

WHY YOU CAN'T FIND A TAXI IN THE RAIN AND OTHER LABOR SUPPLY LESSONS FROM CAB DRIVERS*

HENRY S. FARBER

I replicate and extend the seminal work of Camerer et al. (“Labor Supply of New York City Cabdrivers: One Day at a Time,” *Quarterly Journal of Economics*, 112 [1997], 407–441), who find that the wage elasticity of daily hours of work for New York City taxi drivers is negative and conclude that their labor supply behavior is consistent with reference dependence. In contrast, my analysis of the complete record of all trips taken in NYC taxi cabs from 2009 to 2013 shows that drivers tend to respond positively to unanticipated as well as anticipated increases in earnings opportunities. Additionally, using a discrete choice stopping model, the probability of a shift ending is strongly positively related to hours worked but at best weakly related to income earned. I find substantial heterogeneity across drivers in their elasticities, but the estimated elasticities are generally positive and rarely substantially negative. I find that new drivers with smaller elasticities are more likely to exit the industry, whereas drivers who remain quickly learn to be better optimizers (have positive labor supply elasticities that grow with experience). These results are consistent with the neoclassical optimizing model of labor supply and suggest that consideration of gain-loss utility and income reference dependence is not an important factor in the daily labor supply decisions of taxi drivers. *JEL* Codes: D01, D03, J22.

I. INTRODUCTION

That it is difficult to find a taxi in the rain has been a standard complaint in Manhattan for as long as there have been taxis. If asked why this is the case, the answer from an economist 20 years ago would have been that rainy weather increases the demand for taxi rides and there is no or an insufficiently rapid supply response to meet this transitory demand increase. That answer may have changed in recent years. In their seminal work, Camerer, Babcock, Loewenstein, and Thaler (1997), referred to here as CBLT, present evidence, based on a regression of log daily hours on log average hourly earnings, suggesting that the daily labor supply function of taxi drivers is negatively sloped so that a

*This article is based on my Albert Rees Lecture at the annual meeting of the Society of Labor Economists, May 2, 2014, Arlington, VA. The author thanks participants in workshops at the Federal Reserve Bank of Atlanta, Princeton University, MIT, Harvard University, Brandeis University, Yale University, University of Oxford, University College London, University of Oslo, University of Chicago, Temple University, and Ecole Polytechnique for helpful comments.

© The Author(s) 2015. Published by Oxford University Press, on behalf of President and Fellows of Harvard College. All rights reserved. For Permissions, please email: journals.permissions@oup.com

The Quarterly Journal of Economics (2015), 1975–2026. doi:10.1093/qje/qjv026.

Advance Access publication on July 13, 2015.

transitory change in the wage results in a reduction in hours worked. On this basis, they characterize taxi drivers as having income reference-dependent preferences, which can be summarized simply by saying that workers will set a daily income target and generally work until that target is met. Others have also found that labor supply curves for taxi drivers appear to slope downward, and the consensus of much of this work is that taxi drivers do have reference-dependent preferences (e.g., Chou 2002; Agarwal et al. 2015).¹ This suggests an alternative answer to the question of why it is difficult to find a taxi in the rain: to the extent that drivers have a daily income target and a rain-induced increase in demand increases earnings, drivers will reach their targets sooner and quit driving for the day. The new view, then, is that at least part of the reason you can't find a taxi in the rain is because drivers reach their daily earnings targets quickly and go home so that the demand increase is exacerbated by the resulting decline in supply.²

The question of whether income reference dependence plays a substantial role in labor supply decisions is important, both intellectually and in designing public tax and transfer policies. There is a natural tension between the standard neoclassical optimizing model of labor supply and the model based on reference-dependent preferences. Setting income targets is an inefficient way to earn money because it implies working less on high-wage days and working more on low-wage days. The neoclassical model implies the opposite. Clearly, over a period of days, the neoclassical optimizer works fewer hours than the target earner to earn the same income.³ In this study I use new data, consisting of the complete records of all taxi drivers in New York City over the five-year period 2009–2013 to estimate models of labor supply to determine whether reference-dependent preferences play an important role in explaining labor supply in this setting or whether the standard neoclassical model can account for most of the observed variation. The use of these new data addresses

1. Using different approaches, Crawford and Meng (2011), Doran (2014), and Agarwal et al. (2015) find support for reference-dependent preferences in analyses of labor supply of taxi drivers. Koszegi and Rabin (2006) and Ordoñez et al. (2009) are examples of how this result has been accepted generally.

2. See Ordoñez et al. (2009) for an explicit statement that reference-dependent preferences are part of the explanation for difficulty in finding a taxi in the rain.

3. This is the point of the taxi driver example used by Ordoñez et al. (2009), and the inefficiency of target earnings behavior is highlighted by Camerer (1997).

a weakness of much of the earlier work on taxi drivers, including my own, that the analyses are based on very small convenience samples.

A continuing focus of this comparison of models is the daily hours decisions of taxi drivers, and with some exceptions, the model used is a regression of log daily hours on the log daily wage (the log of daily average hourly earnings). This is why I focus on estimating the slope of the daily hours function in most of what I do here. An alternative empirical approach is to model the stopping decision of a driver, where the driver decides at the completion of each trip whether to continue driving or to end the shift. This discrete-choice stopping decision may be a function of accumulated hours and income on the shift as well as factors that reflect variation in supply and demand factors such as time of day, day of week, and month of year, among other variables. Using this reduced form approach in earlier work (Farber 2005), I found that taxi driver labor supply is best characterized by the neoclassical model and that there is little evidence in support of income reference-dependent preferences.⁴ Following this approach, I supplement my regression analysis of the slope of the labor supply function with a reduced-form discrete choice analysis of the stopping decisions of drivers that uses the new data.

After estimating the basic models of labor supply, I investigate whether drivers differ in their labor supply behavior (are some drivers target earners while others are optimizers?).

4. In a later paper (Farber 2008) I estimated a structural version of the stopping model using the same data on NYC taxi drivers and again found little evidence for reference-dependent preferences. Crawford and Meng (2011), using the same data, estimate a structural stopping model that allows for reference points in both daily income and daily hours, and they conclude that the data are consistent with this dual reference point model. The empirical evidence from other settings using a variety of methods is mixed. Oettinger (1999) examines the labor supply of stadium vendors at baseball games and finds evidence that labor supply on the extensive margin (number of vendors showing up for games) is consistent with the neoclassical model. Fehr and Goette (2007) run a field experiment varying the piece rate paid to bicycle messengers. Their evidence is generally consistent with the neoclassical model in that messengers work more in months with high piece rates. On the other hand, they interpret evidence that messengers work fewer hours per day on days in months with high piece rates as evidence of reference-dependent preferences. See also Nguyen and Leung (2009) for an analysis of labor supply in fisheries and Chang and Gross (2014) for an analysis of labor supply in fruit picking that find evidence consistent with reference-dependent preferences. Camerer (2000) presents a nice survey of evidence from the field on loss aversion and reference dependence in a variety of settings.

I then investigate whether new drivers learn to be better optimizers and whether drivers who not strong optimizers disproportionately quit the industry.

II. SETTING THE STAGE: COMPETING THEORIES OF LABOR SUPPLY

In the standard neoclassical intertemporal model of labor supply, individuals work in period t until the shadow value of time (a function of lifetime wealth/income and increasing in hours in a given period) equals the period t wage rate.⁵ The model implies that there is intertemporal substitution in labor supply across periods so that a transitory increase in the wage rate in period t implies an increase in period t labor supply because the shadow value of time conditional on hours is unaffected. However, a permanent change in the wage will have an offsetting income effect as lifetime wealth increases, increasing the shadow value of time conditional on hours.⁶

The reference dependent model of choice has its roots in the literature on loss-aversion (Kahneman and Tversky 1979; Tversky and Kahneman 1991). In the context of the daily labor supply decisions of taxi drivers, the basic idea of the reference-dependent preference model is that a driver has in mind a particular reference level of daily income, and utility as a function of income is evaluated relative to this reference level. The loss in utility from failing to reach the reference income level by some amount exceeds the gain in utility from exceeding the reference income level by the same amount. In other words, the individual is loss averse. There is a kink in the utility function at the reference income level, with higher marginal utility below the kink and lower marginal utility above the kink.

Consider the following simple model of labor supply with reference-dependent preferences. Individuals facing a wage rate W receive utility from income ($Y = Wh$) and disutility from hours of

5. See MaCurdy (1981) for an early empirical analysis of intertemporal substitution in labor supply. Blundell and MaCurdy (1999) present a relatively recent survey of the literature on labor supply.

6. See Ashenfelter, Doran, and Schaller (2010) for an analysis of the effect of two fare increases (1996 and 2004) on the labor supply of NYC taxi drivers. They find an elasticity of -0.2 in response to these permanent fare increases.

work (h). Individuals have a kink in their utility function at some reference level of income (T):

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

$$(2) \quad U(Y, h) = (1 - \alpha)(Y - T) - \frac{\theta}{1 + \nu} h^{1+\nu} \quad Y \geq T,$$

where the parameter $\alpha > 0$ controls the change in marginal utility at the reference point, θ indexes the disutility of hours, and ν is a parameter related to the elasticity of labor supply. This specification follows the model of Koszegi and Rabin (2006), where utility with reference-dependent preferences is additive in the usual “consumption utility” and in a “gain-loss” utility around the reference point. The model is then based on a neoclassical utility function $((Y - T) - \frac{\theta}{1+\nu} h^{1+\nu})$ augmented with a gain-loss component ($\pm\alpha(Y - T)$).

The purely neoclassical model is the special case where there is no kink in the utility function (no gain-loss utility, $\alpha = 0$). The labor supply function in this case is $h = (\frac{W}{\theta})^{\frac{1}{\nu}}$, and the elasticity of labor supply is $\frac{1}{\nu}$. The extent to which labor supply behavior is neoclassical or influenced by reference dependence (gain-loss utility) depends on both α and the wage as follows: maximizing the utility function in equations (1) and (2) with respect to hours of work yields three distinct labor supply functions depending on the wage.

- (i) For sufficiently low wages ($W < W^*$), the reference point is not relevant because the hours required to reach the reference point at such a low wage yield sufficient disutility of hours that it is optimal to stop on the high marginal utility section of the utility function (short of the reference point). In this region, the labor supply function is neoclassical:

$$(3) \quad h = \left(\frac{(1 + \alpha)W}{\theta} \right)^{\frac{1}{\nu}}$$

with elasticity of labor supply $\frac{1}{\nu} > 0$.

- (ii) For intermediate wage levels ($\bar{W}^* < W < W^{**}$), it is optimal to stop working on reaching the reference income level. This is because the wage is in a range that is high enough to reward working when marginal utility is high

($Y < T$) but too low to reward working when marginal utility is low ($Y \geq T$). In this region, the individual is a target earner with labor supply function

$$(4) \quad h = \frac{T}{W}$$

and elasticity of labor supply is -1 .

- (iii) For sufficiently high wages ($W > W^{**}$), the reference point is not relevant because the wage is sufficiently high that it is optimal to operate on the low marginal utility section of the utility function (beyond the reference point). In this region, the labor supply function is neoclassical:

$$(5) \quad h = \left(\frac{(1-\alpha)W}{\theta} \right)^{\frac{1}{\nu}}$$

with elasticity of labor supply $\frac{1}{\nu} > 0$.

The bounds of the range where reference-dependent preferences are relevant (where gain-loss utility plays a role) are derived from the optimizing behavior of the individuals. Consider first the lower bound, W^* . The value of W^* is defined as the wage at which an individual with the “steep” utility function defined in equation (1) would choose hours so as to earn T . Based on the labor supply function in equation (3), this is

$$(6) \quad W^* = \left(\frac{\theta}{1+\alpha} \right)^{\frac{1}{1+\nu}} T^{\frac{\nu}{1+\nu}}$$

Consider next the upper bound, W^{**} . The value of W^{**} is defined as the wage at which an individual with the “flat” utility function defined in equation (2) would choose hours so as to earn T . Based on the labor supply function in equation (5), this is

$$(7) \quad W^{**} = \left(\frac{\theta}{1-\alpha} \right)^{\frac{1}{1+\nu}} T^{\frac{\nu}{1+\nu}}$$

The ratio of the upper bound to the lower bound is

$$(8) \quad \frac{W^{**}}{W^*} = \left(\frac{1+\alpha}{1-\alpha} \right)^{\left(\frac{1}{1+\nu} \right)},$$

and this is directly related the degree of gain-loss utility (indexed by α) and the labor supply elasticity (measured inversely

by ν). If α is close to 0, indicating little gain-loss utility, the range of wages where the individual is a target earner is very small, and reference dependence is not important. But as α grows, reference dependence and target earnings behavior have more relevance.

Koszegi and Rabin (2006), on whose formulation I rely heavily, suggest that the reference income level will be based on expected income.⁷ Expected income will be driven by the income level generated by the expected wage and the hours choice made by the individual based on the expected wage. Defining reference points as reflecting expected income has important implications, both for thinking about the potential importance of reference dependence in determining labor supply and in designing an empirical analysis that reflects appropriate variation in earnings opportunities.

The Koszegi-Rabin model of expectation-based reference income levels suggests importantly that labor supply, consistent with the neoclassical model, will be positively related to *anticipated* transitory wage changes. They argue that reference dependence (gain-loss utility in their terms) is related only to unanticipated variation in the wage. In periods where high wages are expected, individuals will have higher reference points, implying higher labor supply as the neoclassical model predicts. The prediction of the reference dependent preferences model, that the elasticity of labor supply is -1 , is relevant only with regard to unanticipated transitory wage changes that are close to the expected wage. This limits how much of labor supply behavior can be accounted for by reference dependent preferences and suggests that much of the variability in labor supply is likely to be consistent with the neoclassical model. Later, I decompose variation in average hourly earnings into components that are plausibly interpreted as permanent, anticipated transitory, and unanticipated transitory.

I now use the Koszegi-Rabin formulation of expectation-based reference points to derive the bounds (W^* and W^{**} in equations (6) and (7), respectively) as a function of the expected wage. To simplify the exposition given the multiplicative functional

7. Abeler et al. (2011) present experimental evidence that variation in work effort is consistent with reference points based on expectations. Crawford and Meng (2011) rely on expectation-based reference points in their analysis of taxi driver labor supply.

forms, I work with the logarithms of the wage and of hours and the expectation of the logarithms. Expected log labor supply based on the consumption part of the utility function is

$$(9) \quad E(\ln(h)) = \frac{1}{\nu} E(\ln(W)) - \frac{1}{\nu} \theta,$$

and the implied log income reference point is

$$(10) \quad \begin{aligned} \ln(T) &= E(\ln(h)) + E(\ln(W)) \\ &= \frac{1 + \nu}{\nu} E(\ln(W)) - \frac{1}{\nu} \theta. \end{aligned}$$

The bounds are related to the expected wage through the reference income level. Substituting the expression in equation (10) for the log reference income level into the logarithms of equations (6) and (7) yields particularly simple expressions for the logarithms of the bounds (W^* and W^{**}). These are

$$(11) \quad \ln(W^*) = E(\ln(W)) - \left(\frac{1}{1 + \nu} \right) \ln(1 + \alpha),$$

and

$$(12) \quad \ln(W^{**}) = E(\ln(W)) - \left(\frac{1}{1 + \nu} \right) \ln(1 - \alpha).$$

The likelihood that the realized wage is outside the bounds, yielding neoclassical behavior (positive labor supply elasticity), is a function of how variable the wage is around its expected value. If the wage has only small unanticipated variation (since anticipated variation is built into the reference income level through the expected wage), then behavior in response to unanticipated transitory wage variation will generally look like target earning. On the other hand, if the wage has substantial unanticipated variation, then behavior will look neoclassical.

This formulation of the labor supply model with reference-dependent preferences has a direct empirical prediction: on days when the wage rate unexpectedly varies substantially from its expected value, labor supply will be more likely to vary directly with the wage rate. But on days where the wage rate is relatively close to expectation, hours worked will be more likely to vary inversely with the relatively small unanticipated wage variation.

III. BACKGROUND AND DATA ON TAXI DRIVERS

There are 13,238 taxi medallions in New York City. This number is set by regulation. Roughly speaking, there are two types of medallions.

- (i) Fleet medallions, of which there are 7,664, are attached (literally) to taxis generally operated through a fleet garage and leased on a daily shift basis to individual drivers with hack licenses.⁸ Owners of fleet medallions must own at least two medallions, and each fleet medallion must operate for two shifts of at least nine hours a day. It is not clear if or how the latter requirement is enforced.
- (ii) Independent medallions, of which there are 5,574, are owned by individuals who may own no more than one medallion. A subset of these medallions are “owner-driver” medallions, which have a requirement that a substantial number of shifts in taxis with such medallions be driven by the owner. Other independent medallions have no such restriction, and these taxis may or may not be driven by the owner. In either case, taxis with independent medallions may be leased for shifts to drivers with hack licenses who are not the owner. Again, it is not clear if or how these requirements are enforced.

The standard employment arrangement of New York City cab drivers who do not own their own cabs/medallions is that a driver leases a cab for a fixed period, usually a 12-hour shift. The driver pays a fixed fee for the cab plus fuel, and he keeps 100 percent of the fare income plus tips. The driver is free to work as few or as many hours as he wishes within a 12-hour shift. Thus, the driver internalizes the costs and benefits of working in a way that is largely consistent with an economist’s first-best solution to the agency problem with risk-neutral agents. In a manner of speaking, the employer has “sold the firm to the worker.” Because these drivers are free to set their hours once they have leased a taxi for a shift, analysis of their labor supply is fertile ground for learning about behavioral models.

Taxi drivers earn income only when there is a passenger in the cab. My data cover the 2009–2013 period. Prior to September

8. Less commonly, taxis with fleet medallions are leased on a weekly basis.

4, 2012, income was earned was earned at the rate of \$2.50 for the first one-fifth mile (the “meter drop”) plus \$0.40 per additional one-fifth of a mile when traveling at 12 miles per hour or more plus \$0.40 for each minute when traveling at less than 12 miles per hour (waiting time). From September 4, 2012, through 2013, the rate for additional fifths of a mile and waiting minutes was increased from \$0.40 income to \$0.50. Throughout the period, there was also a night surcharge of \$0.50 per trip between 8 PM and 6 AM and peak-hour weekday surcharge of \$1.00 Monday–Friday between 4 PM and 8 PM.⁹ Clearly, a central factor in earnings is the ease/speed with which new fares are located.

The earlier studies of taxi driver labor supply were based on analysis of relatively small numbers of hand-written “trip sheets” (one per shift) that drivers were required to fill out with information on the fare and trip start and end times and locations.¹⁰ These sheets were difficult to read accurately, and the limited number of sheets available severely constrained analysis.

This situation has changed dramatically. The New York City Taxi and Limousine Commission (TLC), the agency charged with regulating the industry, now requires all taxis to be equipped with electronic devices that record all trip information, including fares, times, and locations. The (currently two) companies that supply these devices report all this information to the TLC on a regular basis, and I have obtained full information for all trips taken in NYC taxi cabs for the five years from 2009 to 2013. These data, called TPEP by the TLC, identify drivers by encrypted hack license number and medallions (cabs) by encrypted medallion number.¹¹

9. There is a flat fare of \$52.00 (\$45.00 prior to September 4, 2012) plus tolls between Manhattan and JFK International Airport in either direction and a surcharge of \$17.50 (\$15.00 prior to September 4, 2012) on trips to Newark Liberty International Airport.

10. An exception is Agarwal et al. (2015), who have administrative data on taxi drivers in Singapore, but these data lack direct information on fares.

11. The TPEP data do not identify fleet and independent medallions or isolate the owner-driver medallions among the independent medallions. It may be that owner-drivers face different incentives regarding labor supply than do drivers who lease their cabs, whether from a fleet owner or an individual. I am working to get additional information that could be used to analyze by labor supply separately for owners and lessees.

There are 170–180 million trips taken during 7–8 million shifts each year in taxi cabs in NYC. About 62,000 drivers had at least one fare in a cab over the five-year period, with about 40,000 drivers in any single year. About 25,000 drivers worked in all five years. In fact, these are more data than I can use efficiently, and much of my analysis is based on a random subsample of $\frac{2}{15}$ of the drivers.

One limitation of the data is that complete information is not available on tip receipts. Data on tips are available only for fares with tips paid by credit card. While the percentage of fares paid by credit card has been increasing over time, about 46 percent of fares in the most recent year (2013) were paid in cash and so there is no tip information. Tips on credit card transactions averaged about 20 percent over the 2009–2013 period. I proceed ignoring tips due to the missing data for cash transactions; this will not cause a problem for my analysis unless variation the rate of tipping is correlated with hourly fare income.¹²

IV. WHY CAN'T YOU FIND A TAXI IN THE RAIN?

I begin with a direct analysis of the question of whether target earnings behavior can account for some of the perceived increased difficulty in finding a cab in the rain. The underlying idea is that the demand for taxis increases when it is raining, which by itself makes it more difficult to find a cab, and that this increase in demand results in higher hourly earnings for taxi drivers. The result is that the drivers reach their daily earning targets sooner and quit for the day, exacerbating the increased difficulty in finding a cab.

Rain may have a number of effects on the market for taxi rides. First, it likely increases demand. This will make it easier for drivers to find passengers and increase hourly income. However, rain may decrease speed due to congestion and a deterioration in general driving conditions, which could result in lower earnings. Rain may also make driving less pleasant, implying a reduction in the number of cabs on the road having nothing to do with target earning behavior.

12. The lack of observability of tips does introduce measurement error in income and the hourly wage, and I address this problem in my analysis. See Haggag and Paci (2014) for an analysis of tipping behavior in NYC taxi cabs in 2009.

For this analysis, I use a random subsample of the drivers in the data described above. These data include all trips taken by $\frac{2}{15}$ of the drivers I see between 2009 and 2013. This subsample includes 116,177,329 trips for 8,802 drivers.

I begin by using the trip-level data to calculate the average of income, time with a passenger in the cab, miles traveled, and number of taxis on the street for each of the 48,824 clock hours in the period from 2009 to 2013. I then merge these data with data on hourly rainfall in Central Park.¹³ I calculate the hourly wage for a given driver in a given hour using the trip-level fare and time information. The hourly wage was computed by dividing each shift into minutes and assigning a “minute wage” to each minute. For minutes during trips, the minute wage is computed as the fare divided by the number of minutes for that trip. For minutes of waiting time (between fares), the minute wage is set to zero. The hourly wage for each clock hour is computed as the sum of the minute wages during that hour. I calculate time with a passenger in the cab and miles traveled during a given clock hour in an analogous fashion. The count of taxis on the street is a count of taxis who had a passenger in the cab for at least one minute during the clock hour.

A necessary condition for target earnings behavior to contribute to difficulty in finding a taxi in the rain is that hourly earnings be higher when it is raining. The first row of Table I contains the coefficient of an hourly indicator for precipitation in Central Park from ordinary least squares (OLS) regression analyses of average log hourly earnings.¹⁴ The estimates in column (1) include no other control variables, while the estimates in column (2) include indicators for hour of day by day of week (167), month of year (11), year (4), the period subsequent to the September 4, 2012, fare increase (1), and major holiday (1).¹⁵ The hourly wage is not significantly correlated with whether it rained in

13. Source: National Weather Service (NWS) of the National Oceanic and Atmospheric Administration (NOAA) and the Network for Environment and Weather Applications, Cornell University.

14. The indicator for precipitation equals 1 if there is any precipitation in Central Park recorded during the hour. The results I present are qualitatively unaffected by using a graduated measure of the quantity of precipitation. Not surprisingly, on those rare days when there is a major storm (e.g., Hurricane Sandy), dramatically fewer taxis are on the street.

15. Major holidays are defined as New Years Day, Easter Sunday, Memorial Day, Fourth of July, Labor Day, Thanksgiving Day, and Christmas Day.

TABLE I
EFFECT OF RAIN ON HOURLY TAXI MARKET OUTCOMES, OLS ESTIMATES

Outcome	(1)	R^2	(2)	R^2
(1) Log hourly earnings	-0.0047 (0.0054)	0.000	-0.0003 (0.0021)	0.879
(2) Log time w/passenger	0.0508 (0.0077)	0.002	0.0475 (0.0028)	0.904
(3) Log miles w/passenger	-0.0306 (0.0046)	0.002	-0.0241 (0.0026)	0.792
(4) Log number of hacks	-0.0671 (0.0149)	0.001	-0.0711 (0.0089)	0.878
Includes other controls?	No		Yes	

Notes. Coefficient of *Precipitation* (= 1 if rain in Central Park during hour). Estimated using data for each of 43,824 hours in the years 2009–2013 derived from trip-level data for the $\frac{2}{15}$ sample of all drivers of NYC taxi cabs. This sample contains 8,802 drivers on 116,117,329 trips. *Precipitation* is an indicator for hours where there is positive precipitation in Central Park. Other controls include indicators for hour of day by day of week (167), week of year (51), year (4), the period subsequent to the September 4, 2012.

Central Park.¹⁶ There is considerable systematic variation in average hourly earnings over the course of the day, week, month, and year as measured by the R -squared of the regression in column (2), but accounting for whether it was raining does not significantly improve the fit of the model.

The finding of no relationship between earnings and rainfall is sufficient to reject the hypothesis that target earnings behavior contributes to the difficulty of finding a taxi in the rain, but it leaves a puzzle: if demand for taxis is higher in the rain, why are earnings not higher? I investigate this by examining the relationship of other measures of taxi activity with rainfall.

The second row of Table I contains OLS estimates of the coefficient of the precipitation indicator from a regression of average log minutes per hour spent with a passenger in the cab. This is a measure of how busy the cabs on the street are (and how easy it is for drivers to find passengers). Taxi occupancy rates are 4.8 percent higher when it is raining, accounting for systematic variation over time (column (2)). Since drivers make money only when passengers are in the cab, it is clear that demand is

16. Consideration of tips does not change this conclusion. Examining the information on tips for credit card transactions, the tip rate (which averages about 20 percent of the fare inclusive of surcharges) is about 0.2 percentage points higher, on average, in hours with rain.

higher relative to supply when it is raining. However, this does not translate into higher earnings for drivers.

One possibility is that traffic and other driving conditions are worse when it is raining, so the taxis drive more slowly. Since income is roughly proportional to miles traveled (with some payment for idle time in traffic), this would imply lower income than would ordinarily result from higher occupancy. The third row of Table I contains OLS estimates of the coefficient of the precipitation indicator from a regression of average log miles travelled per hour with a passenger in the cab. Miles traveled with a passenger are about 2.4 percent lower when it is raining accounting for systematic variation over time (column (2)). That miles traveled per hour with a passenger are lower despite the fact that the occupancy rate is higher is clear evidence that driving conditions are worse in the rain. This is the factor that offsets the increase in demand and results in no change in average hourly earnings when it is raining.

To investigate any supply response to rain, the fourth row of Table I contains OLS estimates of the coefficient of the precipitation indicator from a regression of the log number of hacks that take at least one trip during the hour in question. The number of cabs on the street is about 7.1 percent lower when it is raining. This reduction in supply with no change in average earnings despite the increase in demand likely reflects added disutility of driving when it is raining. Some drivers stop, but this is not because of reaching their income target. Some drivers stop simply because it is less pleasant to drive in the rain and there is no additional benefit in continuing to drive.

The increase in taxi utilization measured by time with a passenger of 4.8 percent is more than offset by the decline in supply of cabs of 7.1 percent. This means that the supply of rides is lower in rainy hours, and any surge in demand is unmet. One logical response would be to have a rain surcharge to encourage an increase in supply.¹⁷

17. An example of real-time adjustment of rates to meet fluctuations in demand is Uber's "surge pricing." For example, if the labor supply elasticity was 0.5, a 14.2 percent "rain surcharge" could get supply back to the dry-weather level. One might want a larger surcharge to meet increased demand and offset slower driving in the rain. Of course, this depends on there being reasonable very short run elasticity of labor supply.

V. ESTIMATING THE WAGE ELASTICITY OF LABOR SUPPLY

V.A. *How Much Wage Variation Is Unanticipated?*

I noted earlier that the role of reference dependence in determining labor supply is limited to the response of labor supply to unanticipated wage variation. Before proceeding with estimation of the labor supply model, I present some evidence on the magnitude of unanticipated wage variation based on the data on average earnings and number of hacks on the road by hour that I used for the analysis of taxis and rain in Section IV.

Table II contains simple variance decompositions for the average log wage and log number of hacks by hour for the 43,824 hours from 2009 to 2013. This decomposition is carried out in two stages. In the first stage, I regress average log hourly earnings ($\ln W$) on a set of year indicators (4) and an indicator for period subsequent to the September 4, 2012, fare increase (1). These variables capture permanent wage variation, and the variance of the predicted values from this regression is my measure of the variance of permanent wage variation. The residuals from this regression include both anticipated and unanticipated transitory wage variation. In the second stage, I regress these residuals on a set of controls including hour of day by day of week (167), week of year (51), and holiday (1).¹⁸ These controls capture anticipated transitory variation in the wage, and the variance of the predicted values from this regression is my measure of the variance of anticipated transitory wage variation. The residuals from this regression capture unanticipated transitory wage variation, and the variance of these residuals is my measure of the variance of unanticipated transitory wage variation.¹⁹

The first row of Table II contains the variance decomposition for the average log wage by hour. Most variation (76.8 percent of total variation) is anticipated transitory variation. This is largely variation by hour-of-day by day-of-week. Of the total variation, 11.1 percent is permanent variation, driven largely by the September 4, 2012, fare increase. The remaining 12.1 percent of

18. See note 15 for a list of major holidays used in defining the major holiday indicator.

19. Of course, drivers do not perform regression analyses to calculate expected earnings opportunities. Although there is likely to be some difference between drivers' calculations of the expected wage and the model I use, the information used in these models (date, time, etc.) are just the sort of factors drivers are likely to use in forecasting earnings opportunities.

TABLE II

VARIANCE DECOMPOSITION OF HOURLY AVERAGES (ACROSS 43,824 HOURS, 2009–2013)

Variable	Total	Permanent anticipated	Transitory anticipated	Transitory Unanticipated
$\ln(W)$	0.0521	0.0058	0.0400	0.0063
$\ln(Hacks)$	0.2716	0.0011	0.2369	0.0337

Notes. The decomposition is based on two stages: (i) OLS regression of each variable on indicators for year (4) and the period subsequent to the September 4, 2012, fare increase (1), and (ii) OLS regression of the residuals from the first regression on a set of controls including hour of day by day of week (167), week of year (51), and holiday (1). “Total” is the raw variance. “Permanent predicted” are the variances of the predicted values from the first regression. “Transitory anticipated” and “Transitory unanticipated” are, respectively, the variances of the predicted values and residuals from the second regression.

total variation is unanticipated transitory variation. This decomposition clearly limits the scope of reference dependence to account for variation in labor supply. About $\frac{7}{8}$ of variation in average hourly earnings is anticipated so that it does not involve gain-loss utility, and its effects on labor supply are consistent with the neoclassical model. Only the remaining $\frac{1}{8}$ of total variation in average hourly earnings is due to unanticipated factors.

The second row of Table II repeats this decomposition for variation in the log number of hacks on the road by hour. Interestingly, the decomposition yields similar results. An even larger share of total variation (87.2 percent) is anticipated transitory variation. Again, this is largely variation by hour-of-day by day-of-week. Almost no variation (0.4 percent of total variation) is due to permanent changes. The remaining 12.4 percent (about $\frac{1}{8}$) of total variation is unanticipated transitory variation.

The clear implication of this simple decomposition is that reference dependence (gain-loss utility) is not relevant for explaining the broad patterns of variation in labor supply, which largely result from anticipated variation in demand by hour of day and day of week. Only about $\frac{1}{8}$ of total variation in the wage and labor supply could potentially be influenced by reference dependence and gain-loss utility.

V.B. Shift Definition and Creation of the Analysis Sample

It was straightforward to define shifts in the earlier small-sample studies because the data were transcribed from hand-written trip sheets, each of which represented a distinct shift. In contrast, the new electronic data on which I rely is simply a

running list of all trips by a particular driver. The assignment of trips to particular shifts in this case is an analytic decision.

The definition of a shift is necessarily subjective. Given the approach here, which is to model the shift-level labor supply decision of a driver (perhaps with some reference income level for the shift), it makes sense to define a shift as composed of all trips that are part of a sequence the driver considers to be a single shift. Absent a clear guide, I define any gap between trips of more than 6 hours (more than 360 minutes) as marking the end of one shift and the beginning of the next.²⁰ Defining shifts in this way yields a sample of 5,047,343 shifts for 8,802 drivers over the period from the $\frac{2}{15}$ random sample of all drivers from 2009 to 2013. Figure I shows the distribution of shift length in hours (truncated, e.g., 8.5 hours is shown as 8). The modal shift duration is in the ninth hour, and shift durations are concentrated between hours 5 and 11 (81.6 percent). Of shifts, 10.7 percent are less than 6 hours, and 7.7 percent of shifts are 12 hours or longer.

Daily leases typically run for 12 hours, with two such shifts per day. Figure II shows the fraction of shifts in my sample that start in each clock hour. There are two daily spikes, with a surge of “day shifts” (44.5 percent) starting in the six-hour segment from 4 AM and 9:59 AM and another surge of “night shifts” (42.4 percent) in the six-hour segment from 2 PM to 7:59 PM. Only a small fraction of shifts (13.1 percent) are not assigned by me as day or night shifts. It is clear from Figure I that drivers who lease their cabs daily do not use the entire 12 hours to which their lease entitles them.²¹ Day shift drivers generally do not start driving at the first available moment, but they must stop no later than when

20. In my first analysis of these data, presented as the Albert Rees Lecture at the 2014 meeting of the Society of Labor Economists, I defined a shift mechanically as the collection of all trips in the 24-hour period between 5 AM one day and 4:59 AM the next. However, this definition clearly does not mirror drivers' conceptions of a shift. For example, a driver who started two 10-hour shifts at 4 AM on two consecutive days will have the hour from 4 AM to 4:59 AM on the second day as the end of the previous day's shift (starting at 5 AM on the first day). This “shift” will look like it lasted 24 hours when in fact it lasted only 11 hours. Although there are some differences between the results I presented in the Rees Lecture and those presented here, the general conclusions are not sensitive to this choice of shift definition.

21. Owner-drivers and those drivers who lease by the week or month from medallion owners with owner-driver medallions may not be constrained to 12-hour shifts. Their constraints depend on whether there is a second driver who leases (or subleases) the cab. Drivers who lease taxis with fleet medallions daily are so constrained.

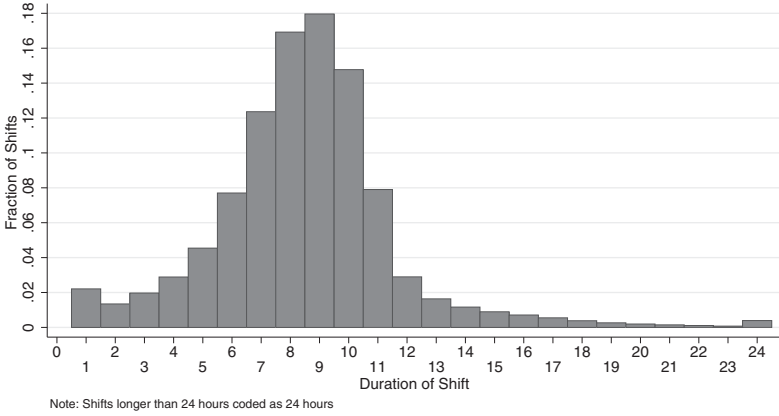


FIGURE I
Distribution of Shift Length in Hours

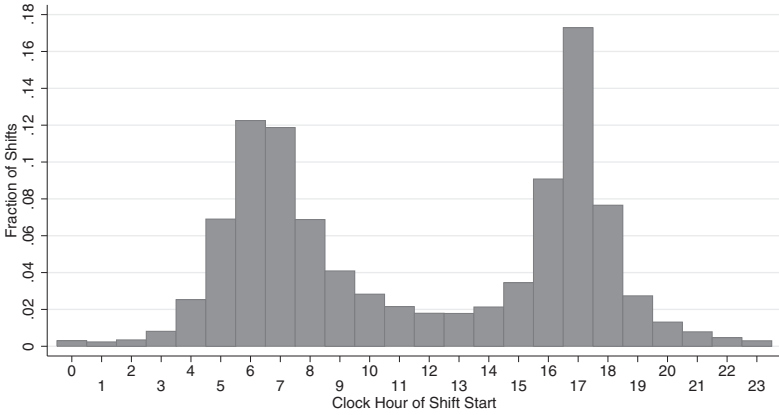


FIGURE II
Distribution of Shift Start Time, by Clock Hour

their shift ends so that the night shift driver can take over. Similarly, the night-shift drivers may take over when the cab is first available, but they generally stop before they are required to do so.

There may be an important difference in the margin on which labor supply can be adjusted on day shifts versus night

shifts. Day shift drivers are more likely to be constrained at the end of their shift when the cab must be turned over to a night shift driver. This implies that day shift drivers may have more freedom to adjust hours by changing their start time rather than their end time. In contrast, night shift drivers may not be able to start early but can adjust hours by changing their end time. As demand declines late at night, many of these drivers stop before their 12-hour lease ends. These patterns are a natural consequence of the two-shift structure of the day.

The difference in the active margin of decision making between day and night shifts is important for analyzing labor supply. Day shift drivers often select hours before information on unanticipated daily earnings opportunities is revealed. In contrast, night shift drivers can experience the evolution of earnings opportunities and decide when to quit for the day. If labor supply is affected by unanticipated transitory variation in earnings opportunities and if information about these earnings opportunities are learned by the driver only after he has started the shift, then night shift drivers will have more opportunity than day shift drivers to adjust their labor supply in response. Operationally, this suggests that estimated labor supply elasticities could be larger for night shift drivers.

Table III contains mean hours, income, and average hourly earnings by shift type.²² Day shifts are longer than night shifts by about 0.7 hour, but more money is earned on night shifts (about \$13 more). Average hourly earnings (the wage) is about \$3.73 higher on the night shift.²³ The fact that average hourly earnings are higher on night shifts is reflected in higher caps set by the TLC on daily taxi lease rates for night shifts than for day shifts. Additionally, total earnings and average hourly earnings vary substantially through the week, particularly on night shifts, and this is reflected in higher lease caps for night shifts later in the week.²⁴

22. Income is defined as the sum of fare income and surcharges. Tip income is not included.

23. All differences in means across shifts are statistically significant (p -value < 10^{-100}).

24. The lease rate cap for day shifts has been \$115 since October 2012. The lease rate cap for night shifts since October 2012 ranges from \$128 Sunday–Tuesday to \$142 Thursday–Saturday. See Farber (2014) for a more detailed breakdown of earnings and hours by shift and day of week.

TABLE III
MEAN HOURS, INCOME, AND AVERAGE HOURLY EARNINGS BY SHIFT

Shift	Hours	Income	Wage	# Shifts
Day	9.48 (3.97)	248.41 (106.03)	26.39 (5.75)	2,247,417
Night	8.78 (3.05)	262.03 (93.06)	30.13 (6.55)	2,137,499
Unassigned	8.30 (5.33)	228.12 (138.46)	28.26 (8.47)	662,427

Notes. Day shifts start between 4 AM and 9:59 AM. Night shifts start between 2 PM and 7:59 PM. Shifts starting at other times are unassigned. Standard deviations are in parentheses.

Given the sharp differences in labor supply and earnings patterns for day shift and night shift drivers and the potential differences in available information regarding earnings opportunities when making labor supply decisions, I analyze the labor supply of day shift and night shift drivers separately in what follows.

V.C. New Estimates of the Labor Supply Elasticity

I begin by presenting OLS estimates of the labor supply elasticity based on my sample of over 5 million shifts from 2009 to 2013. These are regressions of log shift duration on log average hourly earnings during the shift and other variables as noted below. I use four samples: (i) all shifts, (ii) day shifts, (iii) night shifts, and (iv) other (unclassified) shifts. These samples are based on the $\frac{2}{15}$ random sample of drivers described already, which includes 5,047,343 shifts for 8,802 drivers over the 2009–2013 period. For each sample, I estimate three specifications: (i) no controls; (ii) a set of controls including indicators for day of week, calendar week, year, the period subsequent to the September 4, 2012, fare increase, and major holiday; and (iii) additionally including driver fixed effects.²⁵ These controls account for anticipated wage variation and leave the average hourly earnings measure to account for unanticipated transitory variation in earnings opportunities. I present the estimated coefficient of log average hourly earnings, which is interpreted as the wage elasticity of labor supply.

25. See note 15 for a list of major holidays used in defining the major holiday indicator.

TABLE IV
WAGE ELASTICITY, OLS REGRESSION OF AVERAGE LOG DAILY HOURS, BY SHIFT

Model	Controls	Driver FEs	Elasticity all shifts	Elasticity day shifts	Elasticity night shifts	Elasticity other shifts
(1)	No	No	0.0159 (0.0154)	0.0485 (0.0177)	-0.0017 (0.0169)	0.0738 (0.0220)
(2)	Yes	No	-0.0034 (0.0177)	0.0505 (0.0203)	-0.0784 (0.0210)	0.0606 (0.0240)
(3)	Yes	Yes	-0.1002 (0.0089)	-0.0615 (0.0109)	-0.1487 (0.0077)	-0.0501 (0.0138)

Notes. Each estimated elasticity is from a separate OLS regression. “Elasticity” is the estimated coefficient of log average hourly earnings from a regression of log shift duration. “Controls” include indicators for day of week (6), calendar week (51), year (4), the period subsequent to the September 4, 2012, fare increase (1), and major holiday (1). Estimated using sample of 5,047,343 shifts for 8,802 drivers from 2009 to 2013. Sample sizes are listed in Table III. Robust standard errors clustered by driver are in parentheses.

Table IV contains these OLS estimates. The estimates for the entire sample, contained in the first column, show elasticities that are small and insignificantly different from zero in the first two specifications. When driver fixed effects are included, the estimated elasticity is negative and statistically significant but relatively small at -0.1. For day shifts, the estimated elasticities are small and positive but statistically significant in the first two specifications. When driver fixed effects are included, the estimated elasticity is again negative and statistically significant though small. The pattern for night shifts is that the elasticities are significantly negative when the controls are added. The estimated elasticity when driver fixed effects are included is more negative than for the day shift. The estimates for the unclassified (other) shifts are very close to those for day shifts. Although I do find some negative elasticities, none approach -1 as suggested by a target earnings model.²⁶

As CBLT recognize, OLS estimates of the elasticity may well be downward biased due to “division bias” if there is any specification or measurement error. This is because the key explanatory variable, average hourly earnings, is calculated as the ratio of daily income to daily hours and daily hours is the dependent

26. My OLS elasticities are much smaller in magnitude than those found using OLS by CBLT or Farber (2005). This may reflect a lower level or different pattern of measurement error in my administrative data compared with the transcribed trip sheet data used in the earlier work.

variable. Although the administrative data are likely to have less measurement error than data derived from the imperfectly recorded and transcribed paper trip sheets, it is not error-free. Simple consistency checks of the data show more than a few instances of trips ending before they start and new trips starting before the previous trip ends.²⁷ Additionally, as I mentioned earlier, my income data do not include tips, which surely vary across trips as a proportion of fares (Haggag and Paci 2014).

CBLT address this problem in a sensible way by instrumenting for average hourly earnings of a given driver with measures of the distribution of hourly earnings of other drivers on the same calendar date, and I follow this approach. An obvious choice for an instrument for average hourly earnings is the average log hourly earnings of other drivers on the same day, and this is the measure I use.²⁸

To avoid problems using an instrument derived from the dependent variable in the estimation sample, I use a nonoverlapping randomly selected $\frac{2}{15}$ subset of the drivers to generate the instruments.²⁹ The average of log average hourly earnings of shifts starting on date t in the nonoverlapping sample ($\overline{\ln W}_t$) serves as the instrument for the log average hourly earnings of driver i in my estimation sample for shifts that start on date t ($\ln W_{it}$).

Two conditions need to be satisfied for $\overline{\ln W}_t$ to be a valid instrument. First, $\overline{\ln W}_t$ needs to be strongly correlated with $\ln W_{it}$. While I don't present the first-stage estimates here, this is clearly satisfied here.³⁰ Second, the instrument must satisfy the exclusion restriction. In this case, $\overline{\ln W}_t$ should not be correlated with log hours other than through $\ln W_{it}$. The motivation to use an instrument here is the potential for division bias mainly due to measurement error. Since it is reasonable to

27. I used some simple algorithms to adjust the data to eliminate these inconsistencies and serious outliers.

28. The measures CBLT use as instruments are the 25th, 50th, and 75th percentiles of the other-driver daily earnings distributions. Using their IV approach, CBLT find elasticities that range from -1.313 to -0.926 for their larger samples (a maximum of about 1,000 shifts over a small range of days). Interestingly, these IV elasticities are more negative than those they found using OLS.

29. This sample contains 115,733,041 trips on 5,012,244 shifts for 8,768 drivers.

30. The first-stage t -statistic on the instrument is generally greater than 100, and the coefficient on the instrument in the first stage is generally close to 1.

assume that measurement errors are uncorrelated across drivers, the instrument satisfies the exclusion restriction on these grounds.

There is a more subtle argument for potential violation of the exclusion restriction. To the extent that $\overline{\ln W}_t$ reflects unobserved demand conditions and information about these demand conditions is communicated across drivers, labor supply of individual drivers might be affected directly by this information. To the extent that this effect works through individual average hourly earnings (as it likely would to a substantial degree), this is not a violation of the exclusion restriction. However, to the extent that individual drivers use $\overline{\ln W}_t$ to adjust their labor supply independently and individual average hourly earnings are not perfectly correlated with the overall average, there is scope for violation of the exclusion restriction. This stands as a potential caveat to the instrumental variables (IV) analysis.³¹

The IV estimates of the labor supply elasticity are contained in Table V. The results are striking in comparison with the OLS estimates in Table IV. The estimated elasticities are substantially positive and strongly statistically significant. Adding the control variables raises the estimated elasticity for each sample, but controlling for driver fixed effects does not have much effect. The estimated elasticity on the day shift is about 0.36 while the elasticity on the night shift is about 0.62. The larger elasticity for the night shift is consistent with the observation that drivers on a night shift are more likely than drivers on a day shift to be able to adjust hours mid-shift in response to new information regarding earnings opportunities. Interestingly, the elasticity is even larger on unclassified shifts. It may be that these other shifts are less likely to be worked by lease drivers and more likely to be worked by owner-operators who have more flexibility in selecting hours.

Overall, the evidence presented so far is consistent with the neoclassical optimizing model. The positive estimated elasticities do not support the idea that reference-dependent preferences or target earnings behavior are important factors in taxi driver labor supply decisions.

31. In Section VII, I present estimates of a discrete choice stopping model that does not depend on an instrument, and it yields results consistent with the IV estimates I present here.

TABLE V
WAGE ELASTICITY, IV REGRESSION OF AVERAGE LOG DAILY HOURS, BY SHIFT

Model	Controls	Driver FEs	Elasticity all shifts	Elasticity day shifts	Elasticity night shifts	Elasticity other shifts
(1)	No	No	0.2288 (0.0101)	0.0202 (0.0134)	0.3484 (0.0117)	0.2913 (0.0306)
(2)	Yes	No	0.5709 (0.0100)	0.3683 (0.0119)	0.6182 (0.0132)	0.9383 (0.0329)
(3)	Yes	Yes	0.5890 (0.0099)	0.3672 (0.0112)	0.6344 (0.0124)	0.8751 (0.0281)

Notes. Each estimated elasticity is from a separate IV regression. The instrument for average hourly earnings is the average of average hourly earnings for a nonoverlapping sample of drivers on the same day. “Elasticity” is the estimated coefficient of log average hourly earnings from a regression of log shift duration. “Controls” include indicators for day of week (6), calendar week (51), year (4), the period subsequent to the September 4, 2012, fare increase (1), and major holiday (1). Estimated using sample of 5,047,343 shifts for 8,802 drivers from 2009 to 2013. Sample sizes are listed in Table III. Robust standard errors clustered by driver are in parentheses.

VI. DOES THE LABOR SUPPLY ELASTICITY DEPEND ON HOW THE WAGE COMPARES WITH THE EXPECTED WAGE?

The theoretical discussion in Section II set bounds on the range within which one would expect to find behavior influenced by the gain-loss (reference-dependent) component of utility. If the realized daily wage lies in the range defined by equations (11) and (12), then gain-loss utility may be relevant and target earnings behavior may be observed. Otherwise, the labor supply elasticity with respect to unanticipated transitory wage variation will be positive and determined by the consumption (neoclassical) component of utility. Intuitively, reference dependence is a local phenomenon. If the wage is far lower than what was expected, drivers will find it optimal to stop working before the reference income level is reached, and, if the wage is far higher than what was expected, drivers will find it optimal to continue working after the reference income level is reached.

In this section, I examine how the estimated labor supply elasticity varies with the level of unanticipated wage variation (the absolute deviation of the average daily log wage from its expected value). I calculate the expected log wage for each day using data on mean daily log average hourly earnings for drivers in the nonoverlapping sample that I used to construct the instrument for estimation of the labor supply model in Section V.C. The expected log wage is calculated as the predicted value of log average hourly earnings from an OLS regression of daily average

log average hourly earnings on indicators for day of week, week of year, year, the period after the fare increase of September 4, 2012, and major holiday. I then calculate the difference between observed average daily log average hourly earnings in the nonoverlapping sample and the predicted value. This difference is what I use as the deviation of the average daily log wage from its expectation.

Across the 1,826 days in the sample, the average deviation is 0 by construction. Interestingly, the deviations appear relatively small. The average of the absolute deviation is 0.033, and the interquartile range of the absolute deviation runs from 0.011 to 0.043. The 90th percentile of the absolute deviation is 0.067 and the 95th percentile is 0.093. In other words, less than 5 percent of the days considered have an observed deviation from expected average hourly earnings of 10 percent or more.

The bounds on what is a sufficiently small deviation from the expected log wage are defined in equations (11) and (12). These bounds depend on the importance of gain-loss utility in the utility function (equations (1) and (2)), which is controlled by the parameter α and by the neoclassical labor supply elasticity, which is controlled by the parameter ν . The parameter α is directly related to the coefficient of loss aversion (λ) used in the behavioral economics literature. The coefficient of loss aversion is defined as the ratio of marginal utility below the reference point to marginal utility above the reference point. In the utility specification used here (equations (1) and (2)), the coefficient of loss aversion is $\lambda = \frac{(1+\alpha)}{(1-\alpha)}$.

The bounds defined in equations (11) and (12) combined with the definition of λ as a function of α in the previous paragraph imply that size of the range of deviations where reference dependence is relevant is

$$(13) \quad \ln(W^{**}) - \ln(W^*) = \frac{\ln \lambda}{1 + \nu},$$

where ν is the inverse labor supply elasticity. Existing evidence, mostly from laboratory experiments, suggests that the coefficient of loss aversion (λ) is in the range of 1.5 to 2.5.³² Assuming an elasticity of labor supply of $\frac{1}{\nu} = 0.5$ and $\lambda = 1.5$ implies a range of deviations where reference dependent

32. See, for example, Tversky and Kahneman (1991), Abdellaoui, Bleichrodt, and Paraschiv (2007), and Tovar (2009).

preferences are relevant of 0.135 (roughly ± 0.0675). Alternatively, assuming an elasticity of labor supply of $\frac{1}{\nu} = 0.5$ and $\lambda = 2.5$ implies a range of 0.3 (roughly ± 0.15). A larger labor supply elasticity would lead to a larger range.

Even the narrower of the bounds I calculate here (± 0.0675) are quite substantial relative to the observed range of unanticipated variation in daily average hourly earnings. Only 101 of the 1,826 days had unanticipated variation in the average log wage outside this range. In other words, virtually all days (about 94.5 percent) saw unanticipated wage variation small enough to imply reference dependence and target earnings behavior at these reasonable parameter values. Intuitively, because most of the observed unanticipated variation in the wage is likely within the range where target earnings behavior would be relevant if taxi drivers, in fact, had reference-dependent preferences with a coefficient of loss aversion in range of existing estimates (λ in the 1.5 to 2.5 range), strong hints of such behavior ought to be observed in the full-sample estimates in the form of negative labor supply elasticities. Thus, the finding that the estimates of the labor supply elasticity are strongly positive suggests that reference dependence is not playing a large role in taxi driver decision making regarding labor supply (λ close to 1).

These calculations notwithstanding, I investigate how the estimated elasticity varies with the daily level of unanticipated wage variation by estimating separate labor supply functions for days where the absolute deviation between average log hourly earnings and expected log average hourly earnings is very small or is larger. I split the sample into three subgroups: (i) days with an absolute deviation in the bottom 25 percent of days (an absolute log wage deviation less than 0.009), (ii) days with an absolute deviation in the second quartile of days (an absolute log wage deviation between 0.009 and 0.01834), and (iii) days with an absolute deviation above the median of days (an absolute log wage deviation larger than 0.01834).³³ My view is that absolute deviations from the expected wage in the first two groups are so small relative to the calculated bounds that they

33. The distribution considered uses the date the shift started as the operative date. No adjustment is made for the fact that some shifts span calendar days, and the same distribution is used for all shifts, regardless of whether they are day or night shifts.

should provide the reference-dependent preference model a fair chance to exhibit predictive power.

Table VI contains IV estimates of labor supply elasticities for subsamples of the sample of shifts I have been using broken down by days where the absolute deviation the average of log average hourly earnings from its daily expected value is lower or higher than the threshold described. These estimates are calculated first using all shifts and then separately for day shifts and night shifts. The results in the first row of the table, for days on which average wages are relatively close to their predicted value (the 25 percent of days with smallest absolute deviation), show small elasticities that are not significantly different from zero for either day shifts or night shifts. The estimated elasticities on days with intermediate levels of difference between average wages and their predicted values (from the 25th to the 50th percentile of the distribution across days), contained in the second row of the table, are positive, larger, and statistically significant. Finally, the third row contains estimates for the days with relatively larger deviations of average wages from their predicted values (above the median), and these elasticities are quite large, ranging from 0.38 for day shifts to 0.63 for night shifts. As before, elasticities are uniformly larger on night shifts than day shifts.

The finding that elasticities are smaller on days with small deviations than on days with large deviations suggests that there may be some reference dependence on days with small deviations of the wage from its expected value. However, the fact that the smaller estimated elasticities are apparent only for values of the absolute deviation of the average log earnings from its expected value smaller than (generously) the median value of 0.18 implies that the coefficient of loss aversion would be very small. Returning to the expression for the range of relevant deviations in equation (13), a range for $\ln(W^{**}) - \ln(W^*)$ of 0.036 (twice 0.018) and a labor supply elasticity of 0.5 implies value of the coefficient of loss aversion of $\lambda = 1.11$. This is considerably smaller than existing estimates of the coefficient of loss aversion in other settings (note 32), implying about an 11 percent increase in marginal utility at the reference point.

The results of this analysis suggests a tension between the strength of reference-dependent preferences (the coefficient of loss aversion) and the range of wage variation over which such preferences are relevant. The estimates suggest that the

TABLE VI
WAGE ELASTICITY, IV REGRESSION OF AVERAGE LOG DAILY HOURS, BY SHIFT

Sample	Absolute deviation percentile	Elasticity all shifts	Elasticity day shifts	Elasticity night shifts
(1)	0–25 (<i>N</i> = 1,287,951)	0.1398 (0.0668)	−0.1045 (0.1040)	0.1095 (0.0751)
(2)	25–50 (<i>N</i> = 1,274,650)	0.4249 (0.0310)	0.1463 (0.0414)	0.5625 (0.0384)
(3)	50–100 (<i>N</i> = 2,484,742)	0.5809 (0.0102)	0.3784 (0.0121)	0.6268 (0.0135)

Notes. Subsamples by absolute deviation of average log daily wage from expected value. The 25th percentile and median across days of the absolute deviation of the average log daily wage from its expected value are 0.00896 and 0.01834, respectively. The expected value is the predicted value from a regression of average log hourly earnings on indicators for day of week, week of year, year, the high fare period (on or after September 4, 2012), and major holiday. Each estimated elasticity is from a separate IV regression. The instrument for average log hourly earnings is the average of average log hourly earnings for a nonoverlapping sample of drivers on the same day. “Elasticity” is the estimated coefficient of log average hourly earnings from a regression of log shift duration which additionally includes the set of variables listed in the note to Table V. The listed sample sizes are for the “all shifts” samples based on the underlying sample of 5,047,343 shifts for 8,802 drivers from 2009 to 2013. Robust standard errors clustered by driver are in parentheses.

elasticity is smaller on days when the absolute deviation between average wages and the expected wage is small. Although this is consistent with reference dependence, the magnitudes are such that there cannot be a particularly sharp kink in the utility function at the reference point. In other words, to the extent there is reference dependence in this setting, the coefficient of loss aversion is quite small, which limits that the importance of these preferences for labor supply behavior.

VII. THE DISCRETE CHOICE STOPPING MODEL

While estimating the slope of the relationship between hours of work and average hourly earnings is informative about models of labor supply, modeling the number of hours worked on a particular day in this way is made difficult by the fact that the available wage at any point in time not constant (or monotone decreasing or increasing) within a given day due to variation in demand during the day. In this situation it may be appropriate to model the labor supply decision as a dynamic discrete choice problem where the end of each fare is a decision point for the driver: based on a comparison of the marginal utility of stopping with the marginal utility of continuing to drive, the driver can continue to

work or the driver can end the shift. A reduced-form version of this comparison at any point in the shift should account for variation in prospective earnings opportunities (to account for variation in the benefit of continuing), hours worked to that point (to account for increasing disutility of work during a shift), and other variables that could affect preferences for work (e.g., hour of day). To the extent there are income reference-dependent preferences or important daily income effects, the reduced-form comparison should also include a measure of income earned to that point on the shift.

The neoclassical labor supply model and the model with income reference dependence have sharp and contrasting predictions for this modeling approach. The neoclassical model implies that the probability of ending a shift after a given trip will be positively related to accumulated hours (conditional on accumulated income) and unrelated to accumulated income (conditional on accumulated hours and assuming income effects are zero with regard to daily wage variation). In contrast, the model with preferences that are dependent on an income reference level implies that the probability of ending a shift after a given trip will be positively related to accumulated income (conditional on accumulated hours) and less strongly related to accumulated hours (conditional on accumulated income).³⁴

Following Farber (2005), a reasonable approximate solution to the dynamic stopping problem can be implemented empirically as a simple discrete choice problem. At any point τ during the shift, a driver can calculate the forward-looking expected optimal stopping point, τ^* . The optimal stopping point may be a function of many factors, including hours worked so far on the shift and expectations about future earnings possibilities. If daily income effects are important, the optimal stopping point may also be a function of income earned so far on the shift. A driver will stop at τ if $\tau \geq \tau^*$ so that $\tau - \tau^* \geq 0$.

34. The dual reference point approach suggested by Koszegi and Rabin (2006) and implemented by Crawford and Meng (2011) suggests that behavior will be governed by both an income target (like the standard reference dependent preferences model) and by an hours target (with implications more like the neoclassical model). I discuss the implications of my estimates in the context of this model later in this section.

A reduced-form representation of $R(\tau) = \tau - \tau^*$ on a given shift is

$$(14) \quad R_{it}(\tau) = h_{\tau}\gamma_1 + y_{\tau}\gamma_2 + X_{it}\beta + \mu_i + \epsilon_{it\tau},$$

where i indexes the particular driver, t indexes time (hour of day, day of week, week of year, year), μ_i is a driver fixed effect, and ϵ_i is a random component. The quantity h_{τ} is a vector of indicators for ranges of hours worked at τ on the shift, y_{τ} is a vector of indicators for ranges of income earned at τ on the shift, and X measures other factors affecting the determination of the optimal stopping time and the comparison with τ . Elements of the vector X_{it} include sets of fixed effects for hour of day by day of week (168), week of year (52), and year (5) as well as indicators for the period subsequent to the September 4, 2012, fare increase and major holiday. These are included to capture systematic variation in earnings opportunities and the utility of leisure from continuing to drive.

The standard discrete choice model implies that the individual stops driving at τ if $R_{it}(\tau) \geq 0$, and, with appropriate distributional assumptions this implies a typical probit or logit specification based on the latent variable defined in equation (14). However, the presence of thousands of individual driver fixed effects makes estimation of a latent variable model impractical, and I proceed using a linear probability model.³⁵

I estimate the model separately for day and night shifts for the $\frac{2}{15}$ sample of drivers used earlier. I drop the shortest 1 percent and longest 1 percent of shifts. The resulting sample contains 51,021,936 trips for 6,312 drivers on 2,201,443 day shifts and 49,661,892 trips for 6,380 drivers on 2,116,675 night shifts.³⁶ Both day and night shifts average about 23 trips, with day shifts averaging 8.8 hours and night shifts averaging 8.2 hours.

35. The problem with the logit or probit model in this case is not the usual problem of lack of sufficient observations within each group (driver) for consistency. The problem is simply the scale of estimating a probit or logit model with tens of millions of observations and thousands of coefficients. I have estimated probit versions of the model without the driver fixed effects but with all of the other variables, and the results (in terms of marginal effects of the variables on the probability of stopping) are very close to those found using a linear probability model with the same variables.

36. Of the 7,578 drivers in the sample, 5,114 have driven both day and night shifts and so are represented in both samples.

Drivers earned an average of \$244 per day shift and \$261 per night shift.

Table VII contains estimates of the marginal effects of accumulated income and hours on the probability of a shift ending after a specific trip relative the base (lowest) categories from the linear probability model specified above. Columns (1) and (2) contain estimates of the effects of accumulated income on day and night shifts, respectively. As shift income accumulates (conditional on hours) during day shifts (column (1)), the probability of stopping increases slowly by 6 percentage points at \$300 relative to less than \$100. However, there is no increase in the probability of stopping as income accumulates during night shifts (column (2)), with the stopping probability at \$300 no different than the stopping probability at less than \$100 (conditional on hours).

Columns (3) and (4) of Table VII contain the analogous results for accumulation of hours (conditional on income). These show sharp increases in the probability of stopping with accumulated hours on both day and night shifts. On day shifts (column (3)), as hours accumulate (conditional on income), the probability of stopping increases by 12 percentage points from early in the shift to hour 11. The effect of hours is even larger on night shifts (column (4)), where the probability of stopping increases by 25 percentage points from early in the shift to hour 11.

Next, I quantify the extent to which shocks to income and hours affect the probability of shift ending. I start by using the estimates of the models to predict the probability that a shift continues (the survival probability) for each category of income and hours.³⁷ I then compute the survival probabilities for each of two counterfactuals. In the first counterfactual, I shock income by increasing the observed fare on each trip by 20 percent (holding trip time constant). I recalculate accumulated income and compare the predicted survival probability by hour with that predicted by the base (observed) data. It is important to note that this is a shock to income but not to the expected wage. The expected wage is accounted for by the date-time variables and other controls in the model. To the extent that income reference

37. The survival probability is constructed by taking 1 minus the predicted probability of stopping after each trip and creating the running product of this quantity during each shift.

TABLE VII
MARGINAL EFFECTS OF INCOME AND HOURS ON PROBABILITY OF ENDING SHIFT (LINEAR PROBABILITY MODEL)

Income (\$)	(1) Day shift	(2) Night shift	Hours	(3) Day shift	(4) Night shift
100–149	0.0001 (0.0003)	–0.0045 (0.0003)	3–5	0.0020 (0.0004)	–0.0049 (0.0003)
150–199	0.0044 (0.0006)	–0.0077 (0.0005)	6	0.0001 (0.0007)	0.0007 (0.0006)
200–224	0.0157 (0.0010)	–0.0062 (0.0007)	7	0.0034 (0.0011)	0.0223 (0.0010)
225–249	0.0264 (0.0013)	–0.0046 (0.0008)	8	0.0281 (0.0017)	0.0536 (0.0016)
250–274	0.0389 (0.0017)	–0.0042 (0.0011)	9	0.0750 (0.0025)	0.0897 (0.0022)
275–299	0.0506 (0.0020)	–0.0033 (0.0013)	10	0.1210 (0.0035)	0.1603 (0.0031)
300–349	0.0596 (0.0024)	–0.0027 (0.0017)	11	0.1236 (0.0050)	0.2563 (0.0051)
350–399	0.0607 (0.0028)	0.0011 (0.0024)	12	0.1004 (0.0078)	0.2573 (0.0142)
≥ 400	0.0702 (0.0034)	0.0101 (0.0035)	≥ 13	0.1093 (0.0050)	0.2406 (0.0063)

Notes. Based on estimates of two linear probability models for the probability of stopping: day shifts (columns (1) and (3)) and night shifts (columns (2) and (4)). The base category for income is \$0–99 and the base category for hours is 0–2. Both models additionally include sets of fixed effects for driver, hour of the day by day of the week (168), week of the year (52), and year (5) as well as indicators for the period subsequent to the September 4, 2012, fare increase and major holiday. Robust standard errors clustered by driver are in parentheses. See text for information on sample size and composition.

dependence, and perhaps income targeting, is important, survival probabilities at given hours should be lower due to the income shock. Without income reference dependence (or important daily income effects), survival probabilities should be unaffected by the income shock.

Figure III contains separate plots for day and night shifts of the effect on shift survival of the 20 percent uniform income shock. This shows a small reduction in the survival probability throughout day shifts (about 2.4 percentage points by 12 hours) and virtually no reduction (0.4 percentage points by 12 hours) on night shifts.

In the second counterfactual, I shock hours by increasing the observed time on and between each trip by 20 percent (holding fare income constant). I then recalculate accumulated hours and compare the predicted survival probability by income category

Shift Survival Probability -- Income Counterfactual vs. Observed

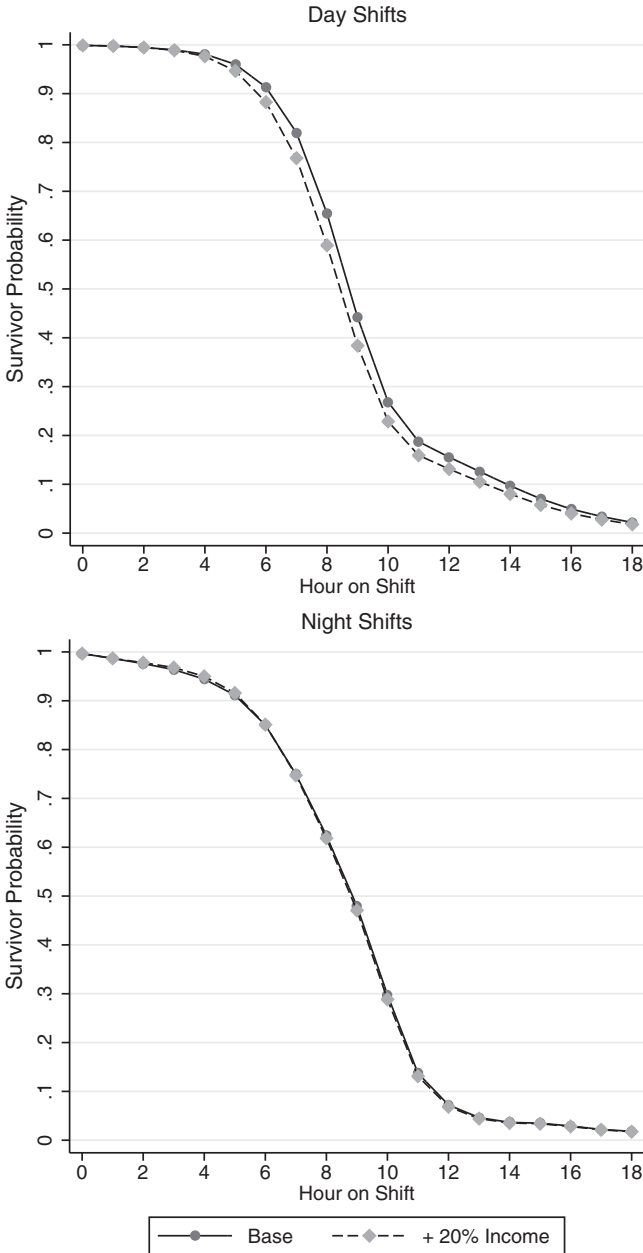


FIGURE III

Effect of 20 Percent Income Shock on Shift Survival

with that predicted by the base (observed) data. Note that this is a shock to hours worked but not to the expected wage, which is accounted for by the date-time variables and other controls in the model. With income reference dependence, the prediction is that the increase in hours worked at a given income level will not substantially affect the likelihood of a shift surviving to earn that amount. The neoclassical model predicts that the increase in hours required to earn a given amount will be associated with reduced likelihood of a shift surviving to earn that amount.

Figure IV contains separate plots for day and night shifts of the effect on shift survival of the 20 percent uniform hours shock. This shows a substantial reduction in the survival probability throughout day and night shifts. The survival probability to \$300–349 when hours are 20 percent higher is about 7 percentage points lower on day shifts and 22 percentage points lower on night shifts.

The conclusion from this analysis complements the results of the estimation of the slope of the labor supply function. The evidence is largely consistent with the neoclassical model. Hours are the important driver in the stopping decision, with some evidence of a marginal effect of income (and, hence, income reference points) on the stopping decision on day shifts only.

To investigate how accumulated hours and income affect drivers' stopping decisions in a less constrained way, I estimated a version of the stopping model that interacts the accumulated hours and income indicators. In other words, rather than the 10 income and 10 hours categories listed in Table VII, this version of the model included the full set of interactions. The model additionally includes driver fixed effects as well as the other controls listed in the note to Table VII.³⁸ Figure V contains separate plots for day shifts (top panel) and night shifts (bottom panel) of the estimated marginal effect of each hours-income category on the probability of a shift ending. For each hours-income category, the size of the square marker is proportional to the magnitude of the estimated marginal effect on the stopping probability

38. To limit the computational burden, this model is estimated using a $\frac{1}{15}$ random sample of drivers rather than the $\frac{2}{15}$ random sample used in the earlier estimation. To investigate whether any differences I might find are due to the difference in sample, I reestimated the model shown in Table VII using this smaller sample, and the results are virtually identical to those shown in the table.

Shift Survival Probability -- Hours Counterfactual vs. Observed

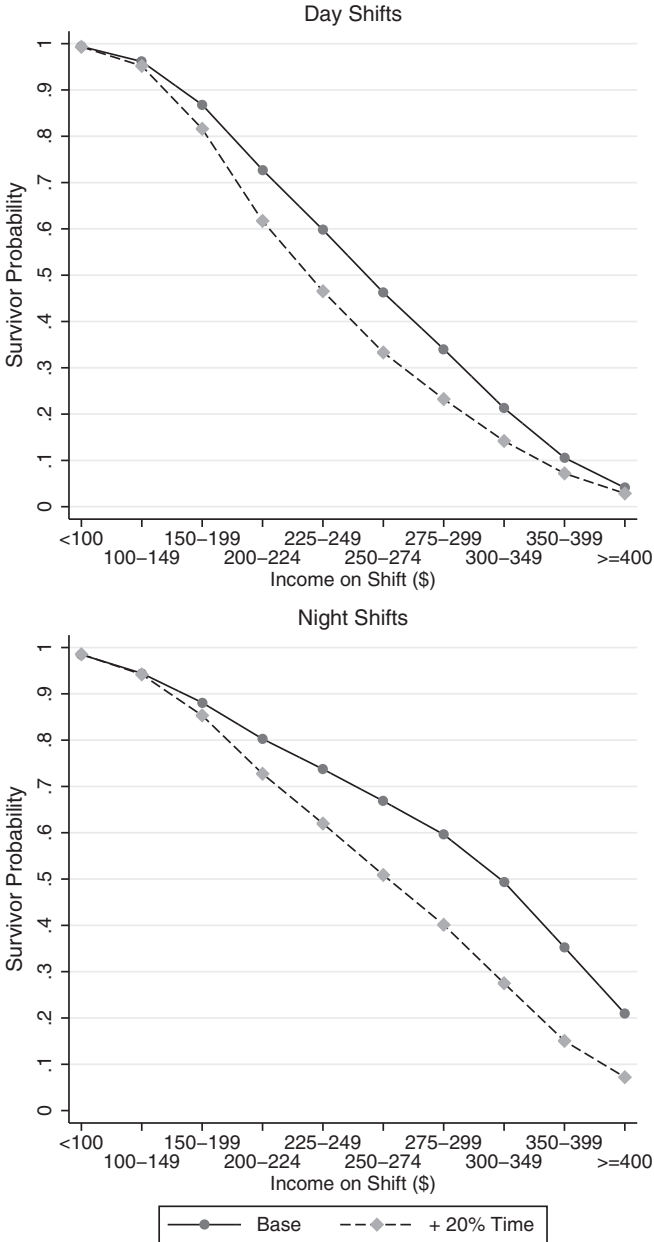
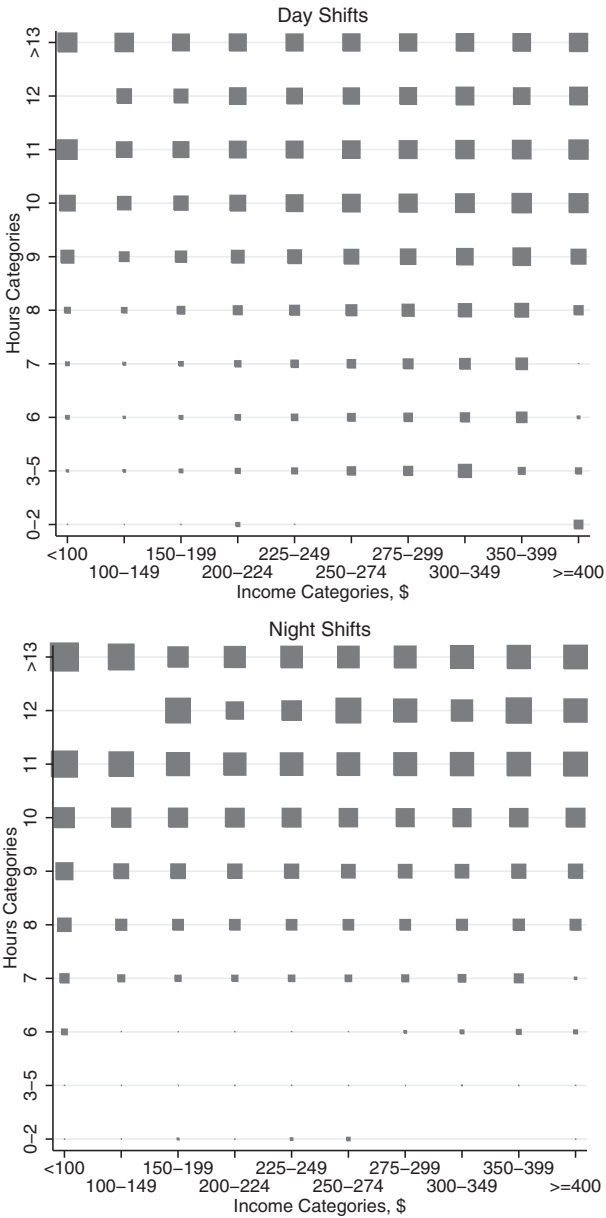


FIGURE IV

Effect of 20 Percent Hours Shock on Shift Survival



Note: Size of marker is proportional to marginal effect of hours-income on probability of shift ending.

FIGURE V

Marginal Effects of Hours and Income on Shift End Probability

relative to the lowest income-hours category (less than \$100, 0–2 hours).³⁹

The estimates for day shifts, in the left panel of Figure V, show that the probability of stopping rises with hours within income categories (reading south to north within a column). The probability of stopping also rises somewhat with income within hours categories but less sharply than the rise with hours (reading west to east within a row). This suggests that there may be some role for daily income reference points on day shifts but that the central determinant of a driver's stopping decision is accumulated hours. The estimates for night shifts, in the right panel of the figure, show a clearer pattern. The probability of stopping on night shifts rises sharply with hours within income categories but the stopping probability is roughly unaffected by income within hours categories. These patterns suggest a limited role, at best, for income reference dependence.

Analyses of taxi driver labor supply by Crawford and Meng (2011) and by Agarwal et al. (2015) posit reference points in daily income and daily hours, with loss aversion to income below an income reference point and hours above an hours reference point. The idea, based on Koszegi and Rabin (2006), is that there are two "domains of losses": (i) if income is below the income reference point then the individual is within the domain of losses for income and has a higher marginal utility of income and (ii) if hours are above the hours reference point then the individual is within the domain of losses for hours and has a higher marginal disutility of work (a higher marginal utility of leisure). While the theory does not have unambiguous predictions on how the decision to end a shift depends on income and hours, the model does suggest in rough terms that the likelihood of stopping (i) will be lowest when income is below the income reference point and hours are below the hours reference point (within the domain of losses in income and outside the domain of losses in hours), (ii) will be highest when income is above the income reference point and hours are above the hours reference point (outside the domain

39. The missing points (five each on day shifts and night shifts) are for hours-income combinations that are observed for 10 or fewer trips (out of a total of approximately 25 million trips on each of day shifts and night shifts). These are all for unlikely combinations of hours and income (very low income with long hours or very high income with short hours).

of losses in income and within the domain of losses in hours), and (iii) will be intermediate in the other cases.

With regard to the estimates shown in Figure V, this is weakly consistent with the pattern for day shifts (left panel), where the probability of stopping increases from southwest to northeast as both income and hours increase. However, as I noted earlier, the strongest relationship is in the hours dimension (south to north) so that the probability of stopping is increasing strongly in hours and weakly in income. The pattern for night shifts (right panel) does not provide support for dual reference point model because, while the stopping probabilities increase from south to north, indicating a strong increase with hours at all income levels, the stopping probabilities are flat going west to east, indicating no relationship between the stopping probability and accumulated income at each level of hours.

I also used the estimates from the model with income-hours interactions to recalculate the income and hours counterfactuals shown earlier, and the results (not shown here) are virtually identical to those shown in Figures III and IV. Shift survival probabilities and, by extension, shift durations are unaffected by a 20 percent income shock while income earned is substantially lower when hours at a given income level are 20 percent higher. These patterns provide additional evidence consistent with the conclusions I drew from Figures III, IV, and V.

To summarize, the pattern of results for variation in the stopping probability with hours and income and for the two counterfactuals is consistent with the predictions of the neoclassical model and with my findings based on estimating the labor supply elasticity. The pattern of results is not consistent with the basic income reference dependent preferences model in that (i) the stopping probability varies only a small amount with income conditional on hours on day shifts and does not vary at all with income conditional on hours on night shifts and (ii) there is virtually no effect of the counterfactual income shock on labor supply. The pattern of results is consistent with reference dependence in daily hours in that (i) the stopping probability increases substantially with hours conditional on income on both day shifts and night shifts and (ii) income is lower on shifts where counterfactually hours are higher at a given income level. However, it is difficult to see how this differs from the pattern predicted by the neoclassical model.

VIII. DO DIFFERENT DRIVERS USE DIFFERENT MODELS?

Economists typically assume a single model of behavior applies to all agents and estimate the parameters of said model. The scale of the data available here makes it feasible to estimate separate labor supply models for individual drivers. It may be that some drivers exhibit reference dependent preferences (substantial negative labor supply elasticities) and others are optimizers (positive labor supply elasticities).⁴⁰

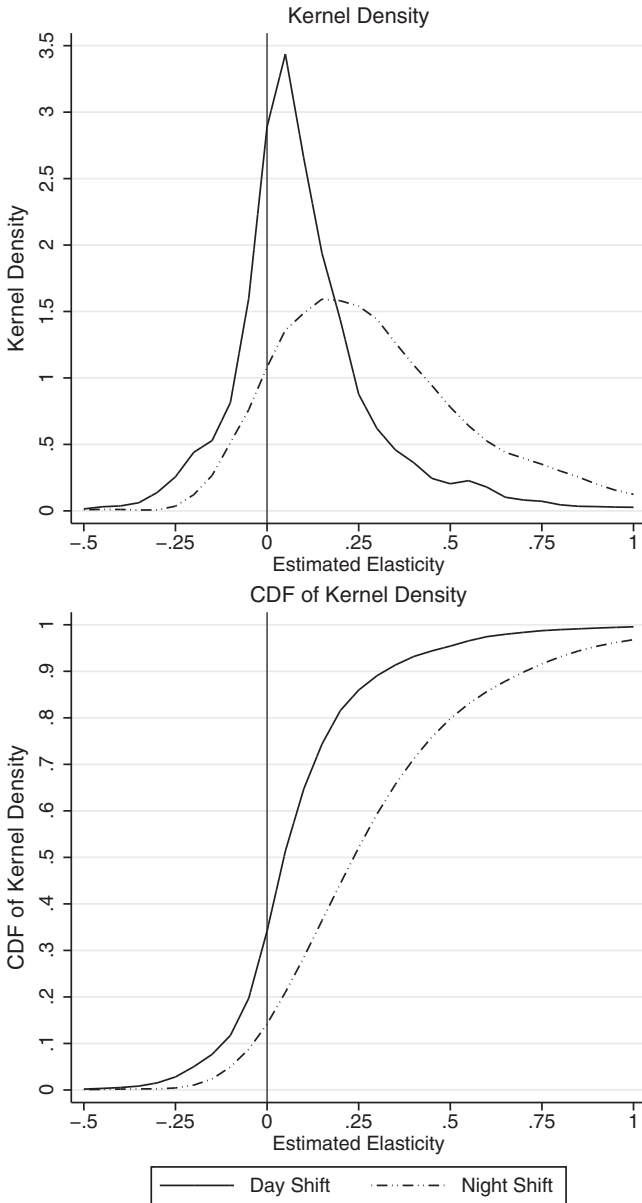
In this section, I return to estimating the slope of the relationship between log hours worked and the log wage and estimate separate labor supply models by driver for the large number of drivers who are observed on a substantial number of shifts. Given the difference in decision making margins and estimated elasticities for day shifts and night shifts, I make a distinction between “day shift drivers” and “night shift drivers.” I classify drivers who work at least 750 day shifts between 2009 and 2013 as day shift drivers and drivers who work at least 750 night shifts between 2009 and 2013 as night shift drivers. There are 1,267 day shift drivers and 1,205 night shift drivers in the $\frac{2}{15}$ random sample. Though most drivers are observed working different shifts over the period, drivers do tend to specialize. For example, there are only two drivers who are classified as both day and night shift. The mean number of night shifts for day shift drivers is 46.6 and the mean number of day shifts for night shift drivers is 55.8. I proceed estimating separate labor supply models for the day shifts of the day shift drivers and for the night shifts of the night shift drivers.

The top panel of Figure VI contains kernel density estimates of the distributions across day shift and night shift drivers of the wage elasticity of labor supply estimated from separate IV regression models for each driver.⁴¹ The density estimates are weighted

40. Doran (2014) presents some estimates that allow for a driver-specific relationship accumulated between income and the probability of stopping for a sample of 66 drivers who are observed for at least 500 trips (about 20 shifts). He finds that about half of these drivers respond to a positive transitory income shock with an increase in the likelihood of stopping, implying a negative wage elasticity of labor supply.

41. Each model, a regression for each driver of log shift hours on log average shift hourly earnings, also includes indicators for day of week, week of year, year, the period subsequent to the September 4, 2012, fare increase, and major holiday. The instrument used is as described earlier and used in Table V: the average of average hourly earnings for a nonoverlapping sample of drivers on the same day and shift.

Kernel Density and CDF, IV Estimates of Individual Elasticities



Note: Weighted by inverse sampling variance of estimates

FIGURE VI

Kernel Density Estimates of Distribution of Estimated Elasticities over Individual Drivers, Separately for Day Shift and Night Shift Drivers

by the inverse sampling variance of the individual estimated elasticities. The bottom panel of the figure contains the cumulative distribution function implied by the kernel density estimates. It is clear from the figure that there is substantial variation across drivers in estimated labor supply elasticities which reinforces the value of having sufficient data and variation to estimate separate models.⁴²

Examination of the distributions of elasticities summarized in Figure VI yield two strong conclusions. First, there is little evidence of a substantial number of individual drivers having the strongly negative labor supply elasticities implied by reference dependence. Only 0.12 percent of day shift drivers and 0.04 percent of night shift drivers are estimated to have elasticities less than -0.5 and less than 2.1 percent of day shift drivers and 0.24 percent of night shift drivers are estimated to have elasticities less than -0.25 . About 25 percent of day shift drivers and 11 percent of night shift drivers are estimated to have negative elasticities.⁴³ The average elasticity (weighted by inverse sampling variance) is 0.105 for day shift drivers and 0.321 for night shift drivers.

The second conclusion from Figure VI is that night shift drivers generally have larger elasticities than day shift drivers. In fact, the plot of the two CDFs indicates that the distribution of elasticities for night shift drivers stochastically dominates the distribution for day shift drivers. The generally larger elasticities for night shift drivers is consistent with these drivers being more likely to be able to adjust hours to unanticipated changes in earnings opportunities revealed during the shift. Note that with reference dependence, these same considerations would imply that elasticities would be more strongly negative for night shift drivers as they could work longer when wages are unexpectedly low to reach the daily earnings reference point while drivers on day shifts may have more limited ability to extend their shifts.

42. The variance of these estimated distributions are overstated as measures of the variance of the distribution of the underlying elasticities due to the fact that the sampling errors in estimating the elasticities is included. Given the positive mean of the distribution of elasticities, this implies that the fraction negative is overstated.

43. These proportions are calculated weighted by the inverse sampling variance of the estimated elasticities.

IX. DO DRIVERS LEARN TO OPTIMIZE?

It is clear that optimizing behavior dominates reference dependence in the sense that individuals who set labor supply optimally will earn more money working fewer hours than individuals who set labor supply as a result of reference dependence. Drivers who are optimizers will work relatively more hours on days when the wage is high, whereas drivers with reference dependent preferences will work relatively more on days when the wage is low. It may be that experience helps taxi drivers take advantage of high wage days, by teaching them not only to behave as optimizers rather than target earners but also to take better advantage of earnings opportunities by modifying their driving strategies.⁴⁴

Given the importance of labor income and the time commitment of taxi drivers, the value of learning how to optimize well is potentially very large in this industry.⁴⁵ To investigate the importance of learning, I analyze the labor supply of new taxi drivers and how it changes as they accumulate experience. The first step is to identify new drivers. This would be relatively straightforward if the actual hack license numbers of the drivers were available since these numbers are assigned in sequence. However, the hack license numbers are available only in encrypted form. As a result, I identify drivers as new if they were not observed driving for some period from January 2009 forward.

To determine a reasonable period of nonobservation to consider a driver a new entrant, I analyzed the labor supply patterns over time in the $\frac{2}{15}$ random subsample. It turns out that there is a fair amount of entry, exit, and reentry among taxi drivers. For example, about 27 percent of the 8,802 drivers in my sample did not drive for at least one three-month period and then returned to driving. About 14 percent did not drive for a six-month period and

44. Learning to be an optimizer is not a new idea. Camerer et al. (1997) present evidence suggesting that the more experienced taxi drivers in their sample do not exhibit target earnings behavior. List (2003) studied two markets for collectibles and found that experience resulted in a reduction in the importance of the endowment effect. This is related to reference dependence in that both the endowment effect and reference dependence flow from loss aversion.

45. Haggag, McManus, and Paci (2014) find substantial learning among New York City taxi drivers about how to find fares efficiently.

returned to driving. After not driving for a year, about 5 percent of drivers returned, after 18 months of not driving about 3 percent of drivers returned, and after 2 years of not driving, about 1.5 percent of drivers returned.

Conservatively, I define new drivers to be those who are not observed driving for a full year at the start of my data. In other words, while my data start in January 2009, I use the 2009 data only to identify drivers who are observed driving at some point in the 2010–2013 period but did not drive in 2009. Without a better method to identify new drivers, I live with the likelihood that about 5 percent of drivers I classify as new are in fact experienced to a greater or lesser extent. The restriction to drivers who did not have any shifts in 2009 eliminates about 64 percent of drivers and about 87 percent of shifts.

I use the full sample of drivers who did not have any shifts in 2009 (new drivers) for the analysis of learning, rather than only those in the $\frac{2}{15}$ random subsample I have been using. The sample of new entrants contains 4,814,278 shifts for 24,114 drivers. However, fully 7,038 (29 percent) of these drivers are observed for only a single shift, and 4,555 of these single shifts have only a single trip. Although it is possible that 29 percent of new drivers go through the procedure of getting a hack license only to quit after a single trip or shift, it may also be that these are simply data anomalies. In what follows, I delete the 7,038 drivers and their associated 7,038 shifts from the analysis. The remaining analysis sample contains 4,807,240 shifts for 17,076 drivers.

I define experience based on number of shifts driven rather than by the passage of time, and I divide experience into 10 categories. I then estimate separate labor supply models for each of the categories. Table VIII contains average characteristics of shifts by experience category. There are some interesting patterns. Average hours worked increases early with experience (from weeks 1 to 3) then declines after six months. Income per shift increases sharply between weeks 1 and 2 and continues to increase as experience accumulates. Average hourly earnings increases sharply with experience, by about 22 percent between the first week and year 3 or later. Finally, it appears that new drivers are relatively more likely than more experienced drivers to be on day shifts and, conversely, more experienced drivers are relatively more likely than less experienced drivers to be on night shifts. As I show in Table III, average hourly earnings are

TABLE VIII
MEAN SHIFT CHARACTERISTICS, BY EXPERIENCE LEVEL

Experience	Hours	Income	AHE	Day	Night	<i>N</i>
Week 1	9.06 (0.01)	227.08 (0.34)	24.97 (0.02)	0.53 (0.00)	0.39 (0.00)	112,387
Week 2	9.55 (0.01)	252.85 (0.36)	26.56 (0.02)	0.53 (0.00)	0.41 (0.00)	102,625
Week 3	9.61 (0.01)	258.80 (0.37)	27.05 (0.02)	0.51 (0.00)	0.42 (0.00)	98,241
Week 4	9.63 (0.01)	262.16 (0.37)	27.38 (0.02)	0.50 (0.00)	0.43 (0.00)	95,217
Month 2	9.61 (0.01)	266.28 (0.19)	27.87 (0.01)	0.49 (0.00)	0.43 (0.00)	358,312
Month 3	9.52 (0.01)	268.54 (0.18)	28.42 (0.01)	0.47 (0.00)	0.45 (0.00)	406,779
Months 4–6	9.44 (0.00)	271.84 (0.12)	29.05 (0.01)	0.45 (0.00)	0.46 (0.00)	891,679
Months 7–12	9.29 (0.00)	273.80 (0.10)	29.77 (0.01)	0.44 (0.00)	0.47 (0.00)	1,243,343
Year 2	9.03 (0.00)	277.16 (0.10)	31.01 (0.01)	0.44 (0.00)	0.46 (0.00)	1,216,563
≥Year 3	8.70 (0.01)	275.25 (0.22)	31.97 (0.01)	0.46 (0.00)	0.43 (0.00)	282,094

Notes. Means from the set of all shifts from 2010 to 2013 for drivers first observed in 2010 or later. Standard deviations of means in parentheses. Means for day and night are fraction of shifts that are day shifts and night shifts, respectively. The column headed AHE contains mean average hours earnings. The last column (*N*) contains the number of shifts for each experience level. I define month 2 as 5–8 weeks, month 3 as 9–13 weeks, months 4–6 as 27–52 weeks, year 2 as 53–104 weeks, and ≥ year 3 as at least 105 weeks. A week is defined as seven shifts, rounding up.

higher on night shifts, but further analysis (not shown here) suggests that movement toward night shifts accounts for very little of the growth in average hourly earnings with experience.⁴⁶

What I am interested in here is whether drivers learn to be better optimizers as they gain experience. Evidence for this would be labor supply elasticities that are less likely to be negative and are on average larger positive values as experience accumulates. I estimated separate IV regressions for each of the 10 experience categories using the model with controls described in the note to

46. Haggag, McManus, and Paci (2014) use the 2009 TPEP data to investigate learning by New York City taxi drivers, and their results suggest that experienced drivers are more productive because they are able to earn more after ending a trip in an area and at a time where new earnings opportunities are less promising.

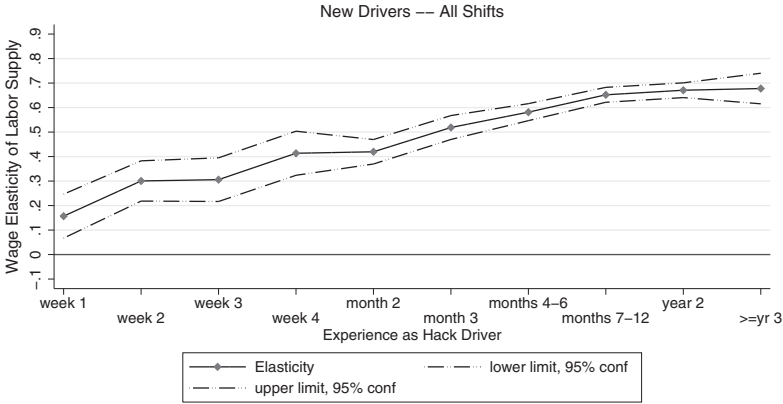


FIGURE VII
Wage Elasticity of Labor Supply, IV Estimates by Experience

Table V and without driver fixed effects.⁴⁷ Figure VII contains a plot of the estimated elasticities for each experience group with the associated 95 percent confidence intervals. The estimated elasticity in the first week is positive and grows steadily from 0.3 in week 2 to about 0.6 after six months and to about 0.7 in the third and later years.⁴⁸

The conclusion from this analysis is that drivers become more responsive to earnings opportunities as they accumulate experience but there is no evidence, even early, of the negative elasticities implied by income reference dependence. The general pattern of elasticities increasing with experience suggests that drivers do learn to become better optimizers.

X. SELECTION: DO INEFFICIENT DRIVERS QUIT DRIVING TAXIS?

Sample selection is an alternative explanation for the patterns of variation of outcomes with experience found in the previous section. Drivers who do not have a substantial positive labor supply

47. I do not include driver fixed effects when estimating separately by experience category results because inclusion of these effects in the models (whose estimates are shown in Table V) had only marginal effects on the estimated elasticities.

48. Farber (2014) contains plots of labor supply elasticities by experience group separately for day and night shifts. These show growth in elasticity with experience in the first two years from 0.1 to 0.45 on day shifts and from 0.15 to 0.75 on night shifts.

elasticity with respect to unanticipated wage changes will find it harder to earn money as a taxi driver. Such drivers (those drivers with small positive labor supply elasticities as well as those drivers with income reference-dependent preferences yielding negative labor supply elasticities) may be relatively likely to quit the business (stop being taxi drivers). The result will be a sample of drivers that becomes progressively composed of drivers with substantial positive labor supply elasticities as experience accumulates while no particular driver has a growing labor supply elasticity.

A key question is whether drivers who exit early have negative or smaller positive labor supply elasticities than do those who do not exit early, and I turn to an analysis of this question. I define exit understanding that a gap in driving does not necessarily imply exit. I use the same algorithm I used earlier in determining who is a new driver to calculate the fraction of new drivers with a gap of various sizes who were observed driving earlier (before the gap). After dropping the 7,038 new drivers who are observed only for a single shift, about 25.5 percent of the remaining 17,076 drivers in my sample of new drivers did not drive for at least one three-month period and then returned to driving. About 12.1 percent did not drive for a six-month period and returned to driving. After not driving for a year, 4.8 percent of drivers returned. Based on these tabulations, I require a full year of observation subsequent to the last observed shift for a driver to classify that driver as having exited. For this reason, I drop from the analysis 4,399 drivers who entered the industry on or after January 1, 2013. There are 12,677 drivers who entered between January 2010 and December 2012, of whom 7,064 are observed driving in December 2013.

Table IX contains IV estimates of labor supply elasticities from a model that interacts log average hourly earnings with a set of indicators for longevity of the driver (total number of shifts recorded for the driver). To focus on differences in elasticity at the start of the driving career, this analysis uses only shifts early in the career even for drivers who are observed for much longer. The first three columns contain estimates of elasticities on the first 12 shifts recorded for each driver.⁴⁹ These elasticities reflect the responsiveness of labor supply to unanticipated wage variation at the very beginning of a driver's experience. The estimates show

49. Drivers who are observed for fewer than 12 shifts necessarily have fewer than 12 observations.

TABLE IX
WAGE ELASTICITY, IV REGRESSION OF AVERAGE LOG DAILY HOURS, BY LONGEVITY
(TOTAL NUMBER OF SHIFTS AS TAXI DRIVER)

Longevity (total shifts)	Number drivers	First 12 Shifts			First 30 Shifts		
		All shifts	Day shifts	Night shifts	All shifts	Day shifts	Night shifts
2–12	1,707	0.0807 (0.0441)	0.0361 (0.0477)	0.0965 (0.0499)	0.1752 (0.0304)	0.1027 (0.0336)	0.1719 (0.0367)
13–30	612	0.2036 (0.0416)	0.1063 (0.0457)	0.2065 (0.0474)	0.2912 (0.0282)	0.1674 (0.0321)	0.2771 (0.0347)
31–60	627	0.2101 (0.0419)	0.1086 (0.0462)	0.2213 (0.0469)	0.3021 (0.0284)	0.1713 (0.0323)	0.2953 (0.0340)
61–90	500	0.2111 (0.0420)	0.1123 (0.0466)	0.2201 (0.0471)	0.3025 (0.0284)	0.1734 (0.0326)	0.2957 (0.0339)
91–150	798	0.2299 (0.0411)	0.1270 (0.0450)	0.2367 (0.0467)	0.3195 (0.0277)	0.1898 (0.0313)	0.3092 (0.0337)
151–300	2,160	0.2262 (0.0411)	0.1240 (0.0452)	0.2381 (0.0466)	0.3188 (0.0276)	0.1903 (0.0313)	0.3114 (0.0337)
≥ 301	6,273	0.2287 (0.0412)	0.1245 (0.0451)	0.2375 (0.0468)	0.3228 (0.0276)	0.1921 (0.0311)	0.3124 (0.0337)
# Drivers		12,677	8,722	7,318	12,677	9,159	8,091
# Shifts		139,690	74,253	54,878	331,434	171,103	135,871

Notes. Each column presents estimated elasticities from a single IV regression. The elasticities the coefficients of the interaction of $\ln W$ with a set of indicators for total shifts observed for each driver. The instrument set is the interaction for average hourly earnings is the average of average hourly earnings for a nonoverlapping sample of drivers on the same day with the set of indicators for total shifts observed for each driver. All models include a set of indicators for day of week (6), calendar week (51), year (3), the period subsequent to the September 4, 2012, fare increase (1), and major holiday (1). Robust standard errors clustered by driver are in parentheses.

that drivers who will quit driving with 12 or fewer shifts (13.5 percent of new drivers) have a statistically significantly lower elasticity than drivers in any category with a greater number of total shifts (p -value ≤ 0.01 for all tests). These patterns hold for day shifts and night shifts considered separately as well as for all shifts. Note further that considering up to the first 12 shifts, elasticities are larger on night shifts than on day shifts at every longevity, but these differences by shift are not significantly different from zero at conventional levels.

The next three columns of Table IX contain estimates of elasticities on up to the first 30 shifts recorded for each driver.⁵⁰ The pattern of results is qualitatively identical to those for the

50. Again, drivers who are observed for fewer than 30 shifts necessarily have fewer than 30 observations.

first 12 shifts. Drivers who will quit driving with 12 or fewer shifts have a statistically significantly lower elasticity on than drivers in any category with a greater level of total shifts (p -value $\leq .01$ for all tests). When considering up to the first 30 shifts, night shift elasticities are significantly greater than day shift elasticities at every longevity (p -value $< .05$ for all tests). I repeated this analysis for the first 60 shifts, and the pattern of results is the same (not shown here).

To summarize, there is no evidence of the negative elasticities associated with reference dependence even for drivers who quit after a small number of shifts, but it does appear that drivers who quit the business quickly (having driven no more than 12 shifts in total) are less responsive to unanticipated wage changes.

The estimates in Table IX show that the estimated elasticities for drivers who remain for the long term, while significantly positive, are relatively small (in the 0.1–0.25 range) early in their experience. This suggests that the selective exit of low-elasticity new drivers is not an important factor in the growth of elasticity with experience shown in Figure VII. To get at this directly, I computed IV estimates of the wage elasticities by experience for a sample longer-term drivers. To identify the longer-term drivers, I note that most exit happens very quickly, with 10.7 percent driving 7 or fewer shifts, 13.5 percent driving 12 or fewer shifts, and 17.9 percent driving 28 or fewer shifts.⁵¹ The rate of exit slows down substantially after 28 shifts, and I use this value to demarcate longer-term drivers.

Figure VIII contains a plot of the IV estimates of wage elasticities by experience for longer-term drivers, defined as those who are observed to drive more than 28 shifts. These estimates show the same pattern of increasing labor supply elasticity that was seen in Figure VII for the combination of long-term and short-term drivers.⁵² The conclusion is that drivers, in fact, learn to optimize and that the increase in elasticity with experience is not an artifact of selection.

To summarize, there is clear evidence that drivers with low labor supply elasticity tend to exit the industry early. Short-term

51. This tabulation is based on the 12,677 new drivers who entered between January 1, 2010, and December 31, 2012, and who drive for more than a single shift.

52. Farber (2014) contains plots of labor supply elasticities by experience group for longer-term drivers separately for day and night shifts. These show growth in elasticity with experience in the first two years similar to that found for the combination of long-term and short-term drivers. See note 48.

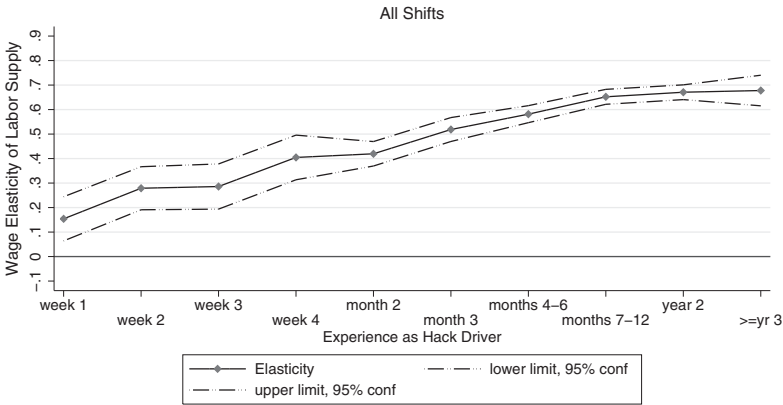


FIGURE VIII

Wage Elasticity of Labor Supply, IV Estimates by Experience (New Drivers With More Than 28 Shifts)

drivers (those who exit with 12 or fewer total shifts) exhibit significantly lower elasticity of labor supply than do longer-term drivers at the same (low) level of experience. The evidence on longer-term drivers, that the labor supply elasticity grows with experience over the first few years of driving, is consistent with the view that these drivers optimize over time by learning to adjust their daily labor supply positively in response to unanticipated variation in daily earnings opportunities. There is no evidence in these data that the labor supply decisions of drivers, either those who exit early or those who are longer-term, are substantially influenced by income reference dependence.

XI. FINAL REMARKS

Following on the work of Camerer et al. (1997), I evaluated the role of reference dependence versus neoclassical optimizing behavior in determining daily hours decisions of taxi driver in New York City using the complete driving records of all NYC taxi drivers over the 2009–2013 period. The high-level conclusion is that there is little evidence that income reference dependence is an important factor determining the labor supply of NYC taxi drivers. This conclusion rests on several pieces of evidence:

- Koszegi and Rabin’s (2006) model of expectations-based reference points implies that reference dependence is

relevant only for unanticipated wage movements in the wage while individuals respond neoclassically to anticipated wage movements. However, only about one-eighth of total hourly wage variation is unanticipated. Thus, most variation in wages (the seven-eighths that is anticipated) affects labor supply neoclassically.

- IV estimates of labor supply elasticities, based on a very large sample of about 5 million shifts for 8,802 drivers, are strongly positive and range from 0.4 to 0.8, depending on shift. There is no evidence of negative labor supply elasticities or target earnings behavior (a prediction of the income reference dependence model).
- The model further predicts that income reference dependence will be relevant when unanticipated transitory wage variation is small and not relevant when this variation is large. Although I do find that the elasticity of labor supply is smaller on shifts (particularly day shifts) with very little unanticipated wage movement, the range of unanticipated wage variation over which these smaller elasticities are found is so narrow that it implies only a very limited degree of loss aversion (a coefficient of loss aversion of about 1.1).
- Estimation of a discrete choice model of the probability of stopping after a trip (ending a shift) yields the clear result that the stopping probably is at best weakly correlated with accumulated income but strongly positively related to accumulated hours worked.
- Allowing individual drivers to have their own labor supply elasticities yields a range of estimated elasticities of labor supply, but the large majority of elasticities are estimated to be positive, and only a tiny fraction of drivers are estimated to have substantial negative elasticities.

Reference dependence and target earnings behavior leads to inefficient labor supply decisions. Responding positively to earnings opportunities, as the neoclassical model implies, is efficient in the sense that more money is earned when the return to working (the wage) is higher. This raises two questions regarding efficiency of work decisions that I have addressed here. First, do new drivers learn to take advantage of strong earnings

opportunities by working more on high-wage days and less on low-wage days? Second, do new drivers who start with negative or small positive labor supply elasticities quit the business? The answer to both questions is yes. The estimated labor supply elasticity grows substantially with experience, and new drivers with small labor supply elasticities are more likely to quit. Those who remain have growing elasticity with experience. In other words, new drivers who continue in the industry learn how to be better optimizers (respond more positively to wage variation). There is no evidence of negative labor supply elasticities on average for new drivers even at the start.

To summarize, the overall pattern is clear. Drivers tend to respond positively to both anticipated and unanticipated increases in earnings opportunities. This is consistent with the neo-classical optimizing model of labor supply, and I conclude that income reference dependence does not play an important role in determining the labor supply of taxi drivers in New York City.

PRINCETON UNIVERSITY

REFERENCES

- Abdellaoui, Mohammed, Han Bleichrodt, and Corina Paraschiv, "Loss Aversion under Prospect Theory: A Parameter-Free Measurement," *Management Science*, 53 (2007), 1659–1674.
- Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman, "Reference Points and Effort Provision," *American Economic Review*, 101 (2011), 470–492.
- Agarwal, Sumit, Mio Diao, Jessica Pan, and Tien Foo Sing, "Are Singaporean Cabdrivers Target Earners?," Mimeo, University of Singapore, 2015.
- Ashenfelter, Orley, Kirk Doran, and Bruce Schaller, "A Shred of Credible Evidence on the Long-Run Elasticity of Labour Supply," *Economica*, 77 (2010), 637–650.
- Blundell, Richard, and Thomas MaCurdy, "Labor Supply: A Review of Alternative Approaches," in *Handbook of Labor Economics*, vol. 3, Orley Ashenfelter and David Card, eds. (Amsterdam: Elsevier, 1999).
- Camerer, Colin, "Taxi Drivers and Beauty Contests," *Engineering & Science*, 60 (1997), 10–19.
- , "Prospect Theory in the Wild: Evidence from the Field," in *Choices, Values, and Frames*, Daniel Kahneman and Amos Tversky, eds. (Cambridge: Cambridge University Press, 2000).
- 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 (1997), 407–441.
- Chang, Tom, and Tal Gross, "How Many Pears Would a Pear Packer Pack if a Pear Packer Could Pack Pears at Quasi-Exogenously Varying Piece Rates?," *Journal of Economic Behavior and Organization*, 99 (2014), 1–17.
- Chou, Yuan K., "Testing Alternative Models of Labor Supply: Evidence from Taxi Drivers in Singapore," *Singapore Economic Review*, 47 (2002), 17–47.
- Crawford, Vincent P., and Juanjuan Meng, "New York City Cab Drivers Labor Supply Revisited: Reference-Dependent Preference with Rational

- Expectations Targets for Hours and Income,” *American Economic Review*, 101 (2011), 1912–1932.
- Doran, Kirk, “Are Long-Term Wage Elasticities of Labor Supply More Negative than Short-Term Ones?,” *Economic Letters*, 122 (2014), 208–210.
- Farber, Henry S., “Is Tomorrow Another Day? The Labor Supply of New York City Cab Drivers,” *Journal of Political Economy*, 113 (2005), 46–82.
- , “Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers,” *American Economic Review*, 98 (2008), 1069–1082.
- , “Why You Can’t Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers,” Working Paper No. 583, Industrial Relations Section, Princeton University, October 2014.
- Fehr, Ernst, and Lorenz Goette, “Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment,” *American Economic Review*, 97 (2007), 298–317.
- Haggag, Kareem, Brian McManus, and Giovanni Paci, “Learning by Driving: Productivity Improvements by New York City Taxi Drivers,” June 2014, available at <http://ssrn.com/abstract=2456733>.
- Haggag, Kareem, and Giovanni Paci, “Default Tips,” *American Economic Journal: Applied Economics*, 6 (2014), 1–19.
- Kahneman, Daniel, and Amos Tversky, “Prospect Theory: An Analysis of Decision under Risk,” *Econometrica*, 47 (1979), 263–292.
- Koszegi, Botond, and Matthew Rabin, “A Model of Reference-Dependent Preferences,” *Quarterly Journal of Economics*, 121 (2006), 1133–1165.
- List, John A., “Does Market Experience Eliminate Market Anomalies?,” *Quarterly Journal of Economics*, 118 (2003), 41–71.
- MaCurdy, Thomas E., “An Empirical Model of Labor Supply in a Life-Cycle Setting,” *Journal of Political Economy*, 89 (1981), 1059–1085.
- Nguyen, Quang, and Pingsun Leung, “Revenue Targeting in Fisheries: The Case of Hawaii Longline Fishery,” MPRA Paper No. 13846, March 2009, available at <http://mpra.ub.uni-muenchen.de/13846>.
- Oettinger, Gerald S., “An Empirical Analysis of the Daily Labor Supply of Stadium Vendors,” *Journal of Political Economy*, 107 (1999), 360–392.
- Ordoñez, Lisa, Maurice E. Schweitzer, Adam D. Galinsky, and Max H. Bazerman, “Goals Gone Wild: The Systematic Side Effects of Over-Prescribing Goal Setting,” *Academy of Management Perspectives*, 23 (2009), 6–16.
- Tovar, Patricia, “The Effects of Loss Aversion on Trade Policy: Theory and Evidence,” *Journal of International Economics*, 78 (2009), 154–167.
- Tversky, Amos, and Daniel Kahneman, “Loss Aversion in Riskless Choice: A Reference-Dependent Model,” *Quarterly Journal of Economics*, 106 (1991), 1039–1061.