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

Vincent P. Crawford and Juanjuan Meng
University of California, San Diego

UC Berkeley Psychology and Economics/Labor Economics Seminar
9 September 2008

Institute of Economics, Academia Sinica, Taipei, Seminar
7 October 2008

University of Oxford Seminar
20 November 2008

Introduction

In the absence of large income effects, the neoclassical model of labor supply predicts a positive wage elasticity of hours.

But recent studies of workers who choose their own hours reach ambiguous conclusions on this issue:

- Camerer, Babcock, Loewenstein, and Thaler (1997 *QJE*) found a strongly negative elasticity of hours with respect to realized daily earnings for New York City cabdrivers, especially for inexperienced drivers.

- Farber (2005 *JPE*, 2008 *AER*), analyzing new data on a different set of New York City cabdrivers, found a similarly negative relationship.

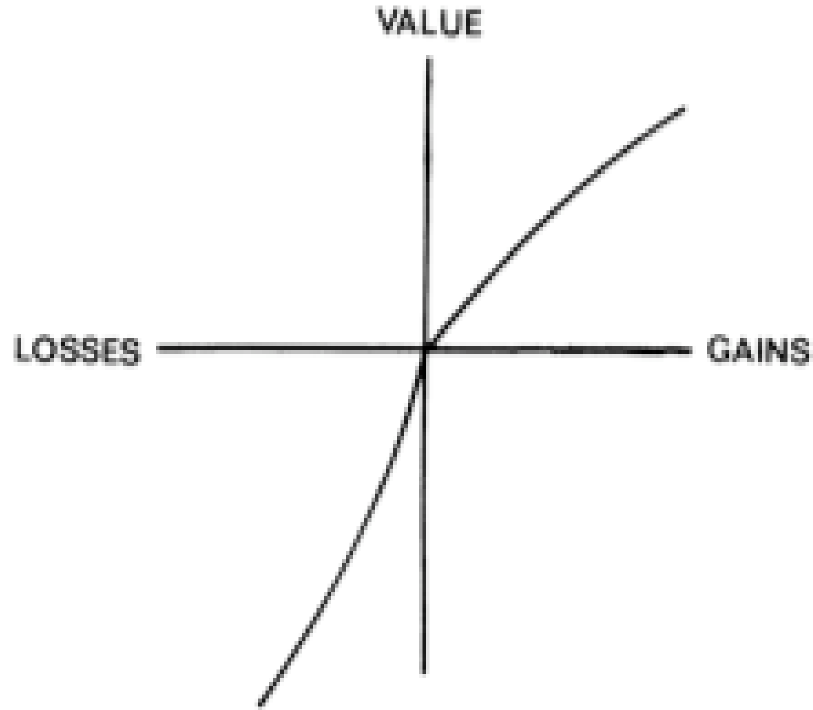
To explain their results Camerer et al. informally proposed a model in which drivers have daily income targets and work until the target is reached.

They therefore tend to work less on days when realized earnings per hour (the natural analog of the wage, which we shall call it) are high.

Camerer et al.'s explanation is in the spirit of Kahneman and Tversky's (1979 *Econometrica*) Prospect Theory, in which:

- A person's preferences respond not only to income 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 above it ("gains").



In the proposed explanation, the reference point is a daily income target.

Loss aversion creates a kink that tends to make realized income respond to the target as well as the wage, and bunch around the target.

As a result, realized hours have little or none of the positive wage elasticity predicted by a neoclassical model.

As Farber (2008 *AER*) 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 (2005 *JPE*) found only mixed evidence for income targeting:

- Before controlling for driver fixed effects, the probability of stopping work is significantly related to income realized on a given day, but

- Driver fixed effects and other relevant controls render this effect statistically insignificant, and

- The probability of stopping is significantly related to cumulative hours.

Further, other studies of workers who choose their hours have found positive relationships between *expected* earnings and labor supply, as suggested by the neoclassical model:

- Oettinger (1999 *JPE*) finds that stadium vendors are more likely to go to work on days when their wage can be expected to be higher; and

- Fehr and Goette (2007 *AER*) find that bicycle messengers sign up for more shifts when their commissions are experimentally increased.

Farber (2008 *AER*) reexamines the evidence, using his 2005 dataset to estimate a structural model explicitly derived from reference-dependence, with daily income targets.

He estimates drivers' income targets as latent variables with driver-specific means and driver-independent variance.

He assumes, mainly for computational reasons, that both mean and variance of income are constant across days of the week, thus allowing the target to vary across days for a given driver, but only as a random effect.

(This assumption is strongly rejected in the data, with Fridays', Saturdays', and Sundays' mean incomes systematically higher than those of other days.

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 income target to vary across days of the week.)

Farber (2008 *AER*) finds that a sufficiently rich parameterization of his reference-dependent model has a better fit than a standard neoclassical specification.

The estimated probability of stopping increases significantly and substantially once the income target is reached.

But his income targeting 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 effects in drivers' targets are large with high estimated variances, from which Farber (2008 *AER*) concludes that the targets are too unstable and imprecisely estimated to yield a useful reference-dependent model of labor supply:

> "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."

Kőszegi and Rabin (2006 *QJE*) recently 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.

In Kőszegi and Rabin's theory as applied to cabdrivers' labor supply:

- A driver's preferences reflect both the standard consumption utility of income and leisure and reference-dependent "gain-loss" utility, with their relative importance tuned by an estimated parameter.

- A driver has a daily target for hours as well as income, and as in Farber's model he is loss-averse, but working longer than the hours target is now a loss, just as earning less than the income target is.

- Most importantly for our analysis, the targets are endogenized by setting them equal to a driver's theoretical rational expectations of hours and income (Kőszegi and Rabin's notion of "preferred personal equilibrium").

As Kőszegi and Rabin (2006, Section V) suggest, their model's treatment of the targets as rational expectations and its distinction between the effects of anticipated and unanticipated wage increases has the potential to reconcile:

- The negative wage elasticity of hours found by Camerer et al. (1997 *QJE*) and Farber (2005 *JPE*, 2008 *AER*).

- The positive relationships between *expected* earnings and labor supply found by Oettinger (1999 *JPE*), Fehr and Goette (2007 *AER*), and others.

Our paper reconsiders whether reference-dependent preferences allow an empirically 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.

We operationalize those expectations 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 as Kőszegi and Rabin's (2006) theory suggests.

We show that a Kőszegi and Rabin-style model can:

- Reconcile the negative wage elasticity of hours found by Camerer et al. and Farber with the positive relationships between *expected* earnings and labor supply found by Oettinger, Fehr and Goette, and others.

- Reconcile the smoothness of the relationship between stopping probability and realized income Farber found.

And (despite Farber's negative conclusion) it can:

- Yield estimates of the targets that are stable and sufficiently precisely estimated to yield a useful reference-dependent model of labor supply.

Outline

1. Remarks on neoclassical versus reference-dependent models of labor supply and econometric testing

2. Adapting Kőszegi and Rabin's model to cabdrivers' labor supply

3. Econometric estimates of linear and nonlinear probit models of the probability of stopping as in Farber's (2005) analysis

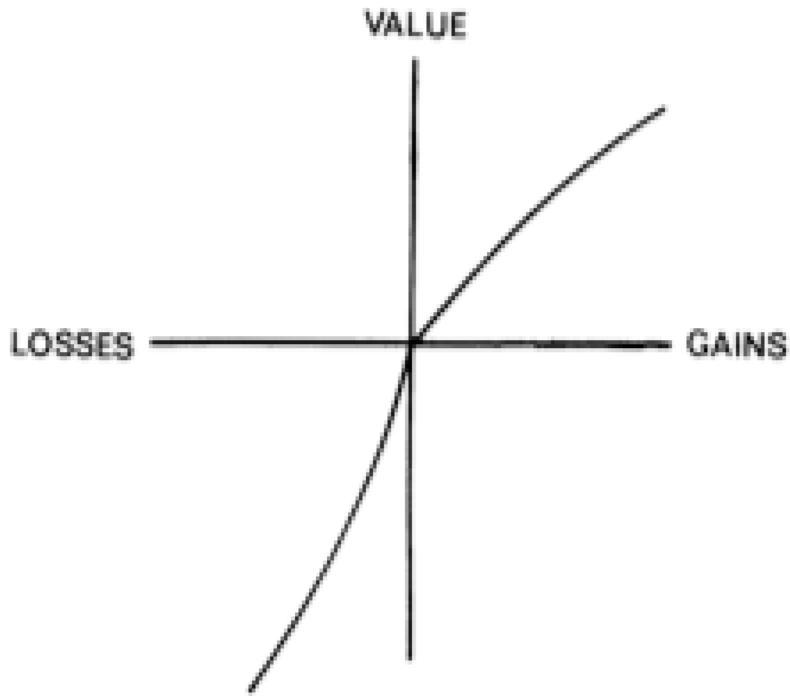
4. Econometric estimates of a structural reference-dependent model as in Farber's (2008) analysis, with changes suggested by Kőszegi and Rabin's Model

5. Conclusion

Remarks on neoclassical versus reference-dependent models of labor supply and econometric testing

How do Kőszegi and Rabin’s and our models relate to standard neoclassical models of labor supply?

And what new issues do they raise in econometric testing?

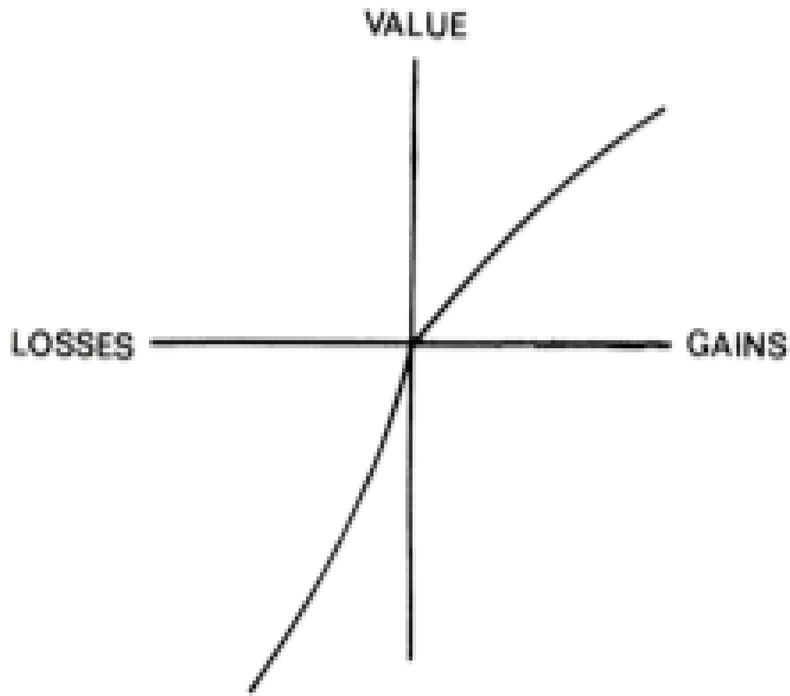


Prospect Theory departs from neoclassical theory in three main ways:

(a) Reference-dependence and loss aversion (people are less sensitive to changes above their reference point ("gains") than below it ("losses")).

(b) "Diminishing sensitivity" (concavity for gains but convexity for losses).

(c) "Nonlinear probability weighting" (overweighting small probabilities).



(a) Reference-dependence expands the domain of preferences to include the reference point, but is consistent with the standard notion of rationality as choice consistency.

(b) Diminishing sensitivity is unfamiliar and may make the objective function nonconcave, but it is fully consistent with rationality.

(c) Nonlinear probability weighting is plainly inconsistent with rationality.

We follow Kőszegi and Rabin in keeping reference-dependence, for which there is a great deal of evidence, but dropping diminishing sensitivity and nonlinear probability weighting (for which there are also evidence, but less).

Thus our models are fully consistent with rationality, with concave objective functions.

The only important deviation from a neoclassical model is adding reference points to income and leisure in the domain of preferences.

With regard to econometric testing, the kink at the reference point is not important per se.

What is important is that there is something that varies independently of income and leisure to which preferences (as revealed by choices) systematically respond.

The very large body of experimental evidence on reference-dependence and loss aversion starting with Kahneman and Tversky (1979) strongly suggests that deviations from neoclassical preferences are common, and that almost all of them are in the direction of loss rather than gain aversion.

Further, people’s sensitivities to changes in income or leisure above their reference points (gains) are roughly half as large as people’s sensitivities to changes in income or leisure below their reference points (losses).

One could test a model that allows reference-dependence even without a specification that links reference points to data, either taking a nonparametric approach or, like Farber, taking a structural approach that treats the reference points as latent variables.

However, in Farber's dataset his latent-variable approach causes computational problems, which were what led him to conclude that the income targets in his model are too unstable and imprecisely estimated to yield a useful reference-dependent model of labor supply.

The plausible additional structure we add by treating the reference points as rational-expectations to some extent avoids those problems, and allows us to test the model by looking for systematic, predictable shifts in preferences associated with the reference points.

This yields estimates of the targets that are stable and sufficiently precisely estimated to yield a useful reference-dependent model of labor supply.

Adapting Kőszegi and Rabin's model to cabdrivers' labor supply

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

I and H denote his income earned and hours worked that day.

I^r and H^r denote his income and hours targets for the day.

His total utility, $V(I, H|I^r,H^r)$, is 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 ($0 \leq \eta \leq 1$):

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

(1)-(2) incorporate several of Kőszegi and Rabin's provisional assumptions:

- 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.

- Gain-loss utility is also separable, determined component by component by differences between realized and target consumption utilities.

- Gain-loss utility is a linear function of those utility differences, ruling out Prospect Theory's "diminishing sensitivity" as in a leading case Kőszegi and Rabin sometimes focus on (their Assumption A3').

- Losses have a constant weight λ relative to gains, "the coefficient of loss aversion," which is the same for income and hours. Empirically, $\lambda \approx 2$ to 3.

(1)-(2) depart from Kőszegi and Rabin in treating drivers' targets as deterministic point expectations, a natural simplification given that our model (unlike theirs) makes explicit allowance for errors and therefore can have gains and losses even with point expectations. (This may exaggerate the effect of loss aversion, and if anything it biases the comparison against a reference-dependent model and in favor of a neoclassical model.)

We follow Kőszegi and Rabin in equating the income and hours targets I^r and H^r to drivers' rational expectations, proxied as explained below.

If gain-loss utility has small weight, Kőszegi and Rabin's model approaches a neoclassical model, with standard implications for labor supply.

Even when gain-loss utility has large weight, the standard implications carry over for changes in the wage that are perfectly anticipated.

But when realized wages deviate from expected, the probability of stopping is more strongly influenced by hours or income, depending on which target is reached first, and the model deviates from a neoclassical model.

When the wage is lower than expected the hours target tends to be reached first, hours have a stronger influence on stopping probability, and the wage elasticity of labor supply is pushed toward zero.

But when the wage is higher than expected the income target tends to be reached first, and its stronger influence on stopping probability can make even a driver who values income but is "rational" in the generalized sense of Prospect Theory have a negative wage elasticity.

Because the elasticity of labor supply is negative in the former wage regime but near zero in the latter, the aggregate 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.

Further, the heterogeneity of realized wages yields a smooth aggregate relationship between stopping probability and realized income.

Thus, Kőszegi and Rabin's model can also reconcile Farber's finding that aggregate stopping probabilities are significantly related to hours but not income with a negative aggregate wage elasticity of hours.

Because drivers' earnings are determined randomly rather than by a known wage rate, they must form expectations after each trip about their earnings per hour if they continue work that day.

Farber (2005) argues 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 option value) 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, we simplify to illustrate our model's possibilities by assuming that drivers extrapolate their daily income linearly, assuming a constant expected hourly wage rate w^a and ignoring option value.

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, thus allowing the targets to vary across days of the week, but ignoring sampling variation for simplicity.

Given that $\lambda \geq 1$ our model allows a simple characterization of a driver's optimal stopping decision with targets for hours as well as income.

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 $I = w^a H$.

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. (This depends on ruling out "diminishing sensitivity".)

Thus the driver's decision is characterized by a first-order condition, generalized to allow kinks at the reference points: He continues if and only if the anticipated wage w^a exceeds the relevant marginal rate of substitution.

Table 1 lists the marginal rates of substitution in the four possible gain-loss regions, expressed as hours disutility costs of an additional unit of income.

Table 1. Marginal Rates of Substitution with Reference-Dependent Preferences		
	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*)	$-[U_2'(H)/U_1'(I)]/[1-\eta+\eta\lambda]$	$-U_2'(H)/U_1'(I)$

When hours and income are both in the gains or 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.

(On boundaries, marginal rates of substitution are replaced by generalized derivatives whose left- and right-hand values equal the interior values.)

When hours and income are in opposite domains, the marginal rate of substitution equals the consumption-utility trade-off times either $(1 - \eta + \eta\lambda)$ (> 1 when $\lambda > 1$) or $1/(1 - \eta + \eta\lambda)$.

(The tradeoff favors work more than the neoclassical tradeoff in the income loss/hours gain domain, but less in the hours loss/income gain domain.)

Figure 1 illustrates the driver's optimal stopping decision when $w^a > w^e$, so realized income is higher than expected, income target is reached before hours target ($w^a = I/H > w^e$ and $H = H^r = I^r/w^e$ imply $I = w^a H = w^a I^r/w^e > I^r$).

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 $(H_0, I_0) = (0, 0)$, and anticipates moving along a sample line $I = w^a H$ with constant w^a.

As time passes he heads northwest along a random but monotone path, which is approximately continuous (the average trip length is 12 minutes).

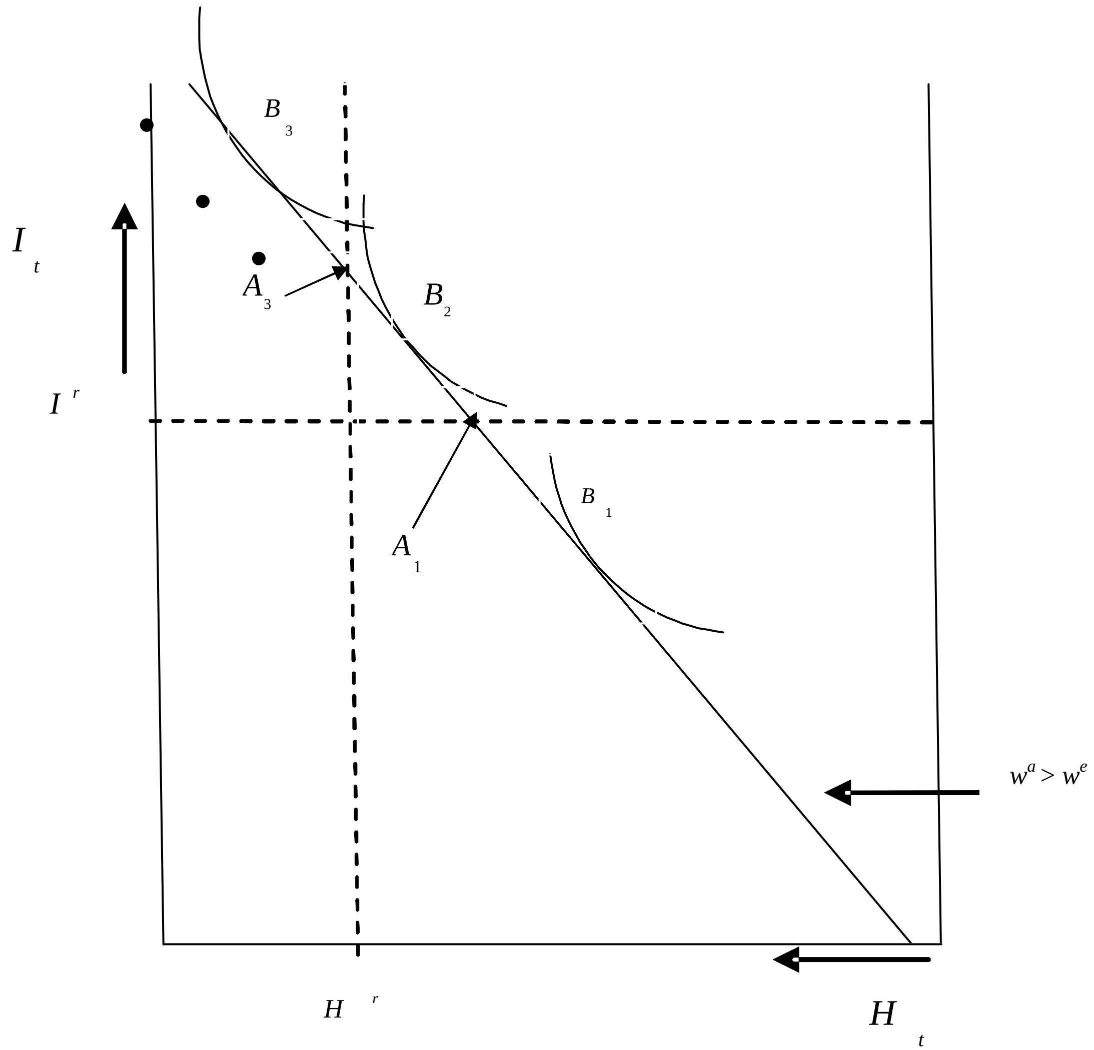


Figure 1: A Reference-dependent Driver's Stopping Decision

The three indifference curves with tangency points B_1, B_2, and B_3 represent possible alternative income-hours trade-offs for consumption utility.

Starting at $(I_0, H_0) = (0, 0)$ in the income-loss/hours-gain domain, the driver continues working as long as the anticipated wage w^a exceeds the hours disutility cost of an additional unit of income.

In this domain the comparison favors working more than the neoclassical one; but for a given w^a the tradeoff becomes (weakly) less and less favorable as income and hours accumulate.

If the hours disutility cost of income rises to w^a before the driver reaches his first target—with $w^a > w^e$, income—he stops at a point between B_1 and A_1, where B_1 maximizes consumption utility 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 line segment from B_1 to A_1.)

If the hours disutility cost of income remains below w^a until he reaches his income target, he compares w^a with the cost in the domain he is entering—income-gain/hours-gain—and stops if and when the new hours disutility cost of income, $-U_2'(H)/U_1'(I)$, exceeds w^a; and so on.

Whether or not $w^a > w^e$, a driver who extrapolates income linearly anticipates passing through a series of domains such that the hours disutility cost of income weakly increases as hours and income accumulate—reflecting the concavity of reference-dependent utility in *I* and *H*.

Thus, given our strong assumptions about the driver's expectations, the decision characterized here is globally optimal. (With more realistic assumptions, the conclusions would be similar but messier.)

Econometric estimates

Our econometric estimates use Farber's (2005, 2008) data and closely follow his econometric strategies.

But instead of treating income targets as latent variables, we treat them as rational expectations.

We operationalize these expectations by using average sample realizations of income and hours as proxies for them, dealing with simultaneity problems 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 as Kőszegi and Rabin's (2006) theory suggests.

Data description

Farber's data are now posted on the *AER* website with Farber (2008). Farber (2005) describes his data cleaning and relevant statistics.

The data are converted from 584 trip sheets recorded by 21 drivers from June 2000 to May 2001.

Trip sheets contain information about starting/ending time/location and fare (excluding tips) for each trip.

Based on Farber's classification of hours into driving hours, waiting hours and break hours, we use only driving and waiting hours. (The results are similar when break time is included.)

Farber also collected data about weather conditions for control purposes.

Drivers lease their cabs weekly, so are free to choose hours day by day.

Because each driver's starting and ending hours vary widely, and 11 of 21 work some night and some day shifts, subleasing seems unlikely.

Econometric estimates of linear and nonlinear probit models of the probability of stopping as in Farber's (2005) analysis

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).

If a driver forms his expectations by extrapolating earnings approximately linearly, he tends to reach his income target first when his realized wage at the end of a day, w^a, is higher than expected.

Accordingly, using our rational-expectations proxy for his expected wage, we estimate linear and nonlinear models that parallel Farber's, but splitting the sample, day by day, according to whether or not a driver's realized wage at the end of a day, w^a, is higher than w^e, his expected wage, proxied by his full-sample mean for that day of the week.

(Splitting the sample in a way that depends on partly endogenous variables creates potential simultaneity problems; but a number of robustness checks, reported in the paper, suggest that they are not important here.)

For a wide class of reference-dependent models, including our structural model, the probability of stopping work increases sharply at the first-reached target and again at the second-reached target.

By contrast, in a neoclassical model, the targets per se have no effect.

Thus, our reference-dependent model predicts large differences in stopping probabilities across the two wage regimes, independent of structural details.

This sharply distinguishes it from a neoclassical model even if our proxy for expectations is imperfect, and without invoking most structural restrictions.

Linear Probits

Table 2 reports the marginal probability effects from the estimation of the probit model with linear effects. (Table A1 in the paper's online appendix reports the marginal effects for the model with the full set of controls.)

In each panel, the left-hand column replicates Farber's pooled-sample estimates; center and right-hand columns report our split-sample estimates.

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)
Snip for (3)…									
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

In judging the results in Table 2, the following observations are relevant:

- A neoclassical model would predict that hours have an influence on stopping probability that varies smoothly with realized income on any given day, whether or not the wage is higher than expected.

- A pure income-targeting model such as Farber's 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 strongly influenced by realized income when the wage is higher than expected, so the income target is reached first, with a jump again when the hours target is reached; and that the probability of stopping is strongly influenced by hours when the wage is lower than expected, so the hours target is reached first, with a jump again when the income target is reached.

These predictions are qualitatively robust to imperfections in our sample-splitting criterion.

In comparing the results in Table 2 with these models' predictions, both the magnitudes and significance levels of the coefficient estimates matter.

In the left-most panel (1), only total hours, total waiting hours, total break hours and income at trip end are used to explain the stopping probability.

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*** **(0.009)**	**0.005** **(0.009)**	**0.016 **** **(0.007)**	0.010 *** (0.003)	0.003 (0.004)	0.011*** (0.008)	0.009* (0.006)	0.002 (0.005)	0.011*** (0.002)
Waiting hours	0.010** (0.003)	0.007 (0.007)	0.016 *** (0.001)	0.001 (0.009)	0.001 (0.012)	0.002 (0.004)	0.003 (0.010)	0.003 (0.012)	0.005*** (0.003)
Break hours	0.006 ** (0.003)	0.005 *** (0.001)	0.004 (0.008)	-0.003 (0.006)	-0.006 (0.009)	-0.003 (0.004)	-0.002 (0.007)	-0.004 (0.009)	-0.002 (0.001)
Income/100	**0.053 ***** **(0.000)**	**0.076 ***** **(0.007)**	**0.055 ***** **(0.007)**	0.013 (0.010)	0.045*** (0.019)	0.009 (0.024)	0.010 ** (0.005)	0.042*** (0.019)	0.002 (0.011)
Snip for (3)…									
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

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.

But in our split-sample estimates with these controls, when $w^a \geq w^e$ the effect of hours is insignificant, but the effect of income is large and highly significant. But when $w^a < w^e$, the effect of income remains important and hours also becomes significant.

Here the patterns of magnitudes and significance levels are inconsistent with the neoclassical model because the targets have large effects.

They are also inconsistent with Farber's pure income-targeting model because hours has a strong and significant effect when income is lower than expected but an insignificant effect when income is higher than expected.

The patterns are generally consistent with our model, but in this case they do not completely support it because when the wage is lower than expected, the coefficient of income is also significant (as well as hours, as our model predicts), and the income coefficient is larger than the hours coefficient.

In the center panel (2) of Table 2, we control for driver heterogeneity, day of the week, and hour of the day.

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)
Snip for (3)…									
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

In the pooled sample, with these controls, income has 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 ($w^a \geq w^e$), hours has a small marginal 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, while the coefficient that corresponds to the marginal effect of hours becomes significant, and the magnitude of the marginal effect increases from 0.3% to 1.1%.

Here the pattern is clearly inconsistent with a neoclassical or pure income-targeting model, but fully consistent with our reference-dependent model.

In the right-most panel (3) of Table 2 (next slide) we also control 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. (In this case our estimates fully replicate Farber's point estimates, but not his standard errors.)

The pattern in the split-sample estimates with these controls is again clearly inconsistent with a neoclassical or pure income-targeting model, but fully consistent with our reference-dependent model:

- Income but not hours significantly affects the stopping probability when the wage is higher than expected, and

- hours but not income significantly affects the stopping probability when the wage is lower than expected.

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 further that because the wage elasticity of labor supply tends to be negative when the driver reaches his income target first ($w^a \geq w^e$) but tends to be near zero when the driver reaches his hours target first ($w^a < w^e$), the aggregate wage elasticity is likely to be negative.

Thus, the model can resolve the apparent contradiction between the positive incentive to work created by an anticipated wage increase with a negative aggregate wage elasticity.

Further, because the two wage regimes have roughly equal weights in the sample, the heterogeneity of realized wages yields the observed smooth aggregate relationship between stopping probability and realized income.

Nonlinear Probits

Farber (2005) also estimated a nonlinear 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.

We replicate Farber's results for this much more flexible specification, and then re-do the estimates with the sample split as before.

Table 3 reports the marginal probability effects from the estimation of the nonlinear probit model.

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 coefficient estimates and likelihood ratio tests of the hypotheses that the marginal effects of all income or hours groups are zero.

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	

The results for the pooled sample are consistent with Farber's:

- Hours categories have marginal effects that are jointly significantly different from zero, but income categories do not.

- The effects of hours categories vary widely, but the effect of income categories in the pooled sample is smooth, with few effects differing significantly from the baseline income category of $150 - $174.

But when the sample is split, the results change dramatically:

- In both the center panel, which reports the results for $w^a \geq w^e$, and the right-most panel, which reports the results for $w^a < w^e$, the effects of all income and hours categories are jointly significant.

- Mean income categories and high-end hours categories have the predominant influences on the stopping probability when $w^a \geq w^e$, but high-end income and mean hours categories are predominant when $w^a < w^e$: the first-reached target has a larger effect on stopping probability.

Figures 2 and 3 graph the estimated probabilities of stopping against hours and income categories, with both pooled and split sample results.

Figure 2: Probability of Stopping: Marginal Effect of Hours

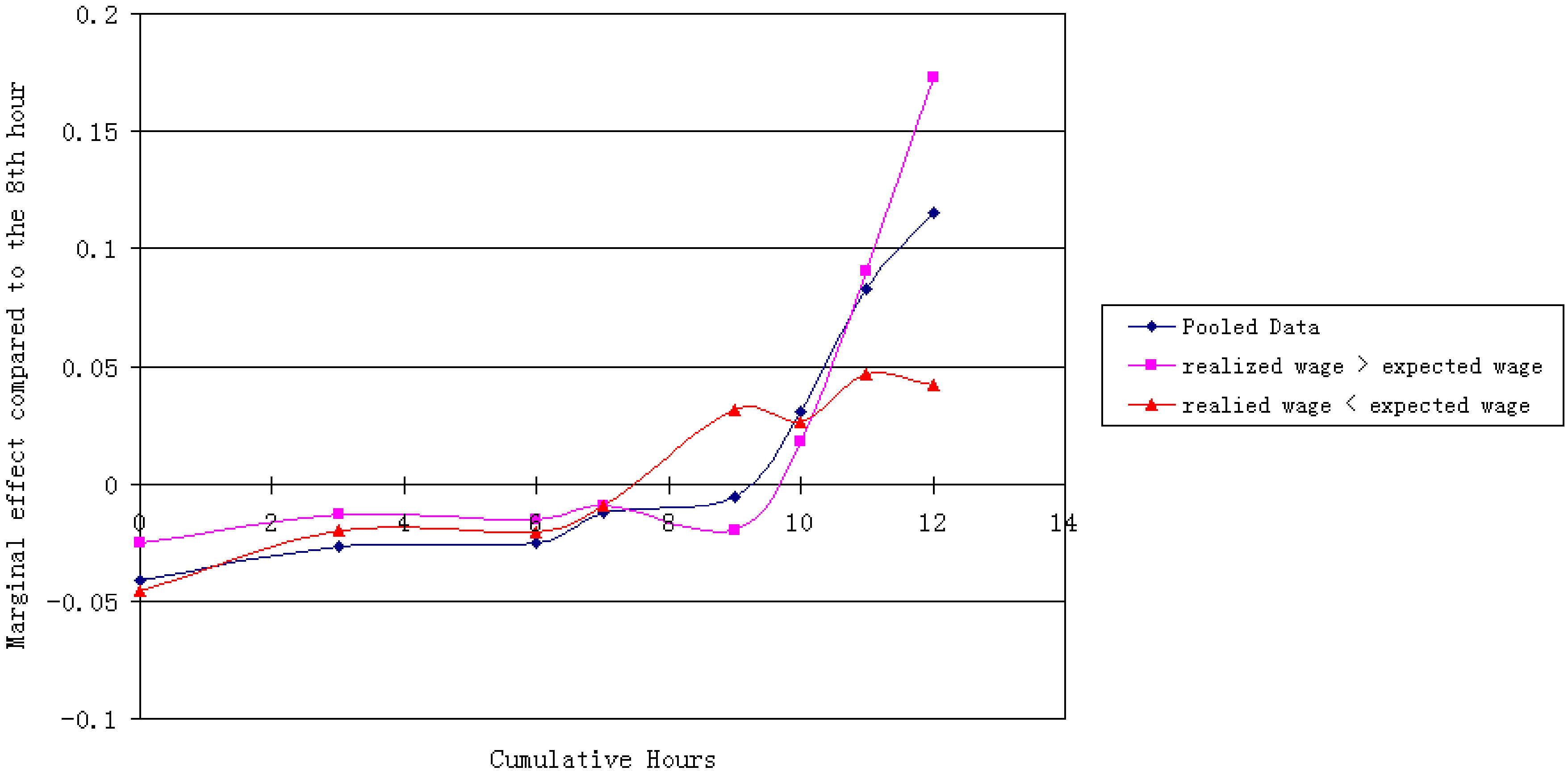
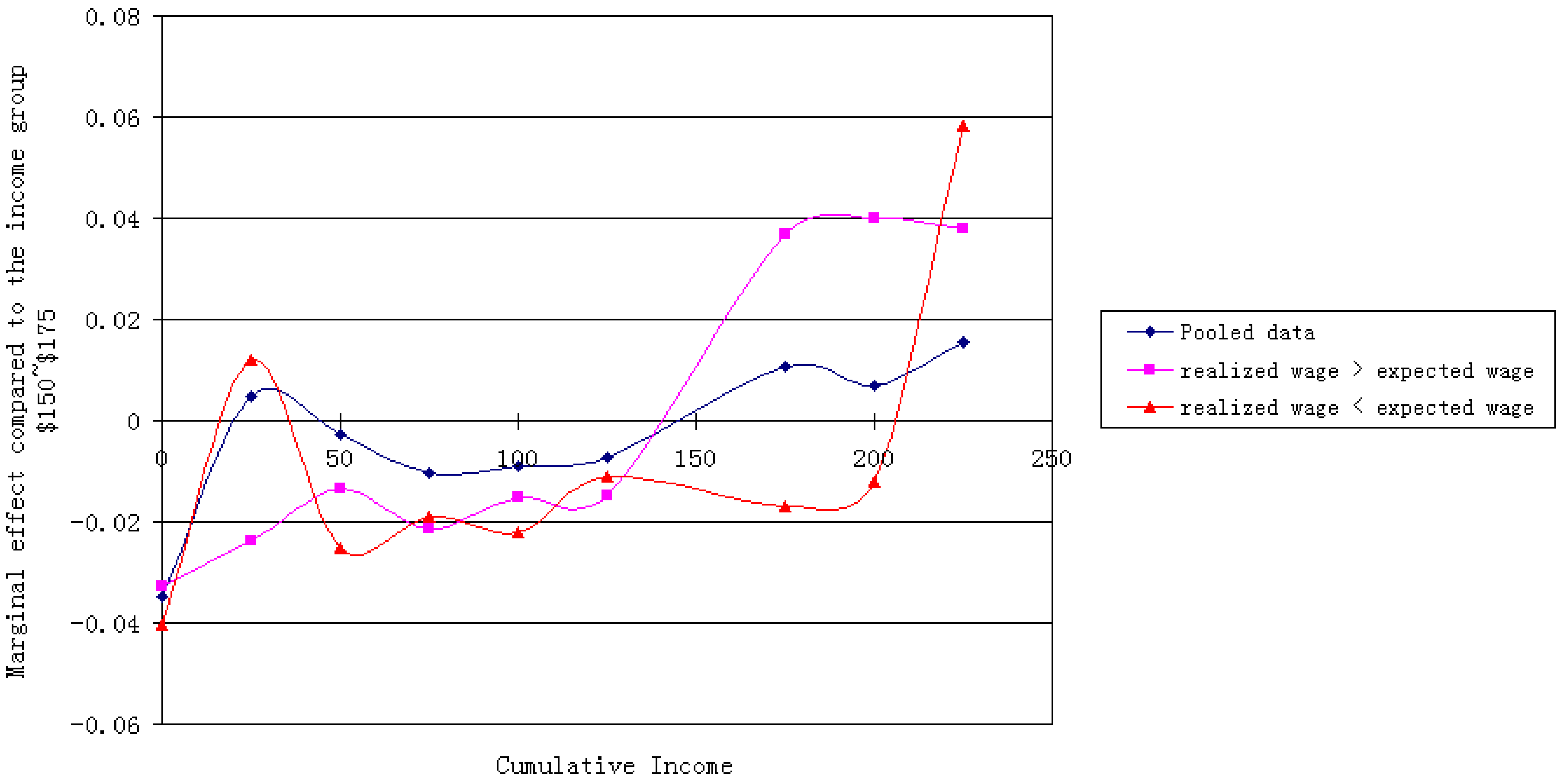


Figure 3: Probability of Stopping:Marginal Effect of Income



In both figures the marginal effects of hours categories are highly nonlinear. When realized wage is higher (lower) than expected, the probability of stopping increases first in response to income (hours) and then hours (income), closely matching our theory's predictions for the split samples:

- When the realized wage is lower than expected ($w^a < w^e$), so the hours target is reached before the income target, the probability of stopping first jumps at the mean (our estimated target) hours category (8~10 hours, Figure 2), then again at high-end income categories (> $200 in Figure 3).

- But when the realized wage is higher than expected ($w^a \geq w^e$), so that the income target is reached first, the probability of stopping first jumps at the mean (our estimated target) income category ($150-$200 in Figure 3), then again at high-end hours categories (> 10 hours in Figure 2).

- In Figure 2, between the 7th and 10th hours, the marginal effects of hours on stopping probability when $w^a < w^e$ are higher than when $w^a \geq w^e$; while after the 10th hour, the marginal effects when $w^a \geq w^e$ rise dramatically.

- In Figure 3, the marginal effect of income on stopping probability increases dramatically when income reaches $125 - $150 when $w^a \geq w^e$, but only when income reaches $200 - $225 when $w^a < w^e$.

Our split-sample estimates are inconsistent with the neoclassical model and Farber's income-targeting model.

Instead they strongly support our reference-dependent model's predictions, even with minimal structural restrictions and an imperfect proxy for targets.

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.

Econometric estimates of a structural reference-dependent model as in Farber's (2008) analysis

Farber (2008) estimates a structural reference-dependent model with income targeting.

We use the full sample to estimate a structural model that parallels Farber's, but as suggested by Kőszegi and Rabin's (2006) theory, adapted as follows:

- a driver's preferences reflect both the standard consumption utility of Income and leisure and reference-dependent "gain-loss" utility, with their relative importance tuned by an estimated parameter;

- a driver has a daily target for hours as well as income, and as in Farber's model he is loss-averse, but working longer than the hours target is now a loss, just as earning less than the income target is;

- most importantly for our analysis, instead of treating the targets as latent variables as Farber did, we endogenize them by setting them equal to a driver's rational expectations of hours and income, proxied as above by their natural sample analogs.

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. (1997) found with Farber's (2008) finding that in the full sample, stopping probabilities are significantly related to hours but not income.

More generally, it tests the model's potential to give an empirically as well as theoretically useful account of drivers' labor supply.

Recall that we specify the preferences of a given driver during his shift on a given day, with *I* and *H* denoting his income earned and hours worked that day and I^r and H^r denoting his income and hours targets for the day.

His total utility, $V(I, H|I^r,H^r)$, is 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 – η and η ($0 \leq \eta \leq 1$):

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

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 inverse of the wage elasticity.

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

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

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

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.

The likelihood function is (see the online appendix for details):

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

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 the appendix.

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 or Table 1, where reference-dependence introduces kinks whose magnitudes are determined by $1 - \eta + \eta\lambda$.

(But $1 - \eta + \eta\lambda$ is directly comparable to estimates of λ in most other models, which assume that $\eta = 1$ (so there is only gain-loss utility).)

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 $0 \leq \eta \leq 1$, 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, we interpret them as a driver's rational expectations, and proxy them via the averages of their natural sample analogs.

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.

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.

We also consider a "sophisticated" 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 also confirms our main messages.

Table 4 columns 1 and 3 report estimates for naïve and sophisticated models (referring to how trip-level expectations $E(f_{t+1})$ and $E(h_{t+1})$ are formed), 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 Table 4 in each case we consider alternative models of how drivers form the expectations $E(f_{t+1})$ 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 variables into account ("sophisticated"). (The online appendix gives more detail.)

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***
	(0.132)	(0.113)	(0.086)	(0.164)
θ	0.219*	0.176	0.090	0.022
	(0.119)	(0.147)	(0.133)	(0.078)
ρ	0.128***	0.363***	0.390	1.122
	(0.025)	(0.119)	(0.334)	(1.232)
c	0.001	0.020	-0.051	-0.024***
	(0.043)	(0.051)	(0.049)	(0.020)
σ	0.069	0.101	0.204**	0.179***
	(0.043)	(0.064)	(0.085)	(0.032)
Observations	13461	13461	13461	13461
Log-likelihood	-1687.8105	-1762.426	-1696.6684	-1761.2436

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 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.

(Since $\eta = 1$ is imposed in most other estimates, our estimates of $1 - \eta + \eta\lambda$ can be directly compared with most other estimates 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				
	η			
λ	(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

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.

The estimates imply comparatively little bunching around the targets (Table D1 from the appendix, reproduced below), 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, the first-reached target has a larger effect than the second-reached target.

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

Despite the varying influence of the targets on stopping probabilities, the heterogeneity of realized wages again yields a smooth aggregate relationship between stopping probability and realized income.

Thus, the model reconciles 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).

When $1 – \eta + \eta\lambda$ is large enough, and with a significant number of observations where drivers' stopping decisions are more heavily influenced by their income than hours targets, the model will yield a negative aggregate wage elasticity of hours.

Table 6 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.

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	0.022	0.025	0.023	0.025
At income target	0.161	0.124	0.165	0.130
In between two targets	0.115	0.102	0.134	0.120
At hours target	0.238	0.167	0.233	0.166
Above hours target	0.287	0.234	0.278	0.227
$w^a \leq w^e$				
Before hours target	0.022	0.024	0.031	0.030
At hours target	0.139	0.134	0.149	0.135
In between two targets	0.168	0.164	0.178	0.149
At income target	0.266	0.245	0.282	0.260
Above income target	0.283	0.234	0.305	0.254

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.

Conclusion

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 model.

They also reconcile the negative wage elasticity of hours found by Camerer et al., Farber, and others with the aggregate smoothness of the relationship between stopping probability and realized income Farber found, and with the positive relationships between *expected* earnings and labor supply found by Oettinger, Fehr and Goette, and others.

Finally, our structural model avoids Farber's (2008) criticism that drivers' 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 is plausibly and precisely estimated, robust to the specification of proxies for expectations, and comfortably within the range that indicates reference-dependence.

The model avoids Farber's criticism 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.

Further, 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 via rational expectations greatly reduces the difference in degrees of freedom.

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.