) model of expectations-based reference dependence provides a potential explanation for the recency effect. The model improves upon the classical outcome-based formulation of utility by adding a role for expectations. Under this framework, drivers update their expectations about earnings on a given day as they work and acquire new information, and the lagged expectation determines the reference point. Due to the lag with which the reference point adjusts, additional recent earnings increase the probability that a driver stops working. Adopting a discrete view of how the reference point adjusts would produce a stark contrast between the most recently accumulated earnings, which the reference point does not incorporate, and any earlier earnings, which the reference point does incorporate. The data instead show a gradual decline in the influence of less recent earnings on stopping decisions.

To explain the results, we propose a model of reference-dependent preferences with an adaptive reference point. The model extends the framework of Kőszegi and Rabin (2009) by allowing for a gradual adjustment of the reference point. This formalization provides a way of capturing the following intuition which conveys our main findings: people overreact to surprises, as they work less in response to higher accumulated earnings, but surprises wear out over time, so that quitting depends to a greater extent on more recent earnings. Our formulation nests the neoclassical model as well as static income targeting. At one extreme lies a reference point that adjusts instantaneously, which produces behavior that coincides with the neoclassical prediction, and at the other extreme lies a fixed reference point. While neither extreme case permits stronger reactions to more recent experiences, the intermediate case does. To illustrate, a gain in earnings causes the reference point to drift upward over time, so an earlier gain is less likely than a more recent gain to make a driver exceed his income target. In general, the slow-adjusting reference point exhibits some degree of persistence or stickiness, with decreasing weights on lagged values of the reference point. This backward-looking feature causes the updated reference point to incorporate earlier earnings to a greater extent than more recent earnings and thus to account for the gradually diminishing effect of earnings on quitting.

To quantify the speed of adaptation of the reference point, we estimate a structural model of daily labor supply. The model, based on Crawford and Meng (2011), enables us to examine alternative specifications of the reference point in more detail. Crawford and Meng (2011) provide evidence that drivers' expectations of daily earnings as determined by past outcomes serve as the reference point, supporting the Kőszegi and Rabin (2006) notion that the reference point coincides with the lagged expectation. While the adaptive reference point we propose reduces to their beginning-of-day expectation of daily earnings in the special case that the reference point does not adjust within the day, maximum likelihood estimates of the model confirm that an intermediate degree of adaptation provides a better fit for the patterns in the data. As an alternative specification of the reference point that also accommodates within-day updating, we consider forward-looking reference points based on a one-period lag of expectations under different definitions of the lag (e.g., previous trip instead of previous day). However, we find that earlier lags remain important for explaining the patterns we observe. This paper thus contributes to a growing literature on structural behavioral economics (DellaVigna, 2018).

A dynamic view of reference dependence helps to reconcile some of the conflicting perspectives in the literature on labor supply. Empirical work on daily labor supply in a variety of settings using different methods yields mixed results.[5] We do not focus on daily wage elasticities (Camerer et al., 1997; Farber, 2015) and instead document a mechanical bias that arises when estimating them (see Appendix B). Some authors use field experiments to investigate the relationship between accumulated daily earnings and quitting decisions. Andersen et al. (2018) conclude that unexpected cash windfalls in the morning do not affect the labor supply of vendors in their Betel Nut Experiment in India, and Dupas et al. (2018) examine the effects of unexpected cash windfalls on bicycle-taxi drivers in Kenya who have daily cash needs, though neither paper demonstrates the role of variation in the timing of earnings. The field experiment on Swiss bicycle messengers by Fehr and Goette (2007) obtains a positive labor-supply elasticity, consistent with both reference dependence and a neoclassical explanation, but individual-level measures of loss aversion correlate with effort reductions, suggesting a role for reference dependence. By emphasizing reference-point adaptation, our model describes the extent to which workers exhibit neoclassical behavior through the speed of adjustment.

Our investigation of labor supply also reveals lessons about models of reference dependence. Following Kőszegi and Rabin (2006), much of the literature examines forward-looking expectations-based reference points.[6] However, existing empirical tests of the model tend to assume a particular view of what constitutes the reference point, including how quickly the reference point adapts to experimental manipulations in the context of lab studies. This paper evaluates a reference point based on a one-period lag of expectations but instead finds support for further history dependence in modeling the reference point. Post et al. (2008) on risky choice in a game show also illustrates the role of reference-point adjustment and path dependence in prospect theory, though with a much more limited specification. Some recent lab evidence

[5]Papers finding negative wage elasticities include Ashenfelter et al. (2010), Doran (2014), and Schmidt (2017) on taxi drivers in NYC; Chang and Gross (2014) on pear packers in California; Agarwal et al. (2015) on taxi drivers in Singapore; and Nguyen and Leung (2015) on swordfish fishermen in Hawaii. Papers finding positive wage elasticities include Jonason and Wållgren (2013) on taxi drivers in Stockholm and Stafford (2015) on lobster fishermen in Florida.

[6]See Fehr and Goette (2007), Pope and Schweitzer (2011), Card and Dahl (2011), and Bartling et al. (2015) for field evidence of expectations-based reference dependence; but also see conflicting lab evidence from Heffetz and List (2014), Gneezy et al. (2017), and Goette et al. (forthcoming).

(e.g., Gill and Prowse 2012 and Song 2016) also explicitly considers the speed of reference-point adjustment. DellaVigna et al. (2017) on job search evaluates a forward-looking reference point based on a one-period lag of expectations, but the patterns in the data demonstrate a clear need for an alternative formulation. In a financial context, Baucells et al. (2011) and Meng and Weng (2017) also provide perspectives on reference-point formation.

The paper proceeds as follows. The next section provides background information on the institutional context and describes the data. Section 2 analyzes the impact of accumulated daily earnings on labor supply and discusses some possible explanations for the income effect. Section 3 presents a model of loss aversion with adaptive reference points along with structural estimates. Section 4 concludes.

1 Data

1.1 Background

Our study uses trip-level data provided by the New York City Taxi and Limousine Commission (TLC) for every fare served by NYC medallion taxicabs in 2013. The "trip sheets" consist of detailed information about each fare, including anonymized identification numbers for the driver and car, start and end times for each trip, pick-up and drop-off locations, tips paid by credit card, and the fare charged. These data are collected and transmitted electronically in accordance with the Taxicab Passenger Enhancements Project (TPEP). Haggag and Paci (2014) and Farber (2015) provide further details about the data, with the former using data from 2009 and the latter using data from 2009–2013.

Prior to TPEP, cabdrivers were required to fill out trip sheets by hand to record and store information on paper about each fare. By 2008, all medallion taxicabs in NYC had implemented a series of technology-based service improvements (e.g., credit/debit card payment systems, passenger information monitors, and text messaging between the TLC and drivers) due to a March 2004 mandate by the TLC Board of Commissioners, which also led to automated trip sheet data collection. Relative to the earlier handwritten trip sheets, the electronically transmitted data also include Global Positioning System (GPS) coordinates for pick-up and drop-off locations, available for over 98 percent of the data.

For each trip at the standard city rate (i.e., within the city limit), the meter computes the fare by combining a base rate of \$2.50, any surcharges, and an incremental charge of \$0.50 for each unit of distance (0.2 miles at a speed of at least 12 miles per hour) or time (60 seconds when the cab is not in motion or is traveling at less than 12 miles per hour).

Appendix Figure 3 depicts the average number of cabs that are on the road working during each minute of the day. The systematic drops in the number of cabs available in the early morning and early evening reflect the common institutional arrangement whereby two drivers share the same cab (typically switching at 5 AM and 5 PM).[7] The TLC regulates the maximum amount that can be charged to lease a cab for a twelve-hour shift, with a "lease cap" of roughly \$130 depending on the day of the week and the time of the shift.[8]

In addition to institutional constraints, weather can potentially affect labor-supply decisions. Our study uses minute-level weather data (temperature, precipitation, and wind speeds) from the National Centers for Environmental Information collected at five locations around NYC. We match each trip from the TPEP data with the weather conditions at the closest station during the minute when the trip ends.

1.2 Descriptive Statistics

The raw data consist of information on about 41,000 unique drivers and 14,000 taxicabs taking around 173 million trips in 2013. To study cabdrivers' labor-supply decisions, we group trips into *shifts*. We define a shift as a sequence of consecutive

[7]For additional details, see Frechette et al. (2016) as well as the January 2011 article by Michael M. Grynbaum in the New York Times: `http://www.nytimes.com/2011/01/12/nyregion/12taxi.html`.

[8]During our sample period, the lease caps for standard vehicles were \$115 for all AM shifts, \$125 for Sunday-Tuesday PM shifts, \$130 for Wednesday PM shifts, and \$139 for Thursday-Saturday PM shifts. The lease caps for hybrid vehicles are \$3 higher. Cabs can also be leased on a weekly basis, with a lease cap that is about six-sevenths of the sum of the daily lease caps.

trips that are not more than six hours apart from each other (Haggag and Paci, 2014). In other words, we infer that a driver ends a shift after a given trip if the driver does not pick up any more passengers within the next six hours. As in Farber (2005), we define a break as a long waiting time between fares: at least 30 minutes between a fare that ends in Manhattan and a fare that starts in Manhattan; at least 60 minutes between fares that start or end outside Manhattan but do not end at an airport; or at least 90 minutes between a fare that ends at an airport and the next fare. Our analysis of daily income effects focuses on cumulative earnings, which we define at the trip level as the sum of fare earnings (excluding tips) from the beginning of the shift to the end of the current trip. After eliminating shifts with missing or inconsistent information (see Appendix A), over 5.8 million shifts by over 37,000 drivers remain, comprising over $1.5 billion in transactions for cab fares.

Appendix Table 1 provides summary statistics at the trip level and at the shift level. Over 85 percent of all trips start and end in Manhattan, and the median ride takes 10 minutes. Cabdrivers earn a median fare of about $9.50, with 90 percent of fares falling below $22. We observe tips for the 54 percent of fares paid using a credit card.

Figure 1 displays the fraction of shifts starting at each hour of the day as well as the distribution of work hours. A typical shift consists of 22 trips, with 75 percent of shifts exceeding 7.2 hours. On average, a driver spends about 64 percent of their time with a passenger in the cab, 26 percent of their time searching for the next passenger, and 13 percent of their time on break.

The market wage varies considerably throughout the day. For each minute that a driver spends searching for or riding with passengers, we define the driver's per-minute wage as the ratio of the fare they earn for that trip to the number of minutes spent working (i.e., searching and riding). We define the market wage in each minute as the average of the per-minute wages of all drivers working during that minute. Appendix Figures 1 to 3 depict the average wage and number of drivers working during each minute of the day. Cabdrivers earn an average wage of about $31 per hour, which amounts to a gross income (excluding tips) of about $280 per shift, from which drivers may pay leasing fees and gasoline costs. The highest wages occur during the two hours with the lowest number of drivers working, which correspond to the transitions between AM and PM shifts each day. Appendix B documents that the pattern of wages throughout the day can bias estimates of the daily wage elasticity.

Figure 2 investigates the predictability of hourly wages. Residualizing hourly wages on a set of time effects (an interaction between the hour of day and day of week, the week of the year, and an indicator for federal holidays) and weather effects, the figure shows a positive autocorrelation.

2 Tests of Income Effects

2.1 Stopping Model

We start by examining the marginal effect of accumulated fare earnings on the probability of ending a shift. We model the decision of a driver at the end of each trip to stop working or to continue working (Farber, 2005). After completing t trips and accumulating y_{int} in fares after a total of h_{int} hours, driver i decides to end shift n when the cost of additional effort exceeds the continuation value. Letting d_{int} indicate the decision to stop working, we express the probability that driver i ends shift t at trip n as

$$\Pr(d_{int} = 1) = f(h_{int}) + \gamma(h_{int})y_{int} + X_{int}\beta(h_{int}) + \mu_i(h_{int}) + \epsilon_{int}, \qquad \text{(TT)}$$

where $f(\cdot)$ represents the baseline hazard; X consists of controls that can potentially be related to variation in earnings opportunities from continuing to work, such as location, time, and weather; and μ absorbs differences in drivers' baseline stopping tendencies.[9] This discrete-choice problem represents a reduced form of a forward-looking dynamic optimization model based on hours worked so far on the shift, expectations about future earnings possibilities, and other variables that could affect

[9]Equation (TT) has the form of a non-parametric additive hazards model (Aalen, 1989). See Martinussen and Scheike (2006) for a textbook treatment.

preferences for work.[10] Equation (TT) allows for a flexible and driver-specific hazard of stopping as well as a time-dependent relationship between each of the covariates and the stopping probability.

The term $\gamma(h)$ represents the effect of an additional dollar of accumulated daily earnings on the probability of ending a shift for a driver who finishes a trip after h hours of work. A positive effect of accumulated earnings on quitting suggests the presence of a daily income effect under the assumption that cumulative daily earnings are uncorrelated with unobserved determinants of the value of stopping (such as effort or fatigue) or the value of continuing (such as future earnings opportunities) conditional on the full set of time-varying covariates, which Section 2.4 discusses in more detail.

We use local linear regression techniques to estimate the baseline hazard and the time-varying coefficients in Equation (TT). For any given time h, the associated parameter estimates solve a separate weighted least squares problem

$$\min_{\alpha,\beta,\gamma,\mu} \sum_{i,n,t} w(h - h_{int})(d_{int} - (\alpha h_{int} + \gamma y_{int} + X_{int}\beta + \mu_i))^2.$$

with weights $w(\cdot)$ (Cleveland and Devlin, 1988). Using uniform weights, the coefficients at any time h represent the fit of a linear model to a localized subset of the data. The results we report in Section 2.2 use uniform weights over a 10-minute window of time during the shift.[11]

[10]Also see Farber (2008) and Farber (2015). Buchholz et al. (2018) provides a method for estimating a dynamic optimal stopping model, applied to the labor-supply decisions of cabdrivers. However, for convex disutility of effort, this trip-by-trip stopping rule is consistent with maximizing the static objective function as long as the wage rate $y_n'(h_n)$ does not increase too rapidly. The pattern in Appendix Figure 2 suggests that a stopping model that does not explicitly incorporate option value may not apply to PM shifts on weekends, a point that we revisit in Section 2.2 when discussing the empirical results.

[11]We find that varying the window (e.g., to 5 or 30 minutes) or using a local quadratic fit results in similar estimates for $\gamma(h)$.

For comparison, we also replicate the specifications from previous work:

$$\Pr(d_{int}=1)=\alpha h_{int}+\gamma y_{int}+X_{int}\beta+\mu_i+\epsilon_{int} \tag{F-1}$$

$$\Pr(d_{int}=1)=\sum_j \alpha_j \mathbf{1}_{\{h_{int}\in H_j\}}+\sum_k \gamma_k \mathbf{1}_{\{y_{int}\in Y_k\}}+X_{int}\beta+\mu_i+\epsilon_{int} \tag{F-2}$$

$$\Pr(d_{int}=1)=\sum_{j,k} \delta_{jk} \mathbf{1}_{\{h_{int}\in H_j\}}\mathbf{1}_{\{y_{int}\in Y_k\}}+X_{int}\beta+\mu_i+\epsilon_{int}, \tag{F-3}$$

where H_j and Y_k form a partition of hours and income, respectively. Farber (2005) estimates Equations (F-1) and (F-2), and Farber (2015) estimates Equations (F-2) and (F-3).[12] While Equations (F-1) to (F-3) impose that for any pair of drivers one of them has a uniformly higher or lower predicted probability of stopping at the end of any given trip conditional on the other covariates, Equation (TT) accommodates a driver-specific relationship between hours and the probability of stopping. Similarly, Equations (F-1) to (F-3) may suggest that drivers are more likely to stop at 4 PM, when it rains, or when a trip ends near the taxi garage regardless of how many hours they have worked, whereas Equation (TT) allows the marginal effect of each variable on the probability of stopping to vary flexibly throughout the shift.

Appendix C.2 conducts an empirical Monte Carlo exercise which validates the approach in Equation (TT) and demonstrates that the functional-form assumptions in Equations (F-1) to (F-3) can lead to spurious results. For example, even under neoclassical behavior such as quitting after reaching a target number of hours irrespective of income, estimates of γ in Equations (F-1) to (F-3) can be statistically significant, either positive or negative. Intuitively, due to the positive correlation between accumulated income and hours of work, a misspecified functional form for the relationship between hours and the stopping probability can cause the model to incorrectly attribute part of the effect of hours to earnings.

[12] As in Farber (2015), we take H to partition the shift at 3, 6, 7, 8, 9, 10, 11, 12, and 13 hours and Y at 100, 150, 200, 225, 250, 275, 300, 350, and 400 dollars.

2.2 Estimates of the Stopping Model

Table 1 presents in Panel A our estimates of the elasticity of stopping at 8.5 hours (approximately the median stopping time) with respect to accumulated fare earnings. The strategy for estimating the quitting response to additional accumulated earnings follows Farber (2005) in using variation in earnings conditional on an extensive set of covariates that capture the value of stopping (hours worked so far on the shift) and the value of continuing (expectations about future earnings possibilities). Appendix D.1 provides more detail on variation in earnings, and Section 2.4 discusses a supplementary analysis that uses speeds to instrument for earnings.

The first column reports estimates from our preferred specification, based on Equation (TT). Each row corresponds to a more comprehensive set of controls than the previous one. All specifications consist of controls for minutes spent working, including indicators for whether the driver spends time with passengers in each hour. The specification in the first row, with no additional controls, shows a small but statistically significant relationship between cumulative daily earnings and stopping probabilities. If drivers with higher average earnings tend to work more, then using across-driver variation likely underestimates the relationship between accumulated earnings and quitting. The second row shows that the inclusion of driver fixed effects strengthens the estimated effect considerably. Relatedly, to the extent that drivers work more on days with higher expected wages, failing to control for time effects may also understate the magnitude of the income effect. Adding an interaction between clock hour and day of week as well as indicators for day of year in the third row indeed results in a larger estimate. Since drivers may end their shifts with higher probability when a trip ends in a convenient location coinciding with higher accumulated earnings (e.g., near the driver's home or the cab garage in one of the outer boroughs), the fourth row adds fixed effects for the 195 Neighborhood Tabulation Areas (NTA) in NYC (see Haggag et al., 2017), which decreases the estimated effect to a 3.1 percent increase in the probability of ending a shift at 8.5 hours (0.44 percentage-point increase relative to a baseline stopping probability of 13.2 percent) in response to a 10 percent ($26) increase in cumulative earnings. The elasticity estimate remains stable around 0.3

after adding a set of weather controls (indicators for precipitation, temperature above 80 degrees Fahrenheit, temperature below 30 degrees Fahrenheit, and wind speed on the Beaufort scale), measured in the minute when a trip ends, in the last row.

The remaining columns in Panel A report estimates based on Equations (F-1) to (F-3).[13] While the control variables tend to influence the estimates in the directions discussed above, the magnitudes differ across specifications. Since Farber (2015) does not use location or weather controls in estimating the stopping model, the third row of the third column corresponds to the primary specification that Farber (2015) uses for counterfactual analysis and reports an effect that exceeds our preferred estimate by over 40 percent. The less constrained specification in Farber (2015) corresponds to the third row of the fourth column, which reports a smaller but imprecisely estimated effect that does not significantly differ from our preferred estimate. Under the full set of controls, the point estimates across specifications suggest that the probability of ending a shift at 8.5 hours increases by between 1.2 and 3.8 percent in response to a 10 percent increase in cumulative earnings.

Panel B provides a direct comparison with previous papers. Farber (2005) reports a statistically insignificant effect of earnings on quitting (reproduced in Panel B), though the point estimate implies that a 10 percent increase in cumulative earnings corresponds to a 1.2 percent increase in the probability of ending a shift.[14] The confidence interval also encompasses the estimates from all specifications with the full set of controls in Panel A, including our preferred estimate. Farber (2015), using a specification analogous to Equation (F-2), reports a separate estimate for day shifts (start between 4 AM and 10 AM) and night shifts (start between 2 PM and 8 PM) and finds sizable income effects only for day shifts (9.5 percent increase in the probability of stopping at 8.5 hours in response to 10 percent higher cumulative earnings).[15]

[13]Since Equations (F-2) and (F-3) coarsen income into bins of at least $25, we compute the effect of a 10 percent increase in income on the probability of stopping after earning $260 in 8.5 hours and scale this to obtain the elasticity.

[14]Table 5 in Farber (2005) reports that an additional $100 increases the probability of ending a shift by 0.011 percentage points at 8.5 hours under the full set of controls. With a mean income of $161.33, an additional 10 percent in earnings corresponds to a $16.13 \cdot 0.011 \approx 0.18$ percentage-point increase in the probability of stopping relative to a baseline of 14.67 percent.

[15]See Table VII of Farber (2015).

Appendix Table 5 replicates Table 1 for day shifts and night shifts and confirms this pattern for the Farber (2015) specifications (columns 3 and 4), finds the opposite pattern for the main Farber (2005) specification (column 2), and shows that our preferred approach (column 1) instead yields very similar estimates for day shifts and night shifts. Overall, Equation (TT) provides evidence of a modest-sized daily income effect, comparable with but more consistent than previous estimates.

Figure 3 shows that the magnitude of the income effect from Table 1, evaluated at 8.5 hours of work, persists throughout the shift using Equation (TT) with the full set of controls. The figure plots the probability of stopping (left axis) and the percentage-point change in the stopping probability in response to a 10 percent increase in earnings (right axis) every fifteen minutes over a six-hour period, roughly corresponding to the 10$^{\text{th}}$ and 90$^{\text{th}}$ percentile of the distribution of stopping times. A clear, stark relationship appears between hours of work and the probability of ending a shift, consistent with the prediction of the neoclassical model as Farber (2015) highlights. In addition, the elasticity of stopping with respect to earnings remains significant and consistent throughout the shift as the average stopping probability increases from 2 percent to 28 percent. The figure shows an elasticity of about one-third during the hours leading up to the median stopping time and about one-fourth in the hours that follow.

Appendix Figure 8 plots the percent change in the probability of stopping estimated on four separate groups of shifts: day-weekday shifts, night-weekday shifts, day-weekend shifts (day shifts on Saturday and Sunday), and night-weekend shifts (night shifts on Friday and Saturday). While day shifts and night-weekday shifts show evidence of income effects consistent with our estimates from the full sample, night-weekend shifts stand out with the opposite result during the hours when drivers are most likely to end their shifts. As Appendix C.1 points out, the trip-by-trip interpretation of the stopping model relies on the assumption that the option value of continuing to drive is sufficiently small (or that drivers ignore option value), and the pattern in Appendix Figure 2 suggests that this assumption may not be reasonable for night-weekend shifts, when wages rise substantially and predictably over time. This observation explains a potential discrepancy with the results in Farber (2015) that hours of work are roughly unaffected by income during night shifts.

2.3 The Role of Timing

We proceed to test whether drivers exhibit stronger responses to more recent experiences. Relaxing the implicit assumption that money is fungible within the day, we augment Equation (TT) to express the probability of stopping as

$$\Pr(d_{int} = 1) = f_j(h_{int}) + \sum_{\ell} \gamma^{\ell}(h_{int}) y_{int}^{\ell} + X_{int}\beta(h_{int}) + \mu_i(h_{int}) + \epsilon_{int}, \qquad (1)$$

where y_{int}^{ℓ} denotes fare earnings accumulated in hour ℓ of the shift. If drivers compare their cumulative daily earnings with a fixed target, then we would expect to find that the impact of an additional dollar on the probability of ending a shift does not depend on when the dollar was accumulated (i.e., γ^{ℓ} is independent of ℓ).

Figure 4 plots the estimated percent change in the probability of ending a shift at 8.5 hours in response to an additional 10 percent ($26) in income earned at various times in the shift.[16] While Table 1 shows that a 10 percent increase in cumulative earnings corresponds to a 3 percent increase in the probability of ending a shift at 8.5 hours, earnings from the first four hours of the shift have little or no effect on the decision to stop after 8.5 hours. If the additional earnings arrive in the eighth hour of the shift, then our estimates imply a 10.2 percent increase in the probability of stopping (1.35 percentage-point increase relative to a baseline stopping probability of 13.2 percent). The effect gradually diminishes for earnings accumulated earlier. An additional dollar accumulated in the eighth hour increases the probability of stopping by an order of magnitude more than an additional dollar accumulated four hours earlier. For comparison, an additional 10 minutes of work (median trip duration) increases the probability of ending a shift by over 7 percent. The positive autocorrelation in earnings from Figure 2 suggests that higher recent earnings should, if anything, be associated with a higher value of continuing to work. However, the probability of stopping depends on the path of earnings in the opposite direction of

[16] Appendix Table 8 presents analogous results based on extensions of Equations (F-1) to (F-3).

this prediction.

Figure 5 shows that the recency pattern from Figure 4, evaluated at 8.5 hours of work, persists throughout the shift. The columns of the figure correspond to different times during the shift between 5.5 and 11 hours. Each column depicts the effect that an additional 10 percent of cumulative income, accumulated in different hours corresponding to the rows, has on the probability of stopping. Appendix Figure 6 shows that the same pattern holds for shifts that start at different hours. Appendix Table 6 shows similar qualitative patterns for groups of drivers with lower or higher variability of daily hours, with stronger effects for the latter group. Beyond the result that stopping decisions do not respond to earnings accumulated early in a shift, Appendix Table 7 (column 1) shows no significant relationship between the probability of ending a shift and earnings on the previous day.

2.4 Alternative Explanations

We interpret the increase in the probability of ending a shift in response to higher cumulative fare earnings as evidence of a daily income effect exhibited by a reduction in labor supply. This section addresses potential challenges to our modeling assumptions. We consider the possibility that cumulative earnings are correlated with unobserved determinants of the stopping decision such as effort or fatigue, or that cumulative earnings convey information about future earnings opportunities. In addition, we assess whether the relationship between cumulative earnings and stopping arises due to other factors such as option value or inexperience. Appendix D elaborates on these considerations and also examines liquidity constraints and the measurement of work hours.

Effort A potential concern with our interpretation would arise if additional earnings coincide with an increase in the intensity or difficulty of work. We suggest three ways of addressing this: by constructing a proxy for effort, by applying an instrumental-variable (IV) strategy, and by analyzing income effects and recency effects at different hours of the shift.

First, although our dataset does not contain a direct measure of effort, we use as a proxy how quickly a driver finds the next passenger. Drivers likely earn more money during shifts that have a higher fraction of the time riding with passengers. Despite this mechanical effect, the data show only a weak relationship between daily earnings and the fraction of working time spent searching for passengers (correlation of -0.10). This weak correlation may arise because of drivers reducing their effort (e.g., taking quick breaks) in response to increased demand, as Appendix D.6 discusses in more detail. Moreover, survey evidence suggests that driving passengers to a specific destination requires less effort than driving while searching for potential passengers (Camerer et al., 1997). This further suggests that drivers exert less effort on high earning days due to spending relatively less time searching for passengers. We thus find little scope for the reduction in hours to represent an increase in the intensity of effort rather than a response to additional accumulated earnings.

Second, we use an IV strategy to address the possible correlation between cumulative fare earnings and unobserved determinants of the decision to end a shift such as effort or fatigue. The analysis in Appendix D.1 instruments for earnings based on speed and restricts to trips that stay within Manhattan, where variation in earnings plausibly arises due to traffic conditions unrelated to the drivers' decisions to exert additional effort. If anything, we might expect drivers to exert *less* effort when facing smooth-flowing traffic and thus to be less likely to quit. Despite this, the estimates in Appendix Table 9 still show quitting responses to earnings in the past several hours and not earlier in the shift.

Third, Figures 3 and 5 display consistent and sizable income effects and recency effects *throughout* the shift, which poses a difficulty for fatigue-based explanations. If fatigue poses a confound for estimating the effect of cumulative earnings, we would expect much larger magnitudes of the income effect in the later hours of a shift (e.g., after working 10 hours compared to 8.5 hours) insofar as drivers face an increasing marginal disutility of effort. Even in the earlier hours of a shift, an additional 10 percent in cumulative earnings corresponds to an increase in stopping probability of at least 2.5 percent and a significantly stronger response to recent earnings. Also note that Equation (TT) accommodates a driver-specific relationship between work hours and the probability of stopping as well as a flexible relationship between the effect of work hours on the probability of stopping, which mitigates the scope for the estimated effect of additional earnings to reflect a response to fatigue.

Learning about Future Earnings Another potential concern with interpreting the effect of earnings on stopping behavior would arise if accumulated earnings convey additional information about future opportunities, either within the same shift or across shifts.

If higher cumulative earnings or higher recent earnings indicate lower expected earnings from continuing conditional on all the covariates, then the estimated relationship between earnings and quitting would overstate the income effect. The pattern in Figure 2, however, suggests the opposite.

Likewise if higher earnings correlate with plentiful opportunities on the next day, then drivers may engage in intertemporal substitution, quitting during times of high earnings to conserve energy for the next shift. The evidence in Appendix D.2 suggests a limited role for this channel, as earnings do not appear predictive of market conditions on subsequent days.

Option Value A driver who explicitly solves the dynamic optimization problem may exhibit a low probability of ending a shift in response to low cumulative daily earnings if he has information that wages will rise later in the shift. Not fully incorporating option value in the model could potentially pose concerns if the rate of increase in the wage exceeds that of the monetary equivalent of the disutility of effort. Column (2) of Appendix Table 7 restricts our analysis to trips on Friday and Saturday after 5 PM, when the typical wage profile is nonincreasing, and exhibits a significant relationship between recent earnings and quitting.

Experience Based on findings in related settings, one may hypothesize that the positive relationship between earnings and stopping reflects a failure to optimize by inexperienced drivers. Camerer et al. (1997) present evidence that more experienced drivers exhibit more positive wage elasticities of labor supply, which Farber (2015) corroborates. Recent work by Haggag et al. (2017) documents using the TPEP data from 2009 that productivity differences between new and experienced drivers vanish after 17 to 62 shifts (depending on the difficulty of the situations). Given that performance improves quickly with experience, drivers might also learn to supply labor more efficiently by ignoring daily earnings.

Appendix D.5 considers the possibility of heterogeneity in income effects based on experience. We classify drivers as *new* if we first observe them in our data on or after April 1 (Haggag et al., 2017). As Appendix Table 11 shows, we do not find any evidence of larger quitting responses to additional earnings for new drivers. In addition to this across-driver definition of experience, we also consider a within-driver definition of experience in Appendix Table 10 and do not find significant differences as new drivers gain more experience.

Measurement of Hours At least two issues arise when measuring work hours in this setting, which Appendix D.6 discusses in more detail. First, the data do not distinguish between a driver who ends a shift immediately after dropping off their last passenger and a driver who spends time searching for another fare unsuccessfully. This would only pose a concern if a driver tends to face greater difficulties in finding passengers towards the end of a shift in which the driver earns more, which is unlikely. Second, the data do not contain an explicit measure of break times. A positive correlation between earnings and unobserved tiredness, and thus quitting behavior, could arise if drivers take fewer breaks in response to additional earnings. However, we find no support for this in the data.

3 Structural Model of Reference Dependence

This section examines what our evidence of excess sensitivity of daily labor-supply decisions to recent earnings implies about reference-point formation. While various different models can potentially formalize how reference points influence behavior, this section focuses on models based on prospect theory, as existing work invokes reference dependence and loss aversion to explain income-targeting behavior. Appendix F provides a complementary investigation of how models based on salience (Bordalo et al., 2015) account for the evidence. Even without taking a strong stance on a particular account of how the reference level influences decisions, our findings suggest that the reference level must adjust within a day. We therefore aim to develop and assess alternative formulations of the reference point.

3.1 Daily Labor Supply Model Setup

We model the stopping decision to depend on the next trip's expected fare $\mathbb{E}_t[f_{t+1}]$ and duration $\mathbb{E}_t[h_{t+1}]$ (Crawford and Meng, 2011). A driver with cumulative earnings I_t and hours of work H_t at the end of trip t decides to end a shift if the driver expects that completing an additional trip results in lower utility, i.e.,

$$\mathbb{E}_t[v(I_{t+1}, H_{t+1})] - v(I_t, H_t) + \varepsilon_t < 0, \tag{2}$$

where $I_{t+1} = I_t + \mathbb{E}_t[f_{t+1}]$, $H_{t+1} = H_t + \mathbb{E}_t[h_{t+1}]$, and ε is an error term. The neoclassical model in Appendix C.1 posits that the marginal utility of lifetime income—and hence labor supply—does not vary in response to small, within-day changes in wealth. To capture this, the objective function v takes the form

$$\begin{aligned} v(I_t, H_t) &= v_I(I_t) + v_H(H_t) \\ &= I_t - \frac{\psi}{1+\nu} H_t^{1+\nu}, \end{aligned} \tag{3}$$

where ψ parameterizes the disutility of work and ν is the elasticity parameter. While the objective function depends explicitly on cumulative daily earnings I_t, quasi-linearity implies that earnings do not affect the decision to end a shift.

Explaining the results in Section 2 requires non-trivial within-day changes in the marginal utility of income. Appendix E.1 presents a back-of-the-envelope calculation showing that the daily income effect implies an implausibly high degree of risk aversion of over 100. The leading explanation in the literature for the mixed evidence on behavior in daily labor-supply decisions comes from the theory of reference-dependent preferences due to Kőszegi and Rabin (2006) (see Crawford and Meng 2011 and a survey of the earlier work in DellaVigna 2009). In their model, utility depends not only on a standard outcome-based consumption component but also on a gain-loss component which captures how decision makers assess choices relative to a reference point. Our primary analysis of loss aversion involves a simplified version of the model following the implementation from Farber (2008) and Crawford and Meng (2011). The objective function of the driver takes the form

$$v^{LA}(I_t, H_t) = (1-\eta)v(I_t, H_t) + \eta n(I_t \mid I_t^r), \tag{4}$$

where I^r denotes the reference level for income (i.e., the driver's expected earnings for the shift), η determines the relative weight on gain-loss utility, and the gain-loss utility is given by

$$n(I \mid I^r) = \Big(\mathbf{1}_{\{I>I^r\}} + \lambda \mathbf{1}_{\{I<I^r\}}\Big)(I - I^r),$$

where $\lambda \geq 1$ parameterizes the degree of loss aversion. This coincides with the neoclassical model when there is no difference in utility from gains and losses (i.e., $\lambda = 1$ or $\eta = 0$). Compared to the specification from Crawford and Meng (2011), Equation (4) does not include a reference level for hours of work. In addition, this formulation makes two simplifying assumptions about the general gain-loss component of utility from Kőszegi and Rabin (2006). First, the reference levels represent a driver's point expectations for income and hours on a given shift, abstracting from stochasticity whereby the reference levels represent the full distribution of potential earnings for that particular shift. Second, the piecewise-linear gain-loss function rules out diminishing sensitivity, the observation that decision makers experience smaller marginal changes in gain-loss sensations further away from their reference levels. The main results below do not substantively change when we relax each of these assumptions in Section 3.4.

3.2 Specifications of the Reference Point

The model of loss aversion can produce a daily income effect through a decrease in the marginal utility of income at a given reference point. However, predictions about how the probability of ending a shift responds to the timing of earnings depend crucially on how we specify the reference point.

One class of specifications consists of forward-looking reference points based on the lagged expectation. Kőszegi and Rabin (2006) posit that rational expectations endogenously determine the reference point. They make an important distinction between how beliefs and preferences adjust to new information: the model does not require expectations to adjust slowly, but reference points depend on the lagged expectation and thus do not change instantaneously with new information.

In the absence of a theoretical account of how quickly the reference point adjusts, we consider a range of possibilities. A lag sufficiently long that reference points do not adjust results in a fixed reference point as in the original view of income targeting (Camerer et al., 1997). With a vanishingly short lag, reference points fully adjust to new information, removing the influence of the reference point on decisions and resulting in neoclassical behavior. In general, we take the reference point in period τ as the expectation held in period $\tau - 1$, for some definition of a period. The daily-level income target from (Crawford and Meng, 2011), which we denote I_0^r, lies between the two extremes above, as does a reference point that updates every hour or every trip.

Another class of specifications consists of backward-looking reference points based on past experiences or outcomes. Bowman et al. (1999) and DellaVigna et al. (2017), for instance, consider reference points based on past consumption or income. Post et al. (2008) presents a dynamic model for the reference point which separates the effect of initial expectations from the effect of the most recent outcomes. In the daily labor supply setting, a reference point along these lines preserves some of the advantages of a forward-looking reference point, such as explaining why reference dependence does not require that higher wages generically lead to lower effort.

We model the adaptive reference point as a convex combination of the lagged reference point and the reference point that would obtain if new information were fully incorporated. Let Δ_t denote the difference between realized and expected earnings given the duration of trip t, i.e., $\Delta_t = f_t - \mathbb{E}_{t-1}[f_t]\frac{h_t}{\mathbb{E}_{t-1}[h_t]}$. Then the expectation of daily earnings after trip t is given by $E_t = I_0^r + \sum_{\tau=1}^{t} \Delta_\tau$. We define the updated reference point as

$$I_t^r = \theta I_{t-1}^r + (1-\theta)E_t, \tag{5}$$

where $0 \leq \theta \leq 1$, with $\theta = 1$ corresponding to a reference point that does not adjust within the day and $\theta = 0$ corresponding to a reference point that adjusts instantaneously. Rewriting this recursive formulation, we express the updated reference point as

$$I_t^r = \theta^t I_0^r + (1-\theta)\sum_{\tau=1}^{t} \theta^{t-\tau} E_\tau,$$

which highlights that the adaptive reference point consists of a weighted average of multiple lagged values of expectations. Rewriting the updated reference point as

$$I_t^r = I_0^r + \sum_{\tau=1}^{t}(1-\theta^{t+1-\tau})\Delta_\tau.$$

highlights that the reference point incorporates less recent earnings to a greater extent, consistent with the idea that reference points take time to adjust in response to recent changes in expectations.

Explaining both the daily income effect and the recency effect requires a slow-adjusting reference point within the day. While our specification nests the static reference point ($\theta = 1$) as well as a reference point that adjusts instantaneously ($\theta = 0$), both of these extreme cases imply the fungibility of money within a shift, and only the intermediate case with $0 < \theta < 1$ permits a violation of fungibility. A reference point that does not evolve within the day can account for a labor-supply response to earnings but eliminates the scope for more recent experiences to have a stronger influence on stopping decisions. With a reference point that adjusts within a shift, a more recent gain may make a driver more likely to exceed his income target than an earlier gain because the reference point takes time to adjust. However, a reference point that adjusts instantaneously cannot explain the sensitivity of labor-supply decisions to daily earnings because deviations from expectations no longer bring cumulative daily earnings closer to or further from the reference point.

Appendix E.2 provides suggestive evidence that drivers use an updated income target which incorporates more information from earlier in the shift. We estimate the reduced-form stopping model allowing for jumps at the shift-level expected income target I_0^r and targets that differentially incorporate gains and losses accumulated in the earliest several hours or latest several hours of the shift. Across specifications, we find that the probability of stopping significantly increases when income passes the target that updates more in response to earlier experiences. The remainder of this section investigates the determination of the reference point more formally by estimating the structural model above.

3.3 Estimation and Identification

Using the data on stopping decisions, we estimate the models via maximum likelihood under the various specifications of the reference point. Given the stopping rule in Equation (2), we follow Crawford and Meng (2011) in assuming that $\varepsilon_t = x_t\beta + \xi_t$, where $x_t\beta$ captures the effect of control variables and ξ_t are independent and normally distributed with mean zero and variance σ^2. This yields likelihood functions of the form

$$\sum \log \Phi\left(\frac{v^{LA}(I_t, H_t) - \mathbb{E}_t\Big[v^{LA}(I_{t+1}, H_{t+1})\Big]}{\sigma}\right), \tag{6}$$

where Φ denotes the standard normal cumulative distribution function. The control variables consist of the time, weather, and location controls from Section 2. To accommodate a more flexible relationship between hours and quitting, we allow the disutility of effort to take a separate value on each half-hour interval of the shift. Although the model imposes the same vector of parameters across shifts for the disutility of effort, the true disutility of effort might vary with expected hours for a given shift. To minimize the need for introducing additional parameters, we restrict the sample by removing shifts in the top and bottom quartile of the distribution of expected hours.[17]

We begin with the adaptive reference point from Equation (5). To proxy for drivers' initial expectations I_0^r, we follow Crawford and Meng (2011) in using the sample average of income by driver and day of week (excluding the current shift). To set the updating term Δ_t, we define trip-level expectations of fares f_t based on a regression of next-trip fare on the time and weather controls from Figure 2 and the location controls from Table 1 (and similarly for hours h_t).[18] As Appendix E.3 shows, the parameters η and λ are not separately identifiable under this specification of the reference point. This occurs because the decision maker takes the reference point as exogenous to their choice, and thus behavior depends only on the ratio between utility from losses and gains $L = 1 + (\lambda - 1)\eta$.

Although we estimate the parameters jointly, the following describes some of the key sources of identification. Variation in work hours, expected wages from continuing, cumulative income, and the timing of income contributes to the identification of the disutility of effort ψ, elasticity parameter ν, coefficient of loss aversion L, and speed of adjustment θ, respectively. Appendix Figure 10 elucidates the link between the structural parameters and these sources of variation in the data.

We estimate the parameters and obtain standard errors using subsampling (Politis and Romano, 1994). The estimates in Table 2 obtain from 230 subsamples (without replacement) of 150,000 observations each.

Column (1) reports estimates under the restriction $L = 1$, corresponding to the case without loss aversion. Column (2) allows for loss aversion relative to a static reference point ($\theta = 1$), which coincides with the model that Crawford and Meng (2011) estimate. The parameter values generally fall within a comparable range to the analogous specification in Crawford and Meng (2011) (see column 5 of their Table 4). A likelihood ratio test rejects the null hypothesis of no loss aversion.

Relaxing the assumption of a static reference point, column (3) presents estimates from the full specification with loss aversion relative to the adaptive reference point.

[17] We proxy for expectations about hours in using the sample average of hours by driver and day of week, excluding the current shift (Crawford and Meng, 2011).

[18] We use seemingly unrelated regressions following Crawford and Meng (2011).

Table 2 reports results consistent with a violation of fungibility, highlighting the importance of within-day adjustments. The estimate of θ differs significantly from 0 and 1, and a likelihood ratio test rejects the restriction to a static reference point. The remaining parameters do not substantially differ from their counterparts in column (2). The magnitude of our estimate for the speed of adjustment θ depends on the definition of a period. Since we take each period to be a trip, the point estimate implies that the reference point adjusts immediately to incorporate about 17 percent of a shock to earnings. Estimating the speed of reference point adjustment at a lower frequency (e.g., defining a period as an hour instead of a trip) would result in a smaller magnitude of θ. Within an hour, the reference point incorporates about 40 percent of the shock, and within four hours only about 10 percent of the shock remains unincorporated.

To assess how well the model fits the data, Appendix Figure 11 compares the observed and predicted recency pattern of the income effect. While the stronger effects of recently accumulated earnings serve as the motivation for developing a model in which the reference point adjusts, our estimation approach does not explicitly target these moments. The figure plots the predicted effect of an additional 10 percent ($26) in income earned at various times in the shift on the probability of ending a shift at 8.5 hours, with the estimates from Figure 4 providing a benchmark for comparison. Adaptive reference points, as we expect, lead to stronger labor-supply reductions in response to more recent experiences. Moreover, the consistency between the predicted income effects and the observed magnitudes in the data provides support for the model of adaptive reference points.

3.4 Variants of the Model

This section discusses several variants of the model. We begin with alternate specifications of the reference point based on the lagged expectation of earnings, ranging from a fixed reference point to one that updates each trip. We then consider different versions of the model which relax the simplifying assumptions in Section 3.1.

Alternative Specifications of the Reference Point Appendix Table 13 presents estimates from a wide range of specifications of the reference point based on the lagged expectation of earnings. The first column considers a fixed reference point, which we take to equal the driver's average earnings across all shifts. This could either represent an ad-hoc model of daily income targeting in which the marginal utility of income declines substantially around the level of average daily earnings as Camerer et al. (1997) suggest, or a model in which the reference point updates sufficiently slowly that it does not vary. The second column defines the reference point as drivers' expectations at the end of the previous day. This corresponds to the reference point in Crawford and Meng (2011), as well as the static reference point in column (2) of Table 2. The remaining specifications allow for within-day updating of the reference point. We estimate how expectations evolve throughout the day using Equation (1) with total fare earnings for the shift as the outcome variable. Columns (3) and (4) define updated expectations with a lag of an hour and a trip, respectively. In all cases, a likelihood ratio test rejects the null hypothesis of no loss aversion.

The reference points based on lagged expectations of earnings predict a pattern of income effects that do not match the observed recency pattern in the data. As Appendix Figure 12 illustrates, these specifications of the reference point produce a stark contrast between the most recently accumulated earnings and any earlier earnings because the updated reference point incorporates fully the latter but not at all the former. The data instead show a gradual relationship between the timing of additional earnings and the probability of ending a shift.

Alternative Specifications of the Model As Section 3.1 discusses, we make several simplifying assumptions when introducing the model of reference dependence. Relaxing these simplifying assumptions in the model of loss aversion does not change our conclusions about the importance of adaptive reference points.

First, the model focuses only on reference dependence in earnings. Kőszegi and Rabin (2006) posit that loss aversion operates in all dimensions of utility, and Crawford and Meng (2011) implement the model by including a reference point for hours in the cabdriver's objective function. Under this view, drivers experience losses from working longer than their “hours target,” analogous to the losses from earning less than their “income target,” with the same coefficient of loss aversion on both dimensions. An hours target seems particularly difficult to disentangle from neoclassical behavior because drivers may, for example, form commitments that coincide with their expectations about work hours. Nevertheless, Appendix Table 14 presents results from one specification with a common coefficient of loss aversion for income and hours and another specification with a separate parameter for each. We find similar magnitudes for the speed of adjustment across the various specifications of the reference-dependence model. The estimates reveal a significant degree of loss aversion on both dimensions, interpreting the magnitude of loss aversion over the hours dimension poses difficulties due to a potential relationship between drivers' expectations and any commitments they may have that coincide with planned work hours.

Second, as in Crawford and Meng (2011), our income target I^r consists only of a point expectation and the gain-loss function does not exhibit the diminishing-sensitivity feature of prospect theory. Appendix Table 16 re-estimates the model by adding curvature in the gain-loss function, and Appendix Table 15 allows for stochastic reference points that capture the distribution of potential earnings. In both cases, the results continue to reject a reference point that does not adjust within a day (i.e., $\theta = 1$).

4 Discussion

This paper examines reactions to recent experiences in the context of daily labor-supply decisions. We conclude by discussing the implications of our findings for research on foundations and applications of reference-dependent behavior.

Our results suggest an important role for future work in developing the behavioral foundations of and characterizing the optimality properties of adaptation. Loss aversion may reflect a deep-rooted preference, as in the evolutionary account of Galor and Savitskiy (2018). Under this view, adaptation may serve as a rule of thumb that effectively brings drivers closer to neoclassical behavior, and the lag in the adaptive expectations process leads to inefficiency. Alternatively, as Camerer et al. (1997) suggest, daily income targeting may occur because it helps mitigate self-control problems much like other forms of mental accounting. Recent theoretical work (Koch and Nafziger, 2016; Hsiaw, 2018) formalizes this idea and describes conditions under which decision makers optimally evaluate outcomes separately rather than jointly. Using narrow brackets or setting incremental goals provides incentive effects to counteract present bias. Setting broader goals partially insures decision makers against losses, allowing for benefits from risk pooling, at the cost of weakening incentives. Koch and Nafziger (2016) show that their conclusions about the optimality of narrow bracketing also hold for settings in which the reference point adjusts quickly to changes in individuals' expectations when new information arrives. In light of our findings, future work can explore the role of history dependence in goal setting. Moreover, our evidence against fixed daily income targets suggests a new perspective on bracketing behavior as being less rigid than existing models permit.

Understanding reference dependence through the notion of adaptation may prove useful in other settings as well. In our framework, the speed of adjustment of the reference point governs the extent to which decision makers exhibit neoclassical behavior. This view suggests recasting debates about whether reference points influence behavior into the exercise of empirically analyzing the determinants of reference points. A systematic characterization of how quickly reference points adjust in different settings would shed further light on the sources and implications of reference dependence. As self-employment and contract work become increasingly prevalent, further research can investigate the role of adaptive reference points in flexible work relationships. Our results on excess sensitivity to recent earnings may also have implications outside the domain of labor supply, such as in understanding how the timing of information or payments affects consumption and savings decisions. This could potentially inform the design of effective policy instruments for providing fiscal stimulus or encouraging retirement savings. Future work can explore the influence of adaptive reference points in these as well as other field settings.

References

Aalen, Odd O., "A Linear Regression Model for the Analysis of Life Times," *Statistics in medicine*, 1989, *8* (8), 907–925.

Agarwal, Sumit, Mi Diao, Jessica Pan, and Tien Foo Sing, "Labor Supply Decisions of Singaporean Cab Drivers," *Mimeo*, 2015.

Andersen, Steffen, Alec Brandon, Uri Gneezy, and John A. List, "Toward an Understanding of Reference-Dependent Labor Supply: Theory and Evidence from a Field Experiment," *Mimeo*, 2018.

Ashenfelter, Orley, Kirk Doran, and Bruce Schaller, "A Shred of Credible Evidence on the Long-run Elasticity of Labour Supply," *Economica*, 2010, *77* (308), 637–650.

Bartling, Björn, Leif Brandes, and Daniel Schunk, "Expectations as Reference Points: Field Evidence from Professional Soccer," *Management Science*, 2015, *61* (11), 2646–2661.

Baucells, Manel, Martin Weber, and Frank Welfens, "Reference-Point Formation and Updating," *Management Science*, 2011, *57* (3), 506–519.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Memory, Attention and Choice," *Mimeo*, 2015.

Bowman, David, Deborah Minehart, and Matthew Rabin, "Loss aversion in a consumption–savings model," *Journal of Economic Behavior & Organization*, 1999, *38* (2), 155–178.

Buchholz, Nicholas, Matthew Shum, and Haiqing Xu, "Dynamic Labor Supply of Taxicab Drivers: a Semiparametric Optimal Stopping Model," *Mimeo*, 2018.

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler, "Labor Supply of New York City Cabdrivers: One Day at a Time," *Quarterly Journal of Economics*, 1997, *112* (2), 407–441.

Card, David and Gordon B. Dahl, "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior," *Quarterly Journal of Economics*, 2011, *126* (1), 103–143.

Chang, Tom and Tal Gross, "How Many Pears Would a Pear Packer Pack if a Pear Packer Could Pack Pears at Quasi-Exogenously Varying Piece Rates?," *Journal of Economic Behavior & Organization*, 2014, *99*, 1–17.

Chou, Yuan K., "Testing Alternative Models of Labour Supply: Evidence from Taxi Drivers in Singapore," *Singapore Economic Review*, 2002, *47* (01), 17–47.

Cleveland, William S. and Susan J. Devlin, "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting," *Journal of the American statistical association*, 1988, *83* (403), 596–610.

Crawford, Vincent P. and Juanjuan Meng, "New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income," *American Economic Review*, 2011, *101* (5), 1912–1932.

DellaVigna, Stefano, "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 2009, *47* (2), 315–72.

_ , "Structural Behavioral Economics," in Stefano DellaVigna Doug Bernheim and David Laibson, eds., *Handbook of Behavioral Economics*, Vol. 1, Elsevier, 2018.

_ , Attila Lindner, Balázs Reizer, and Johannes F. Schmieder, "Reference-Dependent Job Search: Evidence from Hungary," *Quarterly Journal of Economics*, 2017, *132* (4), 1969–2018.

Doran, Kirk, "Are long-term wage elasticities of labor supply more negative than short-term ones?," *Economics Letters*, 2014, *122* (2), 208–210.

Dupas, Pascaline, Jonathan Robinson, and Santiago Saavedra, "The Daily Grind: Cash Needs, Labor Supply and Self-Control," *Mimeo*, 2018.

Farber, Henry S., "Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers," *Journal of Political Economy*, 2005, *113* (1), 46–82.

_ , "Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers," *American Economic Review*, 2008, *98* (3), 1069–1082.

_ , "Why You Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers," *Quarterly Journal of Economics*, 2015, *130* (4), 1975–2026.

Fehr, Ernst and Lorenz Goette, "Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment," *American Economic Review*, 2007, *97* (1), 298–317.

Frechette, Guillaume R., Alessandro Lizzeri, and Tobias Salz, "Frictions in a Competitive, Regulated Market Evidence from Taxis," *Mimeo*, 2016.

Galor, Oded and Viacheslav Savitskiy, "Climatic Roots of Loss Aversion," *Mimeo*, 2018.

Gill, David and Victoria Prowse, "A Structural Analysis of Disappointment Aversion in a Real Effort Competition," *American Economic Review*, 2012, *102* (1), 469–503.

Gneezy, Uri, Lorenz Goette, Charles Sprenger, and Florian Zimmermann, "The Limits of Expectations-Based Reference Dependence," *Journal of the European Economic Association*, 2017, *15* (4), 861–876.

Goette, Lorenz, Annette Cerulli-Harms, and Charles Sprenger, "Randomizing Endowments: An Experimental Study of Rational Expectations and Reference-Dependent Preferences," *American Economic Journal: Microeconomics*, forthcoming.

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

_ , Brian McManus, and Giovanni Paci, "Learning by Driving: Productivity Improvements by New York City Taxi Drivers," *American Economic Journal: Applied Economics*, 2017, *9* (1), 70–95.

Heffetz, Ori and John A. List, "Is the Endowment Effect an Expectations Effect?," *Journal of the European Economic Association*, 2014, *12* (5), 1396–1422.

Hsiaw, Alice, "Goal Bracketing and Self-Control," *Games and Economic Behavior*, 2018.

Jonason, David and Joel Wållgren, "Labor Supply of Stockholm Cab Drivers: Revisiting the Evidence," *Mimeo*, 2013.

Kahneman, Daniel and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 1979, *47* (2), 263–291.

Kőszegi, Botond and Matthew Rabin, "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, 2006, *121* (4), 1133–1165.

_ **and** _ , "Reference-Dependent Consumption Plans," *American Economic Review*, 2009, *99* (3), 909–936.

Koch, Alexander K. and Julia Nafziger, "Goals and bracketing under mental accounting," *Journal of Economic Theory*, 2016, *162*, 305–351.

Martinussen, Torben and Thomas H. Scheike, *Dynamic Regression Models for Survival Data*, Springer Science & Business Media, 2006.

Meng, Juanjuan and Xi Weng, "Can Prospect Theory Explain the Disposition Effect? A New Perspective on Reference Points," *Management Science*, 2017.

Morgul, Ender Faruk and Kaan Ozbay, "Revisiting Labor Supply of New York City Taxi Drivers: Empirical Evidence from Large-scale Taxi Data," *Transportation Research Board*, 2014.

Nguyen, Quang and PinngSun Leung, "Revenue Targeting in Fisheries: The Case of Hawaii Longline Fishery," *Environment and Development Economics*, 2015.

Politis, Dimitris N. and Joseph P. Romano, "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions," *Annals of Statistics*, 1994, pp. 2031–2050.

Pope, Devin G. and Maurice E. Schweitzer, "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes," *American Economic Review*, 2011, *101* (1), 129–157.

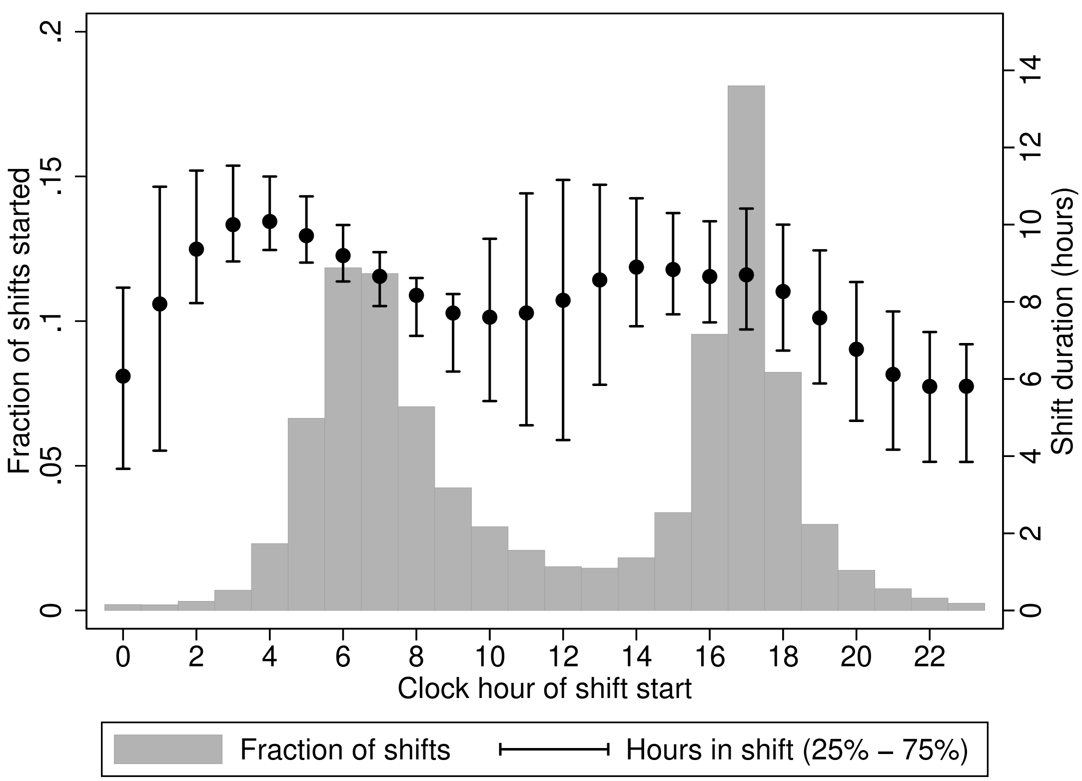
Post, Thierry, Martijn J. Van den Assem, Guido Baltussen, and Richard H. Thaler, "Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show," *American Economic Review*, 2008, *98* (1), 38–71.

Schmidt, Marc-Antoine, "Identifying Behavioral Responses in Labor Supply: Idiosyncratic Shocks vs. Market-Level Shocks," *Mimeo*, 2017.

Song, Changcheng, "An Experiment on Reference Points and Expectations," *Mimeo*, 2016.

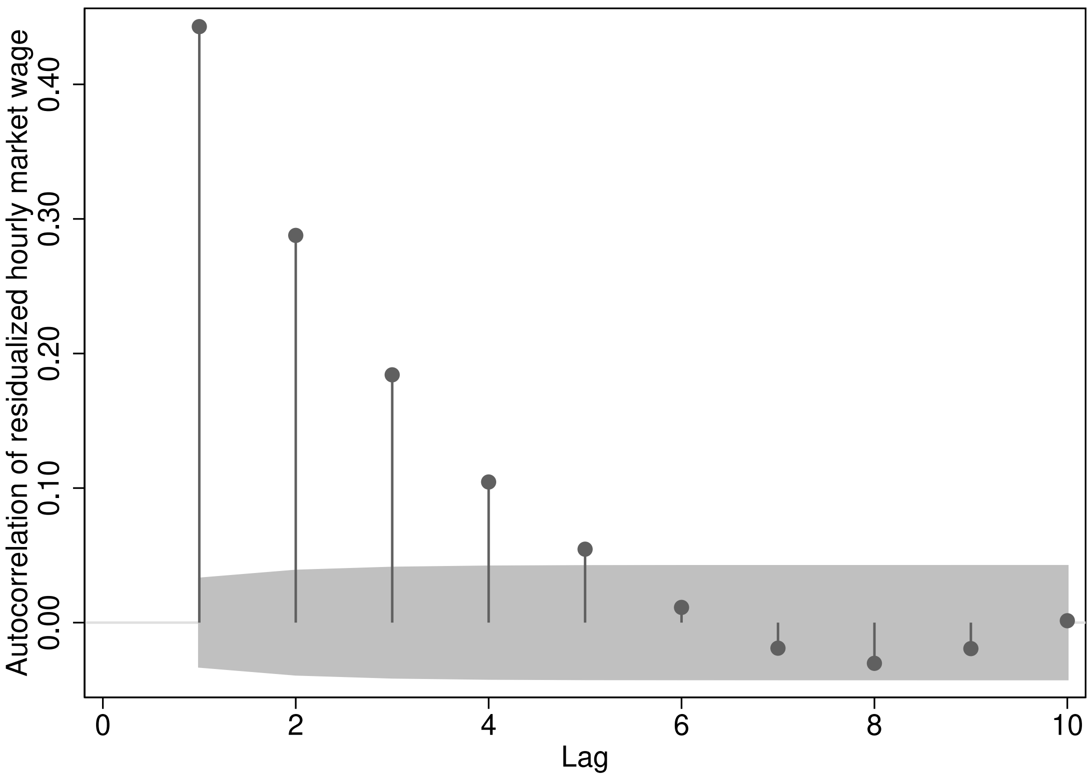
Stafford, Tess M., "What Do Fishermen Tell Us That Taxi Drivers Do Not? An Empirical Investigation of Labor Supply," *Journal of Labor Economics*, 2015, *33* (3), 683–710.

Figure 1: Shift-level summary statistics



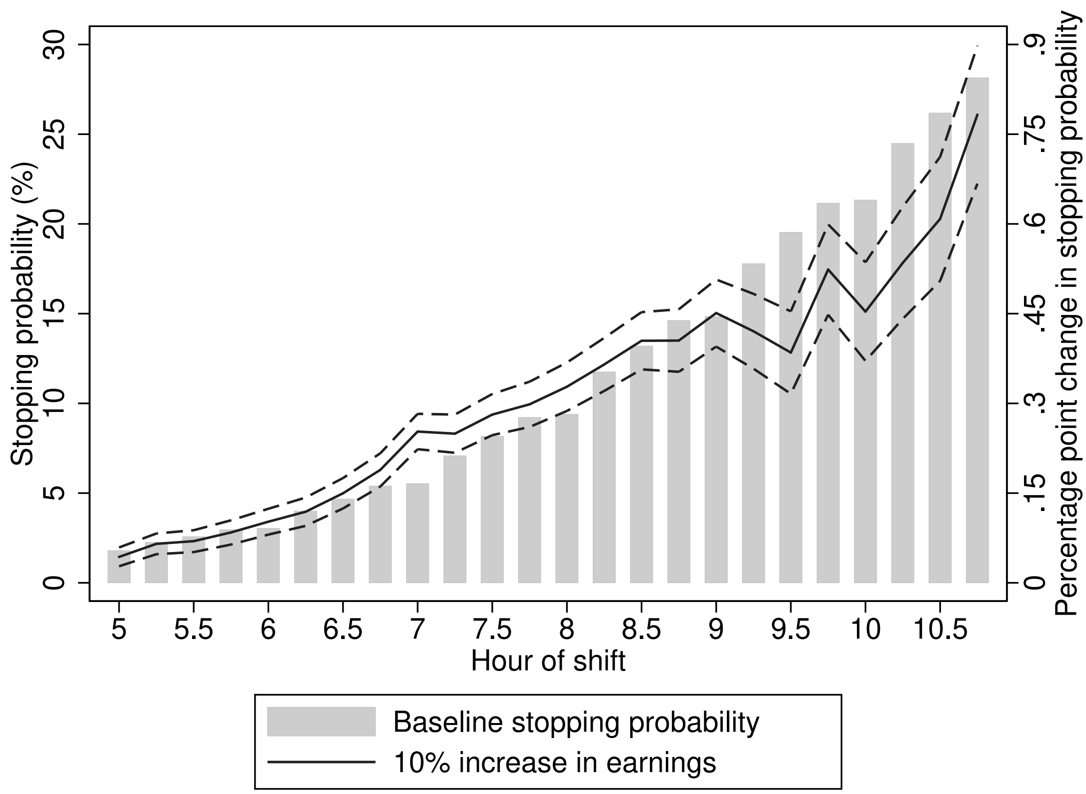
Note: The histogram depicts the distribution of shifts by the clock hour of when the shift starts between hour 0 and hour 23. For each clock hour, the distribution of duration of shifts starting at that hour is depicted by the bar graph, with the mean and interquartile range.

Figure 2: Autocorrelation of residualized hourly market wage



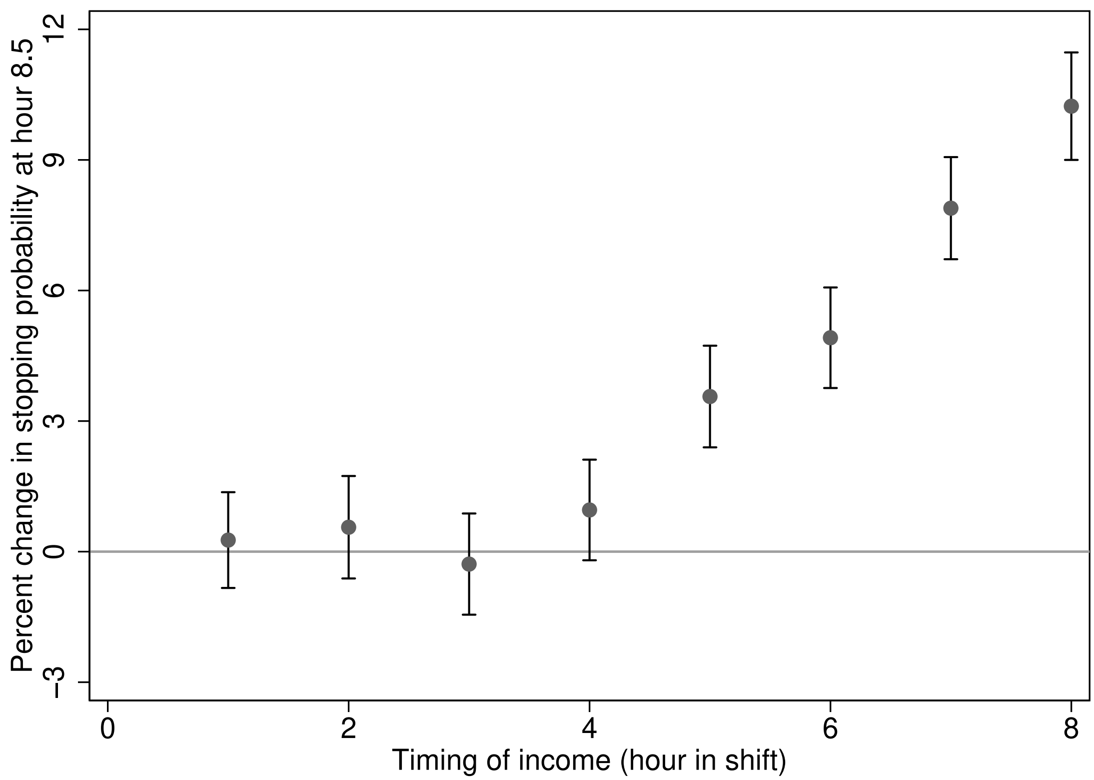
Note: The figure depicts the autocorrelation of hourly market wages indexed by hour of the calendar year 2013. The hourly market wage is the sum of the minute market wage in each hour, with the minute market wage computed as in Appendix Figure 1. The hourly market wage is residualized from a regression on a set of time and weather effects: an interaction between the hour of day and day of week, the week of the year, an indicator for federal holidays, an indicator for whether it rains during that hour, and indicators for high (over 80 degrees Fahrenheit) and low (under 30 degrees Fahrenheit) average hourly temperature. The shaded region denotes a Bartlett 95-percent confidence band.

Figure 3: Stopping model estimates: Income effect throughout the shift



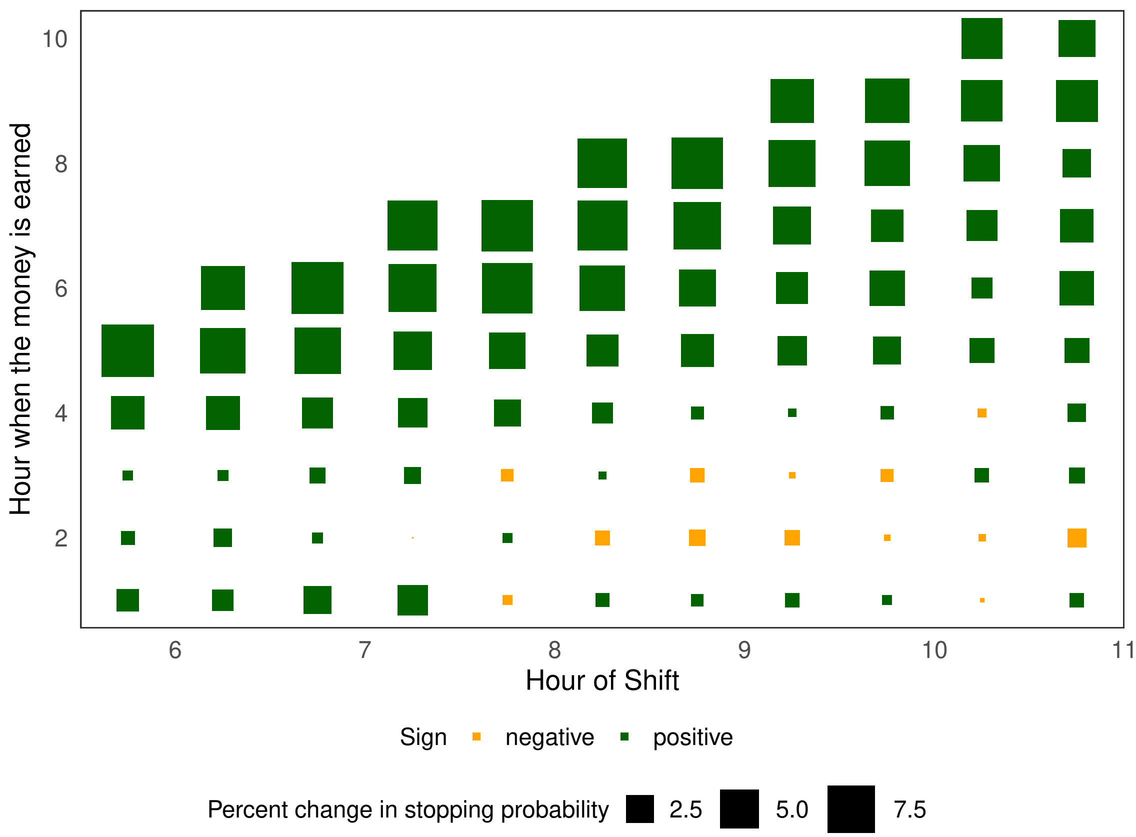
Note: The bars, corresponding to the scale on the left, show the probability that a driver ends a shift after completing a trip at the specified number of hours. The solid lines, corresponding to the scale on the right, depict the percentage point change in the probability of stopping, evaluated at various times throughout the shift, in response to a 10 percent increase in earnings. Estimates obtain from Equation (TT) with the full set of controls (see Table 1 for details) and fixed effects for 37,492 drivers. The dashed lines represent the 95-percent confidence interval, with standard errors adjusted for clustering at the driver level.

Figure 4: Stopping model estimates: Elasticity of stopping at 8.5 hours with respect to income—By timing of income



Note: The figure depicts the percent change in the probability of ending a shift at 8.5 hours (baseline stopping probability of 13.2 percent) in response to a 10 percent ($26) increase in earnings accumulated at different times in the shift. Estimates obtain from Equation (1) with the full set of controls (see Table 1 for details) and fixed effects for 37,166 drivers.

Figure 5: Stopping model estimates: Elasticity of stopping at different hours of the shift with respect to income—By timing of income



Note: The figure depicts the effect of an additional 10 percent in earnings accumulated at different times in the shift (vertical axis) on the probability of stopping at various times throughout the shift (horizontal axis). Each square has area proportional to the corresponding percent change in the probability of stopping. Estimates obtain from Equation (1) with the full set of controls (see Table 1 for details) and fixed effects for 37,460 drivers.

Table 1: Elasticity of stopping at 8.5 hours with respect to income

Panel A: Effect of 1% increase in accumulated earnings

	(TT)	(F-1)	(F-2)	(F-3)
Controlling for				
Hours	0.0636	0.7853	0.0041	-0.2385
	(0.0272)	(0.0143)	(0.0495)	(0.0931)
& Drivers	0.4838	0.8748	0.0754	-0.1999
	(0.0187)	(0.0123)	(0.0486)	(0.0915)
& Time	0.8304	0.6689	0.4902	0.2494
	(0.0176)	(0.0132)	(0.0468)	(0.0879)
& Location	0.3355	0.2078	0.3764	0.1186
	(0.0169)	(0.0132)	(0.0461)	(0.0865)
& Weather	0.3336	0.2074	0.3759	0.1183
	(0.0169)	(0.0132)	(0.0461)	(0.0865)

Panel B: Comparison with previous estimates (95% confidence interval)

This paper:	0.3336	(0.3005, 0.3667)
Farber 2005:	0.12	(-0.20, 0.44)
Farber 2015:	0.9456 (day), 0.0727 (night)	

Note: Panel A reports in each cell an estimate of the percent change in the probability of ending a shift at 8.5 hours in response to a 1 percent increase in cumulative earnings. The columns corresponds to the specifications in Equations (TT) to (F-3), respectively. All specifications include controls for minutes spent working, including indicators for whether the driver spends time with passengers in each hour. Time controls include fixed effects for hour of day by day of week and for day of year. Location controls consist of neighborhood fixed effects. Weather controls consist of indicators for precipitation, wind speed, and temperature in the minute that a trip ends. Drivers denotes fixed effects for the anonymized license numbers. Standard errors reported in parentheses are adjusted for clustering at the driver level. Panel B reports calculations based on Table 5 of Farber (2005) and Table VII of Farber (2015). The sample consists of over 37,000 drivers; see Appendix A for further details.

Table 2: Maximum likelihood estimates

	(1)	(2)	(3)
Disutility of effort ψ	0.1105	0.1932	0.1531
	(0.0697)	(0.1480)	(0.0900)
Elasticity ν	0.8585	0.7902	0.8477
	(0.2641)	(0.2582)	(0.2378)
Loss aversion L		1.9992	2.6074
		(0.2364)	(0.2968)
Adjustment θ			0.8227
			(0.0640)
Error term distribution σ	0.2487	0.4077	0.4009
	(0.0289)	(0.0914)	(0.0640)
Test $\lambda = 1$		$<$0.001	$<$0.001
Test $\theta = 1$			$<$0.001

Note: This table presents maximum likelihood estimates of Equation (6) with the adaptive reference point given by Equation (5). The estimation sample consists of 35.5 million trips from over 37,000 drivers. Column (1) corresponds to the restriction $L = 1$, column (2) corresponds to the restriction $\theta = 1$, and column (3) presents the full specification. We report the estimated disutility of effort parameter that applies to trips that occur between hour 8.5 and hour 9 of the shift. The last two rows contain p-values from likelihood ratio tests of the following null hypotheses: (i) the model without loss aversion, and (ii) the model with a static reference point.