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]

University of California, San Diego

First version 23 July 2008; this version 16 July 2009

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 or reference-dependent preferences. Following KR, our model has targets for hours as well as income, both determined by rational expectations. Estimated with Farber’s data, our model 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 his 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 California, San Diego. 9500 Gilman Drive, La Jolla, CA 92093-0508 (e-mail: vcrawfor@dss.ucsd.edu); 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, Colin Camerer, David Card, Julie Cullen, Stefano Dellavigna, Kirk Doran, Henry Farber, Uri Gneezy, Roger Gordon, 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 decisions 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 realized earnings, especially for inexperienced drivers. 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 realized earnings per hour (the natural analog of the wage in this setting, which we call the "wage" from now on) early in the day are high.

Camerer et al.'s 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, loss aversion creates a kink that tends to make realized daily income bunch around the target. To the extent that such bunching occurs, hours will have negative wage elasticity.

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 fully converged on the extent to which the evidence supports reference-dependence.[2] Much also depends on its scope and the details of its structure. If reference-

[2] KR (2006) and Farber (2008) survey some of the empirical literature. Of particular note are a field study by Gerald S. Oettinger (1999), which found increased 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 (presumably unanticipated) increases in wage. In a field experiment, Ernst Fehr and Lorenz Goette (2007) 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 it was reference-dependent.

dependence were limited to inexperienced workers or unanticipated changes, its direct relevance to most policy questions would be small.[3] 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 reference-dependence with daily income targeting that goes beyond the informal explanations in previous empirical work. He then estimated a reduced form, treating the 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.[4] He found that a sufficiently rich parameterization of his targeting model fits better than a neoclassical specification, and that the probability of stopping increases significantly and substantially when the income 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 Farber 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 of income-targeting in most respects but takes a more specific position on how targets are determined. In KR's theory as applied to cabdrivers, a driver's preferences reflect both the standard consumption utility of income and leisure and reference-dependent "gain-loss" utility,

---

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

[4] Constancy across days of the week is violated in the sample, where Thursday's through Sunday's 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.

with their relative importance tuned by a parameter. As in Farber's model, the 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.[5]

This paper uses Farber's (2005, 2008) data to reconsider the possibility of a useful reference-dependent model of cabdrivers' labor supply, adapting his econometric strategies to estimate models based on KR's (2006) theory.[6]

Section I introduces the model. Following KR, we allow for consumption as well as gain-loss utility and for hours as well as income targets. Importantly, instead of estimating the targets as latent variables as in Farber (2008), we treat them as rational expectations and operationalize them by finding natural sample proxies with limited endogeneity problems as explained below. We also assume for simplicity that the targets are point expectations rather than distributions as in KR.

If the weight of gain-loss utility is small, the model mimics a neoclassical labor-supply model, so that the wage elasticity of hours is normally positive. If the weight of gain-loss utility is large, perfectly anticipated changes in wage still have neoclassical implications because gain-loss utility then drops out of the driver's preferences; but *un*anticipated changes have non-neoclassical implications. In particular, whenever the income target has an important influence on a driver's stopping decision, even a driver who values income but is "rational" in the reference-dependent sense of prospect theory may have a negative wage elasticity of hours, as Camerer et al. found.

Section II reports econometric estimates. In our econometric analyses, 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 to avoid confounding.

In Section II.1 we estimate linear probit models of the probability of stopping as in Farber (2005), but splitting the sample according to whether a driver's earnings early in the day are

[5] 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. (2009).

[6] In an important study that was undertaken independently of ours, Kirk Doran (2009) analyzes field data on another group of New York City cabdrivers, with enough data to reliably estimate individual-level effects. He finds considerable heterogeneity in drivers' behavior, with many drivers clearly reference-dependent (and apparently with expectations-driven targets), and others neoclassical.

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.

In a neoclassical model, it is irrelevant whether early earnings are unexpectedly high or low. In a reference-dependent model, however, this difference matters because high early earnings make a driver more likely to reach his income target before his hours target. In our estimates, when early earnings are high, hours (but not income) has a strong and significant effect on stopping probability; and when they are low this pattern is reversed. This reversal is inconsistent with a neoclassical model, but it is fully consistent with a reference-dependent model in which stopping probability happens to be more strongly influenced by the second target a driver reaches than the first.[7]

Further, 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.[8] 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.

In Section II.2 we use the pooled sample to 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 our proxied targets instead of Farber's estimated targets. The estimated effects of hitting the targets are large and significant, with the signs predicted by a reference-dependent model—so much so that the effects of income and hours come mostly from whether they are above or below their targets, rather than from their levels.

In Section II.3 we use the pooled sample to estimate a structural reference-dependent model in the spirit of Farber's (2008) model, with the changes suggested by KR's theory. Here the specification must take a position on how a driver forms his expectations about the wage, trip by

[7] These 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.

[8] As KR put it (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."

trip during the day. Farber (2005, Section V.C) argued, based on a detailed econometric analysis, 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." Instead he estimated a latent value of continuing (taking the income target into account, and defined as including any option value) and assumed that a driver stops when this value falls short of the cost of continuing.

Although both Camerer et al. (1997) and Farber (2005) found some within-day predictability, we simplify by assuming there is none. We take a driver's expectations about the wage during the day as predetermined rational expectations, proxied in the same way we proxy the targets, namely 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. Like Farber, we also assume that drivers are risk-neutral; but unlike Farber, we assume they ignore option value in their decisions.[9]

In our model the weight of gain-loss utility and the coefficient of loss aversion are not separately identified. However, 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 estimated model again implies significant influences of income and hours targets on stopping probabilities in a pattern consistent with a reference-dependent model but not with any neoclassical model, and it again resolves the puzzles left open by Farber's analyses.

Overall, our results suggest that reference-dependence is an important part of the labor-supply story in Farber's dataset, and using KR's model to take it into account does yield a useful model of cabdrivers' labor supply. The key feature of our analysis, which allows us 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

[9] It seems behaviorally reasonable to ignore option value, as Thierry Post et al. (2008) did. Farber (2008, p. 1073) deals formally with this issue by defining his latent continuation value to include option value; but if option value is truly important, his linear specification of continuation value is unlikely to be appropriate. Farber's (2008) and our treatments of drivers' decisions are both first-order proxies for globally optimal stopping conditions that depend on unobservables, which both yield coherent results despite their imperfections.

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.[10] 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 η (where $0 \leq \eta \leq 1$), as follows:[11]

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

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.[12] Gain-loss utility is also separable, with its components determined by the differences 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 λ relative to gains, "the coefficient of loss aversion," which is assumed to be the same for income and hours.

We follow KR in conceptualizing the income and hours targets I^r and H^r as rational expectations. For simplicity, we assume that they are point expectations.[13] We operationalize

---

[10] 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.

[11] KR (2006, 2007) use a different parameterization, in which consumption utility has weight 1 and gain-loss utility weight η. Our η is a simple transformation of theirs. KR (2009) suggest allowing η to differ for hours and income, but we avoid this complication.

[12] In keeping with the "narrow bracketing" assumption that drivers evaluate consumption and gain-loss utility day by day, $U_1(I)$ should be 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, a restriction Farber (2008) and we impose in our structural analyses. Treating $U_1(\cdot)$ as a von Neumann-Morgenstern utility function, it also suggests that consumption utility is approximately risk-neutral in daily income.

[13] KR's (2006) model treats drivers' targets as distributions, with gain-loss utility defined as the expectation of terms like those in our specification. Because their theory imposes equilibrium and makes no allowance for errors, they need distributions for gains and losses to occur with positive probability. Because our model's error structure and sample variation generate gains and losses in

them via sample proxies with limited endogeneity problems as explained in detail in Section II. We further assume that the driver is approximately risk-neutral in daily income (footnote 12).

We also assume, for expository purposes only, that the driver's predetermined expected wage is a constant, w^e; this can be relaxed without altering our conclusions.[14] With constant expected wage and the universal empirical finding that $\lambda \geq 1$—loss rather than gain aversion—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 alone. The optimal stopping decision 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 given 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 the expected wage w^e exceeds the relevant marginal rate of substitution and stops otherwise.[15] It makes no difference how accurate the driver's expectations are, as long as he is risk-neutral.

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 ($I < I^r$)	$-[U_2'(H)/U_1'(I)]/[1-\eta+\eta\lambda]$	$-U_2'(H)/U_1'(I)$

Table 1 lists the marginal rates of substitution in the interiors of the four possible gain-loss regions, expressed as hours disutility costs of an additional 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 the loss domain, the marginal rate of substitution is the same as for consumption utilities alone, so the stopping decision satisfies the standard neoclassical first-order condition. But when hours and income are in the interiors of opposite domains, the marginal rate

any case, we simplify by treating the targets as point expectations. Point targets may exaggerate the effect of loss aversion, and if anything biases the comparison against a reference-dependent model and in favor of a neoclassical model.

[14] The one important qualification is that 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.

[15] 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.

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, either $(1 - \eta + \eta\lambda)$ or $1/(1 - \eta + \eta\lambda)$. On the 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.

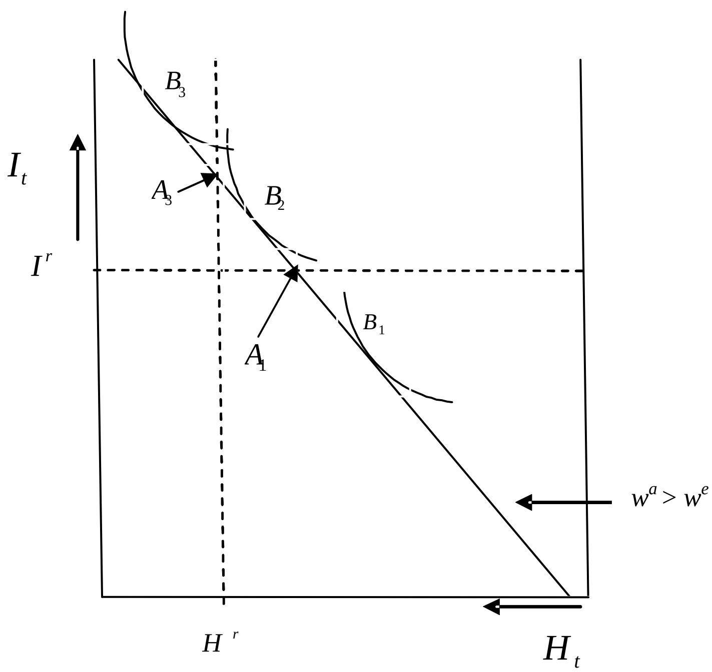

Figure 1: A Reference-dependent Driver's Stopping Decision

Figure 1, in which hours are measured negatively as a "bad," illustrates the driver's optimal stopping decision when the realized wage w^a is constant and $w^a > w^e$, so that realized earnings are higher than expected and the income target is reached before the hours target ($H = H^r$ and $w^a \equiv I/H > I^r/H^r \equiv w^e$ imply $I > I^r$). The constancy of w^a and w^e is only for illustration; the important thing is that realized earnings are higher than expected, so that the income target is reached before the hours target. The case where $w^a < w^e$ is analogous, but with the hours target reached before the income target.

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 $(H_t, I_t) = (0, 0)$. As the hours pass, earnings actually increase along a random but weakly monotone path (not shown), heading northwest. The path is a step function, but as mean trip length is only 12 minutes (Farber (2005, Section V)), the path, I, and H can be treated 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 5), although this matters 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, as long as their expected wages do not 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 $\lambda \geq 1$), comparison with the wage favors working more than the neoclassical comparison. The indifference curves in Figure 1 with tangency points B_1, B_2, and B_3 are 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 $I = w^a H$, and A_1 represents $(I^r/w^a, I^r)$. (The closer η is to one and the larger is $\lambda \geq 1$, the closer the stopping point is to A_1.)

Figure 2 compares the labor-supply curves for a neoclassical driver and a reference-dependent driver with the same consumption utility functions. The solid curve is the neoclassical supply curve, while the dashed curve is the reference-dependent one. The shape of the reference-dependent supply curve depends on which target has a larger influence on the stopping decision, which depends in turn 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. Whenever the wage is to the left of point D, the hours target is reached before the income target, and vice versa.

As Figure 2 illustrates, reference-dependent labor supply is non-monotonic. When the wage is very low, to the left of point A, the higher cost of income losses raises the incentive to work above its neoclassical level (lower left-hand cell of Table 1). 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 wage elasticity of hours is negative. Along segment DE labor supply is determined by the kink at the hours target, which is reached second. Finally, when the wage 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 wages fall close to point D, either along segment CD where hours decrease with increases in the wage because of income targeting, or along segment DE where hours do not change with increases in the wage because of hours targeting.[16]

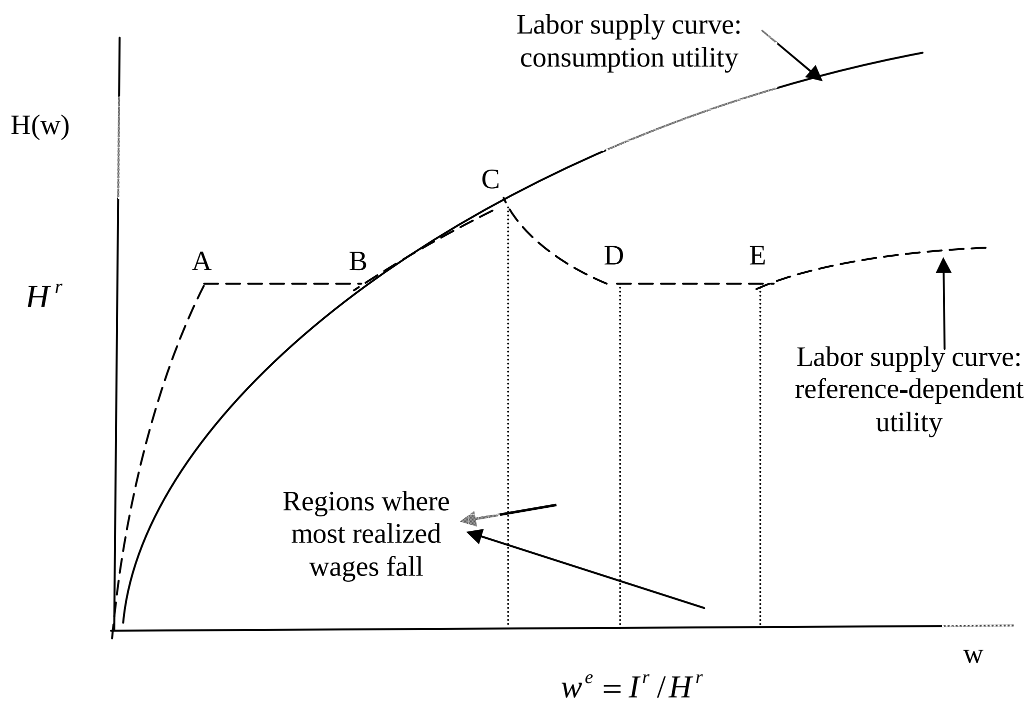

Figure 2: A Reference-dependent Driver's Labor Supply Curve

[16] There are two possible alternatives to the situation depicted in Figure 2. In the first, for wages that reconcile the income and hours targets, the neoclassically 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 neoclassically optimal income and hours exactly equal the targets, as in KR's preferred personal equilibrium. In that case, near where most realized wages 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 not help here.

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.[17]

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.[18] 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 4). 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.[19] 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.

II.1 Linear probit models of the probability of stopping

We begin by estimating linear probit models of the probability of stopping as in Farber (2005), but splitting the sample, shift by shift, according to whether a driver's earnings for the first *x* shifts of the day (or equivalently, the average wage for the first *x* hours, but with no need for the wage 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 Section I's structural model, a driver's

---

[17] Farber generously shared his data with us; and they are 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.

[18] Based on Farber's classification of hours into driving hours, waiting hours and break hours, we use only driving and waiting hours in our hours calculation. The results are similar when break time is included in the hours target and hours worked.

[19] 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.

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 $x = 3$.[20]

Table 2 reports marginal probability effects to maximize comparability with Farber's estimates, but with significance levels computed for the underlying coefficients. (Table A1 in Online Appendix A reports 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 total hours, total waiting hours, total break hours and income at the end of the trip are used to explain the stopping probability. In the pooled-sample estimates with these controls, all coefficients have the expected signs and the effect of hours is significant, but the effect of income is insignificantly different from zero. Waiting and break hours also have insignificant effects. In our split-sample estimates with these controls, when early earnings are higher than expected the effect of hours is large and significant but the effect of income is insignificant. But when early earnings are lower than expected, the effect of income is significant at the 5% level, and the effect of hours is insignificant. This reversal is inconsistent with a neoclassical model, but is fully consistent with a reference-dependent model in which stopping probability happens to be more strongly influenced by the second target a driver reaches than the first, as in Figure 2.

[20] When $x > 3$ the sign pattern of estimated coefficients is preserved, but the coefficients are no longer significantly different than 0, possibly because of the smaller sample size and censoring bias.

Table 2: Marginal Effects on the Probability of Stopping: Linear Probits with Split Samples

	Evaluation Point for Marginal Effect	(1) Pooled data	(1) First hour's earning > expected	(1) First hour's earning < expected	(2) Pooled data	(2) First hour's earning> expected	(2) First hour's earning < expected
Cumulative total hours	8.0	0.022** (0.011)	0.040*** (0.016)	-0.001 (0.010)	.014** (0.008)	.045*** (0.023)	-.006 (0.010)
Cumulative Income/100	1.5	0.029 (0.027)	-0.008 (0.040)	0.071** (0.027)	.011 (0.020)	-.033 (0.052)	.065** (0.035)
Cumulative waiting hours	2.5	0.005 (0.013)	-0.014 (0.016)	0.036*** (0.010)	.002 (0.008)	-.009 (0.019)	.027** (0.013)
Cumulative Break hours	0.5	-0.001 (0.010)	-0.011 (0.013)	0.019 (0.011)	-.005 (0.007)	-.019 (0.019)	.012 (0.011)
Min temp<30	0.0	-	-	-	.002 (0.010)	-.007 (0.030)	-.000 (0.013)
Max temp>80	0.0	-	-	-	-.020 (0.012)	-.056* (0.030)	.002 (0.022)
Hourly rain	0.0	-	-	-	.136 (0.229)	-.997 (0.890)	.329** (0.176)
Daily snow	0.0	-	-	-	-.001 (0.006)	.003 (0.013)	.003 (0.013)
Downtown	0.0	-	-	-	.010 (0.010)	.033 (0.026)	.004 (0.013)
Uptown	0.0	-	-	-	-.002 (0.007)	-.009 (0.018)	-.005 (0.010)
Bronx	0.0	-	-	-	.100 (0.090)	.191 (0.246)	.098 (0.098)
Queens	0.0	-	-	-	.076* (0.060)	.166* (0.118)	.001 (0.056)
Brooklyn	0.0	-	-	-	.115*** (0.050)	.217** (0.125)	.074** (0.052)
Kennedy Airport	0.0	-	-	-	.097*** (0.051)	.279*** (0.130)	.034 (0.043)
LaGuardia Airport	0.0	-	-	-	.101*** (0.046)	.233*** (0.096)	.054 (0.049)
Other	0.0	-	-	-	.184*** (0.107)	.311** (0.163)	.101 (0.140)
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		-1404.905	-688.825	-710.825	-1214.724	- 570.445	-602.904
Pseudo R^2		0.1246	0.1221	0.1333	0.2431	0.2730	0.2649
Observation		8040	3875	4165	8040	3875	4165

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 included in the brackets. We use Farber's evaluation point: after 8 total hours, 2.5 waiting hours, 0.5 break hour 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. 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.

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. The split-sample estimates with these controls are again fully consistent with our reference-dependent model, with hours but not income significant when the wage is higher than expected but income significant at the 5% level and hours insignificant when the wage is lower than expected.

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 such as Farber's (2008) would predict that there is a jump in the probability of stopping when the income target is reached, but that the influence of hours again varies smoothly with realized income. By contrast, our estimates show that the probability of stopping is more strongly influenced by hours when early earnings are higher than expected but by income when they are lower than expected. Our estimates are fully consistent with our reference-dependent model, but inconsistent with the neoclassical model and—because the effect of hours is significant when income is higher than expected but insignificant when income is higher than expected—with Farber's income-targeting model.[21]

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-

[21] 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.

dependent model). They are insignificantly different from 0 when we pool all days of the week (column 2) but large and significant when we disaggregate by day-of-the-week (column 4). This suggests that ignoring day-of-the-week effects is a significant misspecification, which may be one of the reasons why Farber's specification, which imposed constraints across days of the week for computational reasons, yielded different results. The column 4 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 of income has an insignificant, slightly negative effect; while the level of hours has a positive effect, significant only at the 10% level. In this respect, the estimates suggest that the effect of hours may have a nonnegligible "neoclassical" component.

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	0.0	.037*** (.012)	.009 (0.007)	.036*** (.013)	.018*** (0.008)
Cumulative income > income target	0.0	.047*** (.014)	.009 (0.008)	.053*** (.015)	.026*** (0.010)
Cumulative total hours	8.0	.012** (.006)	.008* (0.005)	.011* (.006)	.006* (0.004)
Cumulative Income/100	1.5	-.001 (.017)	.001 (0.014)	.000 (.018)	-.006 (0.012)
Cumulative waiting hours	2.5	.002 (.007)	.003 (0.005)	.006 (.007)	.003 (0.004)
Cumulative Break hours	0.5	.003 (.006)	-.001 (0.004)	.004 (.006)	.001 (0.004)
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		-1934.2673	-1710.3519	-1537.2767	-1355.9858
Pseudo R^2		0.1653	0.2619	0.1679	0.2660
Observation		12979	12979	10337	10337

Note: Significance levels are computed for the underlying coefficients rather than the marginal effects: *10%, **5%, ***1%. Robust standard errors clustered by shift are included in the brackets. We use Farber's evaluation point: after 8 total hours, 2.5 waiting hours, 0.5 break hour 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. 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 again use the sample proxies for drivers' targets introduced at the start of this section. We also need a sample proxy for the wages a driver expects over the course of a day, in the sense of how his expected earnings will vary with hours worked that day. We take a driver's expectations about the wage during the day as predetermined rational expectations, proxied in the same way we proxy the targets, namely by sample averages, driver/day-of-the-week by driver/day-of-the-week, up to but not including the day in question.

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 ρ 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:

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

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 ε is a normal error with mean zero and variance σ^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, σ is identified through $a_{2,it}$, which represents the change in income "gain" relative to the income target.

However, η and λ cannot be separately identified: only $1 - \eta + \eta\lambda$ is identified. This is clear from the likelihood function and from Table 1, where reference-dependence introduces kinks whose magnitudes are determined by $1 - \eta + \eta\lambda$. If $\eta = 0$ the model reduces to a neoclassical model. If $\eta = 1$ the model has only gain-loss utility as was usually assumed before KR (2006), and $1 - \eta + \eta\lambda$ reduces to λ. 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. Column 1 is the baseline, from which columns 2-5 each change one thing at a time. Column 2 checks for robustness to basing targets on sample proxies after as well as before the current shift (but still omitting the current shift; see footnote 19). Column 3 uses a more sophisticated 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 predicted using the current estimation sample. (Table C1 in Online Appendix C reports the trip fares and time estimates whose fitted values are used as proxies for drivers' expectations in those models.[22]) Column 4 rules out differences across days of the week and Column 5 restricts attention to income targeting, in each case as in Farber (2008).

The baseline model yields plausible parameter estimates that confirm and refine the conclusions of Section II.1-2's analyses. The null hypothesis that $1 - \eta + \eta\lambda = 1$ is rejected at the 1% level, ruling out the restriction $\eta = 0$ that reduces the model to a neoclassical model. Our estimate of $1 - \eta + \eta\lambda$ is somewhat lower than most reported estimates of the coefficient of loss

[22] The other variables include day-of-the-week, hour-of-the-day, locations at the end of the trip, weather controls, and realized wage of the day up to the current trip to capture any day-to-day variation known to the drivers but not captured by the constant term. Surprisingly, there is not much variation by time of day, but there is a lot of variation across locations.

aversion, but not implausibly so. There is a precision issue involving the preference parameters, in that in our baseline model, the estimate of θ is significantly different from 0 only at the 5% level, and the estimate of ρ is only significantly different from 0 at the 10% level.

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
$1-\eta+\eta\lambda$	1.715***	1.353***	1.441**	1.182**	5.036
	(.345)	(.158)	(.327)	(.116)	(8.480)
θ	.099**	.073**	.018*	.069*	.051*
	(.062)	(.057)	(.031)	(.086)	(.102)
ρ	.588*	.585*	1.118*	.646*	1.536**
	(.310)	(.312)	(.720)	(.404)	(.704)
σ	.032*	.106***	.115	.089	.789
	(.017)	(.038)	(.093)	(.073)	(1.599)
c	.001	-.008	-.023	.017	.084
	(.015)	(0.048)	(.0604)	(.074)	(.415)
Observations	10337	10337	10337	10337	10337
Log-likelihood	-1363.0367	-1367.4512	-1357.0613	-1375.2973	-1367.6224

Notes: Significance levels *10%, **5%, ***1%. We perform likelihood ratio tests on $1-\eta+\eta\lambda=1$, $\theta=0$ and $\rho=0$ separately and indicate the corresponding significance levels. Control variables include driver fixed effects (18), day of week (6), hour of day (18), location(8), and weather (4). Standard errors are reported in parentheses.

Column 2 shows that basing the targets on sample proxies after as well as before the current shift adds somewhat to precision, and column 3 shows that the results are robust to more sophisticated wage forecasting. Columns 4 and 5 confirm that day-of-the-week differences and hours targeting are both important for detecting the effects of reference-dependence, in that the target effects become smaller and/or insignificant with this variation in specification. Both of these features plainly contribute to our differences from Farber's conclusions.

The five models all have the same number of parameters—a constant term, four structural parameters, and 54 controls.[23] 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

[23] 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—a coefficient of loss aversion as well as heterogeneous income targets—to be fairly compared with a neoclassical model, defining the targets as rational expectations reduces the difference.

constant expectations throughout the shift, Column 1's model, the baseline, fits best.

Table 5: Estimated Optimal Stopping Times (in Hours)

Percentile in the wage distribution	Hourly wage	(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 $\theta = .099$ $\rho = 0.588$ $1 - \eta + \eta\lambda = 1.715$		(2) Use driver and day-of-the-week specific sample averages prior and after the current shift as the income/hours targets and the next-trip the earnings/times expectation $\theta = .073$ $\rho = 0.585$ $1 - \eta + \eta\lambda = 1.353$		(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 $\theta = .018$ $\rho = 1.118$ $1 - \eta + \eta\lambda = 1.441$		(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 $\theta = .069$ $\rho = 0.646$ $1 - \eta + \eta\lambda = 1.182$	
		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%	$19.1	3.06	7.65	5.18	7.80^{H}	8.27	8.27	4.84	6.26
20%	$20.4	3.42	7.80^{H}	5.79	7.80^{H}	8.77	8.77	5.36	6.94
30%	$21.5	3.74	7.80^{H}	6.34	7.80^{H}	9.193	8.37^{I}	5.81	7.52
40%	$22.3	3.98	7.80^{H}	6.75	7.80^{H}	9.50	8.07^{I}	6.15	7.80^{H}
50%	$23.3	4.29	7.73^{I}	7.27	7.73^{I}	9.88	7.80^{H}	6.58	7.73^{I}
60%	$24.3	4.61	7.41^{I}	7.81	7.41^{I}	10.26	7.80^{H}	7.02	7.41^{I}
70%	$25.3	4.93	7.12^{I}	8.37	7.11^{I}	10.63	7.80^{H}	7.47	7.47^{I}
80%	$26.7	5.41	6.74^{I}	9.18	6.74^{I}	11.16	8.05	8.12	7.80^{H}
90%	$28.5	6.04	6.32^{I}	10.26	6.32^{I}	11.83	8.53	4.84	6.26
Correlation of wage and optimal working hours		1	-0.90	1	-0.93	1	-0.24	1	0.73

Note: for illustrative purposes we take the average income ($180) 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 with superscript H/I denotes that the number is bounded by hours/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. The implied stopping times seem reasonable for the models in columns (1), (2), and (4). But column (3), the sophisticated model, yields unrealistically high stopping times due to the low estimated marginal disutility of hours (low θ).[24]

[24] Table D1 in Online Appendix D gives the implied average stopping probabilities for various ranges relative to the targets. The

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 wage increase. 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 wage elasticities will tend to be negative, while the decisions of other drivers on other days will be more heavily influenced by their hours targets, in which case their wage elasticities will be close to zero. When $1 - \eta + \eta\lambda$ is large enough, and with a significant number of observations in the former regime, the model will yield a negative aggregate wage elasticity of hours. To illustrate, Table 5 also reports each specification's implication for the aggregate correlation of wage and optimal working hours, a proxy for the wage elasticity. All models but column (4), which suppresses day-of-the-week differences, have a negative correlation between wage and optimal working hours.

Despite the influence of the targets on stopping probabilities, the heterogeneity of realized wages 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. In this comparatively small sample, there remains some ambiguity about the parameters of consumption utility ρ and θ. But the key function $1 - \eta + \eta\lambda$ 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.

III. Conclusion

Our estimates 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 what determines targets, is likely to yield a reference-dependent model of drivers' labor supply that significantly improves upon the neoclassical model. Of particular note in this context are the recent field study of Doran (2009), and the experimental work of Abeler et al. (2009).

---

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.

References

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. 2009. "Reference Points and Effort Provision." Manuscript, University of Bonn and University of Nottingham.

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: Coefficients for Table 2's Probit Model of the Probability of Stopping with Linear Effects with the Full Set of Controls Used in the Analysis

Table A1: Probability of Stopping: Linear Probits with Split Samples

		(1)			(2)	
	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	.132** (.062)	.222*** (0.075)	-.005 (.068)	0.119* (0.058)	0.236** (0.088)	-0.057 (0.091)
Cumulative Income/100	.174 (.175)	-.046 (.220)	.476** (.210)	0.094 (0.178)	-0.172 (0.264)	0.599* (0.282)
Cumulative waiting hours	.033 (.078)	-.075 (.089)	.243*** (0.075)	0.021 (0.069)	-0.047 (0.098)	0.246** (0.105)
Cumulative Break hours	-.005 (.062)	-.046 (0.220)	.126 (.081)	-0.040 (0.062)	-0.097 (0.090)	0.114 (0.099)
Min temp<30	-	-	-	0.021 (0.082)	-0.037 (0.159)	-0.004 (0.120)
Max temp>80	-	-	-	-0.205 (0.130)	-0.370* (0.195)	0.021 (0.195)
Hourly rain	-	-	-	1.196 (2.001)	-5.183 (4.432)	3.043* (1.454)
Daily snow	-	-	-	-0.006 (0.052)	0.014 (0.066)	0.031 (0.126)
Downtown	-	-	-	0.080 (0.079)	0.156 (0.109)	0.033 (0.117)
Uptown	-	-	-	-0.018 (0.064)	-0.048 (0.096)	-0.051 (0.091)
Bronx	-	-	-	0.573 (0.365)	0.695 (0.697)	0.585 (0.399)
Queens	-	-	-	0.470* (0.273)	0.624* (0.343)	0.012 (0.512)
Brooklyn	-	-	-	0.634** (0.181)	0.770** (0.326)	0.476* (0.233)
Kennedy Airport	-	-	-	0.564** (0.209)	0.934** (0.330)	0.256 (0.273)
LaGuardia Airport	-	-	-	0.579** (0.185)	0.815** (0.255)	0.376 (0.261)
Other	-	-	-	0.880** (0.338)	1.017** (0.405)	0.595 (0.604)
Drivers (21)	No	No	No	Yes	Yes	Yes
Day of week (7)	No	No	No	Yes	Yes	Yes
Hour of day (19)	No	No	No	Yes	Yes	Yes
Log likelihood	-1404.905	-688.825	-710.825	-1214.724	- 570.445	-602.904
Pseudo R^2	0.1246	0.1221	0.1333	0.2431	0.2730	0.2649
Observation	8040	3875	4165	8040	3875	4165

Note: significance level *10%, **5%, ***1%. Robust standard errors clustered by shift are included in the brackets. 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.

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

Given a driver's preferences,

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

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:

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

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 ε is a normal error with mean zero and variance σ^2. The likelihood function can now be written, with i denoting the shift and t the trip within a given shift, as:

(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:

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

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	0.042	0.004	Monday	-0.001	0
	(-0.1)	(-0.022)		(-0.016)	(-0.003)
1	0.024	-0.007	Tuesday	-0.017	0
	(-0.104)	(-0.023)		(-0.016)	(-0.003)
2	-0.076	-0.026	Wednesday	0.003	-0.003
	(-0.111)	(-0.024)		(-0.016)	(-0.003)
3	-	-	Thursday	0.001	0.002
	-	-		(-0.016)	(-0.003)
4	0.198	0.028	Friday	0.017	0
	(0.179)	(0.039)		(-0.015)	(-0.003)
5 - 10	0.027	-0.006	Saturday	-0.015	0.005
	(-0.097)	(-0.021)		(0.016)	(0.003)
11	0.027	-0.012	Mini temp < 30	-0.018	-0.007***
	(-0.098)	(-0.021)		(-0.011)	(-0.002)
12	0.042	-0.006	Max temp > 80	0.01	-0.001
	(-0.098)	(-0.021)		(-0.009)	(-0.002)
13	0.026	-0.002	Hourly rain	0.24	-0.044
	(-0.098)	(-0.021)		(-0.207)	(-0.045)
14	0.019	-0.004	Daily snow	-0.002	-0.001
	(-0.098)	(-0.021)		(-0.006)	(-0.001)
15	-0.002	-0.009	Downtown	-0.057	-0.016
	(-0.098)	(-0.021)		(-0.082)	(-0.015)
16	0.024	0.005	Midtown	-0.098	-0.028*
	(-0.099)	(-0.021)		-(0.082)	(0.015)
17	0.008	-0.007	Uptown	-0.09	-0.027*
	(-0.098)	(-0.021)		(-0.082)	(0.015)
18	-0.011	-0.01	Bronx	-	-0.027
	(-0.098)	(-0.021)		-	(0.023)
19	-0.037	-0.017	Queens	0.267***	0.053***
	(-0.098)	(-0.021)		(-0.102)	(-0.02)
20	0.004	-0.006	Brooklyn	0.104	0.024
	(-0.098)	(-0.021)		(-0.091)	(-0.017)
21	0.001	-0.008	Kennedy Airport	0.544***	0.140***
	(-0.098)	(-0.021)		(-0.092)	(-0.017)
22	-0.01	-0.004	LaGuardia Airport	0.289***	0.084***
	(-0.098)	(-0.021)		(-0.088)	(0.016)
23	0.029	0.002	Others	0.069	-
	(-0.099)	(-0.021)		(-0.106)	-
w +	-0.018	-0.002	Constant	0.374***	0.099***
	(-0.043)	(-0.009)		(-0.128)	0.026
R^2	0.1323	0.1867	0.1323	0.1867	0.1323
Observations	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. + Realized wage of the day up to the current shift

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) 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
Wage in the first hour>expected				
Before income target	.020	.021	.019	.022
At income target	.083	.097	.080	.092
In between two targets	.105	.109	.103	.103
At hours target	.159	.148	.139	.134
Above hours target	.175	.156	.175	.150
Wage in the first hour <expected				
Before hours target	.0180	.0193	.018	.021
At hours target	.081	.086	.094	.094
In between two targets	.106	.109	.113	.119
At income target	.161	.148	.181	.138
Above income target	.188	.180	.187	.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.