Title

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

Authors
Crawford, Vincent P.
Meng, Juanjuan

Publication Date
2008-07-23

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

23 July 2008

Abstract: This paper reconsiders whether cabdrivers' labor supply decisions reflect reference-dependent preferences. Following Botond Kőszegi and Matthew Rabin (2006), we construct a model with targets for hours as well as income, both determined by rational expectations. Estimating using Henry S. Farber's (2005, 2008) data, we show that the reference-dependent model can reconcile his 2005 finding that drivers' stopping probabilities are significantly related to hours but not income with the negative wage elasticity of hours found by Colin Camerer et al. (1997) and Farber (2005, 2008). The model yields sensible estimates that avoid Farber's (2008) criticism that drivers' income targets are too unstable to allow a useful reference-dependent 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, Stefano Dellavigna, Botond Kőszegi, David Miller, Ulrike Malmendier, and Matthew Rabin for insightful comments.

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, the neoclassical model of labor supply predicts a positive wage elasticity of hours. However, Camerer et al. (1997), collecting 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, found a strongly negative elasticity of hours with respect to realized earnings, especially for inexperienced drivers. Farber (2005, 2008), analyzing new data on a different set of New York City cabdrivers, found a similarly negative relationship.

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) is high. This explanation is in the spirit of Kahneman and 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"). In the explanation, a driver's reference point is a daily income target and loss aversion creates a kink that tends to make realized income bunch around the target, so realized hours have little or none of the positive wage elasticity predicted by the neoclassical model.

As Farber (2008, p. 1069) notes, 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."

But Farber finds mixed evidence for income targeting in the empirical literature on labor supply with flexible hours. Farber (2005) found that before controlling for driver fixed effects, the probability of stopping work is significantly related to income realized on a given day, but that driver fixed effects and other relevant controls render this effect statistically insignificant. And other studies of workers who choose their hours have found positive relationships between expected earnings and labor supply, as suggested by the neoclassical model.[2]

Farber (2008) reexamines the evidence, using his 2005 dataset to estimate a model explicitly derived from reference-dependence with daily income targeting that goes beyond the informal explanations that motivated previous work. He estimates 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.[3] He finds that a sufficiently rich parameterization of his reference-dependent model has a better fit than a standard neoclassical specification, and that the probability of stopping increases significantly and substantially once the income target is reached; but that his model cannot reconcile the strong increase in stopping probability at the target with the aggregate smoothness of the relationship between stopping probability and realized income. Further, the random effect turns out to be large but imprecisely estimated, from which he concludes that drivers' income targets are too unstable and imprecisely estimated 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."

Partly in response to Camerer et al.'s (1997) and Farber's (2005, 2008) analyses, Kőszegi and Rabin (2006) developed a theory of reference-dependent preferences that is more general than Farber's in most respects but takes a more specific position on how targets are determined (see also Kőszegi and Rabin (2007, 2009)). In Kőszegi and Rabin's (2006, Section V) theory as applied to cabdrivers' labor supply, a driver has a daily target for hours as well as income. His preferences reflect both the standard consumption utility of income and leisure and

[2] As Kőszegi and Rabin (2006, p. 1150) put it: "While strongly disagreeing about the extent of this behavior and whether it is irrational, Camerer et al. (1997) and Farber [(2005, 2008)] analyzing New York city taxi drivers all find a negative relationship between earnings early in the day and duration of work later in the day. Studies analyzing participation decisions as a function of expected wages, on the other hand, find a positive relationship between earnings and effort: [Gerald S.] Oettinger (1999) finds that stadium vendors are more likely to go to work on days when their wage can be expected to be higher, and [Ernst Fehr and Lorenz Goette (2007)] show bicycle messengers sign up for more shifts when their commission is experimentally increased." (Fehr and Goette (2007) found that the wage elasticity of hours worked was positive but that of work effort was negative. They argued that effort is a more accurate measure of labor supply and concluded that messengers' supply was reference-dependent.)

[3] Constancy across days of the week is a strong restriction. In the sample, Friday has the highest average income ($198.43), while Tuesday has the lowest ($164.31), with Friday's, Saturday's, and Sunday's incomes systematically higher than those of other days. Kruskal-Wallis equality-of-populations rank tests reject the hypotheses that average daily income or hours is constant across days of the week (p-values < 0.0001). Farber includes day-of-the-week dummies in his main specifications of the stopping probability equation, but this turns out to be an imperfect substitute for allowing the mean income target to vary across days of the week.

reference-dependent “gain-loss” utility, with their relative importance tuned by an estimated parameter. As in single-target models like Farber’s (2008), the driver is loss-averse; but working longer than the hours target is a loss, just as earning less than the income target is. Finally, and most importantly for our analysis, reflecting the belief that drivers in steady state learn to predict their daily hours and income with reasonable accuracy, Kőszegi and Rabin endogenize the targets by setting a driver’s targets equal to his theoretical rational expectations of hours and income.[4]

This paper reconsiders whether reference-dependent preferences allow a useful model of cabdrivers’ labor supply, using Farber’s data to estimate a model based on Kőszegi and Rabin’s (2006) theory. We closely follow Farber’s (2005, 2008) econometric strategies, but instead of treating targets as latent variables we treat them as rational expectations, operationalized by using average sample realizations of income and hours as proxies for them. (Proxying the targets by functions of endogenous variables creates some simultaneity problems, which we deal with as explained below.) Further, in the structural estimation that parallels Farber’s (2008) analysis, we allow for consumption as well as gain-loss utility and hours as well as income targets.

Section I introduces our adaptation of Kőszegi and Rabin’s model of reference-dependent preferences to cabdrivers’ daily labor supply decisions. If the weight of gain-loss utility is small, the model approaches a neoclassical model, with standard implications for labor supply. Even when gain-loss utility has larger weight, the standard implications carry over for changes in the wage that are perfectly anticipated, because gain-loss utility then drops out of the driver’s calculation. But when realized wages deviate from expected wages, his probability of stopping is more strongly influenced by hours or income, depending on which target is reached first, and the model’s implications may deviate substantially from those of a neoclassical model. When the realized wage is lower than expected, the hours target tends to be reached before the income target, hours have a stronger influence on the stopping probability than realized income, and the wage elasticity of labor supply is pushed toward zero. But when the realized wage is higher than expected, the income target tends to be reached first, and its stronger influence on the stopping probability can make even a driver who values income but is “rational” in the generalized reference-dependent sense of Prospect Theory have a negative wage elasticity.

[4] There can be multiple expectations that are consistent with the individual’s optimal behavior, given the expectations. Kőszegi and Rabin use a refinement, “preferred personal equilibrium,” to focus on the self-confirming expectations that are best for the individual. Most previous analyses have identified reference points with the status quo, but as Kőszegi and Rabin note most of the evidence does not distinguish these interpretations because expectations are usually close to the status quo. Even so, we shall argue that their rational-expectations view of targets has important substantive implications for modeling cabdrivers’ behavior.

Section II reports our econometric estimates. We begin in Sections II.1-2 by estimating linear and nonlinear probit models of the probability of stopping as in Farber's (2005) analysis, using his data but splitting the sample according to whether a given driver's realized wage on a given day is higher or lower than our sample proxy for his rational-expectations wage. A reference-dependent model like ours predicts large differences in stopping probabilities across these two regimes, independent of the details of the structure. This sharply distinguishes it from a neoclassical model even if our proxy for expectations is imperfect, which allows a robust assessment of the gains from a reference-dependent model, avoiding most restrictions needed for structural estimation.

In our split-sample estimates, when the realized wage is higher than expected, so that the income target is likely to be reached first, the stopping probability is strongly influenced by realized income but not hours; and when the wage is lower than expected so that the hours target is likely to be reached first, the stopping probability is strongly influenced by hours but not income. This qualitative pattern deviates significantly from the predictions of a neoclassical model, but is just as predicted by our reference-dependent model.

Because the wage elasticity of labor supply is negative in the former wage regime but near zero in the latter, the aggregate wage elasticity is likely to be negative. Thus, Kőszegi and Rabin's distinction between anticipated and unanticipated wage increases can resolve the apparent contradiction between the positive incentive to work created by an anticipated wage increase with a negative aggregate wage elasticity.[5] Further, because the two regimes have roughly equal weights in the sample, the heterogeneity of realized wages yields a smooth aggregate relationship between stopping probability and realized income, so the model can also 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).

Section II.3 uses the full sample to estimate a structural reference-dependent model as in Farber (2008), but with the changes suggested by Kőszegi and Rabin's theory described above. Because of the way the weight of gain-loss utility and the coefficient of loss aversion interact in our model, they are not separately identified. However, a simple function of them is identified,

[5] As Kőszegi and Rabin 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."

and our estimates of this function allow inferences that confirm and refine the conclusions of our split-sample analysis. Like our split-sample estimates, our structural estimates imply significant influences of income and hours targets on stopping probabilities in the pattern implied by Kőszegi and Rabin's multi-targeting model. They also reconcile the aggregate smoothness of the relationship between stopping probability and realized income with the negative aggregate wage elasticity of hours. Our structural model avoids Farber's (2008) criticism that drivers' targets are too unstable to allow a useful model of labor supply partly by nesting consumption and gain-loss utility and allowing hours as well as income targets, but mostly by treating the targets as rational expectations estimated from natural sample proxies, rather than as latent variables.

Section III is the conclusion.

I. The Model

This section introduces our model, which adapts Kőszegi and Rabin's (2006) theory of reference-dependent preferences to cabdrivers' labor supply decisions.[6]

Treating each day separately as in all previous analyses, consider the preferences of a given driver during his shift on a given day.[7] 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:[8]

(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 empirical test of Kőszegi and Rabin's theory, for simplicity and parsimony (1)-(2) incorporate some assumptions that they made provisionally: Consumption utility is additively separable across income and hours, with $U_1(\cdot)$ increasing in I,

[6] Kőszegi and Rabin's (2006) model treats drivers' targets as stochastic, with gain-loss utility defined as the expectation of terms like those in our specification. Because their theoretical model makes no allowance for errors, they need stochastic targets for gains and losses to occur. Because the errors that describe sampling variation in our model generate gains and losses even if drivers have point expectations, we simplify their model by treating targets as deterministic. Deterministic targets may exaggerate the effect of loss aversion, and if anything they bias the comparison against Kőszegi and Rabin's model and in favor of the neoclassical model.

[7] 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.

[8] Kőszegi and Rabin (2006, 2007) use a different parameterization, in which consumption utility has weight 1 and gain-loss utility has weight η. Thus our parameter η is a simple transformation of their parameter with the same name. In more recent work, Kőszegi and Rabin (2009) suggest allowing η to differ for hours and income, but we avoid this complication.

$U_2(\cdot)$ decreasing in H, and both concave.[9] Gain-loss utility is also additively separable, determined component by component by the differences between realized and target consumption utilities. As in a leading case Kőszegi and Rabin sometimes 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 the same for income and hours.

We follow Kőszegi and Rabin in equating the income and hours targets I^r and H^r to drivers' rational expectations. Empirically, we proxy drivers' expectations by their natural sample analogs, setting a driver's targets on a given day of the week equal to the analogous full-sample means for that day of the week, thus allowing the targets to vary across days of the week as suggested by the variation of hours and income (footnote 3), but ignoring sampling variation for simplicity.

Because drivers' earnings are determined randomly rather than by a known wage rate, drivers must form expectations after each trip about their earnings per hour if they continue work that day. Drivers face a difficult signal-extraction problem, and Farber (2005, Section V.C) argues, based on a careful and detailed econometric analysis, that hourly earnings are so variable that "predicting hours of work with a model that assumes a fixed hourly wage rate during the day does not seem appropriate." Instead he estimates a value of continuing (defined to include any option value of continuing beyond one more trip) as a latent variable and assumes that a driver's stopping decision is determined by comparing this value to the cost of continuing. Despite Farber's critique, because of the complexity of optimal stopping with hours as well as income targets we illustrate our model's possibilities as simply as possible, by assuming that drivers extrapolate their daily income linearly, assuming a constant expected hourly wage rate w^a and ignoring option value.[10] We further assume that drivers have rational expectations of w^a, which we proxy by their natural sample analogs, the driver's realized daily wages for that day of the week in the full sample.[11]

---

[9] In keeping with the "narrow bracketing" assumption that drivers evaluate consumption and gain-loss utility one day at a time, $U_1(I)$ should be thought of as a reduced form, partly reflecting the future value of income not spent today.

[10] Although the expected wage rate is assumed to be constant, our model and structural estimation allow the realized wage rate to vary. Our assumptions imply that option value is zero, but a richer model predicting the fare and time of the next trip, as considered in Section II.3, might make it positive. Even so, it seems a reasonable approximation to ignore it, as Thierry Post et al. (2008) do.

[11] Farber's (2008) estimation of continuation value as a latent variable and our assumption that drivers have rational expectations and extrapolate income linearly are alternative first-order proxies for globally optimal stopping conditions that depend on unobservables, which both yield coherent results despite their imperfections. As noted above, our proxying the targets by functions of endogenous variables creates simultaneity problems, which are exacerbated by the small samples for some drivers. In Section II.1-2's split-sample estimates our approximation makes little difference, because the theory's implications as tested there are robust to imperfections in the criterion for splitting. In Section II.3's structural estimation simultaneity problems are potentially important. Given the lack of suitable instruments, we consider an alternative proxy using a driver's sample means without allowing day-of-the-week differences, which makes the samples large enough that the simultaneity is negligible and yields similar results.

Given our assumptions about expectations formation 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 targets for hours as well as income, which parallels Farber's (2005, 2008) characterizations. When a driver extrapolates income linearly, his optimal stopping decision maximizes reference-dependent utility $V(I, H|I^r,H^r)$ as in (1) and (2), subject to the linear menu of income-hours combinations $I = w^aH$. 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 anticipated wage w^a exceeds the relevant marginal rate of substitution and stops otherwise.[12]

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 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, in which hours are measured negatively as a "bad," illustrates the driver's optimal stopping decision when $w^a > w^e$, so that realized income is higher than expected and the income target is reached before the hours target ($w^a = I/H > w^e$ and $H = H^r = I^r/w^e$ imply $I = w^aH = w^aI^r/w^e > I^r$). The case where $w^a < w^e$ (not shown), is completely analogous, but with the hours target reached before the income target. Letting I_t and H_t denote income earned and hours worked by the end of trip t, the driver starts in the lower right-hand corner, with $(H_t, I_t) = (0, 0)$, and anticipates moving along a sample line $I = w^aH$ with constant w^a. As the hours pass, his income increases along a random but monotone path (not shown), heading northwest. His realized path is a step function, but because the mean trip length is only about 12 minutes (Farber (2005, Section V)),

---

We also consider an alternative in which drivers' fare and trip time expectations are allowed to depend on time and location as in Farber's (2005, Section V.C) analysis, which confirms the main messages of our basic analysis.

[12] 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.

the path can be approximated as continuous, and *I* and *H* treated as continuous variables.

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

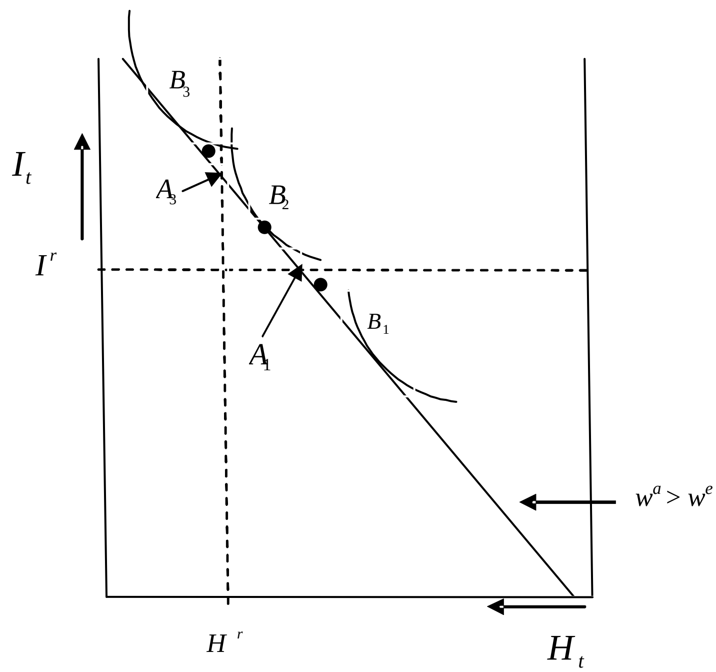


Figure 1: A Reference-dependent Driver's Stopping Decision

The three 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. Starting at $(I_t, H_t) = (0, 0)$, in the income-loss/hours-gain ($I_t < I^r$, $H_t < H^r$) domain, the driver continues working as long as the anticipated wage w^a exceeds the hours disutility cost of an additional unit of income, $-[U_2'(H_t)/U_1'(I_t)]/[1-\eta+\eta\lambda]$ 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 $\lambda \geq 1$), this comparison favors working more than the neoclassical one $w^a \geq -U_2'(H_t)/U_1'(I_t)$.

As hours and income accumulate through the day, either the hours disutility cost of income rises to w^a before the driver reaches his first target— with $w^a > w^e$, income as in Figure 1—and leaves the income-loss/hours-gain domain; or the hours disutility cost remains below w^a until he reaches his income target. In the former case he stops at a point weakly between points B_1 and A_1, where B_1 represents the hours and income that would maximize consumption utility on indifference curve 1, and A_1 represents (I^r/w^a, I^r). Other things equal, the closer η is to one and the larger is $\lambda \geq 1$, the closer the stopping point is to A_1 on the segment from B_1 to A_1.

In the latter case, he compares the cost in the domain he is entering—in this case income-gain/hours-gain ($I_t > I^r$, $H_t < H^r$)—and stops if the new hours disutility cost of income, $-U_2'(H)/U_1'(I)$ from the upper left cell of Table 1, exceeds w^a. In that case, the optimal stopping point is A_1. If, instead, $-U_2'(H)/U_1'(I) < w^a$, the driver continues working. Then either the hours disutility cost rises to w^a before he reaches his second target—with $w^a > w^e$, hours as in Figure 1—or it remains below w^a until he reaches his hours target. In the former case he stops at B_2. In the latter case, he stops if and only if the new hours disutility cost in the income-gain/hours-loss ($I_t > I^r$, $H_t > H^r$) domain he is entering, $-[U_2'(H)/U_1'(I)][1-\eta+\eta\lambda]$ from the upper right cell of Table 1, exceeds w^a. In that case, the optimal stopping point is A_3.

If the hours disutility cost remains less than w^a, the driver continues working. Then, either the cost rises to w^a before he reaches the maximum feasible number of hours, or it remains below w^a. In the latter case he stops at the maximum feasible number of hours. In the former case he stops at a point weakly between points B_3 and A_3, where B_3 represents the income and hours that would maximize consumption utility on indifference curve 3 and A_3 represents (H^r, w^aH^r). Other things equal, the closer η is to one and the larger is $\lambda \geq 1$, the closer the stopping point is to A_3.

Whether or not $w^a > w^e$, a driver who extrapolates income linearly anticipates passing through a sequence of domains such that the hours disutility cost of income weakly increases as hours and income accumulate—a reflection of the concavity of reference-dependent utility in I and H. Thus, given our assumptions about his expectations, the decision characterized here is globally optimal.

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 with rational-expectations proxies for the targets and, in the structural estimation, the other adjustments to the model suggested by Kőszegi and Rabin's (2006) theory.[13]

[13] 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

Farber (2005) estimates the effects of cumulative realized income and hours on the probability of stopping in a probit model, first imposing linearity and then allowing cumulative income and hours to have nonlinear effects (with their marginal effects allowed to differ as they accumulate). We begin in Sections II.1-2 by reporting estimates of linear and nonlinear models that parallel Farber's probit models. But instead of estimating using the full sample as he did, we split the sample, shift by shift, according to whether a driver's realized wage at the end of the day, w^a, is higher or lower than our rational-expectations proxy for his expected wage, w^e, his full-sample mean for that day of the week.[14] If a driver forms his expectations by extrapolating earnings approximately linearly, he tends to reach his income (hours) target first when his realized wage is higher (lower) than expected. Because the reference-dependent model predicts large differences in stopping probabilities across these regimes, splitting the sample allows simple tests that sharply distinguish neoclassical and reference-dependent models. For a wide class of reference-dependent models, including our structural model in Section I, the probability of stopping increases sharply at the first-reached target and again at the second-reached target. By contrast, in a neoclassical model, the targets have no effect. This difference between models is robust to the details of the structural specification and to variations in the specification of expectations. Sample-splitting therefore allows a robust assessment of the gains from a reference-dependent model.[15]

In Section II.3 we use the full sample to estimate a structural reference-dependent model as in Farber (2008), but with positively weighted consumption utility as well as gain-loss utility and hours as well as income targets as suggested by Kőszegi and Rabin's (2006). Our structural model makes no sharp general predictions: Whether the aggregate stopping probability is more strongly influenced by income or hours depends on the estimated parameters and how many shifts have realized income higher than expected. Even so, structural estimation provides an important check on the model's ability to reconcile the negative aggregate wage elasticity of hours Camerer et al.

---

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.

[14] Because some drivers have only one working record for certain days, their expectations and realization are then identical by construction. As noted above, this creates a simultaneity problem, which might bias our estimates of the importance of gain-loss utility. In Section II.1-2's split-sample estimates, whether we assign observations with only one record to high- or low-wage groups does not affect the results. In Section II.3's structural estimation, we consider alternative ways to address this problem.

[15] The difference between reference-dependent and neoclassical models shows up most clearly with Section II.2's nonlinear (time-varying) effects of income and hours. With Section II.1's linear (time-invariant) effects, the model's predictions are ambiguous without further information about how the driver's targets relate to his ideal income and hours. However, we find that enough drivers stop at the first target they reach that income (hours) is significantly related to the probability of stopping when realized wage is higher (lower) than expected wage.

(1997) found with Farber's (2008) finding that in the full sample, stopping probabilities are significantly related to hours but not income, and to give a useful account of drivers' labor supply.

II.1 The Probability of Stopping: Linear Effects

Table 2 reports the marginal probability effects from the estimation of the probit model with linear effects. (Table A1 in Online Appendix A reports the marginal effects for the model with the full set of controls.) In each panel of Table 2, the left-hand column replicates Farber's (2005) pooled-sample estimates; the center and right-hand columns report our split-sample estimates.

In the left-most 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 Farber's pooled-sample estimates with these controls, all coefficients have the expected signs and the effect of income is highly significant, but the effect of hours is small and insignificantly different from zero. Waiting and break hours also have significant effects. By contrast, in our split-sample estimates with these controls, when realized income is higher than expected ($w^a > w^e$) the effect of hours is insignificant, but the effect of income is large and highly significant. But when realized income is lower than expected, the effect of income remains important and hours also becomes significant.

In the center panel of Table 2 we control for driver heterogeneity, day of the week, and hour of the day. In the pooled sample, with these controls, income has only an insignificant effect on the stopping probability, while hours worked has a significant effect, apparently supporting Farber's rejection of his income-targeting model. But in our split-sample estimates with these controls, the results change: When realized income is higher than expected ($w^a > w^e$), hours has an effect insignificantly different from zero, while income has a large and highly significant effect. But when realized income is lower than expected ($w^a < w^e$), income has a small, insignificant effect. The marginal effect of hours remains insignificant, but it increases from 0.3% to 1.1%.

In the right-most panel of Table 2 we control additionally for weather and location. In the pooled sample, with these controls, the estimates are similar to those in the left-most panel, except that hours and income now both have significant effects.[16] The split-sample estimates with these controls fully support our reference-dependent model, with income but not hours significantly affecting the stopping probability when the wage is higher than expected and hours but not income significantly affecting the stopping probability when the wage is lower than expected.

[16] Here our estimates differ from Farber's conclusions regarding significance. Our computations closely replicated Farber's point coefficient estimates, but in this case not his estimated standard errors.

Table 2: Probability of Stopping: Marginal Effects for the Probit Model with Linear Effects

Variable	(1)			(2)			(3)		
	Pooled data	$w^a \geq w^e$	$w^a < w^e$	Pooled data	$w^a \geq w^e$	$w^a < w^e$	Pooled data	$w^a \geq w^e$	$w^a < w^e$
Total hours	0.013*	0.005	0.016 **	0.010 ***	0.003	0.011***	0.009*	0.002	0.011***
	(0.009)	(0.009)	(0.007)	(0.003)	(0.004)	(0.008)	(0.006)	(0.005)	(0.002)
Waiting hours	0.010**	0.007	0.016 ***	0.001	0.001	0.002	0.003	0.003	0.005***
	(0.003)	(0.007)	(0.001)	(0.009)	(0.012)	(0.004)	(0.010)	(0.012)	(0.003)
Break hours	0.006 **	0.005 ***	0.004	-0.003	-0.006	-0.003	-0.002	-0.004	-0.002
	(0.003)	(0.001)	(0.008)	(0.006)	(0.009)	(0.004)	(0.007)	(0.009)	(0.001)
Income/100	0.053 ***	0.076 ***	0.055 ***	0.013	0.045***	0.009	0.010 **	0.042***	0.002
	(0.000)	(0.007)	(0.007)	(0.010)	(0.019)	(0.024)	(0.005)	(0.019)	(0.011)
Min temp<30	-	-	-	-	-	-	.002*	0.007*	-0.002
							(.001)	(0.005)	(0.003)
Max temp>80	-	-	-	-	-	-	-0.015***	-0.014***	-0.011***
							(0.003)	(0.006)	(0.002)
Hourly rain	-	-	-	-	-	-	0.014	-0.104	-0.011
							(0.102)	(0.083)	(0.079)
Daily snow	-	-	-	-	-	-	0.006	-0.004 ***	0.020
							(0.011)	(0.000)	(0.022)
Downtown	-	-	-	-	-	-	0.001	0.006 ***	-0.008***
							(0.001)	(0.000)	(0.005)
Uptown	-	-	-	-	-	-	0.001	0.003	-0.004
							(0.012)	(0.010)	(0.005)
Bronx	-	-	-	-	-	-	0.072***	0.032	0.089*
							(0.005)	(0.075)	(0.093)
Queens	-	-	-	-	-	-	0.043**	0.038***	0.086***
							(0.027)	(0.025)	(0.013)
Brooklyn	-	-	-	-	-	-	0.076***	0.101***	0.046***
							(0.015)	(0.028)	(0.003)
Kennedy Airport	-	-	-	-	-	-	0.054***	0.044***	0.059
							(0.018)	(0.004)	(0.055)
LaGuardia Airport	-	-	-	-	-	-	0.059**	0.078	0.000
							(0.034)	(0.055)	(0.023)
Other	-	-	-	-	-	-	0.130	0.067	0.280*
							(0.138)	(0.121)	(0.180)
Driver dummies	-	-	-	Yes	Yes	Yes	Yes	Yes	Yes
Day of week	-	-	-	Yes	Yes	Yes	Yes	Yes	Yes
Hour of day	-	-	-	Yes	Yes	Yes	Yes	Yes	Yes
Log likelihood	-2039.2	-1148.4	-882.6	-1789.5	-1003.8	-753.4	-1767.5	-988.0	-740.0
Pseudo R^2	0.1516	0.1555	0.1533	0.2555	0.2618	0.2773	0.2647	0.2735	0.2901
Observation	13461	7936	5525	13461	7936	5525	13461	7936	5525

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.

To put these results into perspective, note that a neoclassical model would predict that hours have an influence on the probability of stopping that varies smoothly with realized income on any given day, without regard to whether realized 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 model predicts that the probability of stopping is more strongly influenced by realized income than hours when income is higher than expected but more strongly influenced by hours than income when income is lower than expected, with a jump again when the income target is reached and now another jump when the hours target is reached. Qualitatively, all three models' predictions are invariant to precisely how the sample is split. Our estimates are plainly inconsistent with the neoclassical model and—because hours has a strong and significant effect when income is lower than expected but an effect insignificantly different from zero when income is higher than expected—with Farber's income-targeting model as well. However, our estimates are fully consistent with our reference-dependent model.[17]

II.2 The Probability of Stopping: Nonlinear Effects

In order to capture any nonlinear effects realized income and hours may have upon the probability of stopping, Farber (2005) also estimated a probit model where income and hours are represented by categorical variables over the course of a shift and thereby allowed to have unrestricted nonlinear effects. This much more flexible specification of the probability of stopping gives a more accurate picture of how it is related to income and hours. Here we replicate Farber's results for this specification, and then re-do the estimates with the sample split as before.

Table 3 reports the marginal effects from the estimation of the probit model with nonlinear effects. The left-hand panel replicates Farber's (2005) pooled-sample estimates for comparison, while the center and right-hand columns report our split-sample estimates. For each column, we report marginal effects comparing the probability of stopping of each income and hours category to the baseline groups ($150 - $174 income level and the eighth hour). We also report the underlying coefficient estimates and likelihood ratio tests of the hypotheses that the marginal effects of all income or hours groups are jointly zero.

The results for the pooled sample are consistent with what Farber (2005) found: Hours categories have marginal effects that are jointly significantly different from zero, but income categories do not. Allowing a nonlinear hours effect reveals that the effects of hours categories vary widely. By contrast, the overall effect of income categories in the pooled sample is smooth, with few effects differing significantly from the baseline income category of $150 - $174.

When the sample is split, the results change dramatically: In both the center panel, which reports the results for realized income higher than expected ($w^a > w^e$), and the right-most panel, which reports the results for realized income lower than expected ($w^a < w^e$), the effects of all

[17] One concern is that 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. In the low-wage portion of our split sample, the correlation between wage and hours is -0.04, which favors an hours targeting explanation.

income and hours categories are significant. This pattern reflects very different drivers' behavior across the split samples, with lower income and higher hours categories having the predominant influence on the stopping probability when realized income is higher than expected, but higher income and lower hours categories predominant when realized income is lower than expected.

Table 3: Probability of Stopping: Probit Model with Nonlinear Effects

Variable	Pooled data		$w^a > w^e$		$w^a < w^e$	
	Marginal effects	Coefficients	Marginal effects	Coefficients	Marginal effects	Coefficients
Hour						
< 2	-0.041 ***	-0.835 ***	-0.025	-0.478	-0.046 ***	-0.868 ***
	(0.004)	(0.116)	(0.007)	(0.332)	(0.012)	(0.041)
3 – 5	-0.027 ***	-0.382 ***	-0.013	-0.188	-0.020 **	-0.229 **
	(0.005)	(0.080)	(0.016)	(0.303)	(0.003)	(0.105)
6	-0.025 ***	-0.343 ***	-0.015	-0.230	-0.021*	-0.234 *
	(0.003)	(0.040)	(0.008)	(0.183)	(0.005)	(0.142)
7	-0.012 ***	-0.138 ***	-0.009	-0.125	-0.009	-0.093
	(0.004)	(0.049)	(0.006)	(0.106)	(v.004)	(0.063)
9	-0.006	-0.062	-0.020 **	-0.329 **	0.032 *	0.244 *
	(0.015)	(0.166)	(0.005)	(0.166)	(0.016)	(0.145)
10	0.0304 ***	0.253 ***	0.018 ***	0.185 ***	.026	0.209
	(0.010)	(0.058)	(0.007)	(0.050)	(0.037)	(0.215)
11	0.083*	0.549 *	0.091	0.650 *	0.046 ***	0.334 ***
	(0.059)	(0.295)	(0.089)	(0.392)	(0.009)	(0.100)
> 12	0.116 ***	0.691 ***	0.173 ***	0.982 ***	0.042	0.310
	(0.010)	(0.022)	(0.051)	(0.217)	(0.039)	(0.277)
Income						
< 25	-0.035	-0.580	-0.033	-0.856	-0.041	-0.648
	(0.016)	(0.565)	(0.013)	(0.738)	(0.003)	(0.432)
25 – 49	0.005	0.048	-0.024	-0.441	0.012	0.104
	(0.023)	(0.217)	(0.013)	(0.304)	(0.016)	(0.109)
50 – 74	-0.003	-0.029	-0.014	-0.203	-0.025 ***	-0.302 ***
	(0.016)	(0.175)	(0.018)	(0.309)	(0.002)	(0.061)
75 – 99	-0.010 ***	-0.117 ***	-0.022 ***	-0.375 ***	-0.019 ***	-0.211 ***
	(0.003)	(0.044)	(0.007)	(0.109)	(0.006)	(0.029)
100 – 124	-0.009	-0.102	-0.015	-0.230	-0.022 **	-0.257 **
	(0.009)	(0.119)	(0.013)	(0.218)	(0.002)	(0.106)
125 – 149	-0.007	-0.081	-0.015	-0.228	-0.011 ***	-0.113 ***
	(0.005)	(0.063)	(0.010)	(0.157)	(0.003)	(0.010)
175 – 199	0.011 ***	0.100 ***	0.037 ***	0.340 ***	-0.017	-0.183
	(0.001)	(0.003)	(0.013)	(0.072)	(0.009)	(0.164)
200 – 224	0.007 ***	0.068 ***	0.040 ***	0.363 ***	-0.012 *	-0.126 *
	(0.001)	(0.004)	(0.003)	(0.005)	(0.004)	(0.075)
> 225	0.0156	0.142	0.038	0.348	0.058 ***	0.401 ***
	(0.0322)	(0.268)	(0.038)	(0.247)	(0.007)	(0.094)
p-value (likelihood ratio test)						
All Hours = 0	0.0000		0.0000		0.0574	
All Income = 0	0.1096		0.0123		0.0718	
Observations	13461		7936		5525	
Log-likelihood	-1754.380		-970.929		-735.252	

Notes: Significance levels for marginal effects come from the test for the underlying coefficients. *10%, **5%, ***1%. Robust standard errors allowing for the possibility of correlation within a shift are given in parentheses. Control variables for driver fixed effects, day of week, hour of day, location, and weather, are not reported. As in Farber (2005), fixed effects are weighted equally across days of the week and drivers. We use Farber's evaluation point: at 8 total hours, relative to a baseline income of $150 - $174, on a dry day with moderate temperatures in midtown Manhattan at 2:00 p.m.

Figures 2 and 3 graph the probabilities of stopping against income and hours categories. In Figure 2, the marginal effects of hours categories are highly nonlinear when the realized wage is higher than expected, increasing dramatically after the ninth hour. As one would expect with reference-dependence, between the seventh and tenth hours, the marginal effects when realized income is lower than expected are higher than when it is higher than expected; while after the tenth hour, the marginal effect when realized income is higher than expected rises dramatically.

In Figure 3, drivers again behave very differently in the split samples, especially as they approach their income targets. The marginal effect of income jumps up when income reaches \$125 - \$150 in the high-wage case (presumably as the income target is reached before the hours target) and increases dramatically when income reaches \$200 - \$225 in the low-wage case (presumably as the income target is reached after the hours target). Overall, when realized income is higher (lower) than expected, the probability of stopping increases first in response to income (hours) and then hours (income). Once again, our estimates are inconsistent with the neoclassical model and Farber's income-targeting model, but fully consistent with our reference-dependent model. That our split-sample estimates strongly support the model's predictions even with minimal structural restrictions and an imperfect proxy for targets is cause for confidence.

In the pooled sample, the effects of deviations from expectations that show up so strongly in the split samples largely cancel each other out, yielding the aggregate smoothness of the effect of realized income Farber (2005) found. Looking at Figures 2 and 3 together, when realized income is higher than expected the probability of stopping first rises when drivers reach income \$125 - \$175, as our estimate of the income target is reached; but at this income level hours has no significant effect. However, hours begins to have a strong effect after the tenth hour, as our estimate of the hours target is reached. When realized income is lower than expected the pattern is reversed, with the probability responding first to hours and then to income.

II.3 Structural Estimation

This subsection estimates a structural model, which parallels Farber's (2008) model but with positively weighted consumption as well as gain-loss utility and hours as well as income targets. We follow Farber's (2008) econometric strategy, except that instead of treating the targets as latent variables we treat them as rational expectations, operationalized by using average sample realizations of income and hours as proxies for them.[18]

[18] Based on Farber's classification of hours into driving hours, waiting hours and break hours, we use only driving and waiting

Figure 2: Probability of Stopping: Marginal Effect of Hours

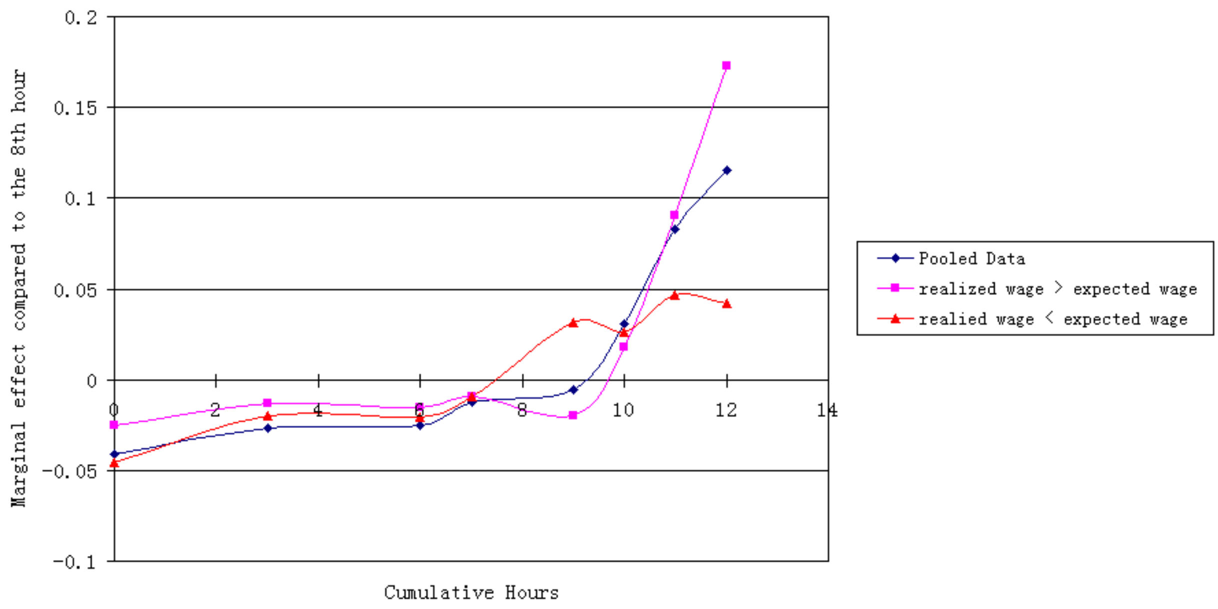
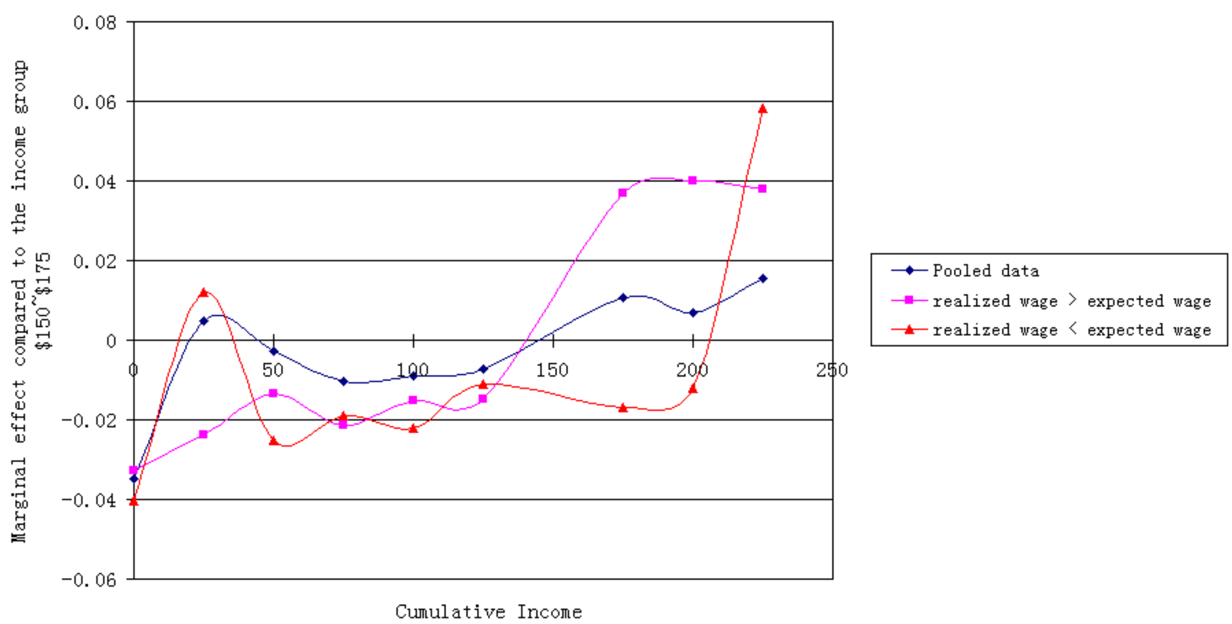


Figure 3: Probability of Stopping:Marginal Effect of Income



hours in our hours calculation. The results are similar when break time is included in the hours target and hours worked.

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. Substituting this 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:

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

where $I_{t+1} = I_t + E(f_{t+1})$ and $H_{t+1} = H_t + E(h_{t+1})$, and $E(f_{t+1})$ and $E(h_{t+1})$ are the next trip's expected fare and time (searching and driving), and ε is a normal error with mean c and variance σ^2.

Online Appendix B gives the details of deriving the likelihood function

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

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 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 $1 - \eta + \eta\lambda$. Although we cannot separately identify η and λ, if we can reject the null hypothesis that $1 - \eta + \eta\lambda = 1$, it follows that $\eta \neq 0$. Further, given the model's restriction that $0 \le \eta \le 1$, our estimates of $1 - \eta + \eta\lambda$ imply lower bounds on λ as explained below.

To make the model operational, we need to specify the shift-level expectations I^r and H^r and the trip-level expectations $E(f_{t+1})$ and $E(h_{t+1})$. As explained above, following Kőszegi and Rabin (2006) we interpret them as a driver's rational expectations, and proxy them via the averages of their natural sample analogs, testing robustness as explained in Section I, footnote 11.

Table 4: Structural Estimates under Alternative Specifications of Expectations				
	(1) Shift (day-of-the-week specific) Trip (naïve)	(2) Shift (general) Trip (naïve)	(3) Shift (day-of-the-week specific) Trip (sophisticated)	(4) Shift (general) Trip (sophisticated)
$1-\eta+\eta\lambda$ +	1.417**	1.254**	2.375***	1.592***
	(.132)	(.113)	(.086)	(.164)
θ	.219*	.176	.090	.022
	(.119)	(.147)	(.133)	(.078)
ρ	.128***	.363***	.390	1.122
	(.025)	(.119)	(.334)	(1.232)
c	.001	.020	-.051	-.024***
	(.043)	(.051)	(.049)	(.020)
σ	.069	.101	.204**	.179***
	(.043)	(.064)	(.085)	(.032)
Observations	13461	13461	13461	13461
Log-likelihood	-1687.8105	-1762.426	-1696.6684	-1761.2436

Notes: Significance levels: *10%, **5%, ***1%. + Significance level corresponds to the null hypothesis that $1-\eta+\eta\lambda=1$. Robust standard errors clustered by shift are given in parentheses. Income variables are divided by 100. Controls include driver fixed effects, day of week, hour of day, location, and weather.

Table 4 columns 1 and 3 report estimates for naïve and sophisticated models (referring to how the trip-level expectations $E(f_{t+1})$ and $E(h_{t+1})$ are formed, as explained shortly), setting I^r and H^r equal to the driver's full-sample averages, day-of-the-week by day-of-the-week ("day-of-the-week specific" in Table 4). Because some drivers have only a few observations for some days of the week, Table 4 columns 2 and 4 report naïve and sophisticated estimates using a second alternative, in which I^r and H^r are aggregated across days of the week, driver by driver ("general"). As a further robustness check, in each case we consider alternative models of how drivers form $E(f_{t+1})$ and $E(h_{t+1})$. Table 4 columns 1 and 2 report estimates for models in which each driver treats trip fares and times as i.i.d. across trips and days, proxied by their average sample realizations, driver by driver ("naïve" in Table 4). Table 4 columns 3 and 4 report estimates for alternative models, in which, in the spirit of Farber's (2005, Section V.C) analysis, drivers form trip-level expectations taking time of day, location, weather, and other relevant variables into account ("sophisticated"). 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.[19]

Table 4's estimates confirm and refine the conclusions of Sections II.1-2's split-sample analyses. The fact that $1-\eta+\eta\lambda$ is significantly greater than one implies that η is significantly different from zero, indicating that the reference-dependent component of drivers' preferences has

[19] The other variables include day-of-the-week and driver dummies, and average hourly wage across drivers for each calendar date 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, drivers, and calendar dates.

positive weight. It also suggests that the coefficient of loss aversion λ is greater than one, with lower bounds ranging from 1.254 to 2.375 across Table 4's alternative specifications, consistent with previous estimates. To get a sense of the possible magnitudes of λ and η, Table 5 reports the values of η implied by our estimates of $1 - \eta + \eta\lambda$ for a range of reasonable values of λ. Different specifications favor gain-loss utility to different degrees, but in general the weight of gain-loss utility is nonnegligible.

Table 5: Illustration of Possible values of λ and η from structural estimation

λ	η			
	(1) Shift (day-of-the-week specific) Trip (naïve)	(2) Shift (general) Trip (naïve)	(3) Shift (day-of-the-week specific) Trip (sophisticated)	(4) Shift (general) Trip (sophisticated)
	$1 - \eta + \eta\lambda$ =1.417	$1 - \eta + \eta\lambda$ =1.254	$1 - \eta + \eta\lambda$ =2.375	$1 - \eta + \eta\lambda$ =1.592
1.5	0.834	0.508	-	-
2	0.417	0.254	-	0.592
2.5	0.278	0.169	0.917	0.395
3	0.209	0.127	0.688	0.296
3.5	0.167	0.102	0.550	0.237
4	0.139	0.085	0.458	0.197
4.5	0.119	0.073	0.393	0.169
5	0.104	0.064	0.344	0.148

Table 6: Estimated Optimal Stopping Times (in Hours)

Percentile in the wage distribution	Hourly wage	(1) Shift (day-of-the-week specific) Trip (naïve)	(2) Shift (general) Trip (naïve)	(3) Shift (day-of-the-week specific) Trip (sophisticated)	(4) Shift (general) Trip (sophisticated)
		θ = 0.219 ρ = 0.128 $1 - \eta + \eta\lambda$ =1.417	θ =0.176 ρ = 0.363 $1 - \eta + \eta\lambda$ =1.254	θ =0.090 ρ = 0.390 $1 - \eta + \eta\lambda$ =2.375	θ = 0.022 ρ = 1.122 $1 - \eta + \eta\lambda$ =1.592
5%	\$17.9	3.150	1.954	6.899	6.899
10%	\$19.1	5.229	2.337	6.899	6.899
25%	\$21.0	6.899	3.034	7.681	7.469
50%	\$23.3	6.899	4.041	6.923	6.923
75%	\$25.9	6.899	5.408	6.899	6.899
90%	\$28.5	6.899	5.660	6.899	6.899
95%	\$30.8	6.899	5.237	6.899	6.899
Correlation of wage and optimal working hours		0.709	0.942	-0.256	-0.257

To illustrate the implications of the estimated utility function parameters under Table 4's alternative specifications, Table 6 presents the optimal stopping times, in hours, implied by our structural estimates of the reference-dependent model for each specification and for representative percentiles of the observed distribution of realized wages. The implied stopping times seem quite reasonable, especially for the sophisticated models, reflecting the lower estimated disutilities of hours for those models.[20]

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). 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 6 also reports each specification's implication for the aggregate correlation of wage and optimal working hours, a proxy for the wage elasticity. The sophisticated cases (columns 3 and 4), with more reasonable estimates of the disutility of hours, imply negative correlations (each close to the aggregate sample correlation of -0.2473). By contrast, the naïve cases, with unreasonably high consumption disutility for hours, imply positive correlations.[21]

Overall, 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'

---

[20] Estimates that allow for differences across days of the week tend to imply lower estimates of the elasticity of the marginal rate of substitution ρ, making the neoclassically optimal stopping times extremely sensitive to the wage. Table D1 in Online Appendix D gives the implied average stopping probabilities for various ranges relative to the targets. The estimates imply comparatively little bunching around the targets, perhaps because consumption utility has almost the same weight as gain-loss utility. Even so, the targets have a very strong influence on the stopping probabilities: As in the nonlinear split-sample estimates (Table 3, Figures 2-3), the second-reached target has a stronger effect than the first-reached target.

[21] It would be ideal if we could calculate the wage elasticity of hours implied by our structural model, to compare with Farber's (2005, Table 3) estimates. But because our structural model only estimates the probability of stopping at the end of each trip, there is no completely sensible way to infer a driver's optimal hours. If we approximate by equating a driver's optimal hours with the trip for which he has the highest probability of stopping, the wage elasticities for the four models in Table 6 are respectively -0.730, -0.702, -0.683, and -0.674, close to Farber's estimates, which range from -0.637 to -0.688. The estimated optimal stopping times in Table 6 provides another possible way to approximate the wage elasticity implied by our model, but this seems less informative because it is sensitive to our simplifying assumption that the daily wage is constant.

expectations, and comfortably within the range that indicates reference-dependent preferences. Our estimates suggest that a more comprehensive investigation of how drivers forecast their income from experience, with larger datasets, will yield a useful model of reference-dependent of driver's labor supply that significantly improves upon the neoclassical model.[22]

III. Conclusion

Although the neoclassical model of labor supply predicts a positive wage elasticity of hours, Camerer et al. (1997) found a strongly negative elasticity of hours with respect to realized earnings for New York City cabdrivers. Farber (2005, 2008), analyzing new data on a different set of New York City cabdrivers, found a similarly negative relationship. To explain these results, Camerer et al. proposed an income-targeting explanation, in the spirit of Prospect Theory, 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 is high, with little of the positive wage elasticity of hours predicted by the neoclassical model.

Farber (2008) was the first to estimate a model explicitly derived from income targeting, using the dataset created for Farber (2005). He estimates 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. He finds that a sufficiently rich parameterization of his income-targeting model has a better fit than a standard neoclassical specification, and that the probability of stopping increases significantly and substantially once the income target is reached; but that his model cannot reconcile the strong increase in stopping probability at the target with the aggregate smoothness of the relationship between stopping probability and realized income. Further, the random effect is large but imprecisely estimated, from which he concludes that drivers' income targets are too unstable and imprecisely estimated to yield a useful reference-dependent model of labor supply.

In this paper we use Farber's data to estimate a model based on Kőszegi and Rabin's (2006) theory of reference-dependent preferences, which is more general than Farber's model in most respects but takes a more specific position on how targets are determined. In the model, a driver

[22] We believe that Farber's (2008) finding to the contrary may have been due to his decision to treat the income target as a latent variable; to the constraints he imposed, partly for computational rather than economic reasons, in estimating it; and to the fact that he estimated his income-targeting model and neoclassical models separately instead of nesting them as suggested by Kőszegi and Rabin's (2006) theory. Further, although Farber 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, our analysis reduces the difference in degrees of freedom by defining the targets via rational expectations.

has a daily reference point or target for hours as well as income. His preferences reflect both the standard consumption utility of income and leisure and gain-loss utility. As in Farber's (2008) income-targeting model, the driver is loss-averse; but working longer than the hours target is now a loss, just as earning less than the income target is. We follow Farber's (2005, 2008) econometric strategies closely, but instead of treating targets as latent variables we treat them as rational expectations, operationalized using average sample realizations of income and hours as proxies.

Estimating linear and nonlinear probit models of the probability of stopping as in Farber's (2005) analysis, but with the sample split according to whether realized income is higher or lower than our proxy for a driver's expected income on a given day, we find very clear evidence of reference-dependence, with the probability of stopping work strongly influenced by realized income (but not hours) when the realized wage is higher than expected, and by hours (but not income) when the wage is lower than expected. The two wage regimes have roughly equal weights, so that the heterogeneity of realized wages yields a smooth aggregate relationship between stopping probability and realized income, as in Farber's (2005) full-sample estimation. But when the wage is higher than expected the wage elasticity of hours is strongly negative, and when it is lower the elasticity is zero. This allows the model to reconcile Farber's finding that in the full sample stopping probabilities are significantly related to hours but not income, with the negative aggregate wage elasticity of hours found by Camerer et al. (1997).

Using the full sample to estimate a structural reference-dependent model as in Farber (2008), but with the modifications suggested by Kőszegi and Rabin's theory, confirms and refines the results of our split-sample analysis. The parameter estimates are sensible; and the key function of the parameters of gain-loss utility is plausibly and precisely estimated, and comfortably within the range that indicates reference-dependent preferences. Finally, the model avoids Farber's (2008) criticism that drivers' targets are too unstable and imprecisely estimated to yield a useful reference-dependent model of labor supply. It does this partly by nesting consumption and gain-loss utility and allowing hours as well as income targets, but mostly by treating the targets as rational expectations estimated from natural sample proxies, rather than as latent variables.

Overall, our estimates suggest that a more comprehensive investigation of how drivers forecast their income from experience, with larger datasets, is likely to yield a reference-dependent model of drivers' labor supply that significantly improves upon the neoclassical model. However, although our results suggest that Kőszegi and Rabin's rational-expectations theory of targets is very promising, like Farber we take the targets as given rather than modeling how they are determined. Pending further analysis of how expectations are formed and adjusted over time, the message of our analysis for labor supply is limited.

References

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.

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(?): ???-???.

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: Probit Model with Linear Effects and Full Set of Controls

Variable	(1)			(2)			(3)		
	Pooled data	$w^a > w^e$	$w^a \leq w^e$	Pooled data	$w^a > w^e$	$w^a \leq w^e$	Pooled data	$w^a > w^e$	$w^a \leq w^e$
Total hours	0.087 * (0.048)	0.037 (0.060)	0.091 ** (0.037)	0.114 *** (0.042)	0.040 (0.046)	0.112 *** (0.043)	0.107 * (0.064)	0.032 (0.059)	0.117 *** (0.013)
Waiting hours	0.067 ** (0.028)	0.053 (0.062)	0.093 *** (0.003)	0.011 (0.095)	0.015 (0.169)	0.025 (0.048)	0.040 (0.113)	0.040 (0.177)	0.058 *** (0.015)
Break hours	0.038 ** (0.016)	0.040 *** (0.002)	0.025 (0.046)	-0.032 (0.067)	-0.075 (0.119)	-0.030 (0.035)	-0.023 (0.076)	-0.053 (0.114)	-0.027 (0.022)
Income/100	0.343 *** (0.041)	0.591 *** (0.056)	0.315 *** (0.052)	0.146 (0.127)	0.604 *** (0.189)	0.097 (0.296)	0.120 ** (0.054)	0.595 *** (0.150)	0.019 (0.122)
Min temp < 30	-	-	-	-	-	-	0.018 (0.014)	0.086 * (0.046)	-0.017 (0.041)
Max temp > 80	-	-	-	-	-	-	-0.212*** (0.042)	-0.262*** (0.079)	-0.131*** (0.023)
Hourly rain	-	-	-	-	-	-	0.165 (1.178)	-1.481 (1.46)	-0.121 (0.912)
Daily snow	-	-	-	-	-	-	0.073 (0.132)	-0.056*** (0.007)	0.221 (0.171)
Downtown	-	-	-	-	-	-	0.005** (0.009)	0.082*** (0.019)	-0.093*** (0.031)
Uptown	-	-	-	-	-	-	0.005 (0.134)	0.039 (0.141)	-0.050 (0.077)
Bronx	-	-	-	-	-	-	0.535 (0.023)	0.332 (0.560)	0.600* (0.322)
Queens	-	-	-	-	-	-	0.363 * (0.175)	0.383*** (0.140)	0.586*** (0.173)
Brooklyn	-	-	-	-	-	-	0.553 (0.073)	0.744*** (0.061)	0.371*** (0.106)
Kennedy Airport	-	-	-	-	-	-	0.434 (0.103)	0.424*** (0.027)	0.450 (0.406)
LaGuardia Airport	-	-	-	-	-	-	0.459 (0.196)	0.630*** (0.226)	0.004 (0.247)
Other	-	-	-	-	-	-	0.795 (0.543)	0.569 (0.636)	1.260 (0.646)
Driver dummies	-	-	-	Yes	Yes	Yes	Yes	Yes	Yes
Day of week	-	-	-	Yes	Yes	Yes	Yes	Yes	Yes
Hour of day	-	-	-	Yes	Yes	Yes	Yes	Yes	Yes
Log likelihood	-2039.2	-1148.4	-882.6	-1789.5	-1003.8	-753.4	-1767.5	-988.0	-740.0
Pseudo R^2	0.1516	0.1555	0.1533	0.2555	0.2618	0.2773	0.2647	0.2735	0.2901
Observation	13461	7936	5525	13461	7936	5525	13461	7936	5525

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

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), and ε is a normal error with mean c and variance σ^2.

The likelihood function can now be written as:

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

where i denotes the shift and t the trip within a given shift, and

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

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

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

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

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

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

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

Online Appendix C: Trip Fares and Time Estimates Whose Fitted Values are Used as Proxies for Drivers’ Expectations in Table 4

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

	Time	Fare	Day of the Week	Time	Fare
Clock hours			Day of the Week		
0	-.079	-.005	Monday	.019**	.002
	(.054)	(.009)		(.008)	(.002)
1	-.060	-.002	Tuesday	.011	.002
	(.056)	(-.002)		(.007)	(.002)
2	-.040	.003	Wednesday	.019***	.003*
	(.059)	(.010)		(.007)	(.002)
3	-.030	-	Thursday	.026***	.006***
	(.065)	-		(.007)	(.002)
4	-	.008	Friday	.019***	.005***
	-	(.015)		(.007)	(.002)
5 - 10	-.040	-.005	Saturday	-	-
	(.053)	(.009)		-	-
11	-.025	-.006	Sunday	.007	.005**
	(.054)	(.009)		(.009)	(.002)
12	-.033	-.006	ID 1	.049**	.003
	(.054)	(-.006)		(.021)	(.005)
13	-.034	-.003	ID 2	.022	.011*
	(.054)	(.009)		(.025)	(.006)
14	-.032	-.002	ID 3	-	-
	(.054)	(.009)		-	-
15	-.046	-.001	ID 4	.027	.001
	(.054)	(.009)		(.020)	(.005)
16	-.060	-.005	ID 5	.070 ***	.013***
	(.054)	(.009)		(.022)	(.005)
17	-.074	-.007	ID 6	.008	.005
	(.053)	(.009)		(.025)	(.006)
18	-.079	.010	ID 7	.036*	.003
	(.053)	(.009)		(.022)	(.005)
19	-.095*	-.012	ID 8	.042**	.002
	(.053)	(.009)		(.020)	(.005)
20	-.069	-.005	ID 9	.013	.001
	(.053)	(.009)		(.021)	(.001)
21	-.091*	-.006	ID 10	.029	.002
	(.053)	(.009)		(.020)	(.005)
22	-.089*	-.005	ID 11	-.006	.002
	(.053)	(.009)		(.026)	(.006)
23	-.060	.002	ID 12	.063***	.009*
	(.054)	(.009)		(.024)	(.005)
Mini temp < 30	-.001	-.002	ID 13	-.027	-.008
	(.006)	(.001)		(.023)	(.005)
Max temp > 80	.023***	.003**	ID 14	.043*	.009
	(.006)	(.001)		(.022)	(.005)
Hourly rain	.003	-.015	ID 15	-.003	-.007
	(.091)	(.021)		(.022)	(.005)

Daily snow	.000	.000	ID 16	.033	.004
	(.004)	(.001)		(.020)	(.005)
Downtown	-.083**	.001	ID 17	.031	.000
	(.041)	(.009)		(.031)	(.005)
Midtown	-.120***	-.007	ID 18	.053***	.008
	(.041)	(.009)		(.019)	(.004)
Uptown	-.103 **	-.005	ID 19	.103***	.021***
	(.041)	(.009)		(.020)	(.005)
Bronx	-	-	ID 20	.012	-.000
	-	-		(.019)	(.004)
Queens	.230***	.086 ***	ID 21	.054***	.003
	(.049)	(.011)		(.021)	(.005
Brooklyn	.097**	.040 ***	Average wage+	-.362***	.005**
	(.046)	(.010)		(.076)	(.017)
Kennedy Airport	.439***	.173***	Constant	.475***	.048***
	(.046)	(.011)		(.072)	(.014)
LaGuardia Airport	.231***	.106***			
	(.043)	(.010)			
Others	-.037	.012			
	(.054)	(.012)			
R^2	0.1257	0.1865		0.1257	0.1865
Observations	12877			12877	

Notes: Significance levels: * 10%, ** 5%, *** 1%. Fare and time (waiting and driving) for the next trip are jointly estimated as seemingly unrelated regressions. + Average hourly wage across drivers for the same calendar date.

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) Shift (day-of-the-week specific) Trip (naïve)	(2) Shift (general) Trip (naïve)	(3) Shift (day-of-the-week specific) Trip (sophisticated)	(4) Shift (general) Trip (sophisticated)
$w^a > w^e$				
Before income target	.022	.025	.023	.025
At income target	.161	.124	.165	.130
In between two targets	.115	.102	.134	.120
At hours target	.238	.167	.233	.166
Above hours target	.287	.234	.278	.227
$w^a \leq w^e$				
Before hours target	.022	.024	.031	.030
At hours target	.139	.134	.149	.135
In between two targets	.168	.164	.178	.149
At income target	.266	.245	.282	.260
Above income target	.283	.234	.305	.254

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.