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

Bonn Economics Research Seminar
23 July 2009

Introduction

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

But in an influential field study of workers who choose their own hours, 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.

(Realized earnings is the natural counterpart of the wage in this setting, henceforth called the “wage”.)

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.

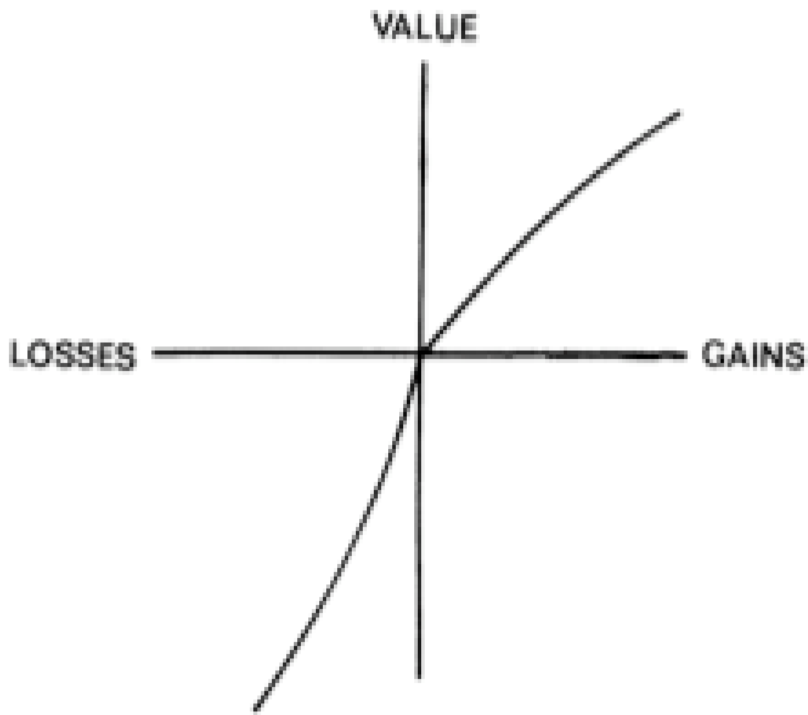
Camerer et al.’s explanation is in the spirit of Kahneman and Tversky’s (1979 *Econometrica*) Prospect Theory, in which:

- 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.

Farber (2008 *AER*) suggests that a finding that labor supply is reference-dependent would have significant policy implications:

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

Although Camerer et al.’s analysis has inspired a number of empirical studies of labor supply, the literature has not yet fully converged on the extent to which the evidence supports reference-dependence.

Much also depends on its scope and the details of its structure.

If reference- dependence were limited to inexperienced workers or unanticipated changes, its direct relevance to most policy questions would be small.

This paper seeks to shed additional light on these issues, building on two recent developments:

- Farber's (2005 *JPE*, 2008 *AER*) empirical analyses of cabdrivers' labor supply.

- Kőszegi and Rabin's (2006 *QJE*; see also 2007 *AER*, 2009 *AER*) theory of reference-dependent preferences.

Farber (2005 *JPE*) collected and analyzed data on the labor supply decisions of a new set of New York City cabdrivers, finding that:

- 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 own 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 Thursdays' through 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*) 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 their 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 also assume for simplicity that the targets are point expectations rather than distributions as in Kőszegi and Rabin's theory.

We operationalize the targets by finding natural sample proxies with limited endogeneity problems.

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 probit models of the probability of stopping as in Farber's (2005) analysis

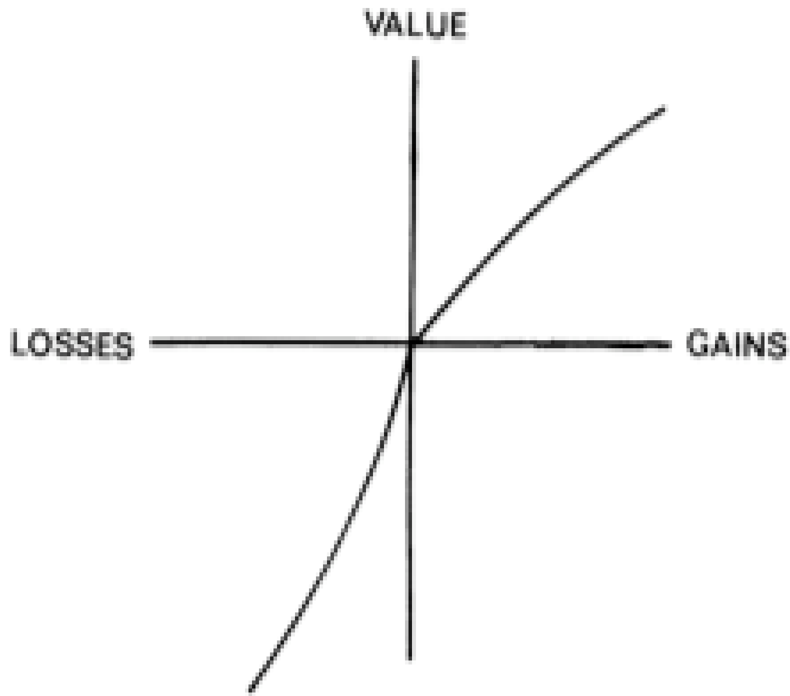
4. Econometric estimates of reduced-form models of the probability of stopping as in Farber's (2008) analysis

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

1. 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?

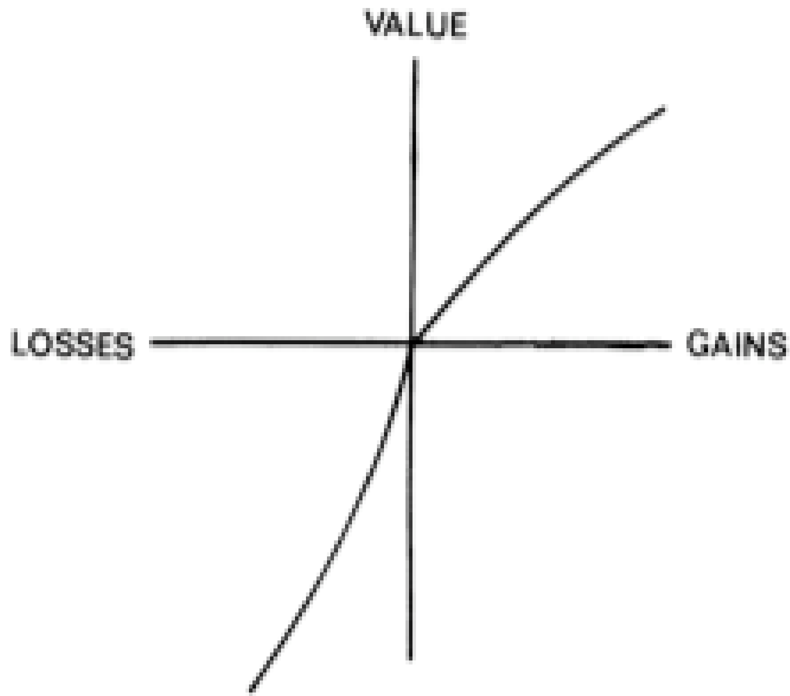


Prospect Theory departs from neoclassical theory in three main ways:

(a) Reference-dependence and loss aversion (people are less sensitive to changes above their target ("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 One or more targets, 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 targets to income and leisure in the domain of preferences.

With regard to econometric testing, the kink at the target 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 targets (gains) are roughly half as large as people’s sensitivities to changes in income or leisure below their targets (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 reduced-form approach that estimates the targets as latent variables.

However, in Farber's dataset estimating the targets 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 targets as proxied 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 targets.

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

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

(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 may be more strongly influenced by hours or income, depending on which target is reached first, and the model deviates from a neoclassical model.

Whenever the income target has an important influence on a driver’s stopping decision, even a driver who values income but is “rational” in the generalized, reference-dependent sense of prospect theory may have a negative wage elasticity of hours, as Camerer et al. found.

To the extent that the hours target has the dominant influence, the wage elasticity of hours will be near zero.

Because the elasticity is negative in one regime but near zero in the other, 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.

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.

Suppose for simplicity that a driver expected the wage to remain constant at w^e.

Then 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 $I = w^e H$.

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^e 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.

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

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

	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.

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 the wage is higher than expected, so the income target is reached before hours target.

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

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 $I = w^e H$.

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

He continues working as long as the anticipated wage w^e exceeds the hours disutility cost of an additional unit of income.

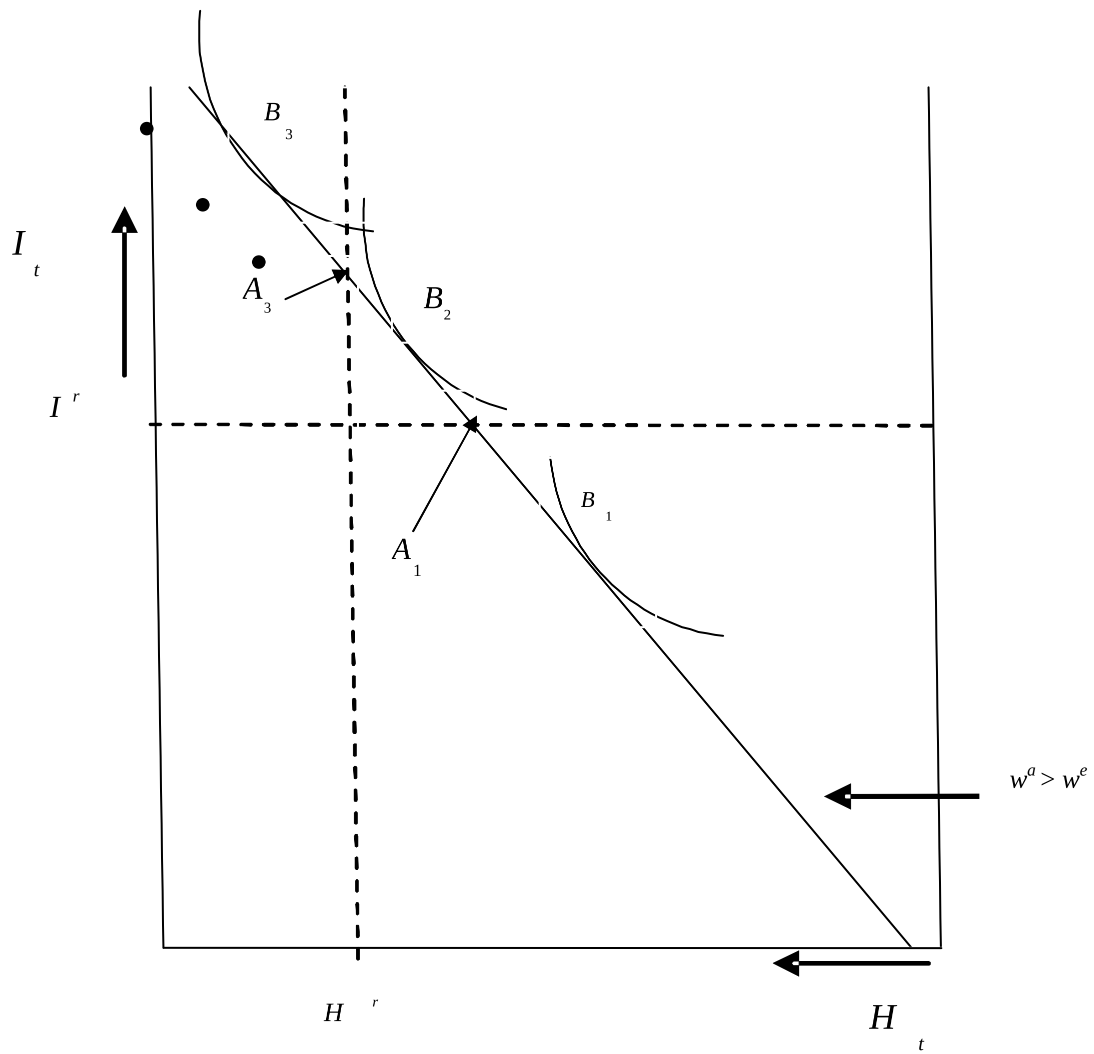


Figure 1: A Reference-dependent Driver’s Stopping Decision

In the initial, income-loss/hours-gain domain, the comparison favors working more than the neoclassical one; but for a given w^e the tradeoff becomes (weakly) less and less favorable as income and hours accumulate.

If the hours disutility cost of income rises to w^e before the driver reaches his first target, income in this case, 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.)

A driver 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 appears globally optimal to him. (With more realistic assumptions, the conclusions would be similar but messier.)

Figure 2 compares the labor-supply curves for a neoclassical driver and a reference-dependent driver with the same consumption utility functions.

The solid curve is the neoclassical supply curve, while the dashed curve is the reference-dependent one.

The shape of the reference-dependent supply curve depends on which target has a larger influence on the stopping decision, which depends in turn on the relation between the neoclassical optimal stopping point (that is, for consumption utility alone) and the targets.

Figure 2 illustrates the case suggested by our estimates:

For wages that reconcile the income and hours targets as at point D, the neoclassically optimal income and hours are higher than the targets, so the driver stops at his second-reached target.

Whenever the wage is to the left of point D, the hours target is reached before the income target, and vice versa.

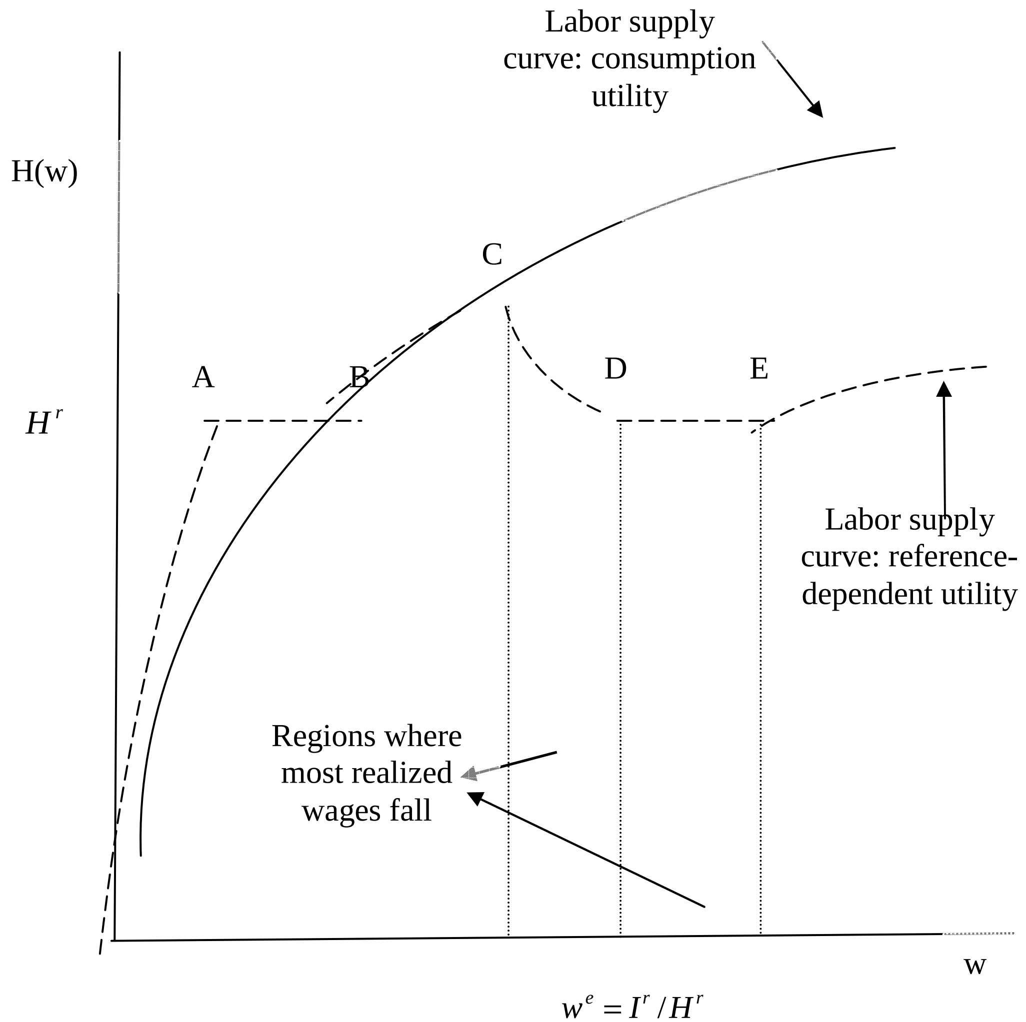


Figure 2: A Reference-dependent Driver's Labor Supply Curve

As Figure 2 illustrates, reference-dependent labor supply is non-monotonic.

When the wage is to the left of point A, the higher cost of income losses raises the incentive to work above its neoclassical level.

Along segment AB labor supply is determined by the kink at the hours target, which is reached first.

Along segment BC the neoclassical optimal stopping point is above the hours but below the income target, so the gain-loss effects cancel out, and reference-dependent and neoclassical labor supply coincide.

Along segment CD labor supply is determined by the kink at the income target, which is reached second, so the wage elasticity of hours is negative.

Along segment DE labor supply is determined by the kink at the hours target, which is reached second.

Finally, when the wage is to the right of point E, the higher cost of hours losses lowers the incentive to work below its neoclassical level.

Most realized wages fall close to point D.

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

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

But instead of treating income targets as estimated latent variables, we proxy drivers' rational point expectations of a day's income and hours, driver/day-of-the-week by driver/day-of-the-week, by their sample averages up to but not including the day in question, ignoring sampling variation.

This avoids confounding from including the current shift's income and hours in the averages, while allowing the targets to vary across days of the week as suggested by the variation of hours and income.

This way of proxying the targets loses observations from the first day-of-the-week shift for each driver because there is no prior information for those shifts.

This is a nonnegligible fraction of the total number of observations (3124 out of 13461). But the criterion for censoring is exogenous and balanced across days of the week and drivers, so should not cause significant bias.

3. Econometric estimates of linear 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.

We estimate linear probit models of the probability of stopping as in Farber (2005), but splitting the sample according to whether a driver's earnings early in the day are higher or lower than his proxied expectations.

In our model as in Farber's, drivers choose only hours, not effort. Thus this "early earnings" criterion should be approximately uncorrelated with errors in the stopping decision, limiting sample-selection bias.

The higher a driver's early earnings, the more likely he is to hit his income target first, simply because early earnings is part of total earnings.

For a wide class of reference-dependent models, including our structural model, a driver's probability of stopping increases at his first-reached target and again (generally by a different amount) at his second-reached target.

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

This difference is robust to variations in the specification of the targets and the details of the structural specification.

Sample-splitting therefore allows a robust assessment of the evidence for reference-dependence, avoiding structural restrictions.

Table 2 reports marginal probability effects, but with significance levels computed for the underlying coefficients.

In each panel, the left-hand column uses the specification of Farber's (2005) pooled-sample estimates, with observations deleted as in our estimates. The center and right-hand columns report our split-sample estimates.

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

In the pooled-sample estimates with these controls, all coefficients have the expected signs and the effect of hours is significant, but the effects of income, waiting, and break hours are insignificantly different from zero.

In our split-sample estimates with these controls, when early earnings are higher than expected the effect of hours is large and significant but the effect of income is insignificant.

But when early earnings are lower than expected, the effect of income is significant at the 5% level, and the effect of hours is insignificant.

This reversal is inconsistent with a neoclassical model, but is fully consistent with a reference-dependent model in which stopping probability happens to be more strongly influenced by the second target a driver reaches than the first, as in Figure 2.

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

		(1)			(2)		
	Evaluation Point for Marginal Effect	Pooled data	First hour's earning > expected	First hour's earning < expected	Pooled data	First hour's earning> expected	First hour's earning < expected
Cumulative total hours	8.0	0.022** (0.011)	0.040*** (0.016)	-0.001 (0.010)	.014** (0.008)	.045*** (0.023)	-.006 (0.010)
Cumulative Income/100	1.5	0.029 (0.027)	-0.008 (0.040)	0.071** (0.027)	.011 (0.020)	-.033 (0.052)	.065** (0.035)
Cumulative waiting hours	2.5	0.005 (0.013)	-0.014 (0.016)	0.036*** (0.010)	.002 (0.008)	-.009 (0.019)	.027** (0.013)
Cumulative Break hours	0.5	-0.001 (0.010)	-0.011 (0.013)	0.019 (0.011)	-.005 (0.007)	-.019 (0.019)	.012 (0.011)
Min temp<30	0.0	-	-	-	.002 (0.010)	-.007 (0.030)	-.000 (0.013)
Max temp>80	0.0	-	-	-	-.020 (0.012)	-.056* (0.030)	.002 (0.022)
Hourly rain	0.0	-	-	-	.136 (0.229)	-.997 (0.890)	.329** (0.176)
Daily snow	0.0	-	-	-	-.001 (0.006)	.003 (0.013)	.003 (0.013)
Downtown	0.0	-	-	-	.010 (0.010)	.033 (0.026)	.004 (0.013)
Uptown	0.0	-	-	-	-.002 (0.007)	-.009 (0.018)	-.005 (0.010)
Bronx	0.0	-	-	-	.100 (0.090)	.191 (0.246)	.098 (0.098)
Queens	0.0	-	-	-	.076* (0.060)	.166* (0.118)	.001 (0.056)
Brooklyn	0.0	-	-	-	.115*** (0.050)	.217** (0.125)	.074** (0.052)
Kennedy Airport	0.0	-	-	-	.097*** (0.051)	.279*** (0.130)	.034 (0.043)
LaGuardia Airport	0.0	-	-	-	.101*** (0.046)	.233*** (0.096)	.054 (0.049)
Other	0.0	-	-	-	.184*** (0.107)	.311** (0.163)	.101 (0.140)
Drivers (21)		No	No	No	Yes	Yes	Yes
Day of week (7)		No	No	No	Yes	Yes	Yes
Hour of day (19)	2:00 p.m.	No	No	No	Yes	Yes	Yes
Log likelihood		-1404.905	-688.825	-710.825	-1214.724	- 570.445	-602.904
Pseudo R^2		0.1246	0.1221	0.1333	0.2431	0.2730	0.2649
Observation		8040	3875	4165	8040	3875	4165

In the right-hand panel we control for driver heterogeneity, day-of-the-week, hour of the day, weather, and location.

In the pooled sample this yields estimates like those in the left-hand panel.

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

Our estimates show that the probability of stopping is more strongly influenced by hours when early earnings are higher than expected but by income when they are lower than expected.

Our estimates are fully consistent with our reference-dependent model, but inconsistent with the neoclassical model and—because the effect of hours is significant when income is higher than expected but insignificant when income is higher than expected—with Farber's income-targeting model.

Further, because the wage elasticity is substantially negative when the income target is the dominant influence on stopping but near zero when the hours target is dominant, the reference-dependent model's distinction between anticipated and unanticipated wage changes can reconcile an anticipated wage increase's positive incentive to work with a negative aggregate wage elasticity of hours.

Finally, with a distribution of realized wages, the model can also reproduce Farber's (2005) findings that aggregate stopping probabilities are significantly related to hours but not realized earnings, and that they respond smoothly to earnings.

4. Econometric estimates of reduced-form models of the probability of stopping as in Farber's (2008) analysis

We next use the pooled sample to estimate a reduced-form model of stopping probability, with dummy variables to measure the increments due to hitting the income and hours targets as in Farber's (2008) Table 2, but with our proxied targets instead of Farber's estimated targets.

Table 3 reports reduced-form estimates of the increments in stopping probability on hitting the estimated income and hours targets.

The estimated coefficients of dummies indicating whether earnings or hours exceeds the targets are positive as in a reference- dependent model.

They are insignificantly different from 0 when we pool days of the week (column 2) but large and significant when we disaggregate by day-of-the-week (column 4). In column 4, the significant effects of income and hours come mainly from whether they are above or below their targets, not levels.

This suggests that ignoring day-of-the-week effects is a misspecification, which may be why Farber's specification yielded different results.

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

	Evaluation point for marginal effect	Using driver specific sample average income and hours prior to the current shift as targets (1)	(2)	Using driver and day-of-the-week specific sample average income and hours prior to the current shift as targets (3)	(4)
Cumulative total hours>hours target	0.0	.037*** (.012)	.009 (0.007)	.036*** (.013)	.018*** (0.008)
Cumulative income > income target	0.0	.047*** (.014)	.009 (0.008)	.053*** (.015)	.026*** (0.010)
Cumulative total hours	8.0	.012** (.006)	.008* (0.005)	.011* (.006)	.006* (0.004)
Cumulative Income/100	1.5	-.001 (.017)	.001 (0.014)	.000 (.018)	-.006 (0.012)
Cumulative waiting hours	2.5	.002 (.007)	.003 (0.005)	.006 (.007)	.003 (0.004)
Cumulative Break hours	0.5	.003 (.006)	-.001 (0.004)	.004 (.006)	.001 (0.004)
Weather (4)		No	Yes	No	Yes
Locations (9)		No	Yes	No	Yes
Drivers (21)		No	Yes	No	Yes
Days of the week (7)		No	Yes	No	Yes
Hour of the day (19)	2:00 p.m.	No	Yes	No	Yes
Log likelihood		-1934.2673	-1710.3519	-1537.2767	-1355.9858
Pseudo R^2		0.1653	0.2619	0.1679	0.2660
Observation		12979	12979	10337	10337

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

We next use the pooled sample to estimate a structural reference-dependent model in the spirit of Farber’s (2008) model, with the changes suggested by KR’s theory.

Here the specification must take a position on how a driver forms his expectations about the wage, trip by trip during the day.

Farber (2005) argued that hourly earnings are so variable and unpredictable that “predicting hours of work with a model that assumes a fixed hourly wage rate during the day does not seem appropriate.”

We assume there is no within-day predictability, and take a driver’s expectations about the wage during the day as predetermined rational expectations, proxied in the same way we proxy the targets.

This is a noisy proxy, but it is not systematically biased, and because it is predetermined it should not cause endogeneity bias.

Like Farber, we also assume that drivers are risk-neutral; but unlike Farber, we assume they ignore option value in their decisions.

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 $1 - \eta$ and η ($0 \le \eta \le 1$):

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

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

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 0 and variance σ^2.

Table 4 reports structural estimates, expanded to identify the effects of different proxies and the reasons for the differences between our and Farber's (2008) results.

Column 1 is the baseline; columns 2-5 each change one thing at a time.

Column 2 checks for robustness to basing targets on sample proxies after as well as before the current shift (but still omitting the current shift).

Column 3 uses a more sophisticated model of next-trip fare/time expectations, using the 3124 observations omitted from the first shifts for each day-of-the-week for each driver, predicted using the current estimation sample.

Column 4 rules out differences across days of the week, as in Farber (2008).

Column 5 restricts attention to income targeting, as in Farber (2008).

Table 4: Structural Estimates under Alternative Specifications of Expectations					
	(1) Use driver and day-of-the-week specific sample averages prior to the current shift as the income/hours targets and the next-trip earnings/times expectation	(2) Use driver and day-of-the-week specific sample averages prior and after the current shift as the income/hours targets and next-trip the earnings/times expectation	(3) Use driver and day-of-the-week specific sample averages prior to the current shift as the income/hours targets and fit the sophisticated next-trip earnings/time expectation	(4) Use driver (without day-of-the-week difference) specific sample averages prior to the current shift as income/hours targets and the next-trip earnings/time expectation	(5) Income target only: use driver and day-of-the-week specific sample averages prior to the current shift as income target and next-trip earnings/time expectation
$1 - \eta + \eta\lambda$	1.715***	1.353***	1.441**	1.182**	5.036
	(.345)	(.158)	(.327)	(.116)	(8.480)
θ	.099**	.073**	.018*	.069*	.051*
	(.062)	(.057)	(.031)	(.086)	(.102)
ρ	.588*	.585*	1.118*	.646*	1.536**
	(.310)	(.312)	(.720)	(.404)	(.704)
σ	.032*	.106***	.115	.089	.789
	(.017)	(.038)	(.093)	(.073)	(1.599)
c	.001	-.008	-.023	.017	.084
	(.015)	(0.048)	(.0604)	(.074)	(.415)
Observations	10337	10337	10337	10337	10337
Log-likelihood	-1363.0367	-1367.4512	-1357.0613	-1375.2973	-1367.6224

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

The baseline model yields plausible parameter estimates that confirm and refine the conclusions of our split-sample probits and reduced-form analyses.

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 $1 - \eta + \eta\lambda$.

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

The null hypothesis that $1 - \eta + \eta\lambda = 1$ is rejected at the 1% level, ruling out the restriction $\eta = 0$ that reduces the model to a neoclassical model.

Our estimate of $1 - \eta + \eta\lambda$ is somewhat lower than most reported estimates of the coefficient of loss aversion, but not implausibly so.

Of the preference parameters, in our baseline model the estimate of θ is significantly different from 0 only at the 5% level, and the estimate of ρ is only significantly different from 0 at the 10% level.

Column 2 shows that basing the targets on sample proxies after as well as before the current shift adds somewhat to precision.

Column 3 shows that the results are robust to more sophisticated wage forecasting.

Columns 4 and 5 confirm that day-of-the-week differences and hours targeting are both important for detecting the effects of reference-dependence, in that the target effects become smaller and/or insignificant with this variation in specification.

Both of these features plainly contribute to our differences from Farber's conclusions.

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

Column 3's model, with drivers sophisticated enough to predict future wages based on location, clock hours, etc., fits best.

Of the remaining four models, all with constant expectations throughout the shift, Column 1's model, the baseline, fits best.

Like our earlier models, our structural model resolves the apparent contradiction between a negative aggregate wage elasticity and the positive incentive to work of an anticipated wage increase.

In our model, the stopping decisions of some drivers, on some days, will be more heavily influenced by their income targets, in which case their wage elasticities will tend to be negative, while the decisions of other drivers on other days will be more heavily influenced by their hours targets, in which case their wage elasticities will be close to zero.

When $1 - \eta + \eta\lambda$ is large enough, and with a significant number of observations in the former regime, the model will yield a negative aggregate wage elasticity of hours.

To illustrate, Table 5 also reports each specification's implication for the aggregate correlation of wage and optimal working hours, a proxy for the wage elasticity.

All models but column (4), which suppresses day-of-the-week differences, have a negative correlation between wage and optimal working hours.

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

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).

Finally, our structural model avoids Farber's (2008) criticism that drivers' estimated targets are too unstable and imprecisely estimated to allow a useful reference-dependent model of labor supply.

In this comparatively small sample, there remains some ambiguity about the parameters of consumption utility ρ and θ. But the key function $1 - \eta + \eta\lambda$ of the parameters of gain-loss utility is plausibly and precisely estimated, robust to the specification of proxies for the targets, and comfortably within the range that indicates reference-dependent preferences.