Reference-Dependent Preferences and Labor Supply:
The Case of New York City Taxi Drivers *

Henry S. Farber

Princeton University

Abstract

Recent theoretical work has focused on the importance of reference-dependent preferences. Typically, this work assumes a discontinuity in marginal utility at some base reference level. Marginal utility below this kink in the utility function is higher than marginal utility above the kink. I develop an empirical model of daily labor supply that incorporates reference-dependent preferences, and I apply this model to data on the daily labor supply of New York City taxi drivers. The estimates suggest that there may be a reference level of income on a given day that affects labor supply. However, the reference level varies substantially from day to day for a given driver. This seriously limits the predictive power of the reference point model and undermines the usefulness of the construct of the reference income level as a determinant of labor supply. [JEL Classification J22]

1 Introduction

Increased attention has been paid in recent years to deviations from the standard neoclassical model of consumer behavior. A substantial segment of this work focuses on reference-dependent preferences where there is a change in the shape of the utility function at some base (reference) level of income or consumption. These models have strong predictions for how responses to changes in prices are affected by the actual level of consumption or income relative to the reference level. A difficulty in bringing this class of models to the data is that the reference level of income or consumption is seldom observed and we only have weak information on what determines the reference point.

There are important reasons to understand how these consideration affect estimation of labor supply elasticities. 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.

In this study, I develop an empirical model of daily labor supply that incorporates reference-dependent preferences but does not require that the reference level of income be observed or defined in advance. I apply this model to data on the daily labor supply of New York City taxi drivers by allowing taxi drivers to have a reference level of daily income. The estimates suggest that, while there may be a reference level of income on a given day that affects labor supply, the reference level varies substantially from day to day. This seriously limits the predictive power of the reference point model and undermines the usefulness of the construct of the reference income level as a determinant of labor supply.

1.1 The Emerging Literature on Labor Supply in Settings with Flexible Hours

There is a very large literature in economics estimating the wage elasticity of labor supply. This literature has been surveyed exhaustively (Killingsworth and Heckman, 1986; Pencavel, 1986; Blundell and MaCurdy, 1999), and a reasonable summary of the findings is that labor

supply elasticities for men are very small and often not significantly different from zero while labor supply elasticities for women are somewhat larger (though considerably less than one). One criticism of this literature is that the standard neoclassical model assumes that workers are free to set their hours in response to changes in the wage or, alternatively, can select a job with the optimal wage-hours combination from a dense joint distribution of jobs.[1] Evidence that neither of these are credible assumptions is that the distribution of hours is quite lumpy, with a substantial fraction of workers reporting usual weekly hours of precisely forty.[2]

There is an emerging literature on labor supply that is not subject to this criticism because it investigates labor supply responses in settings where workers are free to set their labor supply (e.g., Camerer, Babcock, Lowenstein, and Thaler (1997) (CBLT) Oettinger (1999), Fehr and Goette (2002), and Farber (2005)).[3] A By and large, this literature finds substantial positive labor supply elasticities and evidence consistent with the standard neoclassical labor supply model. A notable exception that is directly relevant to my study is the finding of a negative labor supply elasticity among New York City Taxi Drivers by Camerer, et al. (1997). None of this this literature allows directly for the possibility of reference-dependent preferences.[4]

1.2   Reference-Dependent Preferences

The theory reference-dependent preferences has its roots in the concept of loss aversion. Tversky and Kahneman (1991) present an analysis of choice in a riskless framework where

---

[1] Some examples of analyses of labor supply that consider the case where workers are not free to set their hours or to choose an optimal wage–hours combination are Ham(1982), Altonji and Paxson (1988), Kahn and Lang (1991), and Dickens and Lundberg (1993). Ham and Reilly (2002) present an analysis of labor supply that tests an hours restriction model along with intertemporal substitution and implicit contracts models.

[2] This is based on tabulation of the the 2002 merged outgoing rotation group files from the Current Population Survey. It is the case that over fifty percent of workers report 40 hours as their usual weekly work hours. While some of this heaping at 40 hours may be do to the natural tendency to round off by respondents, it is clear that there is substantial bunching of hours.

[3] I survey this literature briefly in Farber (2005).

[4] While Camerer, et al. (1997) interpret their finding of a negative labor supply elasticity as consistent with target earnings behavior, a strong form of reference-dependent preferences, their empirical model does not allow for the special nature of labor supply with reference-dependent preferences. I present a simple model of labor supply with reference-dependent preferences in section 2.

the concavity of utility as a function of income changes at some reference or status quo value of income (Y_R). Marginal utility is higher and increasing in income at incomes below Y_R. Marginal utility lower and decreasing in income at incomes above Y_R. This preference structure implies that, for equivalent absolute dollar gains and losses around Y_R, individuals give up more utility from the dollar loss than they enjoy from the dollar gain.

This core idea has been credited as the explanation for the "endowment effect" (Thaler, 1980) The endowment effect, which has much experimental support, is that individuals appear to value items they own more than identical items owned by others. For example, an easily replicable experiment is to distribute commemorative mugs to a random subgroup of some experimental subject pool, to elicit from those who received the mugs a willingness to accept (WTA) a given amount in exchange for the mug, and to elicit from those who did not receive the mugs a willingness to pay (WTP) a given amount in to purchase a mug. It is a routine finding that the average WTA is substantially greater than the average WTP. Indeed, this relationship between WTA and WTP is considered a test of reference-dependent preferences (Bateman, Munro, Rhodes, Starmer, and Sugden, 1997).[5]

While there has been continued refinement of the theoretical ideas underlying reference-dependent preferences (e.g., Munro and Sugden, 2003; Koszegi and Rabin, 2004), there is relatively little direct field evidence on reference-dependent preferences. One example is Genesove and Mayer (2001), who examine seller behavior in the housing market using the purchase price as the reference point. They find clear evidence that sellers are more risk averse when faced with losses. This is consistent with a utility function that changes concavity at the reference point.

Another example is the work of Camerer, Babcock, Lowenstein, and Thaler (1997), who interpret evidence of a negative labor supply elasticity for New York City taxi drivers as evidence of target earnings behavior. A target earner is defined as an individual who works until earnings reach the target level and then quits. Target earnings behavior is an extreme version of reference-dependent preferences, where, at income levels less than the target/reference level, the marginal utility of income is extremely high and, at income levels greater than the

---

[5] See Kahneman, Knetsch, and Thaler (1991) for a nice discussion of the endowment effect and its relationship to reference dependent preferences.

target/reference level, the marginal utility of income is extremely low. Essentially, there is such a sharp kink in the utility function at the reference/target level, workers always quit with income at the reference/target level. This is clearly non-optimal from a neoclassical perspective since it implies quiting early on days when it is easy to make money and working longer on days when it is harder to make money. Utility would be higher by allocating time in precisely the opposite manner. In an earlier paper (Farber, 2005), I critique the conceptual and econometric approach used by Camerer, et al., and I analyze new data (also used in this study) on the labor supply of New York City taxi drivers using a completely different framework. In that study, I found no evidence of target earnings behavior.

One can also interpret work that considers the effect of relative income or consumption comparisons across individuals on behavior as implicitly using reference-dependent preferences. The idea is that the inter-personal comparisons set up a reference point that introduces an asymmetry between gains and losses relative to this reference point. For example, Neumark and Postlewaite (1998) use relative income concerns to understand the increase in labor supply of married women. As more married women enter the labor force and raise their families' incomes, the reference level of income that families use to judge their wellbeing increases. The result is that more women enter the labor force so that do not suffer the substantial loss in utility that comes from family income below the reference level.

In this study, I provide a new analysis of the labor supply of New York City taxi drivers that allows directly for the possibility of reference-dependent utility (without assuming target earnings behavior) in a consistent way. My analysis suggests that there may be a reference level of income on a given day that affects labor supply but that the reference level for a given driver varies substantially from day to day. This seriously limits the predictive power of the reference point, and undermines the usefulness of the construct of the reference income level as a determinant of labor supply.

2   Reference-Dependent Preferences and Labor Supply

The daily labor supply problem for a taxi driver is essentially an optimal stopping problem. The driver begins his shift, and the decision problem is determining when to stop driving for the day. Where the wage is fixed for the day or where the wage is variable but non-increasing during the day, the labor supply or optimal stopping rule problem with reference-dependent preferences can be solved in a straightforward manner, and I proceed to do so in the succeeding two sub-sections. The optimal stopping rule problem is more complicated in the case where the wage is variable and increasing at at least some points during the day. In this case, a full dynamic programming solution is required. This is quite cumbersome and does not yield tractable solutions.

2.1   Reference-Dependent Labor Supply with a Fixed Wage

A straightforward parametric model of labor supply that incorporates a reference point is based on a utility function of income and hours worked of the form

$$U(Y, h) = \alpha(Y - T) - \frac{\theta}{1 + \nu} h^{1+\nu} \qquad Y < T \qquad (2.1)$$

$$U(Y, h) = (Y - T) - \frac{\theta}{1 + \nu} h^{1+\nu} \qquad Y > T \qquad (2.2)$$

where

- h = hours,
- Y = income = Wh,
- T = reference income,
- α = parameter determining sharpness of "kink" in utility function $(\alpha > 1)$,
- θ = parameter determining disutility of work, and
- ν = elasticity parameter (wage elasticity of labor supply = $1/\nu$).

This utility function is kinked at income level T, with the marginal utility of income higher below the kink $(\alpha > 1)$ than above the kink. The labor supply function implied by this utility function is

Wage	Hours	Elasticity
$W < W^*$	$h = (\frac{\alpha W}{\theta})^{1/\nu}$	$\frac{1}{\nu}$
$W^* < W < W^{**}$	$h = \frac{T}{W}$	-1
$W > W^{**}$	$h = (\frac{W}{\theta})^{1/\nu}$	$\frac{1}{\nu}$

where

$$W^* = (\frac{\theta}{\alpha})^{(\frac{1}{1+\nu})}T^{(\frac{\nu}{1+\nu})} \tag{2.3}$$

and

$$W^{**} = \theta^{(\frac{1}{1+\nu})}T^{(\frac{\nu}{1+\nu})}. \tag{2.4}$$

The labor supply function in this case has an unusual shape. Maximization of this utility function implies a standard upward sloping labor supply function with constant elasticity $1/\nu$ for values of the wage below a critical value W^* and above another critical value W^{**} where $W^* < W^{**}$. When the wage is in between W^* and W^{**}, the labor supply function is downward sloping with elasticity -1. Intuitively, for wages where it is optimal to be "near" the reference level of income individuals act as target earners, adjusting their hours so that income equals the reference level of income ($Y = T$). For much lower wages and for much higher wages, the kink in the utility function at the reference point is not relevant and the individual has a standard upward sloping labor supply function. Thus, reference-dependent utility does not imply a generally negative labor supply elasticity.

2.2 Reference-Dependent Labor Supply with a Non-Increasing Wage

The solution is more complicated when the wage rate varies within the relevant time period as is surely the case with taxi drivers. The approach used in section 2.1 can be modified to incorporate a variable non-increasing wage by defining the marginal wage at time t to be $\gamma W_0(t)$ where $W_0'(t) \leq 0$ and γ is a positive constant. On this basis income after h hours of work is $Y(h) = \int_0^h W(t)dt$. reference-dependent utility in this case is

$$U(Y(h), h) = \alpha(Y(h) - T) - \frac{\theta}{1 + \nu}h^{1+\nu} \qquad Y < T \tag{2.5}$$

$$U(Y(h), h) = (Y(h) - T) - \frac{\theta}{1 + \nu}h^{1+\nu} \qquad Y > T \tag{2.6}$$

6

In order to derive the labor supply function in this case note that the hours required to earn the target (h_T) solves

$$T = \gamma \int_0^{h_T} W_0(t)dt \tag{2.7}$$

Define two functions of hours as

$$h^* = \left(\frac{\alpha \gamma W_0(h)}{\theta}\right)^{1/\nu} \tag{2.8}$$

$$h^{**} = \left(\frac{\gamma W_0(h)}{\theta}\right)^{1/\nu} \tag{2.9}$$

The labor supply function in this case is defined by

Region	Hours	Elasticity
$h^* < h_T$	$h^* = \left(\frac{\alpha \gamma W_0(h)}{\theta}\right)^{1/\nu}$	$\frac{1}{\nu}$
$h^{**} < h_T < h^*$	h_T	-1
$h^{**} > h_T$	$h^{**} = \left(\frac{\gamma W_0(h)}{\theta}\right)^{1/\nu}$	$\frac{1}{\nu}$

In thinking about the shape of this labor supply function, consider proportional shifts in the marginal wage function defined by shifts in γ. The labor supply function in this case has a standard upward slope with constant elasticity $1/\nu$ for values for low values of the marginal wage function (implying $h^* < h_T$) and for high values of the marginal wage function (implying $h^{**} > h_T$). When the marginal wage function is intermediate (implying $h^{**} < h_T < h^*$) the labor supply function is downward sloping with elasticity -1. Intuitively, for marginal wage functions where it is optimal to be "near" the reference level of income individuals act as target earners, adjusting their hours so that income equals the reference level of income $(Y = T)$. For much lower and much higher wage functions, the kink in the utility function at the reference point is not relevant and the individual has a standard upward sloping labor supply function. Thus, as in the fixed-wage case, reference-dependent utility does not imply a generally negative labor supply elasticity.

The key weakness in this approach is the assumption that the marginal wage function is non-increasing. If the marginal wage during the relevant time period (a day) can increase, then this solution may not be appropriate. In fact, as documented by Farber (2005), the marginal wage for New York City taxi drivers does vary during the day with periods of sharply increasing hourly wages.

7

3 An Empirical Model of the Stopping Decision of a Taxi Driver Facing a Varying Non-Monotonic Wage

Consider the end of each passenger trip as an opportunity to stop driving. Without deriving the full dynamic solution to the optimal stopping problem, I can derive a reasonable approximate solution that I can implement empirically as a discrete choice problem. At any point τ during the shift, a driver can calculate the difference between the utility from stopping at τ and the forward-looking expected value of continuing to drive (the continuation value). This difference will be a function of many factors, including hours worked so far on the shift and expectations about future earnings possibilities. The difference will also be a function of income earned and the reference income level.

Define a latent variable S_{ijt} for driver i on shift j after trip t that represents the difference between the utility from stopping at t and the forward-looking expected value of continuing to drive (the continuation value). An empirical representation of this latent variable is

$$S_{ijt} = X_{ijt}\beta + \delta I[Y_{ijt} \geq T_{ij}] + \epsilon_{ijt}, \tag{3.1}$$

where

- T_{ij} represent the reference income level for driver i on shift j,

- Y_{ijt} represent the income level for driver i on shift j at trip t,

- X_{ijt} is a vector of variables that determine the difference between current utility and the continuation value,

- β is a parameter vector to be estimated,

- ϵ_{ijt} is a random component with a standard normal distribution.

- $I[Y_{ijt} \geq T_{ij}]$ is an indicator function that equals one if income exceeds the reference income equals zero otherwise, and

- δ is a positive parameter that represents the increment to the latent variable when income exceeds the reference income.

Driver i on shift j will stop after trip t if $S_{ijt} > 0$ and continue driving if $S_{ijt} \leq 0$. The probability of stopping in this case is

$$P^s_{ijt} = \Phi[X_{ijt}\beta + \delta I[T_{ij} \leq Y_{ijt}]], \tag{3.2}$$

where $\Phi[\cdot]$ is the standard normal cumulative distribution function. This probability can be rewritten as the two conditional probabilities for the cases where income is less than the reference level and income is greater than the reference level. These probabilities are, respectively,

$$P^s_{ijt}|(T_{ij} > Y_{ijt}) = \Phi[X_{ijt}\beta], \tag{3.3}$$

and

$$P^s_{ijt}|(T_{ij} \leq Y_{ijt}) = \Phi[X_{ijt}\beta + \delta]. \tag{3.4}$$

If the reference income (T_{ij}) level is known, then this specification would imply the usual probit model and estimates could be derived by maximum likelihood in the usual way. But the reference income level is not known and needs to be estimated. Suppose the reference income level is

$$T_{ij} = \theta_i + \mu_{ij}, \tag{3.5}$$

where θ_i is an individual mean reference income level and μ_{ij} is a random component, distributed normally with mean 0 and variance σ^2, representing daily deviations from θ_i in the reference income level.

This characterization of the reference income level embodies an important test of the importance of the reference income concept in labor supply. To the extent that the reference income level for a driver varies substantially day-to-day, the concept of a reference income level loses predictive power for and importance in the labor supply decision. Formally, if σ is large relative to θ_i, then the reference income level can vary greatly from one day to the next and we risk a tautology where whenever the worker is observed to stop is where their income exceeded the idiosyncratic reference income for that day. For the reference income concept

to have empirical and theoretical relevance, it must be the case that there is substantial constancy in the reference income level from day to day.

The next step is to define the probability of stopping after a trip in the case where the reference income level is unknown but defined by equation 3.5. The notation is cumbersome, and I will derive the probability of stopping in steps. First note that if T_{ij} is known and driver i stops after trip t, this implies that he did not stop after the first t-1 trips. Thus, if the probability of driver i stopping after some trip k is P_{ijk}^s then the probability of observing driver i stop after the tth trip is

$$P_{ij}^t | T_{ij} = (P_{ijt}^s | T_{ij}) \cdot \prod_{k=1}^{t-1} (1 - P_{ijk}^s | T_{ij}), \tag{3.6}$$

Because T_{ij} is unknown, the unconditional stopping probabilities depend on the distribution of T_{ij} as follows. For a driver who stops after the tth trip, there are t+1 possible intervals for the reference level of income to fall relative to accumulated income after each trip during the shift. The reference level of income may be

- less than Y_{ij1},

- in one of the t-1 intervals $Y_{ij(k-1)} < T_{ij} \leq Y_{ijk}$, or

- above Y_{ijt}.

Using the information on income after each trip on a shift, the unconditional probability of stopping after trip t is

$$
\begin{aligned}
P_{ij}^t &= P_{ij}^t | (T_{ij} \leq Y_{ij1}) \cdot Pr(T_{ij} \leq Y_{ij1}) \\[2ex]
&+ \sum_{h=2}^{t} (P_{ij}^t | (Y_{ij(h-1)} < T_{ij} < Y_{ijh}) \cdot Pr(Y_{ij(k-1)} \leq T_{ij} < Y_{ijk})) \\[2ex]
&+ P_{ij}^t | (T_{ij} > Y_{ijt}) \cdot Pr(T_{ij} > Y_{ijt}). \tag{3.7}
\end{aligned}
$$

It remains to write the probabilities of the reference income falling in each of the t+1 intervals and the stopping probabilities conditional on the reference income falling in each of the t+1 intervals.

The probabilities of reference income falling in each interval follow from the definition of T_{ij} in equation 3.5. These are defined as

- The probability that the reference income level is no greater than income after the first trip is

$$Pr(T_{ij} \leq Y_{ij1}) = \Phi[(Y_{ij1} - \theta_i)/\sigma]. \tag{3.8}$$

- The probability that the reference income level lies in one of the t-1 possible intervals $Y_{ij(k-1)}$ to Y_{ijk} is

$$\begin{aligned} Pr(Y_{ij(k-1)} < T_{ij} \leq Y_{ijk}) &= Pr(T_{ij} \leq Y_{ijk}) - Pr(T_{ij} \leq Y_{ij(k-1)}) \\ &= \Phi[(Y_{ijk} - \theta_i)/\sigma] - \Phi[(Y_{ij(k-1)} - \theta_i)/\sigma]. \end{aligned} \tag{3.9}$$

- The probability that the reference income level is greater than the income after trip t is

$$P_{ij(t+1)}^T = Pr(T_{ij} > Y_{ijt}) = 1 - \Phi[(Y_{ijt} - \theta_i)/\sigma]. \tag{3.10}$$

The probabilities of stopping conditional on the reference income level falling in each interval follow from equations 3.3, 3.4, and 3.6. These are defined as

- The probability of stopping after trip t conditional on the reference income level being less than income after the first trip is

$$P_{ij}^t | (T_{ij} \leq Y_{ij1}) = \Phi[X_{ijt}\beta + \delta] \cdot \prod_{k=1}^{t-1}(1 - \Phi[X_{ijk}\beta + \delta]). \tag{3.11}$$

- The probability of stopping after trip t conditional on the reference income level being in the one of the t-1 possible intervals $Y_{ij(h-1)}$ to Y_{ijh} is

$$P_{ij}^t | (Y_{ij(h-1)} < T_{ij} \leq Y_{ijh}) = \Phi[X_{ijt}\beta + \delta] \cdot \prod_{k=1}^{h-1}(1 - \Phi[X_{ijk}\beta]) \cdot \prod_{k=h}^{t-1}(1 - \Phi[X_{ijk}\beta + \delta]). \tag{3.12}$$

- The probability of stopping after trip t conditional on the reference income level being at least as great as income after trip t is

$$P_{ij}^t | (T_{ij} > Y_{ijt}) = \Phi[X_{ijt}\beta] \cdot \prod_{k=1}^{t-1}(1 - \Phi[X_{ijk}\beta]). \tag{3.13}$$

11

The probabilities defined in equations 3.8-3.13 define the components of the unconditional probability P_{ij}^t, defined in equation 3.7, of driver i observed to end shift j after trip t. If driver i actually stops after trip s on shift j, the unconditional probability of the observed stopping behavior is

$$C_{ij} = P_{ij}^s. \tag{3.14}$$

Assuming each shift for a driver is an independent observation, the likelihood function appropriate to this model is defined as

$$L = \prod_{i=1}^{n} \prod_{j=1}^{m_i} C_{ij}, \tag{3.15}$$

where n denotes the number of drivers in the sample and m_i is the number of shifts for driver i in the sample.

In section 5, I derive maximum likelihood estimates of the parameters of the model (β, θ_i, σ^2, and δ) using a sample of 15 New York City taxi drivers on a total of 538 shifts.

3.1   An Alternative Model without Reference-Dependent Preferences

An alternative to the reference-dependent labor supply model is a more standard probit model measuring the stopping probability. This model would be based on a latent variable for the probability of driver i on shift j stopping after trip t of

$$S_{ijt} = X_{ijt}\beta + \epsilon_{ijt}. \tag{3.16}$$

This latent variable has the same form as that specified for the reference dependent model in equation 3.1 with the exception that there is no discrete shift in the latent variable after some level of income is reached. The probability of stopping after a given trip is

$$P_{ijt}^s = Pr(S_{ijt} > 0) = \Phi[X_{ijt}\beta,] \tag{3.17}$$

and the probability of the shift ending after trip t is the probability that the driver did not stop after the first t-1 trips and did stop after the tth trip. This is

$$C_{ij} = \Phi[X_{ijt}\beta] \cdot \prod_{k=1}^{t-1}(1 - \Phi[X_{ijk}\beta]). \tag{3.18}$$

12

as in equation 3.13. Applying this definition of C_{ij} to equation 3.15, yields the likelihood function for the alternative model.

This is a standard probit specification for the probability of stopping after a given trip, and I used this model in my earlier analysis of intertemporal substitution in the labor supply of taxi drivers (Farber 2005). Holding constant the variables in the X vector, this model is a special case of the reference-dependent preference model where the reference income level is very small (smaller than the first fare) or very large (larger than the last fare) so that there is no discrete change in stopping probability. The standard probit model is also a special case of the reference dependent model where there is no increment to the stopping probability at the reference income level ($\delta = 0$).

3.2   Testing the Reference-Dependent Labor Supply Model

A central goal of the analysis is to determine whether daily reference income levels are an important factor in the labor supply decisions of taxi drivers. Estimates of the models derived in this section can be used to achieve this goal.

One approach to testing the model is to estimate both the standard probit stopping model and the stopping model that incorporates the reference income level. The standard model is nested in the reference income model, and a standard likelihood ratio test can be performed to determine if the standard model can be rejected. However, this test is made difficult by the fact that the model accounts for inter-driver differences by allowing the mean reference income to differ across drivers. The constrained standard probit model allows for no fixed differences across drivers. To the extent that there are such fixed differences in stopping behavior, the standard model may be rejected simply because there are fixed differences that it cannot account for. While inclusion of an additional set of individual fixed effects in the X vector would take care of this, it is asking much of the data to include a set of driver fixed effects in the X vector along with individual means (θ_i) in the reference income equation.

Another test of the model is more subjective and relies on two properties of the estimates of the reference income model: 1) variation across days in a driver's reference income level (determined by σ^2) and 2) the amount by which the probability of stopping is increased after

the reference income level is reached (determined by δ).

For reference-dependent preferences to be a useful construct in the labor supply domain, it must be the case that a given individual's reference income levels are constant over reasonable periods of time. If this were not the case, then variation across days in labor supply could be rationalized as an optimal response to random changes in the reference level of income. The variance (σ^2) of the random component determining reference income is a measure of how much the reference income level varies from day to day. If the variance is small relative to the mean (θ_i), then I can conclude that reference incomes are fairly stable. In this case, knowing θ_i could provide important information regarding when the driver will stop.

In addition to allowing an evaluation of reference-dependent preferences, the inclusion of a stochastic component in the definition of reference income that I use (equation 3.5) insures continuity of the the probability of stopping defined in equations 3.7-3.15 with respect to changes in the individual reference levels (θ_i). The random component allows there to be some non-zero probability that the reference income level on a given day falls into any of the $t+1$ categories defined by the progression of income during the shift. If reference income were fixed at θ_i across days, the likelihood function used to estimate the parameters of the model, including the θ vector, would not meet the usual necessary regularity conditions.

The parameter δ indexes the increment to the stopping probability when shift income is at least as great at the reference level (equation 3.1). Clearly, if the increase in the stopping probability implied by the estimated value of δ is substantial, then the reference level of income has a salient effect on labor supply. Alternatively, if there is little or no increase in the stopping probability implied by the estimated value of δ, then the reference level of income is not an important factor in labor supply.

To summarize, 1) an estimated value of σ that is small relative to the values estimated for the mean reference income levels (the θ vector) and 2) a substantial increment to the stopping probability when the reference income level is reached (as indexed by the estimated value of δ) would be evidence consistent with reference-dependent preferences is playing an important role in determining daily labor supply of the taxi drivers in my sample.

4   Data

The data necessary to carry out my analysis are available on "trip sheets" that drivers fill out during each shift. Each trip sheet lists the driver's name, hack number, and date, along with detail on each trip. The information for each trip includes the start time, start location, end time, end location, and fare. In order to obtain a sample of trip sheets, in the Summer of 2000 my research assistants created a list of taxi leasing companies from the current edition of the New York City Yellow Pages. After contacting more than seventy leasing companies, one was found that was willing to provide trip sheets. We were sent 244 trip sheets for 13 drivers covering various dates over the period from June 1999 through May 2000. We contacted the leasing company again in the Summer of 2001, and we were sent an additional 349 trip sheets for 10 drivers covering various dates over the period from June 2000 through May 2001. Two of the drivers appear in both groups so that I have a total of 593 trip sheets for 21 drivers over the period from June 1999 through May 2001. A few of these trip sheets refer to common dates so that I have data on 584 shifts. Because I am trying to estimate driver-specific mean reference income levels, I dropped six drivers with ten or fewer shifts from my analysis. These six drivers accounted for a total of 46 shifts so that the analysis sample has a total of 538 shifts and 12,187 trips for 15 drivers.

The drivers in my sample lease their cabs weekly for a fee of $575. Each driver pays for his own fuel and keeps all of his fare income and tips.

An unfortunate consequence of receiving the trip sheets in an unsystematic fashion is that I have no information on the number of shifts worked. If a trip sheet is not available for a specific driver on a given day, I cannot determine if that driver did not work on that day or if the trip sheet was simply not provided. This prevents me from examining in any conclusive way inter-day relationships in labor supply.

Completeness of the trip sheets is a concern. Unfortunately, I do not have the shift summary printed by the meter after each shift, which lists the total number of trips, in order to verify the completeness of the trip sheets. However, for several reasons, it is likely that the trip sheets are relatively complete. First, there is no particular disincentive for drivers to avoid listing trips on their trip sheets. The trip sheets are not used for tax or

other financial purposes. More importantly, there are financial incentives working in favor of complete listing of trips. In my informal interviews, I have asked drivers in New York about their trip-sheet practices, and they generally told me that they are careful about filling out the sheets. Some explicitly told me that this was due to the possibility of fines levied due to incomplete trip sheets. Apparently, taxi cabs are stopped by New York City police officers or by Taxi and Limousine Commission inspectors, either randomly or for cause.[6] When stopped, drivers are asked for their trip sheet and a printout of the meter summary to that point. The driver can be fined a substantial amount for each fare that is a shortfall between the number of fares listed on the meter summary and the number of fares listed on the trip sheet. Additionally, from time to time, police request trip sheets as part of the investigation of a crime. In the end, while there is no way to ensure that the trip sheets are complete, I proceed under the assumption that they are.

I performed several regularity checks to insure that the trip sheets are internally consistent, and where they are not, I cleaned the data using a set of reasonable rules. These rules are outlined in detail in the appendix.

I coded the starting and ending locations on the trip sheets into eleven categories. These are Downtown Manhattan (below 14th street), Midtown Manhattan (14th Street to 59th Street), Uptown Manhattan (above 59th Street), The Bronx, Queens, Brooklyn, Staten Island, Kennedy Airport, LaGuardia Airport, Newark Airport, and other. Almost all trips (92 percent) started and ended in Manhattan. My earlier work using these data (Farber, 2005) suggests that there are important differences in stopping behavior between trips ending outside Manhattan and trips ending within Manhattan but no significant differences within these two broad locations. I proceed using an indicator to distinguish Manhattan from other locations.

I additionally collected data from the National Atmospheric and Oceanic Administration (NOAA) on temperature and precipitation in New York City. I collected daily average, minimum, and maximum temperatures, and total daily rainfall and snowfall in Central Park. I also collected hourly rainfall data at LaGuardia Airport.

---

[6] Das Gupta (2002) notes that ". . . the rules governing drivers have become more elaborate and punitive." and that tickets are ". . . zealously issued."

4.1 Shift-Level Summary Statistics

There are a total of 538 shifts and 12,187 trips for 15 drivers in the cleaned sample. Appendix table 1 contains average statistics by shift for each driver. I have data on an average of 27.8 shifts per driver. I have more than 30 shifts for 9 drivers and more than 20 sheets for 11 drivers. Hours worked per day is defined as the sum of driving time (the sum over trips of the time between the trip start time and the trip end time) and waiting time (the sum over trips of the time between the end of the last trip and the start of the current trip). Waiting time is substantial, accounting for 33 percent of working time on average. Break time averages about 52 minutes per shift.

There is substantial variation across drivers in average hours worked per day, with means ranging from 3.89 to 8.62. Still, the majority of the variation in daily work hours is within-driver variation across days. The standard deviation of daily work hours is 2.50. The R-squared from a regression of daily hours on a set of driver fixed effects is 0.152 with a residual root mean squared error of 2.33. Panel A of figure 1 contains a histogram of hours worked for the 584 shifts. The distribution is single-peaked, with the mode at eight hours.

There is also substantial variation across drivers in total fare income per day, with means ranging from \$97.10 to \$228.26.[7] Not surprisingly, daily income co-varies strongly with daily hours with a simple correlation of 0.91. As with hours, the majority of the variation in daily income is within-driver variation across days. The standard deviation of daily income is \$59.82. The R-squared from a regression of daily income on a set of driver fixed effects is 0.153 with a residual root mean squared error of \$55.77. A reference-dependent model where drivers had fixed (or only slightly variable) reference incomes would likely imply that more of the variation in income would be accounted for by driver fixed effects, and this will be considered directly using the estimates of the reference-dependent labor supply model. Panel B of figure 1 contains a kernel density estimate of daily income.[8]

The last column of appendix table 1 contains the daily average for each driver of their

---

[7] Income per day is the sum of fares. Tip income is not measured or accounted for.

[8] All kernel density estimates in this study use the Epanechnikov kernel. The bandwidths are listed in the figures.
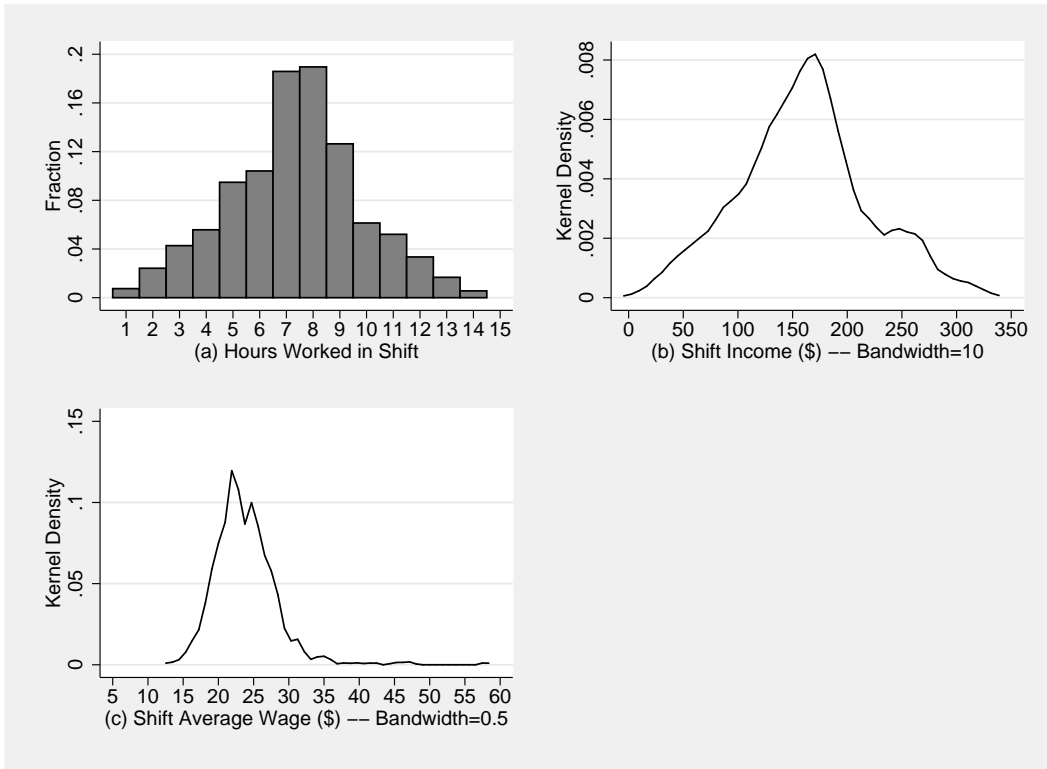
Figure 1: Distributions of Hours, Income, and Average Wage by Shift

hourly wage rate (total income divided by working hours). These show less inter-driver variation, ranging from a low of $21.46 to a high of $25.78. The standard deviation of the daily wage rate is $4.52. Most of this is within-driver variation as the R-squared from a regression of the daily wage on a set of driver fixed effects is 0.082 with a residual root mean squared error of $4.39. Panel C of figure 1 contains a kernel density estimate of the shift average hourly wage.

4.2   Trip-Level Summary Statistics

Figure 2 contains kernel density estimates of the distributions of trip times, waiting times, and fares for the 12,187 trips in my sample. I have truncated these distributions at 60 minutes for trip and waiting times and at $50 for fares. These truncations do not change the shape of the distribution since the kernel density estimates in these upper tails is essentially
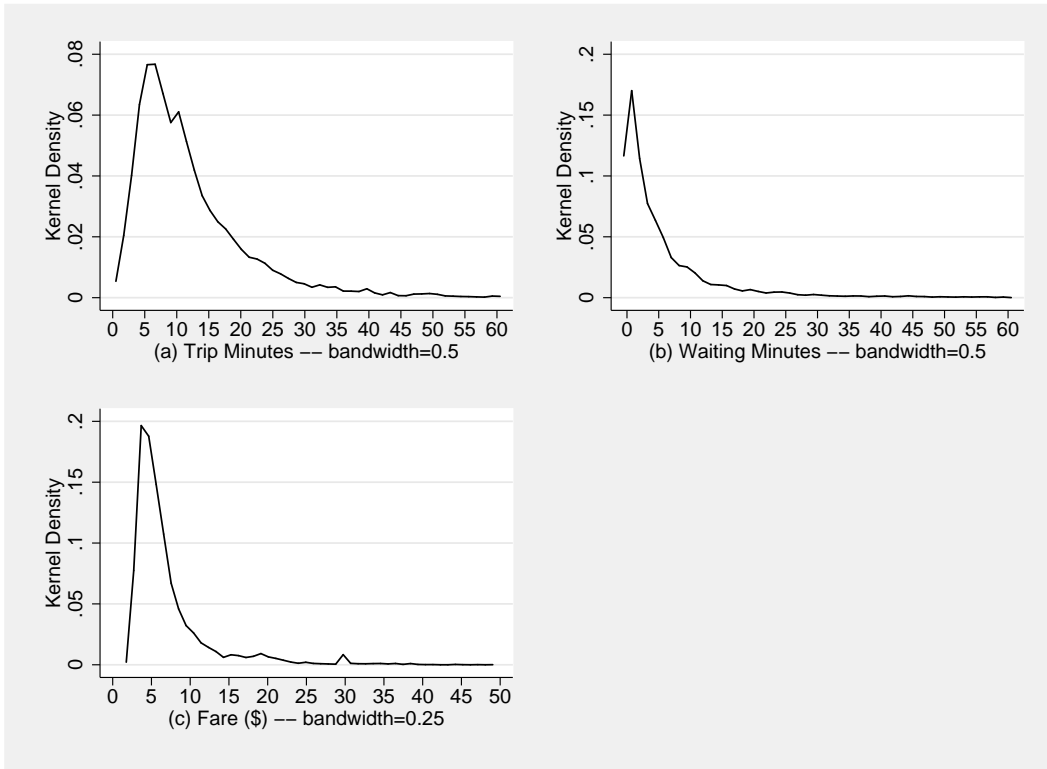
Figure 2: Kernel Density Estimates of Trip Times, Wait Times, and Fares

zero.[9]

Panel A of figure 2 contains a kernel density estimate of the distribution of trip times. Median time per trip is 10 minutes and the mean is 12.1 minutes. That trips are this short reflects the fact that 92.5 percent of the trips in my sample begin and end in Manhattan. The fact that the great majority of trips are short implies that opportunities to stop driving occur frequently.

Panel B of figure 2 contains a kernel density estimate of the distribution of waiting times before each trip. Median waiting time is 3 minutes and the mean waiting time is 6.0 minutes. There are 2640 trips with zero wait time. As I describe in the appendix, I reclassified 316 long wait times between fares as break times, and fares after these breaks are classified as having preceding wait times of zero. Variation in waiting time is a key determinant of earnings opportunities since the earnings during a trip is determined by a consistent formula

---

[9] There are 37 trips with times greater than 60 minutes and 26 waiting times greater than 60 minutes. There are 8 fares greater than $50.

based on time and distance that does not vary other than because of variation in route and traffic.

Panel C of figure 2 contains a kernel density estimate of the distribution of fares. The median fare is $5.30, and the mean fare is $7.06. Once again, the small size of the fares (which exclude tips) is due to the fact that most trips are intra-Manhattan. The average intra-Manhattan fare is $5.93, while the average fare that starts or ends outside Manhattan is $20.76. The small blip at $30 represents the flat rate between Kennedy Airport and Manhattan in force during my sample period. The fact that the great majority of fares are small implies that the first opportunity to stop driving after the reference income is exceeded will generally arise at an income level not greatly in excess of the reference income level.

5 Estimation of the Labor Supply Model

In this section I present maximum likelihood estimates of the parameters of the reference-dependent labor supply model based on the likelihood function defined in equations 3.7-3.15 and on the data described in the previous section. The parameters to be estimated include

- the coefficient vector (β) of the X vector in the stopping probability function (equation 3.1),

- the parameter (δ) indexing increment to the probability of stopping after the reference level of income is reached,

- the vector of individual mean reference income levels (θ), and

- the variance (σ^2) of the random component (μ) in the reference income function (equation 3.5).

In addition to a constant, the specification of the X vector in the stopping probability function includes

- indicators for eight categories of hours worked at trip end,

- six indicators for the day of week,

- indicators for eighteen clock hours at trip end,

- a day-shift indicator,

- an interaction of day-shift with clock hour 3-4PM meant to capture the likelihood that a day-shift driver must turn the car over to a night-shift driver at that time of day,

- four variables measuring weather including daily snowfall, hourly rainfall, high heat (maximum temperature $>= 80$ degrees), and cold (minimum temperature < 30 degrees), and

- an indicator for location outsize of Manhattan.

Table 1 contains estimates of key parameters from the reference-dependent labor supply model. The estimates of β, in the first column, show a strongly increasing probability of stopping with hours worked on the shift. The standard normal latent variable determining the probability of stopping (equation 3.1) increases by 1.5 from the beginning of the shift to 10th hour. Aside from a lower probability of stopping on very hot days, the weather does not appear to be significantly related to the stopping decision. Ending a fare outside of Manhattan results is a significantly larger probability of stopping, probably due to the fact that many drivers outside of Manhattan.

The estimates of the individual mean reference income levels (the θ vector) are estimated fairly precisely and vary significantly across drivers, from a low of \$105.51 for driver 2 to a high of \$324.46 for driver 11. The mean reference income is \$191.42 with a standard deviation across the fifteen drivers of \$61.40. The hypothesis that all drivers have the same mean reference income level can be rejected at any reasonable significance level (p-value $< 10^{-15}$) using a likelihood-ratio test.

With regard to estimates that reflect on the role of reference income levels on labor supply, the evidence is mixed. The estimate of δ, which indexes the change in the probability of stopping once the reference income level for the day is reached, is substantial at 5.63 (single tailed p-value = 0.06). Given the standard normal distribution of the latent variable determining stopping in equation 3.1, this implies that the stopping probability is very close to unity once the reference income level is reached, regardless of when in the shift it is

Table 1: Maximum Likelihood Estimates of Reference-Dependent Stopping Model

Variable	$\hat{\beta}$	Driver	$\hat{\theta}$
Hour $\leq$ 2	-1.046	1	217.71
	(0.173)		(24.57)
Hour 3-5	-0.671	2	105.51
	(0.135)		(16.37)
Hour 6	-0.424	3	166.27
	(0.136)		(10.93)
Hour 7	-0.175	4	173.41
	(0.117)		(19.37)
Hour 9	-0.097	5	157.12
	(0.169)		(9.29)
Hour 10	0.536	6	253.58
	(0.179)		(30.66)
Hour 11	0.485	7	191.49
	(0.218)		(20.29)
Hour $\geq$ 12	0.952	8	205.47
	(0.242)		(12.91)
Min Temp $<$ 30	0.034	9	162.00
	(0.092)		(18.86)
Max Temp $\geq$ 80	-0.269	10	191.42
	(0.130)		(19.48)
Hourly Rain	0.036	11	324.46
	(0.139)		(96.63)
Daily Snow	0.243	12	290.58
	(0.192)		(20.78)
Non-Manhattan	0.405	13	182.66
	(0.142)		(13.72)
Day Shift	-0.446	14	278.48
	(0.151)		(14.54)
End of Day Shift	1.21	15	127.43
	(0.241)		(12.04)
Constant	-1.11		
	(0.185)		
$\hat{\delta}$	5.63	$\hat{\sigma}^2$	2779.58
	(3.63)		(375.8)

Note: The sample includes 12187 trips in 538 shifts for 15 drivers. The model also includes 6 indicators for day of the week, and 18 indicators for hour of the day The hours from 5AM to 10AM have a common fixed effect. Standard errors are reported in parentheses. The maximized log-likelihood value is -1612.6.

reached. This would seem to be strong evidence of the importance of reference-dependent preferences in this context. It implies that, while drivers may or may not stop before their reference income level is reached, they are almost sure to stop once they reach that level.

However, there is substantial daily variation around a given driver's mean reference income level. The estimated variance of 2779.58 implies a daily within-driver standard deviation of $52.72. Based on the assumption of normality, the inter-quartile range for a particular driver with a mean reference income level of $200 is from $164 to $236, and the 95 percent confidence interval is from $97 to $303. On this basis, the predictive power of the reference income level for daily labor supply is likely quite limited.

To the extent that the estimated reference income level is important in determining labor supply, I would expect that the income and hours will be strongly correlated with the estimates of θ. In fact, there is a strong positive correlation between both average income and average hours across shifts for a given driver and that driver's estimated θ_i. These correlations are 0.72 and 0.75 respectively. The strong correlations are due to the fact that it is the variation in average income across shifts for a given driver that is driving the estimates of θ_i. However, the estimates of θ_i are less successful in accounting for variation in income and hours across shifts within drivers.

Consider the following simple regression at the shift level of income on the estimated mean reference income level for the 538 shifts in my sample.

$$Y_{ij} = \alpha_0 + \alpha_1 \hat{\theta}_i + \gamma_{ij}, \tag{5.1}$$

where Y_{ij} is income for driver i on shift j. While the estimate of α_1 is significantly positive at 0.262 (s.e. = 0.038), it is substantially less than one and the R^2 from this regression is only 0.08. Thus, very little of the overall variation in hours across shifts can be accounted for by variation across drivers in the estimated mean reference income level. In fact, most of the variation in income across shifts is within driver variation. A regression of daily income on a set of driver fixed effects yields an R^2 of 0.15.[10]

---

[10] A parallel analysis for the relationship between hours and $\hat{\theta}$ yields the same pattern. The estimate of α_1 is significantly positive at 0.011 (s.e. = 0.0015), the R^2 from this regression is 0.08. A regression of daily hours on a set of driver fixed effects yields an R^2 of 0.15.
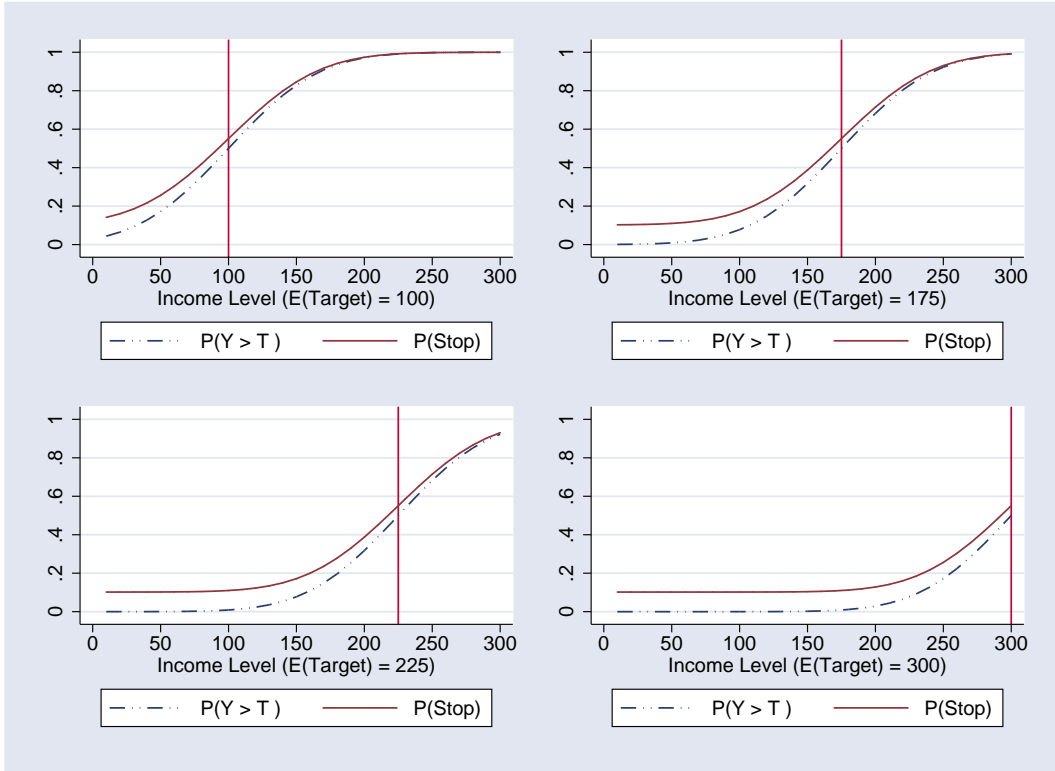
Figure 3: Predicted Probabilities of Income Exceeding the Reference Level and of Stopping, by Mean Target Level: midnight Saturday, clear moderate weather, in 8th hour. Based on parameters in table 1.

In order to determine how differences in the mean reference income level for a driver translates into differences in the probability of stopping, figure 3 contains plots for four different values of the mean reference income level ($100, $175, $225, and $300) of 1) the estimated probability that income exceeds the reference level as a function of income and 2) the estimated probability that a driver ends his shift as a function of income. The probabilities of stopping are calculated using the estimates in table 1 assuming that it is midnight on Saturday in clear moderate weather in the 8th hour of a night shift.

These plots are revealing. The probability that income after trip t exceeds the reference level for driver i (represented by the dashed line) is the CDF of a normal distribution with the indicated mean and variance equal to 2779.58. This is

$$P(Y_{it} > T_i) = \Phi[(Y_{it} - \theta_i)/\sigma] \tag{5.2}$$

24

and is increasing in Y_{it}. Conditional the value of the reference level of income, the probability that driver i stops after trip t at any income level is

$$P_{it}^s = \Phi[X_{it}\beta] \cdot (1 - P(Y_{it} > T_i)) + \Phi[X_{it}\beta + \delta] \cdot P(Y_{it} > T_i). \tag{5.3}$$

Given the large estimated value of δ, it is the case that $\Phi[X_{it}\beta + \delta]$ is approximately one for all observed values of X. Thus, equation 5.3 can be rewritten approximately as

$$P_{it}^s = \Phi[X_{it}\beta] \cdot (1 - P(Y_{it} > T_i)) + P(Y_{it} > T_i) \tag{5.4}$$

It is clear from equation 5.4 that the probability of stopping is greater than the probability that income exceeds the reference level and that the difference between the two is declining in income.[11]

Holding hours worked and other factors fixed, the probability of stopping increases with income. More importantly, the relationship between income and the probability of stopping is strongly related to the mean reference income level. In the upper left panel of figure 3, where the mean reference income level is $100, the probability of stopping rises rapidly, and is almost 0.6 at an income level of $100. At an income of $150, the probability of stopping is approximately 0.8, and at an income of $200 the probability of stopping is approaching 1.0.

At a mean reference income level of $175, in the upper right panel of figure 3, the probability of stopping rises less rapidly, and is only about 0.2 at an income level of $100. At an income of $150, the probability of stopping is approximately 0.4, and at an income of $200 the probability of stopping is approaching 0.6.

In the lower left panel of figure 3, where the mean reference income level is $225, the probability of stopping at lower income levels is small. At an income level of $100, the probability that income exceeds the reference income level is almost zero, and the probability of stopping is less than 0.2. At an income of $150, the probability of stopping has grown only slightly, and at an income of $200 the probability that income exceeds the reference level is approximately 0.3 and the probability of stopping is approaching 0.4. As income increases beyond $200, the probability of stopping increases along with the probability that income exceeds the reference level, both probabilities reaching 0.8 at about $270.

---

[11] The difference between the probability of stopping and the probability that income exceeds the reference income level is the first term in equation 5.4, which goes to zero as income increases.

At the highest mean reference income level of \$300, in the lower right panel of figure 3, the probability that income exceeds the reference level is virtually zero until income is about \$200. As a result the probability of stopping is roughly constant in income until income reaches about \$200. Both probabilities then increase, reaching approximately 0.5 at \$300.

The clear message is that consideration of reference income levels that vary across drivers has important implications for daily labor supply. It is no surprise that drivers differ in their mean reference income level and that hours increase with the mean reference income level. Yet the large amount of variation day-to-day in reference income levels for a given driver raises questions about the interpretation of the pattern of results as consistent with reference-dependent preferences. It may simply be that there is heterogeneity among drivers in their labor supply behavior and that there is day-to-day variation in labor supply having to do with other factors.

5.1 Estimates of an Alternative Model without Reference Dependent Preferences

In section 3.1, I presented a model of taxi driver labor supply that did not rely on reference-dependent preferences. This model implied a straight-forward probit representation for stopping behavior on a shift. The probit likelihood function is implied by the expression for the probability that a shift ends after trip t defined in equation 3.18. The results are this analysis are presented in table 2.

As I noted in section 3.1, the standard probit model is a special case of the reference-dependent model where there is no increment to the stopping probability at the reference income level ($\delta = 0$) or where the reference income level is very small or very large. The first column of table 2 contains estimates of the probit model that constrains there to be no reference-dependent preferences. Like the estimates in table 1, these estimates show a strong positive relationship between hours worked and the probability of stopping. They also show that drivers are more likely to stop after a trip that ends outside Manhattan and that day shift drivers are more likely to stop than night shift drivers.

The constrained model has 17 fewer parameters than the model in table 1 (15 individual

Table 2: Probit Estimates of Stopping Model

Variable	(1)	(2)	(3)	(4)
Hour ≤ 2	-0.776	-0.803	-0.635	-0.690
	(0.116)	(0.134)	(0.177)	(0.195)
Hour 3-5	-0.411	-0.407	-0.309	-0.325
	(0.088)	(0.098)	(0.120)	(0.131)
Hour 6	-0.369	-0.388	-0.334	-0.324
	(0.098)	(0.102)	(0.109)	(0.114)
Hour 7	-0.173	-0.164	-0.136	-0.132
	(0.089)	(0.092)	(0.094)	(0.096)
Hour 9	-0.068	-0.037	-0.063	-0.048
	(0.110)	(0.114)	(0.116)	(0.118)
Hour 10	0.308	0.370	0.321	0.349
	(0.125)	(0.132)	(0.136)	(0.146)
Hour 11	0.319	0.426	0.360	0.390
	(0.155)	(0.163)	(0.169)	(0.180)
Hour ≥ 12	0.798	0.777	0.684	0.738
	(0.161)	(0.170)	(0.182)	(0.188)
Min Temp < 30	-0.055	0.077	0.086	0.082
	(0.062)	(0.073)	(0.074)	(0.074)
Max Temp ≥ 80	-0.109	-0.163	-0.156	-0.151
	(0.073)	(0.082)	(0.082)	(0.082)
Hourly Rain	0.032	0.100	0.103	0.105
	(0.098)	(0.102)	(0.101)	(0.102)
Daily Snow	0.089	0.122	0.112	0.118
	(0.148)	(0.153)	(0.153)	(0.154)
Non-Manhattan	0.553	0.537	0.530	0.533
	(0.086)	(0.088)	(0.088)	(0.088)
Day Shift	0.379	0.049	0.037	0.038
	(0.073)	(0.109)	(0.109)	(0.110)
End of Day Shift	0.595	0.577	0.561	0.572
	(0.181)	(0.201)	(0.201)	(0.205)
Shift Income ÷ 100	---	---	0.151	---
			(0.104)	
Constant	-0.723	-0.946	-1.219	-0.974
	(0.131)	(0.158)	(0.245)	(0.166)
Driver (14)	No	Yes	Yes	Yes
Day-of-Week (6)	Yes	Yes	Yes	Yes
Hour-of-Day (18)	Yes	Yes	Yes	Yes
Accumulated Income (9)	No	No	No	Yes
Log L	-1673.9	-1619.6	-1618.6	-1614.1

Note: The sample includes 12187 trips in 538 shifts for 15 drivers. The hours from 5AM to 10AM have a common fixed effect. The ten income categories in column 4 are <25, 25-49, 50-74, 75-99, 100-124, 125-149, 150-174, 175-199, 200-224, and ≥225. The base category is at midnight Saturday on the night shift in the eighth hour in dry moderate weather. Standard errors are reported in parentheses.

27

reference income levels, δ, and σ^2 are not estimated). A likelihood ratio test based on the log-likelihood values in table 1 and the first column of table 2 clearly rejects the constrained model. However, the constrained model does not allow for any fixed differences across drivers or any direct role for accumulated income. Thus, the rejection of the constrained model should not be taken as evidence that reference-dependent preferences are important.

A standard probit model with driver fixed effects is not nested in the reference-dependent preference model presented in table 1. The appropriate unconstrained model in this case would have driver-specific mean reference income levels along with driver fixed effects in the probability of stopping function. However, as I noted in section 3.2, estimation of two sets of driver fixed effects in the model is asking too much of the data.[12]

In order to get an idea of how well the standard probit model can do in accounting for driver stopping decisions, re-estimated the standard probit model including driver fixed effects. These estimates are contained in the second column of table 2. The log-likelihood value improves from -1673.9 to -1619.6, and the hypothesis that the driver fixed effects are all zero can be rejected at any reasonable level (p-value < 0.0005). The general pattern of results in the fixed-effect model is the same as described for the results in column 1. One exception is that stopping probabilities are not significantly higher on the day shift. Apparently, drivers who choose the day shift generally work fewer hours, but when drivers change shifts they bring their hours with them. It is the case that day shift drivers are more likely to stop in the 3PM and 4PM hours other things equal.

It is worth noting that the maximized log-likelihood of the fixed-effect probit model is substantially lower than than of the reference-dependent utility model (-1619.6 vs. -1612.6). However, there are three fewer parameters in the fixed-effect probit model (no reference income level, no δ, and no σ^2). Because these models are not nested, I do not perform a formal statistical test, but it does appear that the reference-dependent utility model fits the data somewhat better than the standard probit model.

One potential explanation for the superior fit of the reference dependent utility model could be that the standard probit does not allow for differences in accumulated income

---

[12] I did attempt to estimate such a model, and I was not able to find an optimum despite considerable effort.

conditional on hours worked to have an effect on the stopping probability. In order to investigate this, column 3 of table 2 contains estimates of the fixed-effect probit model that includes a measure of accumulated shift income. The t-statistic on the coefficient of income is 1.46 and the log-likelihood value improves from -1619.6 to -1618.6. The implied p-value for a likelihood-ratio test of the hypothesis that the coefficient of income is zero is 0.157. Thus, a linear representation of income does not account for the difference in the fit of the reference-dependent model and the standard probit model.[13]

It may be that including only a linear representation of income in the probit stopping model is too constrained. In order to investigate this, the last column of table 2 contains estimates of the standard probit model including nine indicators for income category. Only the indicator for income less than \$25 is significantly different from zero at conventional levels. The addition of these indicator variables improves the log-likelihood from -1619.6 to -1614.1, but the hypothesis that the coefficients of the income variables are zero cannot be rejected (p-value = 0.276). This specification of the fixed-effect probit model, which has six more parameters than the reference-dependent preferences model, does not fit the data quite as well as the reference-dependent preferences model.

6   Final Remarks

The results of the empirical analysis are intriguing. On the one hand, the reference-dependent utility model appears to fit the data somewhat better than a standard probit stopping model that allows for variation in accumulated income, hours worked, driver fixed effects, and other factors. The reference-dependent utility model builds a particular non-linearity into the relationship of income with the probability of stopping. This is that the probability of stopping increases discretely once a given income level is reached. This particular characterization does seem to contribute mildly to the fit in a parsimonious way.

On the other hand, the estimates imply that reference income levels change substantially

---

[13] I also estimated the reference-dependent utility model including accumulated income in the X vector. This improved the log-likelihood only slightly, from -1612.6 to -1612.4. The t-statistic on the income variable was 0.427. This finding is consistent with the estimates in table 2.

from day to day for a particular driver. Unless this movement is systematic and predictable, a possibility that I have not examined carefully, the usefulness of the reference-dependent utility model for understanding labor supply is limited. Further research into the determinants of temporal variation in daily reference income levels will move us toward an ultimate conclusion regarding the importance of reference-dependent preferences in this context.

7   References

Altonji, Joseph and Christina Paxson, "Labor Supply Preferences, Hours Constraints, and Hours-Wage Trade-offs" *Journal of Labor Economics* 6 (April 1988), pp. 254-276.

Bateman, Ian; Alistair Munro; Bruce Rhodes; Chris Starmer; and Robert Sugden, "A test of the Theory of Reference-Dependent Preferences," *Quarterly Journal of Economics* 112 (May 1997) 479-505.

Blundell, Richard and Thomas MaCurdy, "Labor Supply: A Review of Alternative Approaches," in *Handbook of Labor Economics*, O. Ashenfelter and D. Card, eds. Amsterdam, North Holland, 1999, vol 3a, pp. 1559-1695.

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

Chou, Yuan K. "Testing Alternative Models of Labor Supply: Evidence from Taxi Drivers in Singapore," University of Melbourne, Department of Economics Research Paper No. 768, 2000.

Das Gupta, Monisha. "Of Hardship and Hostility: The Impact of 9/11 on Taxi Drivers in New York City," University of Hawaii at Manoa, undated unpublished mimeo, 2002.

Dickens, William and Shelley Lundberg. "Hours Restrictions and Labor Supply," *International Economic Review*, 34 (February 1993), pp. 169-192.

Farber, Henry S. "Is Tomorrow Another Day? The Labor Supply of New York City Cab Drivers," *Journal of Political Economy* 113 (February 2005) (in press).

Fehr, Ernst and Lorenz Goette. "Inter-temporal Substitution at Work? Evidence from a Field Experiment," mimeo, April 2002.

Ham, John. "Estimation of a Labor Supply Model with Censoring Due to Unemployment and Underemployment," *The Review of Economic Studies* 49 (1982), pp. 335-354.

Ham, John and Kevin T. Reilly. "Testing Intertemporal Substitution, Implicit Contracts, and Hours Restriction Models of the Labor Market Using Micro Data," *American Economic Review* 92 (September 2002), pp. 905-927.

Kahn, Shulamit and Kevin Lang. "The Effect of Hours Constraints on Labor Supply Estimates," *The Review of Economics and Statistics* 93 (1991), 605-611.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias," *Journal of Economic Perspectives* 5 (1993), pp. 193-206.

Koszegi, Botond and Matthew Rabin. "A Model of Reference-Dependent Preferences," University of California, Berkeley, Institute of Business and Economic Research, Department of Economics, Working Paper No. E04-337 (January 2002).

Killingsworth, Mark R. and James J. Heckman. "Female Labor Supply: A Survey," in *Handbook of Labor Economics*, O. Ashenfelter and R. Layard, eds. Amsterdam, North Holland, 1986, vol 1, pp. 103-204.

MaCurdy, Thomas E. "An Empirical Model of Labor Supply in a Life-Cycle Setting," *Journal of Political Economy* 89 (December 1981), pp. 1059-1085.

Munro, Alistair and Robert Sugden. " On the Theory of Reference-Dependent Preferences," *Journal of Economic Behavior and Organization* 50 (2003), 407-428.

Neumark, David and Andrew Postelwaite. "Relative Income Concerns and the Rise in Married Women's Employment," *Journal of Public Economics* 70 (1998), 157-183.

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

Pencavel, John. "Labor Supply of Men: A Survey," in *Handbook of Labor Economics*, O. Ashenfelter and R. Layard, eds. Amseterdam, North Holland, 1986, vol 1, pp. 3-102.

Thaler, Richard. "Toward a Positive Theory of Consumer Choice,"*Journal of Economic Behavior and Organization* 1 (1980), 39-60.

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

Appendix: Procedures for Cleaning Data

I perform several regularity checks to insure that the sheets are internally consistent, and where they are not, I clean the data using a set of reasonable rules. I outline that process here. The main consistency checks are that 1) trips must start no earlier than the end of the previous trip and 2) trips must start before they end (be of positive duration). I also develop

31

a set of rules to impute starting or ending times for trips where such times are missing and fares where fares are missing. For the most part, the trips on my trip sheets satisfy these rules and have non-missing time and fare data. In the small number of cases where these rules are not satisfied or time or fare data are missing (about one percent of trips), I use the following procedures.

- Impute new trip times for 38 trips of zero duration. I start by computing the length of time between the end of the previous trip and the start of the next trip. Clearly, the time of the "current" trip is no greater than this available gap. I compute a predicted trip length based on regression of trip length on the fare and its square (R-squared = 0.68). In almost all cases, this predicted trip length is less than the available gap. I then center the predicted trip length in the available gap and compute the implied adjusted start and end times for the trip. In the handful of cases where the predicted trip length is longer than the gap (never by more than a few minutes), I set the trip length equal to the available gap and compute the adjusted start and end times for the trip accordingly.

- Impute new trip times for 52 cases where one or both of the starting and ending times are missing. Most of these are imputed using the same predicted trip length defined above as a function of the fare. If the start time is missing, it is computed as the end time minus the predicted trip length. Analogously, if the end time is missing, it is computed as the start time plus the predicted trip length. Where both times are missing, they are computed as defined above for trips of zero duration. There are a handful of cases where one or the other of the times are missing and the fare is missing. In these cases, I predict trip length using a complete interaction of starting and ending locations of the trip (R-squared = 0.44).

- Adjust times where they are out of sequence. There are 83 cases where a trip apparently started before the previous trip ended. The mean overlap was 2.7 minutes, and the maximum overlap was 9 minutes. I adjust these by reducing the end time of the previous trip by one-half of the overlap and increasing the start time of the current trip by by one-half of the overlap.

- Impute fares for 45 cases where the fare is missing. I impute fares in these cases using predicted values from a regression of fare on trip length and a complete set of interactions of starting and ending locations (R-squared = 0.83). In the 7 cases where location is missing I predict based on a model using only trip length (R-squared = 0.64).

- Classify as breaks (non-work time during the day) certain long wait times between fares. I classify as breaks 1) 195 periods of at least 30 minutes between a fare that ended in Manhattan and a fare that started in Manhattan, 2) 112 periods of at least 60 minutes between fares that either started or ended outside Manhattan but did not end at either LaGuardia or Kennedy Airport, and 3) 9 periods of at least 120 minutes that ended at LaGuardia or Kennedy Airport.

- Impute starting locations to 144 trips and ending locations to 223 trips with missing or illegible locations. Where the starting location is missing, I assume that the starting location is the same as the ending location of the previous trip. Where the ending location is missing, I assume that the ending location is the same as the starting location of the next trip. In cases where the necessary information for imputation by these rules is not available, I examined the individual strip sheets and made an imputation based on the "nearest" available locations, the size of the fares, and elapsed time of the trip. Since over 96 percent of trips start in Manhattan and over 95 percent of trips end in Manhattan, almost all of the allocated locations are in Manhattan.

- Drop 46 shifts for drivers with 10 or fewer shifts in my sample. This eliminates 1274 trips from the sample, leaving 538 shifts including 12,187 trip for 15 drivers.

Appendix Table 1: Shift Level Summary Statistics, by Driver

Driver	# of Shifts	Average Trips	Working Hours	Driving Hours	Waiting Hours	Break Hours	Total Income	Average Wage
Driver 1	39	23.56	6.85	4.32	2.53	0.90	157.58	23.16
Driver 2	14	12.29	3.89	2.78	1.11	2.41	97.10	25.11
Driver 3	40	22.10	6.28	4.52	1.76	0.39	147.51	23.89
Driver 4	23	16.52	6.46	3.98	2.48	2.11	144.96	23.65
Driver 5	24	22.29	6.47	4.42	2.05	0.74	160.71	25.59
Driver 6	37	25.32	7.78	5.13	2.64	0.86	172.44	22.54
Driver 7	19	25.58	7.17	5.47	1.70	0.54	162.02	23.23
Driver 8	45	20.27	6.35	3.90	2.45	1.65	133.19	21.46
Driver 9	13	19.46	6.15	4.03	2.13	0.55	157.95	25.78
Driver 10	17	21.29	7.06	4.49	2.57	0.64	165.84	23.59
Driver 11	70	25.06	6.84	4.56	2.28	0.93	172.01	25.62
Driver 12	72	28.10	8.53	5.84	2.69	0.60	203.05	24.01
Driver 13	33	17.06	6.91	4.63	2.29	0.97	163.51	23.91
Driver 14	46	24.46	7.10	4.80	2.30	0.67	156.23	22.00
Driver 15	46	19.17	5.32	3.66	1.66	0.24	128.97	24.72
All	538	22.65	6.84	4.58	2.26	0.87	160.03	23.80