# New York City Cabdrivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income

Vincent P. Crawford and Juanjuan Meng[1]

First version 23 July 2008; second revision 9 March 2010

Abstract: This paper proposes a model of cabdrivers' labor supply, building on Henry S. Farber's (2005, 2008) empirical analyses and Botond Kőszegi and Matthew Rabin's (2006; henceforth "KR") theory of reference-dependent preferences. Following KR, our model has targets for hours as well as income, determined by proxied rational expectations. Our model, estimated with Farber's data, reconciles his finding that stopping probabilities are significantly related to hours but not income with Colin Camerer et al.'s (1997) negative "wage" elasticity of hours; and avoids Farber's criticism that estimates of drivers' income targets are too unstable to yield a useful model of labor supply.

[1] Crawford: Department of Economics, University of Oxford, Manor Road, Oxford OX1 3UQ, United Kingdom, and Department of Economics, University of California, San Diego. 9500 Gilman Drive, La Jolla, CA 92093-0508 (e-mail: vincent.crawford@economics.ox.ac.uk); Meng: Department of Economics, University of California, San Diego. 9500 Gilman Drive, La Jolla, CA 92093-0508 (e-mail: jumeng@ucsd.edu). We thank Henry Farber, Colin Camerer, and Linda Babcock for sharing their data, and Nageeb Ali, Dan Benjamin, Camerer, David Card, Julie Cullen, Gordon Dahl, Stefano Dellavigna, Kirk Doran, Farber, Uri Gneezy, Roger Gordon, Nagore Iriberri, Botond Kőszegi, Jaimie Lien, David Miller, Ulrike Malmendier, Ted O'Donoghue, Matthew Rabin, Adam Sanjurjo, Joel Sobel, and the referees for their insightful comments and suggestions.

## New York City Cabdrivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income

In the absence of large income effects, a neoclassical model of labor supply predicts a positive wage elasticity of hours. However, Camerer et al. (1997) collected data on the daily labor supply of New York City cabdrivers, who unlike most workers in modern economies are free to choose their own hours, and found a strongly negative elasticity of hours with respect to their closest analog of a wage, realized earnings per hour.[2]

To explain their results, Camerer et al. informally proposed a model in which drivers have daily income targets and work until the target is reached, and so work less on days when earnings per hour are high. Their explanation is in the spirit of Daniel Kahneman and Amos Tversky's (1979) and Tversky and Kahneman's (1991) prospect theory, in which a person's preferences respond not only to income as usually assumed, but also to a reference point; and there is "loss aversion" in that the person is more sensitive to changes in income below the reference point ("losses") than changes above it ("gains"). If a driver's reference point is a daily income target, then loss aversion creates a kink that tends to make realized daily income bunch around the target, so that hours have a negative elasticity with respect to realized earnings per hour.

Farber (2008, p. 1069) suggests that a finding that labor supply is reference-dependent would have significant policy implications:

> "Evaluation of much government policy regarding tax and transfer programs depends on having reliable estimates of the sensitivity of labor supply to wage rates and income levels. To the extent that individuals' levels of labor supply are the result of optimization with reference-dependent preferences, the usual estimates of wage and income elasticities are likely to be misleading."

Although Camerer et al.'s analysis has inspired a number of empirical studies of labor supply, the literature has not yet converged on the extent to which the evidence supports reference-dependence.[3] Much also depends on reference-dependence's scope and structure: If it

---

[2] In Camerer et al.'s dataset, realized earnings per hour (which they call the "wage") is uncorrelated across days but positively serially correlated within a day, so that high earnings early in a day signal higher earnings later that day, and a neoclassical model predicts a positive elasticity even though realized earnings per hour is not precisely a wage. If instead realized earnings per hour is serially uncorrelated within a day, as Farber (2005) shows is roughly true in his dataset (see however our analysis in Section II.3), then a driver with high early earnings experiences a small change in income but no change in expected wage, and a neoclassical model predicts an elasticity near zero. A neoclassical model could only explain Camerer et al.'s strongly negative elasticities via an implausibly large negative serial correlation of realized earnings per hour.

[3] KR (2006) and Farber (2008) survey some of the empirical literature. Gerald S. Oettinger's (1999) field study found increased

were limited to inexperienced workers or unanticipated changes, its direct relevance to most policy questions would be small.[4] This paper seeks to shed additional light on these issues, building on two recent developments: Farber's (2005, 2008) empirical analyses of cabdrivers' labor supply and KR's (2006; see also 2007, 2009) theory of reference-dependent preferences.

Farber (2005) collected and analyzed data on the labor supply decisions of a new set of New York City cabdrivers. He found that, before controlling for driver fixed effects, the probability of stopping work on a day is significantly related to realized income that day, but that including driver fixed effects and other relevant controls renders this effect statistically insignificant.

Farber (2008) took his 2005 analysis a step further, introducing a structural model based on daily income targeting that goes beyond the informal explanations in previous empirical work. He then estimated a reduced form, treating drivers' income targets as latent variables with driver-specific means and driver-independent variance, both assumed constant across days of the week—thus allowing the target to vary across days for a given driver, but only as a random effect.[5] He found that a sufficiently rich parameterization of his targeting model fits better than a neoclassical model, and that the probability of stopping increases significantly and substantially when the target is reached; but that his model cannot reconcile the increase in stopping probability at the target with the smooth aggregate relationship between stopping probability and realized income. Further, the estimated random effect in the target is large and significantly different from zero, but with a large standard error, which led him to conclude that the targets are too unstable to yield a useful reference-dependent model of labor supply (p. 1078):

> "There is substantial inter-shift variation, however, around the mean reference income level.… To the extent that this represents daily variation in the reference income level for a particular driver, the predictive power of the reference income level for daily labor supply would be quite limited."

KR's (2006) theory of reference-dependent preferences is more general than Farber's (2008) model in most respects, but takes a more specific position on how targets are determined. In

daily participation by stadium vendors on days on which the anticipated wage was higher, as suggested by the neoclassical model, and in seeming contrast to Camerer et al.'s finding of a negative response of hours to (partly unanticipated) increases in wage. Ernst Fehr and Lorenz Goette's (2007) field experiment found increased participation by bicycle messengers, but reduced effort, in response to announced increases in their commission. They argued that effort is a more accurate measure of labor supply and concluded that the supply of effort is reference-dependent.

4 Reference-dependence might still have indirect policy relevance via its influence on the structure of labor relationships.

5 Constancy across days of the week is violated in the sample, where Thursdays' through Sundays' incomes are systematically higher than those of other days, and the hypothesis that income is constant across days of the week is strongly rejected ($p$-value 0.0014, $F$-test with robust standard errors). Farber included day-of-the-week dummies in his main specifications for the stopping probability, but this turns out to be an imperfect substitute for allowing the mean income target to vary across days of the week.

KR's theory applied to cabdrivers, a driver's preferences reflect both the standard consumption utility of income and leisure and reference-dependent "gain-loss" utility, with their relative importance tuned by a parameter. As in Farber's model, a driver is loss-averse; but he has a daily target for hours as well as income, and working longer than the hours target is a loss, just as earning less than the income target is. Finally, KR endogenize the targets by setting a driver's targets equal to his theoretical rational expectations of hours and income, reflecting the belief that drivers in steady state have learned to predict their distributions.[6]

This paper uses Farber's (2005, 2008) data to reconsider the reference-dependence of cabdrivers' labor supply, adapting his econometric strategies to estimate models based on KR's (2006) theory. Section I introduces the model. Following KR, we allow for consumption as well as gain-loss utility and hours as well as income targets; but when we implement the model we follow Farber (2008) in assuming that drivers are risk-neutral in income.

To complete the specification, we must describe how a driver's targets are determined and, for some of our analysis, how he forms his expectations about earnings hour by hour during a day. In an important departure from Farber's approach, we follow KR in conceptualizing drivers' targets and expected earnings as rational expectations, but for simplicity we depart from KR in treating them as point expectations rather than distributions.[7] We operationalize the targets and expected earnings via natural sample proxies with limited endogeneity problems as explained below, for expected earnings assuming for simplicity that earnings per hour are serially uncorrelated within a day (as well as across days).[8] Given this, risk-neutrality in income, and ignoring option value in the stopping decision, a driver's expected hourly earnings are equivalent to a predetermined (though random) daily schedule of time-varying wages.[9]

[6] In theory there can be multiple expectations that are consistent with the individual's optimal behavior, given the expectations. KR use a refinement, "preferred personal equilibrium," to focus on the self-confirming expectations that are best for the individual. Most previous analyses have identified targets with the status quo; but as KR note, most of the available evidence does not distinguish the status quo from expectations, which are usually close to the status quo. Even so, our analysis shows that KR's rational-expectations view of the targets has substantive implications for modeling cabdrivers' labor supply. KR's view of the targets has also been tested and confirmed in laboratory experiments by Johannes Abeler et al. (2010).

[7] Because KR's model has no errors, their distributions are necessary for the existence of deviations from expectations, without which their model reduces to a neoclassical model. Our model has errors and so has deviations even with point expectations.

[8] This simplification is motivated by Farber's (2005) finding, in a detailed econometric analysis of his dataset, of only a weak and insignificant relationship, which led him to argue that hourly earnings are so variable and unpredictable that "predicting hours of work with a model that assumes a fixed hourly wage rate during the day does not seem appropriate." We note however that Camerer et al. did find some within-day predictability of earnings in their dataset. It also seems plausible that drivers on the ground may be able to predict their earnings better than even the most careful econometrics.

[9] Farber (2008) modeled a driver's stopping decision by estimating a daily latent income target and continuation value, assuming that a driver stops working when his continuation value falls below the cost of additional effort. He defined continuation value to include option value; but if option value is truly important, his linear specification of continuation value is unlikely to be appropriate. We simply assume that drivers' decisions ignore option value, as Thierry Post et al. (2008) did, and as seems

If the weight of gain-loss utility is small, our model mimics a neoclassical labor-supply model, so that the elasticity of hours with respect to earnings per hour is normally positive. If the weight of gain-loss utility is large, perfectly anticipated changes in earnings per hour still have this neoclassical implication because gain-loss utility then drops out of a driver's preferences. However, *un*anticipated changes may then have non-neoclassical implications. In particular, when the income target has an important influence on a driver's stopping decision, a driver who values income but is "rational" in the reference-dependent sense of prospect theory will tend to have a negative elasticity of hours with respect to earnings per hour, just as Camerer et al. found.

Section II reports our econometric estimates. In Section II.1 we estimate probit models of the probability of stopping with an index function that is linear in cumulative shift hours and income as in Farber (2005), but splitting the sample according to whether a driver's earnings early in the day are higher or lower than his proxied expectations. This "early earnings" criterion should be approximately uncorrelated with errors in the stopping decision, limiting sample-selection bias. To avoid confounding due to our operationalization of the targets being partly determined by the variables they are used to explain, we proxy drivers' rational point expectations of a day's income and hours, driver/day-of-the-week by driver/day-of-the-week, by their sample averages up to but not including the day in question.

In a neoclassical model, when earnings per hour is serially uncorrelated within a day, it is approximately irrelevant whether early earnings are unexpectedly high or low, because this affects a driver's income but not his expected earnings later in the day, and the income effect is negligible. But in a reference-dependent model, high early earnings make a driver more likely to reach his income target before his hours target, and this has important consequences for behavior. In our estimates drivers' stopping probabilities happen to be more strongly influenced by the second target a driver reaches on a given day than by the first. As a result, when early earnings are high, hours (but not income) has a strong and significant effect on the stopping probability, either because the driver reaches his hours target or because his marginal utility of leisure increases enough to make additional work undesirable. When early earnings are low, this pattern is reversed.[10] Such a reversal is inconsistent with a neoclassical model, in which the targets are

behaviorally reasonable. Farber's (2008) and our treatments of drivers' decisions are both first-order proxies for globally optimal stopping conditions that depend on unobservables, which treatments both yield coherent results, despite their flaws.

[10] Our estimates reverse the patterns of significance from the analogous results in Table 2 of the original version of this paper, Crawford and Meng (2008), suggesting that those results were biased due to the endogeneity of the sample-splitting criterion we used there: whether realized earnings were higher or lower than the full-sample average for a given driver and day-of-the-week.

irrelevant; but it is gracefully explained by a reference-dependent model.[11]

Further, because the elasticity of hours with respect to earnings per hour is substantially negative when the income target is the dominant influence on stopping probability, but near zero when the hours target is dominant, and on a typical day some drivers' earnings are higher than expected and others' lower, KR's distinction between anticipated and unanticipated wage changes can easily reconcile the presumably normally positive incentive to work of an anticipated increase in earnings per hour, with a negative observed aggregate elasticity of hours.[12] Finally, with a distribution of earnings the model can also reproduce Farber's (2005) findings that aggregate stopping probabilities are significantly related to hours but not earnings, but nonetheless respond smoothly to earnings.

In Section II.2 we use the pooled sample to estimate a reduced-form model of the stopping probability, with dummy variables to measure the increments due to hitting the income and hours targets as in Farber's (2008) Table 2 but with our proxied targets instead of Farber's estimated targets. The estimated increments are large and significant, again with a sign pattern strongly suggestive of a reference-dependent model (footnote 11), and the effects of income and hours come mostly from whether they are above or below their targets rather than from levels per se.

In Section II.3 we use the pooled sample to estimate a structural reference-dependent model in the spirit of Farber's (2008) model, again with the changes suggested by KR's theory. In our model the weight of gain-loss utility and the coefficient of loss aversion are not separately identified, but a simple function of them is identified, and its estimated value deviates strongly and significantly from the value implied by a neoclassical model. There is more than enough independent variation of hours and income and our proxies for drivers' targets to identify our model's other behavioral parameters, and to distinguish bunching of realized hours due to targeting from bunching that occurs for conventional neoclassical reasons. The parameter estimates are plausible and generally confirm the conclusions of Section II.1-2's analyses. The

[11] If preferences were homogeneous, as Farber's and our models assume, drivers' stopping probabilities would either all tend to be more strongly influenced by the first target reached on a given day, or all by the second. Thus the pattern of significance in our results is one of the two that are characteristic of a reference-dependent model with homogeneous preferences, and as such is powerful evidence for reference-dependence, even though with heterogeneous preferences other patterns are possible. Kirk Doran (2009), in an important study of yet another group of New York City cabdrivers, with enough data to estimate individual-level effects, finds considerable heterogeneity in drivers' behavior, with some drivers reference-dependent and others neoclassical.

[12] As KR put it (2006, p. 1136): "In line with the empirical results of the target-income literature, our model predicts that when drivers experience unexpectedly high wages in the morning, for any given afternoon wage they are less likely to continue work. Yet expected wage increases will tend to increase both willingness to show up to work, and to drive in the afternoon once there. Our model therefore replicates the key insight of the literature that exceeding a target income might reduce effort. But in addition, it both provides a theory of what these income targets will be, and—through the fundamental distinction between unexpected and expected wages—avoids the unrealistic prediction that generically higher wages will lower effort."

estimated model again implies significant influences of income and hours targets on stopping probabilities, in a pattern that is gracefully explained by a reference-dependent model but inconsistent with a neoclassical model; and resolves the puzzles left open by Farber's analyses.

Our results suggest that reference-dependence is an important part of the labor-supply story in Farber's dataset, and that using KR's model to take it into account does yield a useful model of cabdrivers' labor supply. The key aspect of our analysis, which allows it to avoid Farber's criticism that drivers' estimated targets are too unstable to yield a useful model, is implementing KR's rational-expectations view of drivers' income and hours targets by finding natural sample proxies that limit endogeneity problems, rather than estimating the targets as latent variables.

Section III is the conclusion.

## I. The Model

This section introduces our model of cabdrivers' labor supply decisions.

Treating each day separately as in all previous analyses, consider the preferences of a given driver on a given day.[13] Let $I$ and $H$ denote his income earned and hours worked that day, and let $I^r$ and $H^r$ denote his income and hours targets for the day. We write the driver's total utility, $V(I, H|I^r,H^r)$, as a weighted average of consumption utility $U_1(I) + U_2(H)$ and gain-loss utility $R(I, H|I^r,H^r)$, with weights $1 - \eta$ and $\eta$ (where $0 \leq \eta \leq 1$), as follows:[14]

(1) $$V(I,H \mid I^r,H^r) = (1-\eta)(U_1(I)+U_2(H)) + \eta R(I,H \mid I^r,H^r),$$

where gain-loss utility

(2) $$\begin{aligned} R(I,H \mid I^r,H^r) = {} & 1_{(I-I^r \leq 0)}\lambda(U_1(I)-U_1(I^r)) + 1_{(I-I^r>0)}(U_1(I)-U_1(I^r)) \\ & + 1_{(H-H^r \geq 0)}\lambda(U_2(H)-U_2(H^r)) + 1_{(H-H^r<0)}(U_2(H)-U_2(H^r)). \end{aligned}$$

Because to our knowledge this is the first test of KR's theory, for simplicity and parsimony (1)-(2) incorporate some assumptions KR made provisionally: Consumption utility is additively separable across income and hours, with $U_1(\cdot)$ increasing in $I$, $U_2(\cdot)$ decreasing in $H$, and both concave.[15] Gain-loss utility is also separable, with its components determined by the differences

[13] A driver sometimes works different shifts (day or night) on different days but never more than one a day. Given that drivers seem to form daily targets, it is natural to treat the shift, or equivalently the driver-day combination, as the unit of analysis.

[14] KR (2006, 2007) use a different parameterization, in which consumption utility has weight 1 and gain-loss utility has weight $\eta$. Our $\eta$ is a simple transformation of theirs.

[15] In keeping with the "narrow bracketing" assumption that drivers evaluate consumption and gain-loss utility day by day, $U_1(I)$ is best thought of as a reduced form, including the future value of income not spent today. This suggests that the marginal utility of income is approximately constant and, treating $U_1(\cdot)$ as a von Neumann-Morgenstern utility function, that consumption utility

between realized and target consumption utilities. As in a leading case KR often focus on (their Assumption A3'), gain-loss utility is a linear function of those utility differences, thus ruling out prospect theory's "diminishing sensitivity." Finally, losses have a constant weight $\lambda$ relative to gains, "the coefficient of loss aversion," assumed to be the same for income and hours.[16]

We follow KR in conceptualizing the income and hours targets $I^r$ and $H^r$ as rational expectations, but for simplicity, we assume that they are point expectations.[17] We operationalize them via sample proxies with limited endogeneity problems as explained in Section II. We further assume that the driver is approximately risk-neutral in daily income, so that only its expectation matters to him.

Our model allows a simple characterization of a driver's optimal stopping decision with a target for hours as well as income, which parallels Farber's (2005, 2008) characterization of optimal stopping with income targeting. To simplify this discussion, assume for the moment that the driver has a daily wage in the sense of predetermined daily expected earnings per hour $w^e$ that are constant over time. Further assume that $\lambda \geq 1$, reflecting the almost universal empirical finding that there is loss rather than gain aversion.

The optimal stopping decision then maximizes $V(I, H|I^r,H^r)$ as in (1) and (2), subject to the linear menu of expected income-hours combinations $I = w^eH$. When $U_1(\cdot)$ and $U_2(\cdot)$ are concave, $V(I, H|I^r, H^r)$ is concave in $I$ and $H$ for any targets $I^r$ and $H^r$. Thus the driver's decision is characterized by a first-order condition, generalized to allow kinks at the reference points: He continues if expected earnings per hour exceeds the relevant marginal rate of substitution and stops otherwise.[18] Table 1 lists the marginal rates of substitution in the interiors of the four possible gain-loss regions, expressed as hours disutility costs per unit of income. Under our assumptions that gain-loss utility is additively separable and determined component by component by the difference between realized and target consumption utilities, when hours and income are both in the interior of the gains or loss domain, the marginal rate of substitution is the same as for consumption utilities alone and the stopping decision follows the neoclassical

---

is approximately risk-neutral in daily income, a restriction Farber (2008) and we impose in our structural analyses.

16 This leaves open the question of whether preferences are reference-dependent in both income and hours. Estimates that allow $\lambda$ to differ for income and hours robustly show no significant difference (although the estimated $\lambda$ for hours is always larger than that for income), so in all but Section II.3's structural estimation we assume for simplicity that $\lambda$ is the same for both.

17 This exaggerates the effect of loss aversion, and if anything biases the comparison against a reference-dependent model.

18 If a driver's expected wage varies too much within shift or in response to experience, his optimization problem may become non-convex, in which case optimal stopping requires more foresight than we assume. Further, more general specifications that allow diminishing sensitivity do not imply that $V(I, H|I^r, H^r)$ is everywhere concave in $I$ and $H$. Although they probably still allow an analysis like ours, as do other expectations formation rules, we avoid these complications.

first-order condition. But when hours and income are in the interiors of opposite domains, the marginal rate of substitution equals the consumption-utility trade-off times a factor that reflects the weight of gain-loss utility and the coefficient of loss aversion, $(1 - \eta + \eta\lambda)$ or $1/(1 - \eta + \eta\lambda)$. On boundaries between regions, where $I = I^r$ and/or $H = H^r$, the marginal rates of substitution are replaced by generalized derivatives whose left- and right-hand values equal the interior values.

**Table 1. Marginal Rates of Substitution with Reference-Dependent Preferences by Domain**

| | **Hours gain ($H < H^r$)** | **Hours loss ($H > H^r$)** |
|---|---|---|
| **Income gain ($I > I^r$)** | $-U_2'(H)/U_1'(I)$ | $-[U_2'(H)/U_1'(I)][1-\eta+\eta\lambda]$ |
| **Income loss ($I < I^r$)** | $-[U_2'(H)/U_1'(I)]/[1-\eta+\eta\lambda]$ | $-U_2'(H)/U_1'(I)$ |

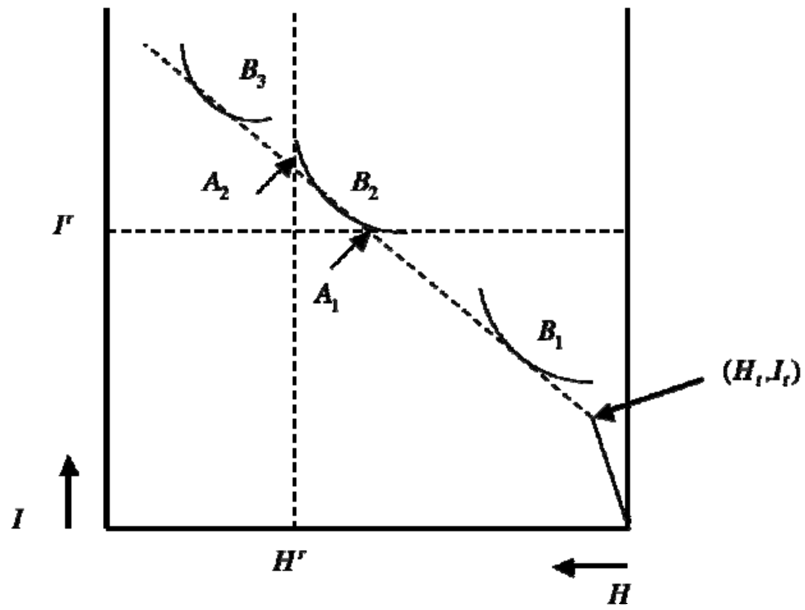


**Figure 1: A Reference-dependent Driver's Stopping Decision when Realized Earnings are Higher than Expected**

Figure 1, in which hours are measured negatively as a "bad," illustrates the driver's optimal stopping decision when, after an initial blip of higher than expected realized earnings, both realized and expected earnings per hour, $w^a$ and $w^e$, remain constant and equal. As a result of the blip, total realized earnings remain higher than initially expected, and the income target is reached before the hours target. We stress that the constancy of $w^e$ and $w^a$ and the fact that there are no surprises after the blip are only for illustration; the important thing is that total realized earnings remain higher than expected. The case where realized earnings are lower than expected and the hours target is reached before the income target is completely analogous.

Letting $I_t$ and $H_t$ denote earnings and hours by the end of trip $t$, the driver starts in the lower right-hand corner with $(H_t, I_t) = (0, 0)$, followed by an initial period of higher than expected realized earnings. Total earnings and hours then increase along a weakly monotone path (not shown), heading northwest. The path is actually a step function, but because mean trip length is only 12 minutes (Farber (2005, Section V)), the path can be treated as smooth and $I$ and $H$ as continuous variables. After any given trip $t$, the driver anticipates moving along a line $I = w^e H$, starting from the current $(H_t, I_t)$. As hours and income accumulate, a driver who continues working passes through a series of domains such that the hours disutility cost of income weakly increases, whichever target is reached first—a reflection of the concavity of $V(I, H|I^r, H^r)$ in $I$ and $H$. The driver considers stopping after each trip, stopping (ignoring option value) when his current expected wage first falls below his current hours disutility cost of income. This myopia may lead the driver to deviate from KR's preferred personal equilibrium (footnote 6), although this can matter only in our structural estimation. The driver stops at a point that appears globally optimal to him, given his myopic expectations. This conclusion extends to drivers who form their expectations in more sophisticated ways, unless their expected earnings vary too much.

For example, in the income-loss/hours-gain ($I_t < I^r$, $H_t < H^r$) domain, the hours disutility cost of income is $-[U_2'(H_t)/U_1'(I_t)]/[1-\eta+\eta\lambda]$ from the lower left cell of Table 1. Because in this domain hours are cheap relative to income ($(1 - \eta + \eta\lambda) \geq 1$ when $0 \leq \eta \leq 1$ and $\lambda \geq 1$), the comparison with expected earnings per hour favors working more than the neoclassical comparison. The indifference curves in Figure 1 with tangency points $B_1$, $B_2$, and $B_3$ represent alternative possible income-hours trade-offs for consumption utility, ignoring gain-loss utility. If a driver stops in the income-loss/hours-gain domain, it will be (ignoring discreteness) at a point weakly between $B_1$ and $A_1$ in the figure, where $B_1$ maximizes consumption utility on indifference curve 1 subject to $I = w^e H$ and $A_1$ represents the point where the income target is reached. (The closer $\eta$ is to one and the larger is $\lambda \geq 1$, other things equal, the closer the stopping point is to $A_1$.)

Figure 2 compares labor-supply curves for a neoclassical and a reference-dependent driver with the same consumption utility functions. The solid curve is the neoclassical supply curve, and the dashed curve is the reference-dependent one. The shape of the reference- dependent curve depends on which target has a larger influence on the stopping decision, which depends on the relation between the neoclassical optimal stopping point (that is, for consumption utility alone) and the targets. Figure 2 illustrates the case suggested by Section II's estimates: For wages that reconcile the income and hours targets as at point D, the neoclassically optimal income and hours are higher than the targets, so the driver stops at his second-reached target. When the wage is to the left of D, the hours target is reached before the income target, and vice versa.

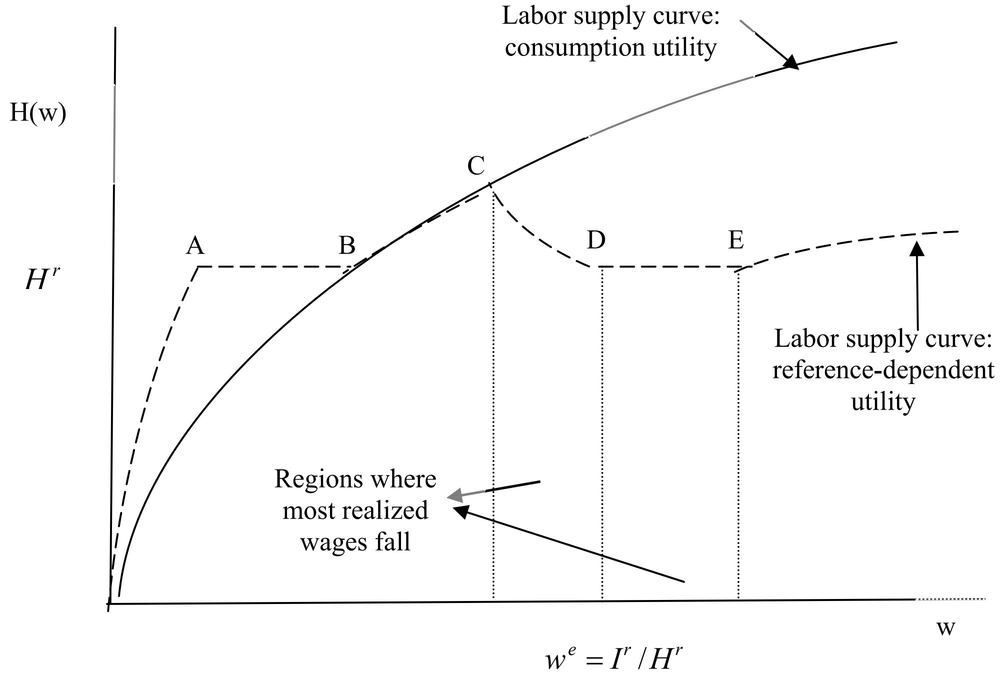


**Figure 2: A Reference-dependent Driver's Labor Supply Curve**

As Figure 2 illustrates, reference-dependent labor supply is non-monotonic. When expected earnings per hour is very low, to the left of point A, the higher cost of income losses raises the incentive to work above its neoclassical level (Table 1's lower left-hand cell). Along segment AB labor supply is determined by the kink at the hours target, which is reached first. Along segment BC the neoclassical optimal stopping point is above the hours but below the income target, so the gain-loss effects cancel out, and reference-dependent and neoclassical labor supply coincide (Table 1's lower right-hand cell). Along segment CD labor supply is determined by the kink at the income target, which is reached second, so that the elasticity of hours with respect to expected earnings per hour is negative. Along segment DE labor supply is determined by the kink at the hours target, which is reached second. (Recall that point D is defined by the wage that is just high enough to reverse which target the driver reaches first.) Finally, when expected earnings per hour is very high, to the right of point E, the higher cost of hours losses lowers the incentive to work below its neoclassical level (Table 1's upper right-hand cell). Most realized earnings fall close to point D, either along segment CD where hours decrease with increases in expected earnings per hour because of income targeting, or along segment DE where hours do not change with increases in expected earnings per hour because of hours targeting.[19]

## II. Econometric Estimates

This section reports econometric estimates of our reference-dependent model of cabdrivers' labor supply. We use Farber's (2005, 2008) data and closely follow his econometric strategies, but instead of treating drivers' targets as latent variables, we treat them as rational expectations and operationalize them via sample proxies with limited endogeneity problems.[20]

Here and in the rest of our econometric analyses, we proxy drivers' point-expectation income and hours targets**,** driver/day-of-the-week by driver/day-of-the-week, via the analogous sample averages up to but not including the day in question, ignoring sampling variation for simplicity.[21] This avoids confounding from including the current shift's income and hours in the averages, while allowing the targets to vary across days of the week as suggested by the variation of hours and income (footnote 5). This way of proxying the targets loses observations from the first day-of-the-week shift for each driver because there is no prior information for those shifts.[22]

---

[19] There are two possible alternatives to the situation depicted in Figure 2. In the first, for earnings that reconcile the income and hours targets, the neoclassical optimal income and hours are lower than the targets, so the driver stops at his first-reached target. This case yields conclusions like Figure 2's with some differences in the details. In the second case, the neoclassical optimal income and hours exactly equal the targets, as in KR's preferred personal equilibrium. In that case, near where most realized earnings per hour fall, stopping would be completely determined by the hours target and the income target would have no effect. Thus, our point-expectations version of preferred personal equilibrium is inconsistent with what we find in Farber's data. This does not prove that KR's distributional preferred personal equilibrium would also be inconsistent, but we suspect it would.

[20] Farber generously shared his data with us, now posted at http://www.e-aer.org/data/june08/20030605_data.zip. His 2005 paper gives a detailed description of the data cleaning and relevant statistics. The data are converted from trip sheets recorded by the drivers. These contain information about starting/ending time/location and fare (excluding tips) for each trip. There are in total 21 drivers and 584 trip sheets, from June 2000 to May 2001. Drivers in the sample all lease their cabs weekly so they are free to choose working hours on a daily basis. Because each driver's starting and ending hours vary widely, and 11 of 21 work some night and some day shifts, subleasing seems unlikely. Farber also collected data about weather conditions for control purposes.

[21] There is some risk of bias in ignoring sampling variation, because sampling error tends to be larger early in the sample period. We take this into account by computing estimates with weights equal to the number of realizations (rescaled to sum to the number of observations in each subsample) that are averaged to calculate the expectations. The results are essentially the same as without weighting, with one exception: In Table 3's estimates, unweighted estimates would yield an income target that is not significant when we do not distinguish day-of-the-week differences (column 2), but weighted estimates make this parameter significant.

[22] For this reason, we cannot make the sample exactly the same as Farber's, who used only the drivers with a minimum of ten shifts. Strictly speaking, our working hypothesis of rational expectations would justify using averages both prior to and after the shift in question (but still excluding the shift itself). This loses fewer observations, but using only prior sample averages is more plausible and yields somewhat cleaner results. The results are similar using averages after as well as before the shift in question. The average within-driver standard deviation of the income target proxies is $34 and that of the hours target proxies is 1.62 hours. Since for most dates there are only a few driver records, we calculate average across-driver standard deviations day-of-the-week

This is a nonnegligible fraction of the total number of observations (3124 out of 13461). But because the criterion for censoring is exogenous and balanced across days of the week and drivers, it should not cause significant bias. When necessary we proxy a driver's expected earnings during the day in the same way, by sample averages, driver/day-of-the-week by driver/day-of-the-week, up to but not including the day in question. This is a noisy proxy, but it is not systematically biased and because it is predetermined it should not cause endogeneity bias.

### II.1 Probit models of the probability of stopping with a linear index function

We begin by estimating probit models of the probability of stopping with an index function that is linear in cumulative shift hours and cumulative shift income as in Farber (2005), but splitting the sample shift by shift according to whether a driver's earnings for the first $x$ hours of the day (or equivalently, average earnings for the first $x$ hours, but with no need for the average to be constant or independent of history) are higher or lower than his proxied expectations. In estimation we include only observations with cumulative working hours higher than $x$.

The higher a driver's early earnings, the more likely he is to hit his income target first, simply because early earnings is part of total earnings and can be viewed as a noisy estimate of it. For a wide class of reference-dependent models, including our structural model, a driver's probability of stopping increases at his first-reached target and again (generally by a different amount) at his second-reached target. By contrast, in a neoclassical model, the targets have no effect. This difference is robust to variations in the specification of the targets and the details of the structural specification. Sample-splitting therefore allows a robust assessment of the evidence for reference-dependence, avoiding most of the restrictions needed for structural estimation.

In our model as in Farber's, drivers choose only hours, not effort. Thus early earnings, unlike total earnings, should be approximately uncorrelated with errors in the stopping decision, and so should avoid most problems of sample selection via endogenous variables.

The larger is $x$ the more accurate the split, but we lose the first $x$ hours of observations from each shift, a nonnegligible fraction of the sample if $x$ is large, risking censoring bias. However, if $x = 1$ we lose only 4 shifts (10 trips) out of a total of 584 shifts, so any bias should be small. We report estimates for $x = 1$, but the results are qualitatively robust to values of $x$ up to $x = 5$.[23]

by day-of-the-week, then average across days-of-the-week. The average standard deviation is $37 for the income target proxies and 2.68 hours for the hours target proxies. Thus, the variation across drivers is indeed larger than that within drivers.

[23] When $x > 5$ the sign pattern of estimated coefficients is preserved, but the coefficients are no longer significantly different than

Table 2 reports marginal probability effects to maximize comparability with Farber's estimates, but with significance levels computed for the underlying coefficients. In each numbered panel, the left-hand column uses the same specification as Farber's (2005) pooled-sample estimates, but with observations deleted as in our split-sample estimates. The center and right-hand columns report our split-sample estimates.

In the left-hand panel, only income and total hours are used to explain the stopping probability.[24] In the pooled-sample estimates with these controls, both coefficients have the expected signs, the effect of hours is significant at the 1% level, and the effect of income is significant at the 10% level. In our split-sample estimates with only these controls, the effect of hours is large and significant whether or not early earnings are higher or lower than expected, but the effect of income is insignificant in either case.

In the right-hand panel we control for driver heterogeneity, day-of-the-week, hour of the day, weather, and location. In the pooled sample this yields estimates like those in the left-hand panel, except that the effect of income is now insignificant even at the 10% level. But in the split-sample estimates with this full set of controls, the effect of hours but not that of income is significant at the 1% level when early earnings are higher than expected, while the effect of income is insignificant even at the 10% level; but the effect of income is significant at the 5% level when early earnings are lower than expected, while the effect of hours is insignificant even at the 10% level.

This reversal of the pattern of significant coefficients depending on whether early earnings are higher than expected is inconsistent with a neoclassical model, but is gracefully explained by a reference-dependent model in which stopping probability is usually more strongly influenced by the second target a driver reaches than the first, as in Figure 2 (footnote 11). Specifically, if the second target reached on a given day normally has the stronger influence, then on good days, when the income target is reached before the hours target, hours has a stronger influence on stopping probability, as in the *** coefficient in the first row of the right-hand panel of Table 2 in the column headed "first hour's earnings > expected". On bad days income then has the stronger influence, as in the ** coefficient in the second row of the right-hand panel. By contrast, if the first target reached on a given day usually had the stronger influence, the pattern of

---

0 in most cases, possibly because of the smaller sample size and censoring bias.

[24] Here we follow Farber (2008) rather than Farber (2005) in using total hours rather than hours broken down into driving hours, waiting hours and break hours, which makes little difference to the results.

significant coefficients would again reverse depending on whether early earnings are higher than expected, but now with a significant influence of income on good days and of hours on bad days. If the population were homogeneous in preferences these would be the only two possible cases, and in that sense the pattern we see is one of two that are characteristic of a reference-dependent model with hours- as well as income- targeting. With heterogeneous preferences other patterns of significance are logically possible, but more "contrived" and so less plausible.

**Table 2: Marginal Effects on the Probability of Stopping: Probit Estimation with Split Samples**

| | | | (1) | | | (2) | |
|---|---|---|---|---|---|---|---|
| | Evaluation Point for Marginal Effect | Pooled data | First hour's earnings > expected | First hour's earnings < expected | Pooled data | First hour's earnings> expected | First hour's earnings < expected |
| Cumulative total hours | 8.0 | .020***<br>(.006) | 0.022***<br>(0.006) | 0.022***<br>(0.008) | 0.009***<br>(0.003) | 0.028***<br>(0.010) | 0.005<br>(0.004) |
| Cumulative Income/100 | 1.5 | 0.035*<br>(.016) | 0.021<br>(0.019) | 0.021<br>(0.027) | 0.020<br>(0.014) | 0.035<br>(0.031) | 0.037**<br>(0.025) |
| Min temp<30 | 0.0 | - | - | - | 0.004*<br>(0.008) | 0.014<br>(0.023) | 0.002<br>(0.010) |
| Max temp>80 | 0.0 | - | - | - | -0.017*<br>(0.010) | -0.004<br>(0.038) | -0.014<br>(0.013) |
| Hourly rain | 0.0 | - | - | - | 0.011<br>(0.164) | -0.945<br>(0.720) | 0.127<br>(0.139) |
| Daily snow | 0.0 | - | - | - | -0.001<br>(0.005) | -0.003<br>(0.010) | -0.028<br>(0.106) |
| Downtown | 0.0 | - | - | - | 0.002<br>(0.008) | 0.005<br>(0.023) | 0.010<br>(0.013) |
| Uptown | 0.0 | - | - | - | -0.002<br>(0.006) | -0.009<br>(0.018) | -0.002<br>(0.008) |
| Bronx | 0.0 | - | - | - | 0.072<br>(0.071) | 0.000<br>(0.075) | 0.056<br>(0.087) |
| Queens | 0.0 | - | - | - | 0.045<br>(0.045) | 0.290<br>(0.188) | 0.044<br>(0.061) |
| Brooklyn | 0.0 | - | - | - | 0.088***<br>(0.041) | 0.187**<br>(0.098) | 0.080**<br>(0.059) |
| Kennedy Airport | 0.0 | - | - | - | 0.076***<br>(0.040) | 0.133**<br>(0.076) | -0.006<br>(0.019) |
| LaGuardia Airport | 0.0 | - | - | - | 0.073***<br>(0.037) | 0.185**<br>(0.138) | 0.001<br>(0.024) |
| Other | 0.0 | - | - | - | 0.148***<br>(0.084) | 0.028**<br>(0.010) | 0.224**<br>(0.189) |
| Drivers (21) | | No | No | No | Yes | Yes | Yes |
| Day of week (7) | | No | No | No | Yes | Yes | Yes |
| Hour of day (19) | 2:00 p.m. | No | No | No | Yes | Yes | Yes |
| Log likelihood | | -1550.452 | -803.93123 | -722.27398 | -1344.8812 | -679.48626 | -607.45459 |
| Pseudo $R^2$ | | 0.1239 | 0.1186 | 0.1278 | 0.2401 | 0.2550 | 0.2664 |
| Observation | | 8958 | 4664 | 4294 | 8958 | 4664 | 4294 |

Note: Standard errors are computed for the marginal effects to maximize comparability with Farber's estimates, but with significance levels computed for the underlying coefficients rather than the marginal effects: *10%, **5%, ***1%. Robust standard errors clustered by shift are assumed. The subsample estimation weights each observation based on the number of realizations in the history (rescaled to sum to the number of observations in each subsample) used to calculate the proxies for expectations (see footnote 21; results for the unweighted estimation are reported in online appendix A, Table A1). We use Farber's evaluation point: after 8 total working hours and $150 earnings on a dry day with moderate temperatures in midtown Manhattan at 2:00 p.m. Driver fixed effects and day of week dummies are equally weighted. For dummy variables, the marginal effect is calculated by the difference between values 0 and 1. Following Farber's suggestion, we do not distinguish between driving hours and wait hours between fares. Among the dummy control variables, only driver fixed effects, hour of the day, day of the week, and certain location controls have effects significantly different from 0.

To put these results into perspective, recall that a neoclassical model would predict that hours have an influence on the probability of stopping that varies smoothly with realized income, without regard to whether income is higher than expected. A pure income-targeting model as in Farber (2008) would predict a jump in the probability of stopping when the income target is reached, but an influence of hours that again varies smoothly with realized income. Our estimates are inconsistent with a neoclassical model and—because the effect of hours is significant when income is higher than expected but insignificant when income is lower than expected—with Farber's income-targeting model.[25] By contrast, our estimates are consistent with our reference-dependent model if the probability of stopping is more strongly influenced by hours when early earnings are higher than expected but by income when lower than expected.

We note again that because the wage elasticity is substantially negative when the income target is the dominant influence on stopping but near zero when the hours target is dominant, the reference-dependent model's distinction between anticipated and unanticipated wage changes can reconcile an anticipated wage increase's positive incentive to work with a negative aggregate wage elasticity of hours. Finally, with a distribution of realized wages, the model can also reproduce Farber's (2005) findings that aggregate stopping probabilities are significantly related to hours but not realized earnings, and that they respond smoothly to earnings.

### II.2 Reduced-form estimates of the probability of stopping

We now estimate a reduced-form model of stopping probability, with dummy variables to measure the increments due to hitting the income and hours targets as in Farber's (2008) Table 2, but with the sample proxies for targets introduced above instead of Farber's estimated targets.

Table 3 reports reduced-form estimates of the increments in stopping probability on hitting the estimated income and hours targets. The estimated coefficients of dummy variables indicating whether earnings or hours exceeds the targets are positive, the sign predicted by a reference-dependent model, and significantly different from 0. The estimates confirm and extend the results from our split-sample probits, in that the significant effects of income and hours come mainly from whether they are above or below their targets rather than from their levels. The level

[25] When the utility cost of hours is highly nonlinear, drivers' neoclassical utility-maximizing choices resemble hours targeting. But neoclassical drivers should still have positive wage elasticity, in contrast to the zero elasticity implied by hours targeting. Further, Section II.3's structural model can closely approximate a neoclassical model with inelastic labor supply, but there is clear evidence that the hours bunching in the sample follows targets that vary by day-of-the-week in a way that is ruled out by a neoclassical model.

of income has a slightly negative, insignificant effect and the level of hours has a positive, significant effect. In this respect the estimates suggest that hours have a nonnegligible neoclassical effect as well as their reference-dependent effect.

**Table 3: Marginal Effects on the Probability of Stopping: Reduced-Form Model Allowing Jumps at the Targets**

| | | Using driver specific sample average income and hours prior to the current shift as targets | | Using driver and day-of-the-week specific sample average income and hours prior to the current shift as targets | |
|---|---|---|---|---|---|
| | Evaluation point for marginal effect | (1) | (2) | (3) | (4) |
| Cumulative total hours>hours target | | 0.036***<br>(0.013) | 0.030***<br>(0.022) | 0.055***<br>(0.016) | 0.038***<br>(0.030) |
| Cumulative income > income target | | 0.058***<br>(0.018) | 0.020*<br>(0.017) | 0.049***<br>(0.017) | 0.011**<br>(0.011) |
| Cumulative total hours | 8.0 | 0.011***<br>(0.005) | 0.007**<br>(0.006) | 0.012***<br>(0.004) | 0.003**<br>(0.003) |
| Cumulative Income/100 | 1.5 | -0.010<br>(0.015) | 0.010<br>(0.016) | -0.012<br>(0.013) | 0.003<br>(0.007) |
| Weather (4) | | No | Yes | No | Yes |
| Locations (9) | | No | Yes | No | Yes |
| Drivers (21) | | No | Yes | No | Yes |
| Days of the week (7) | | No | Yes | No | Yes |
| Hour of the day (19) | 2:00 p.m. | No | Yes | No | Yes |
| Log likelihood | | -1526.9354 | -1367.8075 | -1493.3419 | -1315.2337 |
| Pseudo $R^2$ | | 0.1597 | 0.2472 | 0.1756 | 0.2740 |
| Observation | | 10337 | 10337 | 10337 | 39651 |

Note: Standard errors are computed for the marginal effects to maximize comparability with Farber's estimates, but with significance levels computed for the underlying coefficients rather than the marginal effects: *10%, **5%, ***1%. Robust standard errors clustered by shift are assumed. The estimation weights each observation based on the number of realizations in the history (rescaled to sum to the number of observations in each estimation) used to calculate proxies for expectations (see footnote 21; results for the unweighted estimation are reported in online appendix A, Table A2). We use Farber's evaluation point: after 8 total working hours and $150 earnings on a dry day with moderate temperatures in midtown Manhattan at 2:00 p.m. Driver fixed effects and day of week dummies are equally weighted. For dummy variables, the marginal effect is calculated by the difference between values 0 and 1. As in Farber (2008) (but no Farber (2005), we do not distinguish between driving hours and waiting hours between fares. Among the dummy control variables, only driver fixed effects, hour of the day, day of the week, and certain location controls have effects significantly different from 0.

### II.3 Structural estimation

We now estimate Section I's structural model. Our structural model makes no sharp general predictions. In particular, whether the aggregate stopping probability is more strongly influenced by income or hours depends on estimated parameters and how many shifts have realized income higher than expected. Even so, structural estimation is an important check on the model's ability to give a useful account of drivers' labor supply.

We use the same sample proxies for drivers' targets as before, and we take a driver's expectations about earnings during the day as predetermined rational expectations, proxied by sample averages, driver/day-of-the-week by driver/day-of-the-week, up to but not including the day in question. This proxy is noisy, but it is not a source of endogeneity or other bias.

Section I explains the model. In the structural estimation, as in Farber (2008), we impose the further assumption that consumption utility has the functional form $U(I,H) = I - \frac{\theta}{1+\rho} H^{1+\rho}$, where $\rho$ is the elasticity of the marginal rate of substitution. Thus, the driver has constant marginal utility of income (and is risk-neutral in it, treating $U(\cdot)$ as a von Neumann-Morgenstern utility function), in keeping with the fact that income is storable and the day is a small part of his economic life. However, he is averse to hours as in a standard labor supply model.

Substituting this functional form into (1)-(2) yields:

$$(3)\quad V(I,H \mid I^r,H^r) = (1-\eta)\left[I - \frac{\theta}{1+\rho}H^{1+\rho}\right] + \eta\left[1_{(I-I^r \le 0)}\lambda(I-I^r) + 1_{(I-I^r>0)}(I-I^r)\right]$$
$$-\eta\left[1_{(H-H^r \ge 0)}\lambda\left[\frac{\theta}{1+\rho}H^{1+\rho} - \frac{\theta}{1+\rho}(H^r)^{1+\rho}\right]\right] - \eta\left[1_{(H-H^r<0)}\left[\frac{\theta}{1+\rho}H^{1+\rho} - \frac{\theta}{1+\rho}(H^r)^{1+\rho}\right]\right].$$

Like Farber, we assume that the driver decides to stop at the end of a given trip if and only if his anticipated gain in utility from continuing work for one more trip is negative. Again letting $I_t$ and $H_t$ denote income earned and hours worked by the end of trip $t$, this requires:

$$(4)\qquad E[V(I_{t+1}, H_{t+1}|I^r,H^r)] - V(I_t, H_t|I^r,H^r) + x_t\beta + c + \varepsilon < 0,$$

where $I_{t+1} = I_t + E(f_{t+1})$ and $H_{t+1} = H_t + E(h_{t+1})$, $E(f_{t+1})$ and $E(h_{t+1})$ are the next trip's expected fare and time (searching and driving), $x_t\beta$ include the effect of control variables, $c$ is the constant term, and $\varepsilon$ is a normal error with mean zero and variance $\sigma^2$. We estimate a non-zero constant term to avoid bias, even though theory suggests $c = 0$.

Online Appendix B gives the details of deriving the likelihood function

$$(5)\quad \sum_{i=1}^{584}\sum_{t=i}^{T_i} \ln \Phi[((1-\eta+\eta\lambda)a_{1,it} + a_{2,it} - (1-\eta+\eta\lambda)\frac{\theta}{\rho+1}b_{1,it}(\rho) - \frac{\theta}{\rho+1}b_{2,it}(\rho) + x_t\beta + c)/\sigma],$$

where $i$ refers to the shift and $t$ to the trip within a given shift, and $a_{1,it}, a_{2,it}, b_{1,it}(\rho)$, and $b_{2,it}(\rho)$ are shorthands for components of the right-hand side of (3), as explained in Online Appendix B.

Here, unlike in a standard probit model, $\sigma$ is identified through $a_{2,it}$, which represents the change in income "gain" relative to the income target. However, as is clear from the likelihood function, $\eta$ and $\lambda$ cannot be separately identified: Only $1 - \eta + \eta\lambda$, the factor by which (directly or inversely) the reference-dependent marginal rate of substitution differs from the neoclassical marginal rate of substitution (Table 1) is identified. If $1 - \eta + \eta\lambda = 1$, or equivalently $\eta(\lambda - 1) = 0$, the model reduces to a neoclassical model. This happens trivially if $\eta = 0$ so there is no weight on gain-loss utility, or if $\eta \neq 0$ but $\lambda = 1$ so gains and losses are weighted equally. If $\eta = 1$ the model has only gain-loss utility as was usually assumed before KR (2006), and $1 - \eta + \eta\lambda = \lambda$. In that sense our estimates of $1 - \eta + \eta\lambda$ are directly comparable to most estimates of the coefficient of loss aversion that have been reported in the literature.

Table 4 reports structural estimates, expanded to identify the effects of different proxies and the reasons for the differences between our and Farber's (2008) results, and to allow different coefficients of loss aversion, $\lambda_H$ and $\lambda_I$, for hours and income. Column 1's baseline model yields plausible parameter estimates that confirm and refine the conclusions of Section II.1-2's analyses. For both $\lambda_H$ *and* $\lambda_I$, the null hypothesis that $\eta(\lambda - 1) = 0$ is rejected at the 1% level, ruling out the restrictions $\eta = 0$ or $\lambda = 1$that would reduce the model to a neoclassical model.[26] For both $\lambda_H$ *and* $\lambda_I$, the implied estimate of $1 - \eta + \eta\lambda$ $(= 1 + \eta(\lambda - 1))$ is comparable to most reported estimates of the coefficient of loss aversion. The hypothesis that $\lambda_H = \lambda_I$ cannot be rejected, although the estimated $\lambda_H$ robustly exceeds $\lambda_I$.

Columns 2-5 change one thing at a time from the baseline. Column 2 confirms the robustness of Column 1's results to basing targets on sample proxies after as well as before the current shift (but still omitting the current shift; see footnote 22). Column 3 confirms the robustness of Column 1's results to more sophisticated earnings forecasting, via a model of next-trip fare/time expectations using the 3124 observations omitted from the first shifts for each day-of-the-week for each driver, and estimated using the current sample.[27] Column 4 suggests that Column 1's results are *not* robust to ruling out day-of-the-week differences as in Farber (2008): This restriction obscures the effects of reference-dependence, in that the effects of the targets become smaller and in one case significant only at the 10% level. By contrast, Column 5 suggest that Column 1's results are robust to Farber's (2008) restriction to income- but not hours-targeting.

[26] The estimated standard errors suggest that $\eta(\lambda-1)$ is not significantly different from zero in most specifications based on the Wald Test. Here we use likelihood ratio tests, which give results somewhat different from the Wald Test. There are at least two reasons why the likelihood ratio test might give different results: First, some parameter transformations are needed to facilitate numerical estimation, and the likelihood ratio test is invariant to such transformations under maximum likelihood estimation, but the Wald Test is not invariant. Second, although both test statistics converge to the Chi-square distribution asymptotically, for small samples the likelihood ratio test statistic is closer to the Chi-square distribution used for inference. Because our sample size is quite large, the first reason is probably the more important one.

[27] The other variables include day-of-the-week, hour-of-the-day, locations at the end of the trip, and weather controls. Surprisingly, there is not much variation by time of day, but there is a lot of variation across locations. Online Appendix C, Table C1 reports the trip fares and time estimates whose fitted values are used to proxy drivers' expectations in those models.

**Table 4: Structural Estimates under Alternative Specifications of Expectations**

| | (1) Use driver and day-of-the-week specific sample averages prior to the current shift as the income/hours targets and the next-trip earnings/times expectation | (2) Use driver and day-of-the-week specific sample averages prior and after the current shift as the income/hours targets and next-trip the earnings/times expectation | (3) Use driver and day-of-the-week specific sample averages prior to the current shift as the income/hours targets and fit the sophisticated next-trip earnings/time expectation | (4) Use driver (without day-of-the-week difference) specific sample averages prior to the current shift as income/hours targets and the next-trip earnings/time expectation | (5) Income target only: use driver and day-of-the-week specific sample averages prior to the current shift as income target and next-trip earnings/time expectation |
|---|---|---|---|---|---|
| $\eta(\lambda_H - 1)$ | 1.309*** | 1.886*** | 0.671*** | 0.188*** | - |
| [p-value] | [0.000] | [0.000] | [0.000] | [0.001] | |
| $\eta(\lambda_I - 1)$ | 0.512*** | 0.299** | 0.256*** | 0.111* | 2.007*** |
| [p-value] | [0.001] | [0.041] | [0.002] | [0.057] | [0.000] |
| $\theta$ | 0.035*** | 0.017*** | 0.043*** | 0.152*** | 0.018*** |
| [p-value] | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] |
| $\rho$ | 0.597*** | 0.782*** | 0.566*** | 0.212*** | 1.407*** |
| [p-value] | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] |
| $\sigma^+$ | 0.127 | 0.117 | 0.072 | 0.045 | 0.286 |
| [p-value] | [0.253] | [ 0.104] | [0.996] | [0.280] | [0.484] |
| $c$ | -0.047 | 0.014 | -0.045 | 0.029 | -0.036 |
| [p-value] | [0.710] | [0.929] | [0.825] | [0.755] | [0.902] |
| Test $\lambda_H = \lambda_I$ [p-value] | [0.243] | [0.112] | [0.997] | [0.666] | - |
| Observations | 10337 | 10337 | 10337 | 10337 | 10337 |
| Log-likelihood | -1321.1217 | -1326.3005 | -1312.8993 | -1367.2374 | -1333.0964 |

Notes: Significance levels *10%, **5%, ***1%. We perform likelihood ratio tests on each estimated parameter and indicate the corresponding $p$-values and significance levels. The null-hypothesis is that each parameter equals zero except for the variance estimate where we test $\sigma = 1$. The estimation weights each observation based on the number of realizations in the history (rescaled to sum to the number of observations in each estimation) used to calculate proxies for expectations (see footnote 21; results for the unweighted estimation are reported in online appendix A, Table A3). Control variables include driver fixed effects (18), day of week (6), hour of day (18), location(8), and weather (4).

Table 4's five models all have the same number of parameters except for column 5, which has no loss aversion coefficient for the hours target: a constant term, five structural parameters, and 55 controls.[28] Column 3's model, with drivers sophisticated enough to predict future wages based on location, clock hours, etc., fits best. Of the remaining four models, all with constant expectations throughout the shift, Column 1's model, the baseline, fits best. The likelihood cost of ruling out sophisticated earnings forecasting is nontrivial, though this does not seem to distort the parameter estimates much. Despite Column 5's robustness result, the likelihood cost of ruling out hours-targeting is also nontrivial, as is that of ruling out day-of-the-week differences.

[28] Our proxies for targets and trip-level expectations are either calculated as sample averages or as predicted values with coefficients estimated out of sample, and this choice does not affect the number of parameters. Although Farber (2008) argues that a reference-dependent model has too many degrees of freedom to be fairly compared with a neoclassical model —a loss aversion coefficient and heterogeneous income targets—defining the targets as rational expectations reduces the difference.

| Table 5: Estimated Optimal Stopping Times (in Hours) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Percentile in the earnings per hour distribution | Hourly earnings | (1) Use driver and day-of-the-week specific sample averages prior to the current shift as the income/hours targets and the next-trip earnings/times expectation | | (2) Use driver and day-of-the-week specific sample averages prior and after the current shift as the income/hours targets and next-trip the earnings/times expectation | | (3) Use driver and day-of-the-week specific sample averages prior to the current shift as the income/hours targets and fit the sophisticated next-trip earnings/time expectation | | (4) Use driver (without day-of-the-week difference) specific sample averages prior to the current shift as income/hours targets and the next-trip earnings/time expectation | |
| | | $\theta = 0.035$<br>$\rho = 0.597$<br>$\eta(\lambda_H - 1) = 1.309$<br>$\eta(\lambda_I - 1) = 0.512$ | | $\theta = 0.017$<br>$\rho = 0.782$<br>$\eta(\lambda_H - 1) = 1.886$<br>$\eta(\lambda_I - 1) = 0.299$ | | $\theta = 0.043$<br>$\rho = 0.566$<br>$\eta(\lambda_H - 1) = 0.671$<br>$\eta(\lambda_I - 1) = 0.256$ | | $\theta = 0.152$<br>$\rho = 0.212$<br>$\eta(\lambda_H - 1) = 0.188$<br>$\eta(\lambda_I - 1) = 0.111$ | |
| | | Neoclassical optimal working hours | Reference-dependent optimal working hours | Neoclassical optimal working hours | Reference-dependent optimal working hours | Neoclassical optimal working hours | Reference-dependent optimal working hours | Neoclassical optimal working hours | Reference-dependent optimal working hours |
| 10% | \$16.7 | 13.70 | $9.58^{I}$ | 18.57 | $9.58^{I}$ | 10.99 | $9.58^{I}$ | 1.56 | 3.51 |
| 20% | \$18.4 | 16.11 | $8.70^{I}$ | 21.02 | $8.70^{I}$ | 13.05 | $8.70^{I}$ | 2.46 | 5.55 |
| 30% | \$19.4 | 17.61 | $8.25^{I}$ | 22.50 | $8.25^{I}$ | 14.32 | $8.25^{I}$ | 3.16 | 7.12 |
| 40% | \$20.3 | 19.00 | $7.88^{I}$ | 23.84 | $7.88^{I}$ | 15.52 | $7.88^{I}$ | 3.91 | $7.88^{I}$ |
| 50% | \$21.3 | 20.59 | $7.80^{H}$ | 25.35 | $7.80^{H}$ | 16.90 | $7.80^{H}$ | 4.91 | $7.80^{H}$ |
| 60% | \$22.0 | 21.74 | $7.80^{H}$ | 26.42 | $7.80^{H}$ | 17.89 | $7.80^{H}$ | 5.72 | $7.80^{H}$ |
| 70% | \$22.8 | 23.08 | $7.80^{H}$ | 27.66 | $7.80^{H}$ | 19.05 | $7.80^{H}$ | 6.77 | $7.80^{H}$ |
| 80% | \$23.8 | 24.80 | $7.80^{H}$ | 29.22 | $7.80^{H}$ | 20.55 | 8.30 | 8.29 | $7.80^{H}$ |
| 90% | \$25.3 | 27.48 | $7.80^{H}$ | 31.59 | 8.10 | 22.90 | 9.20 | 11.06 | $7.80^{H}$ |
| Correlation of earnings per hour and optimal working hours | | 0.99 | -0.83 | 0.99 | -0.75 | 0.99 | -0.27 | 0.97 | 0.80 |

Note: For illustrative purposes we take the average income (\$160) and working hours (7.8) in the estimation sample as income and hours targets to determine the optimal working hours given the estimated coefficients. For each model, we calculate both the neoclassical optimal working hours based on the estimated functional form of the consumption utility, and the reference-dependent optimal working hours based on both the consumption utility and the gain-loss utility. Optimal working hours superscripted H or I denotes that the number is bounded by the hours or income target.

To illustrate the implications of the estimated utility function parameters under Table 4's alternative specifications, Table 5 presents the optimal stopping times implied by our estimates of the structural reference-dependent model for each specification and for representative percentiles of the observed distribution of realized wages, with "neoclassical" optimal working hours for comparison, computed from the estimated parameters using consumption utility only.[29]

[29] Note that this comparison is not completely fair to the neoclassical model, because we do not reestimate the parameters of the

The implied reference-dependent stopping times seem reasonable for all four models. However, for the models of Column 1 and 2 neoclassical working hours are very high, consistent with the interpretation of Section II.1's sample-splitting results that the neoclassical optimal income and hours are higher than the targets, making the second-reached target more important. By contrast, for the model of Column 4 the neoclassical optimal solution ranges from below to above the targets as earnings per hour vary, so labor supply is driven by neoclassical considerations for low earnings but by the hours target for high earnings; in aggregate the correlation between earnings per hour and optimal working hours is positive.

Like Section II.1's probits, our structural model resolves the apparent contradiction between a negative aggregate wage elasticity and the positive incentive to work of an anticipated increase in expected earnings per hour. In our model the stopping decisions of some drivers, on some days, will be more heavily influenced by their income targets, in which case their earnings elasticities will be negative, while the decisions of other drivers on other days will be more heavily influenced by their hours targets, with elasticities close to zero. When $\eta(\lambda - 1)$ is large enough, and with a significant number of observations in the former regime, the model will yield a negative aggregate elasticity. To illustrate, Table 5 also reports each specification's implication for the aggregate correlation of earnings and optimal working hours, a proxy for the elasticity. All reference-dependent models but column (4), which suppresses day-of-the-week differences, have a negative correlation between earnings per hour and optimal working hours.

Despite the influence of the targets on stopping probabilities, the heterogeneity of realized earnings yields a smooth aggregate relationship between stopping probability and realized income, so the model can reconcile Farber's (2005) finding that aggregate stopping probabilities are significantly related to hours but not income with a negative aggregate wage elasticity of hours as found by Camerer et al. (1997).

Finally, our structural model avoids Farber's (2008) criticism that drivers' estimated targets are too unstable and imprecisely estimated to allow a useful reference-dependent model of labor supply. The key function $\eta(\lambda - 1)$ of the parameters of gain-loss utility is plausibly and precisely estimated, robust to the specification of proxies for drivers' expectations, and comfortably within the range that indicates reference-dependent preferences.

---

consumption utility function constraining $\eta = 1$, which might yield more reasonable estimates of the neoclassical optimal working hours. Online Appendix D, Table D1 gives the implied average stopping probabilities for various ranges relative to the targets. Our estimates imply comparatively little bunching around the targets, Even so, the targets have a very strong influence on the stopping probabilities, and the second-reached target has a stronger effect than the first-reached target.

## III. Conclusion

In this paper we have proposed and estimated a model of cabdrivers' labor supply based on KR's theory of reference-dependent preferences, with targets for hours as well as income, both determined by proxied rational expectations. Our analysis builds on Farber's (2005, 2008) empirical analyses, which allowed income- but not hours-targeting and treated the targets as latent variables.

Our model, estimated with Farber's data, suggests that reference-dependence is an important part of the labor-supply story in his dataset, and that using KR's model to take it into account does yield a useful model of cabdrivers' labor supply. Overall, our results suggest that a more comprehensive investigation of the behavior of cabdrivers and other workers with similar choice sets, with larger datasets and more careful modeling of targets, will yield a reference-dependent model of labor supply that significantly improves upon the neoclassical model.

## References

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. 2010. "Reference Points and Effort Provision." *American Economic Review*, 100, in press.

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

Crawford, Vincent P., and Juanjuan Meng. 2008. "New York City Cabdrivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income." U. C. S. D. Discussion Paper 2008-03, http://repositories.cdlib.org/ucsdecon/2008-03/.

Doran, Kirk. 2009. "Reference Points, Expectations, and Heterogeneous Daily Labor Supply." Manuscript, University of Notre Dame.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, 47(2): 263-292.

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

Farber, Henry S. 2008. "Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers." *American Economic Review*, 98(2): 1069-1082.

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

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

Kőszegi, Botond, and Matthew Rabin. 2007. "Reference-Dependent Risk Attitudes." *American Economic Review*, 97(4): 1047-1073.

Kőszegi, Botond and Matthew Rabin. 2009. "Reference-Dependent Consumption Plans," *American Economic Review*, 99(3): 909-936.

Oettinger, Gerald S. 1999. "An Empirical Analysis of the Daily Labor Supply of Stadium Vendors." *Journal of Political Economy*, 107(2): 360-392.

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

Tversky, Amos, and Daniel Kahneman. 1991. "Loss Aversion in Riskless Choice: A Reference-Dependent Model." *Quarterly Journal of Economics*, 106(4): 1039-1061.

## Online Appendix A: Estimations of Tables 2-4 without weights.

**Table A1: Marginal Effects on the Probability of Stopping: Probit Estimation with Split Samples**

| | | (1) | | | (2) | | |
|---|---|---|---|---|---|---|---|
| | Evaluation Point for Marginal Effect | Pooled data | First hour's earning > expected | First hour's earning < expected | Pooled data | First hour's earning> expected | First hour's earning <= expected |
| Cumulative total hours | 8.0 | .020***<br>(.006) | 0.025***<br>(0.006) | 0.017**<br>(0.009) | 0.009***<br>(0.003) | 0.026***<br>(0.009) | 0.004<br>(0.004) |
| Cumulative Income/100 | 1.5 | 0.035*<br>(.016) | 0.030<br>(0.020) | 0.037<br>(0.026) | 0.020<br>(0.014) | 0.034<br>(0.028) | 0.029*<br>(0.022) |
| Min temp<30 | 0.0 | - | - | - | 0.004*<br>(0.008) | 0.011<br>(0.021) | 0.011<br>(0.011) |
| Max temp>80 | 0.0 | - | - | - | -0.017*<br>(0.010) | -0.041<br>(0.025) | -0.017<br>(0.012) |
| Hourly rain | 0.0 | - | - | - | 0.011<br>(0.164) | -0.520<br>(0.718) | 0.098<br>(0.122) |
| Daily snow | 0.0 | - | - | - | -0.001<br>(0.005) | -0.004<br>(0.010) | -0.084<br>(0.094) |
| Downtown | 0.0 | - | - | - | 0.002<br>(0.008) | 0.009<br>(0.019) | -0.002<br>(0.009) |
| Uptown | 0.0 | - | - | - | -0.002<br>(0.006) | -0.006<br>(0.015) | -0.003<br>(0.007) |
| Bronx | 0.0 | - | | | 0.072<br>(0.071) | 0.099***<br>(0.030) | .148**<br>(.122) |
| Queens | 0.0 | - | - | - | 0.045<br>(0.045) | 0.040<br>(0.087) | 0.060<br>(0.067) |
| Brooklyn | 0.0 | - | - | - | 0.088***<br>(0.041) | 0.140*<br>(0.107) | 0.070**<br>(0.050) |
| Kennedy Airport | 0.0 | - | - | - | 0.076***<br>(0.040) | 0.184***<br>(0.086) | 0.038<br>(0.041) |
| LaGuardia Airport | 0.0 | - | - | - | 0.073***<br>(0.037) | 0.176***<br>(0.079) | 0.007<br>(0.028) |
| Other | 0.0 | - | - | - | 0.148***<br>(0.084) | 0.233**<br>(0.132) | 0.082<br>(0.105) |
| Drivers (21) | | No | No | No | Yes | Yes | Yes |
| Day of week (7) | | No | No | No | Yes | Yes | Yes |
| Hour of day (19) | 2:00 p.m. | No | No | No | Yes | Yes | Yes |
| Log likelihood | | -1550.452 | -806.30573 | -742.87617 | -1344.8812 | -683.58849 | -628.45562 |
| Pseudo $R^2$ | | 0.1239 | 0.1314 | 0.1172 | 0.2401 | 0.2636 | 0.2532 |
| Observation | | 8958 | 4664 | 4294 | 8958 | 4664 | 4294 |

Note: Standard errors are computed for the marginal effects to maximize comparability with Farber's estimates, but with significance levels computed for the underlying coefficients rather than the marginal effects: *10%, **5%, ***1%. Robust standard errors clustered by shift are assumed. We use Farber's evaluation point: after 8 total working hours and $150 earnings on a dry day with moderate temperatures in midtown Manhattan at 2:00 p.m. Driver fixed effects and day of week dummies are equally weighted. For dummy variables, the marginal effect is calculated by the difference between values 0 and 1. As in Farber (2008), we do not distinguish between driving hours and waiting time between fares. Among the dummy control variables, only driver fixed effects, hour of the day, day of the week, and certain location controls have effects significantly different from 0.

**Table A2: Marginal Effects on the Probability of Stopping: Reduced-Form Model Allowing Jumps at the Targets**

| | Evaluation point for marginal effect | Using driver specific sample average income and hours prior to the current shift as targets (1) | (2) | Using driver and day-of-the-week specific sample average income and hours prior to the current shift as targets (3) | (4) |
|---|---|---|---|---|---|
| Cumulative total hours>hours target | | 0.040***<br>(0.013) | 0.065***<br>(0.031) | 0.047***<br>(0.014) | 0.109***<br>(0.039) |
| Cumulative income > income target | | 0.052***<br>(0.06) | 0.024<br>(0.025) | 0.043***<br>(0.015) | 0.038*<br>(0.024) |
| Cumulative total hours | 8.0 | 0.012***<br>(0.004) | 0.019***<br>(0.009) | 0.013***<br>(0.004) | 0.016***<br>(0.008) |
| Cumulative Income/100 | 1.5 | -0.007<br>(0.013) | 0.024<br>(0.035) | -0.001<br>(0.015) | 0.006<br>(0.029) |
| Weather (4) | | No | Yes | No | Yes |
| Locations (9) | | No | Yes | No | Yes |
| Drivers (21) | | No | Yes | No | Yes |
| Days of the week (7) | | No | Yes | No | Yes |
| Hour of the day (19) | 2:00 p.m. | No | Yes | No | Yes |
| Log likelihood | | -1546.1866 | -1369.5477 | -1535.036 | -1349.809 |
| Pseudo $R^2$ | | 0.1630 | 0.2587 | 0.1691 | 0.2693 |
| Observation | | 10337 | 10337 | 10337 | 10337 |

Note: Standard errors are computed for the marginal effects to maximize comparability with Farber's estimates, but with significance levels computed for the underlying coefficients rather than the marginal effects: *10%, **5%, ***1%. Robust standard errors clustered by shift are assumed. We use Farber's evaluation point: after 8 total working hours and $150 earnings on a dry day with moderate temperatures in midtown Manhattan at 2:00 p.m. Driver fixed effects and day of week dummies are equally weighted. For dummy variables, the marginal effect is calculated by the difference between values 0 and 1. As in Farber (2008), we do not distinguish between driving hours and waiting time between fares. Among the dummy control variables, only driver fixed effects, hour of the day, day of the week, and certain location controls have effects significantly different from 0.

| **Table A3: Structural Estimates under Alternative Specifications of Expectations** | | | | | |
|---|---|---|---|---|---|
| | (1) Use driver and day-of-the-week specific sample averages prior to the current shift as the income/hours targets and the next-trip earnings/times expectation | (2) Use driver and day-of-the-week specific sample averages prior and after the current shift as the income/hours targets and next-trip the earnings/times expectation | (3) Use driver and day-of-the-week specific sample averages prior to the current shift as the income/hours targets and fit the sophisticated next-trip earnings/time expectation | (4) Use driver (without day-of-the-week difference) specific sample averages prior to the current shift as income/hours targets and the next-trip earnings/time expectation | (5) Income target only: use driver and day-of-the-week specific sample averages prior to the current shift as income target and next-trip earnings/time expectation |
| $\eta(\lambda_H - 1)$ | 2.338*** | 4.327*** | 0.872*** | 0.237*** | - |
| [p-value] | [0.000] | [0.000] | [0.000] | [0.000] | |
| $\eta(\lambda_I - 1)$ | 0.631*** | 0.610*** | 0.267*** | 0.044* | 3.163*** |
| [p-value] | [0.004] | [0.000] | [0.008] | [0.0594] | [0.000] |
| $\theta$ | 0.015*** | 0.020*** | 0.018*** | 0.099*** | 0.014*** |
| [p-value] | [0.000] | [0.000] | [0.000] | [0.000] | [.000] |
| $\rho$ | 0.839*** | 0.483 | 0.883*** | 0.258*** | 1.645*** |
| [p-value] | [0.003] | [0.403] | [0.000] | [0.000] | [0.000] |
| $\sigma$ | 0.196 | 0.185 | 0.096 | 0.040 | 0.539*** |
| [p-value+] | [ 0.168] | [0.293] | [0.996] | [0.105] | [0.757] |
| $c$ | 0.007 | 0.006 | -0.012 | 0.134 | 0.138 |
| [p-value] | [ 0.954] | [0.958] | [0.998] | [0.782] | [0.719] |
| Test $\lambda_H = \lambda_I$ [p-value] | [0.214] | [0.177] | [0.996] | [0.204] | - |
| Observations | 10337 | 10337 | 10337 | 10337 | 10337 |
| Log-likelihood | -1360.9672 | -1361.711 | -1351.4242 | -1368.8756 | -1371.8068 |

Notes: Significance levels *10%, **5%, ***1%. We perform likelihood ratio test on each estimated parameter and indicate the corresponding p-value and significance level. +The null-hypothesis is that each parameter equals zero except for the variance estimate where we test $\sigma = 1$. Control variables include driver fixed effects (18), day of week (6), hour of day (18), location(8), and weather (4).

**Online Appendix B: Derivation of the Likelihood Function in the Structural Estimation**

Given a driver's preferences,

$$\text{(B1)}\quad V(I,H\,|\,I^r,H^r)=(1-\eta)\left[I-\frac{\theta}{1+\rho}H^{1+\rho}\right]+\eta\left[1_{(I-I^r\le 0)}\lambda(I-I^r)+1_{(I-I^r>0)}(I-I^r)\right]$$

$$-\eta\left[1_{(H-H^r\ge 0)}\lambda\left[\frac{\theta}{1+\rho}H^{1+\rho}-\frac{\theta}{1+\rho}(H^r)^{1+\rho}\right]\right]-\eta\left[1_{(H-H^r<0)}\left[\frac{\theta}{1+\rho}H^{1+\rho}-\frac{\theta}{1+\rho}(H^r)^{1+\rho}\right]\right].$$

We assume the driver decides to stop at the end of a given trip if and only if his anticipated gain in utility from continuing work for one more trip is negative. Again letting $I_t$ and $H_t$ denote income earned and hours worked by the end of trip $t$, this requires:

$$\text{(B2)}\qquad E[V(I_{t+1},H_{t+1}|I^r,H^r)]-V(I_t,H_t|I^r,H^r)+\varepsilon<0,$$

where $I_{t+1}=I_t+E(f_{t+1})$ and $H_{t+1}=H_t+E(h_{t+1})$, $E(f_{t+1})$ and $E(h_{t+1})$ are the next trip's expected fare and time (searching and driving), $x_t\beta$ include the effect of control variables, $c$ is the constant term, and $\varepsilon$ is a normal error with mean zero and variance $\sigma^2$. The likelihood function can now be written, with $i$ denoting the shift and $t$ the trip within a given shift, as:

$$\text{(B3)}\sum_{i=1}^{584}\sum_{t=i}^{T_i}\ln\Phi[((1-\eta)(A_{it}-\frac{\theta}{\rho+1}B_{it}(\rho))+\eta(\lambda a_{1,it}+a_{2,it}-\lambda\frac{\theta}{\rho+1}b_{1,it}(\rho)-\frac{\theta}{\rho+1}b_{2,it}(\rho))+x_t\beta+c)/\sigma]$$

$$A_{it}=I_{i,t+1}-I_{i,t}\,.$$

$$B_{it}(\rho)=H_{i,t+1}^{\rho+1}-H_{i,t}^{\rho+1}\,.$$

$$a_{1,it}=1_{(I_{i,t+1}-I_i^r\le 0)}(I_{i,t+1}-I_i^r)-1_{(I_{i,t}-I_i^r\le 0)}(I_{i,t}-I_i^r)\,.$$

$$a_{2,it}=1_{(I_{i,t+1}-I_i^r>0)}(I_{i,t+1}-I_i^r)-1_{(I_{i,t}-I_i^r>0)}(I_{i,t}-I_i^r)\,.$$

$$b_{1,it}(\rho)=1_{(H_{i,t+1}-H_i^r\ge 0)}(H_{i,t+1}^{\rho+1}-(H_i^r)^{\rho+1})-1_{(H_{i,t}-H_i^r\ge 0)}(H_{i,t}^{\rho+1}-(H_i^r)^{\rho+1})\,.$$

$$b_{2,it}(\rho)=1_{(H_{i,t+1}-H_i^r<0)}(H_{i,t+1}^{\rho+1}-(H_i^r)^{\rho+1})-1_{(H_{i,t}-H_i^r<0)}(H_{i,t}^{\rho+1}-(H_i^r)^{\rho+1})\,.$$

Note that

$A_{it}=a_{1,it}+a_{2,it}$ and

$B_{it}=b_{1,it}(\rho)+b_{2,it}(\rho)$.

Substituting these equations yields a reduced form for the likelihood function:

$$\text{(B4)}\quad\sum_{i=1}^{584}\sum_{t=i}^{T_i}\ln\Phi[((1-\eta+\eta\lambda)a_{1,it}+a_{2,it}-(1-\eta+\eta\lambda)\frac{\theta}{\rho+1}b_{1,it}(\rho)-\frac{\theta}{\rho+1}b_{2,it}(\rho)+x_t\beta+c)/\sigma].$$

## Online Appendix C: Trip Fares and Time Estimates Whose Fitted Values are Used as Proxies for Drivers' Expectations in Table 4, column 3

**Table C1: Trip Fares and Time Estimates Whose Fitted Values Are Used as Proxies for Drivers' Sophisticated Expectations in Table 4**

| | Time | Fare | | Time | Fare |
|---|---|---|---|---|---|
| Clock hours | | | Day of the Week | | |
| 0 | -0.100 | 0.006 | Monday | 0.017 | 0.000 |
| | (0.228) | (0.022) | | (0.025) | (0.000) |
| 1 | -0.121 | -0.005 | Tuesday | -0.007 | 0.001 |
| | (0.231) | (0.022) | | (0.023) | (0.003) |
| 2 | -0.255 | -0.025 | Wednesday | -0.012 | -0.002 |
| | (0.239) | (0.024) | | (0.023) | (0.004) |
| 3 | -0.193 | 0.000 | Thursday | 0.013 | 0.004 |
| | (0.265) | (0.000) | | (0.023) | (0.004) |
| 4 | 0.000 | 0.026 | Friday | -0.003 | -0.000 |
| | (0.000) | (0.039) | | (0.023) | (0.003) |
| 5 - 10 | -0.022 | -0.006 | Saturday | 0.038* | 0.006* |
| | (0.226) | (0.021) | | (0.022) | (0.003) |
| 11 | -0.022 | -0.011 | Mini temp < 30 | 0.016 | 0.000 |
| | (0.227) | (0.022) | | (0.027) | (0.004) |
| 12 | 0.026 | -0.005 | Max temp > 80 | 0.019 | -0.002 |
| | (0.227) | (0.022) | | (0.023) | (0.003) |
| 13 | -0.032 | -0.001 | Hourly rain | -0.147 | -0.073 |
| | (0.227) | (0.021) | | (0.317) | (0.046) |
| 14 | -0.074 | -0.003 | Daily snow | 0.006 | 0.000 |
| | (0.227) | (0.021) | | (0.010) | (0.001) |
| 15 | -0.084 | -0.005 | Downtown | -0.025 | 0.013 |
| | (0.227) | (0.021) | | (0.121) | (0.018) |
| 16 | -0.074 | 0.007 | Midtown | -0.066 | 0.001 |
| | (0.227) | (0.022) | | (0.120) | (0.018) |
| 17 | -0.132 | -0.006 | Uptown | -0.036 | 0.003 |
| | (0.226) | (0.021) | | (0.121) | (0.018) |
| 18 | -0.152 | -0.010 | Bronx | 0.000 | 0.000 |
| | (0.226) | (0.021) | | (0.000) | (0.000) |
| 19 | -0.189 | -0.016 | Queens | 0.337** | 0.080*** |
| | (0.226) | (0.021) | | (0.151) | (0.022) |
| 20 | -0.137 | -0.006 | Brooklyn | 0.180 | 0.052*** |
| | (0.226) | (0.021) | | (0.135) | (0.020) |
| 21 | -0.160 | -0.008 | Kennedy Airport | 0.645*** | 0.164*** |
| | (0.226) | (0.021) | | (0.136) | (0.020) |
| 22 | -0.177 | -0.004 | LaGuardia Airport | 0.333** | 0.110*** |
| | (0.226) | (0.021) | | (0.130) | (0.019) |
| 23 | -0.128 | 0.003 | Others | 0.154 | 0.030 |
| | (0.226) | (0.021) | | (0.156) | (0.023) |
| Constant | 0.307 | 0.051* | Driver dummy 21 | Yes | Yes |
| | (0.260) | (0.029) | R2 | 0.122 | 0.202 |
| Observations | 2989 | 2989 | 2989 | 2989 | 2989 |

Notes: Significance levels: * 10%, ** 5%, *** 1%. Fare and time (waiting and driving) for the next trip are jointly estimated as seemingly unrelated regressions.

## Online Appendix D: Implied Average Probabilities of Stopping for Various Ranges

**Table D1. Implied Average Probabilities of Stopping for Various Ranges Relative to the Targets**

| | (1)<br>Use driver and day-of-the-week specific sample averages prior to the current shift as the income/hours targets and the next-trip earnings/times expectation | (2)<br>Use driver and day-of-the-week specific sample averages prior and after the current shift as the income/hours targets and next-trip the earnings/times expectation | (3)<br>Use driver and day-of-the-week specific sample averages prior to the current shift as the income/hours targets and fit the sophisticated next-trip earnings/time expectation | (4)<br>Use driver (without day-of-the-week difference) specific sample averages prior to the current shift as income/hours targets and the next-trip earnings/time expectation |
|---|---|---|---|---|
| *Wage in the first hour > expected* | | | | |
| Before income target | 0.020 | 0.021 | 0.019 | 0.022 |
| At income target | 0.083 | 0.097 | 0.080 | 0.092 |
| In between two targets | 0.105 | 0.109 | 0.103 | 0.103 |
| At hours target | 0.159 | 0.148 | 0.139 | 0.134 |
| Above hours target | 0.175 | 0.156 | 0.175 | 0.150 |
| *Wage in the first hour < expected* | | | | |
| Before hours target | 0.0180 | 0.0193 | 0.018 | 0.021 |
| At hours target | 0.081 | 0.086 | 0.094 | 0.094 |
| In between two targets | 0.106 | 0.109 | 0.113 | 0.119 |
| At income target | 0.161 | 0.148 | 0.181 | 0.138 |
| Above income target | 0.188 | 0.180 | 0.187 | 0.164 |

Note: The probability of each range is calculated from the average predicted probabilities of trips. A range is two-sided with tolerance 0.1: before target means < 0.95×target; at target means > 0.95×target but < 1.05×target; and above target means > 1.05×target. The probabilities are first computed for each driver and range and then averaged across drivers within each range, hence do not sum to one.