Revisiting Labor Supply of New York City Taxi Drivers: Empirical Evidence from Large-scale Taxi Data

Ender Faruk Morgul, M.Sc. (Corresponding Author)
Graduate Research Assistant, Department of Civil & Urban Engineering,
Polytechnic School of Engineering,
Center for Urban Science and Progress (CUSP), New York University (NYU)
One MetroTech Center, 19th Floor, Brooklyn, NY 11201, U.S.A.
Tel: 1-(646)-997-0531
Email: efm279@nyu.edu

Kaan Ozbay, Ph.D.
Professor, Department of Civil & Urban Engineering,
Polytechnic School of Engineering,
Center for Urban Science and Progress (CUSP), New York University (NYU)
One MetroTech Center, 19th Floor, Brooklyn, NY 11201, U.S.A.
Tel: +1 (646) 997 0552
Email: kaan.ozbay@nyu.edu

Word count: 5,974 words + 2 Figure (500 words) + 5 Tables (1,250 words) = 7,724 words
Abstract: 221 words.

Paper Submitted for Presentation and Publication at the Transportation Research Board’s 94th Annual Meeting, Washington, D.C., 2015
Re-submission Date: November 15, 2014
ABSTRACT

Taxicab activity patterns in urbanized regions have been subject to major changes over the past few years, especially after the introduction of online taxi-hailing applications. Demand-and-supply characteristics of taxicab services remain as a significant question to be analyzed when deciding on transportation policy improvements in big cities. On the other hand, labor supply theories that seek to explain possible income-targeting behavior in transitory wage changes have also been tested with taxicab driver data. However, empirical work on investigating the relation between income and work hours have resulted in conflicting findings mainly due to methodological differences and, to some extent, limited number of observations. One of the main limitations of taxi demand-supply questions have been lack of sufficient and reliable data. This paper presents an empirical assessment of taxicab drivers’ labor supply using a novel large-scale data source from New York City. Electronically collected data provides detailed information about work hours of drivers and collected fares. The methodological framework employed in this paper was mainly borrowed from the previous literature and the findings were compared with results from earlier studies. Some of the results using the large-scale data were found to deviate from these earlier findings. Moreover, seasonal variations in labor supply response to transitory wage changes were empirically identified. The results of this study provide empirical support for the income-targeting hypothesis.

Keywords: Urban Taxi Traffic; Big Taxi Data; Labor Supply of Taxi Drivers; Income-Target Behavior; Reference-Dependent Preferences
INTRODUCTION AND BACKGROUND

Taxicabs, as a paratransit mode of travel, are vital components of urban mobility systems. Although highly controlled by law and regulations, taxicab service involves dynamic human-market interactions since the system operates mostly in a passenger demand-based manner (1). Moreover, newly emerging services of online taxi hailing, such as Uber and Hailo, have incredible potential to force changes in the existing taxicab regulations and to create further competitiveness in the taxicab market (2). Unlike traditional taxicab dispatching practice, online-taxi hailing applications allow passengers to request pick-ups directly from the drivers whose geo-location information is shown on a map and the passengers are usually charged lower rates compared to standard taxicabs (3). Moreover, these services have innovative ways for meeting taxicab supply during times when the demand is higher than regular conditions (i.e. in the rain). For example, Uber uses “surge pricing” which adjusts fares real-time to ensure enough number of cabs available to meet service reliability (4). With the increasing development of mobile technologies, undersupply of taxicabs in major cities and growing market penetration of the new online services, policy changes seem inevitable in the near future (5).

Decision-makers must precisely analyze demand-and-supply relationship when deciding on new policies in the presence of the above changes on the taxi demand and supply conditions. Observed data has a significant use for the process of policy making, and while evaluating the aftermath of a policy implementation (6). On the other hand, Big Data, -i.e. similar to the electronically collected taxicab data-, that has become available in several major cities including New York City (NYC) can revolutionize the way econometric analyses are conducted just due to the sheer size of this new data set. Detailed and almost complete information, that was once impossible to obtain, can be extracted, visualized and interpreted from these new data sources including the geographical distribution of taxicab activity, and driver/passenger behavior. For example some researchers used similar taxicab data to explore taxicab driver’s airport pick-up decisions (7), or travel time variability analysis (8).

Big transportation data provides many opportunities for better understanding of long standing economic theories that are also increasingly becoming important to decision-makers in the field of transportation. For example, when cities would like to make decisions in terms of the number of taxi medallions or total number of taxis, they also need to understand the relationship between driver wages and labor supply. In fact, over the last two decades the case of taxicab drivers has been of interest to several economists in order to test competing theories of labor supply in response to transitory wage changes. There are quite a few reasons that make the profession of taxicab driving attractive for researchers. First, taxi drivers have a unique flexibility to choose how many hours to work in a day. Therefore the total daily work hours largely vary across days. Second, hourly wages are not fixed or directly correlated with the work hours. Although customers are usually charged on a per mile basis, for hire taxi drivers spend less time for searching customers on busy days that influence their average hourly earnings. Third, mainly due to demand shocks such as adverse weather conditions or unavailability of alternative modes (i.e. subway...
breakdowns), the wages are highly correlated within days (9; 10). All these particular features of
taxicab driving allow for an exceptional test environment to observe the changes in labor supply
in response to temporary wage fluctuations.

Neoclassical labor supply model, which was first introduced by Lucas and Rapping (11),
predicts a positive response to transitory changes in wages. According to this approach, a driver is
expected to work longer hours as average hourly earnings get higher in shorter horizons. Camerer
et al. (9) rejected this idea by showing empirical evidence for negative wage elasticities from NYC
taxicab data. The results were explained by the drivers’ daily targets, which was mainly inspired
by reference dependent preferences as explained in the Prospect Theory (12; 13). In particular,
based on the sample data they used, Camerer et al. (9) claimed that drivers have daily income
targets of which they work until they reach and quit working even if they are earning unusually
high wages per hour.

The findings presented in Camerer et al. (9) ignited a new research direction on target-
income hypothesis. Although there is no definitive agreement on the extent of reference
dependence, such as the role of experience of the driver on target setting, or policy implications,
several theoretical and empirical contributions have been presented in the following years. Chou
(10) conducted an empirical study based on Singapore taxicab driver data and found negative wage
elasticities similar to Camerer et al. (9). Farber (14) used both one of Camerer et al. (9)’s three
data sets and a new set of taxicab driver data from NYC and found contrasting results to Camerer
et al. (9)’s. The difference in these new findings was explained by the alternative estimation
methods used and the more advanced way for measuring daily wage rate. A probit model was
developed to estimate the stopping probability and when driver fixed effects were included,
significant relation was found between the cumulative worked hours, however insignificant
relation found between cumulative earnings and the probability of stopping. The results in Farber
(14) were, therefore, in more favor of the neoclassical approach (when there are no income effects)
rather than target-income setting approach. In a following study, Farber (15) developed a labor
supply model with reference dependence which assumes a continuous utility function with a kink
(i.e. the point where the marginal utility of income change discontinuously) to represent target
income from where on marginal utility of income is lower. In the empirical example, a reduced
form model was tested in which target income levels were assumed as latent variables with driver
specific means. Using the same data set from NYC taxicab drivers in his previous study, Farber
(15) reached two significant conclusions: 1) Drivers had a higher probability of stopping when
they reached their daily income target. 2) Daily income targets of drivers were not fixed over the
days of the week and yet varied significantly. This latter finding of unstable daily targets was
discussed to be a major obstacle for developing a convenient labor supply model for income-
targeting.

Kőszegi and Rabin (16) formulated a general model for reference-dependent preferences
in which target levels were endogenously determined based on rational expectations. The
generalized reference-dependent utility function was the linear addition of the standard outcome-
based consumption utility and a gain-loss utility which was measured relative to reference points. One of the applications for the developed theory was tested on taxicab drivers. The theoretical model not only assumed an expected income target but also took work effort into account, therefore daily target for work hours were also considered as a decisive factor in labor supply. The application for taxicab drivers agreed with target setting behavior, when earnings of drivers in the morning were high, they tend to work less in the afternoon for any given hourly average wage.

Crawford and Meng (17) conducted an empirical analysis using a combined model based on Farber (15) and Kőszegi and Rabin (16). Stopping probabilities were estimated using Farber (14)’s data while reference dependence was addressed by both income and work hour targets as it was in Kőszegi and Rabin (16). The estimation results showed work hours were significantly related to stopping probability, whereas income levels were not. The findings also did not agree with Farber (15)’s finding that individual driver targets were not stable to develop a reference-dependent labor supply model. In this study, Crawford and Meng (17) indeed showed robust estimation of parameters that account for target levels.

FIGURE 1: A Reference-Dependent Driver who Stops at the Second Target Level He Reaches: Income Target on a Bad Day, Work Hour Target on a Good Day (17).

Figure 1 is adopted from Crawford and Meng (17) to provide a simple illustration of a targeting behavior assuming linear consumption utility. The bold lines represent the budget
constraints of which the line for a “good day” (i.e. higher total income) lies above the “bad day”
line. The dashed lines showing the “income target” and the “work hour target” divides the work-
day into four domains depending on the gain or loss of income and work hours. The driver starts
working in the lower-right corner of the graph which falls into the income-loss/work hours-loss
domain. Earnings and hours increase in the upper-left direction and the driver seeks to enter the
gain domains by reaching one or both of his targets. For every time step in the “hours” axis the
maker makes a decision to continue on his expected wage line or to stop driving. The driver stops
whenever his expected wage falls below his hours disutility cost of income (i.e. the intersection
points with the budget constraint lines). For the shown case, an income targeting driver’s optimal
stopping decision was assumed to be at the second target point that he reaches. For the “good day”
case the driver continues to drive until he reaches his work hour target and for the “bad day” case
he decides to stop at the income target point. Different specifications are possible using non-linear
consumption utility or various target responses – e.g. quitting after reaching the first target, which
are mainly shaped by individual driver characteristics.

Doran (18) investigated the wage elasticities for different length periods of time periods.
Using a new NYC taxi driver data, three models were considered, -i.e. neoclassical model, daily
income target as in Camerer et al. (9) and expectation based income targeting as introduced in
Kőszegi and Rabin (16). Larger negative wage elasticity in short term wage changes were found,
compared to long term wage changes. The results confirmed that neoclassical model’s inability to
explain negative wage elasticities in short-term wage increase. Camerer et al. (9)’s approach
adequately addressed the short-term effect and more advanced expectation based income targeting
approach as proposed by Kőszegi and Rabin (16) outperformed the other two models by
accounting for the differences in both short and long term wage changes.

One of the main limitations of all of the abovementioned studies is the sample sizes used
for empirical applications. Table 1 gives a summary of data that were used in these past empirical
studies with the number of taxi trips and drivers. Most of the studies were based on NYC data, and
the largest dataset used contains only a limited number of trip sheet data from 712 drivers.
Considering more than 12,000 taxicabs operating in the study region for the time period of data
collection, the sample size is quite small and its representativeness of the overall taxicab driver
behavior can be questionable (19). For example, as pointed out by Farber (14), parameters that are
used for model estimations in Camerer et al. (9) include average wage of other drivers for the same
day and same shift. However, if there are not enough observations (for some cases no observations)
for certain time periods of interest, the use of such parameters could lead to biased results.

Therefore, one of the main motivations for revisiting the theoretical contributions is the
availability of the entire data set of taxi trip information thanks to technological advances in vehicle
tracking, payment options and passenger information devices. We present empirical evidence from
a total of four-month period data that were obtained from more than 30,000 taxi drivers in NYC.
Since it was commonly specified in the literature that model selection and estimation
methodologies have significant impact on the findings, this study uses similar methodologies with identical parameter assumptions in Camerer et al. (9) and Farber (14).

### TABLE 1: Data Size in Selected Empirical Literature

<table>
<thead>
<tr>
<th>Study</th>
<th>Data</th>
<th>Study Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camerer et al. (9)</td>
<td>Three data sets: TRIP (70 trips; 13 drivers)</td>
<td>New York City</td>
</tr>
<tr>
<td></td>
<td>TLC1 (1,044 trips; 484 drivers)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TLC2 (712 trips; 712 drivers)</td>
<td></td>
</tr>
<tr>
<td>Chou (10)</td>
<td>Survey data from 92 drivers.</td>
<td>Singapore</td>
</tr>
<tr>
<td>Farber (14), Farber (15),</td>
<td>13,461 trips, 21 drivers.</td>
<td>New York City</td>
</tr>
<tr>
<td>Crawford and Meng (17)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**DATA**

In order to improve passenger service quality, TLC of New York City started the Taxi Passenger Enhancement Program (TPEP). As part of TPEP, the usage of technologically advanced devices was mandated for all taxicabs, such as on-board credit card payment gadgets and passenger information monitors that shows taxicab’s current location in the network. In 2008, the process of equipping the entire yellow taxi fleet operating in New York City was completed (20). The new system provided by TPEP has given decision-makers the opportunity to monitor the movement of entire taxi traffic in an efficient and almost fully automated way. For example, traditional paper trip-sheets were replaced with automated vehicle locator (VCL) technology that records the GPS coordinates of the pick-up and drop-off points with time stamps.

The 2013 taxi data were provided to the research team via a FOIL (The Freedom of Information Law) request in two parts. First part included the trip related information and second part gave the fare details. The two datasets had same number of entries and using driver information and trip pickup and drop-off coordinate information as key variables, two datasets were merged.

The drivers were identified in the dataset using a hack license number and a medallion number both of which were anonymized. Trip durations were provided in seconds and trip distances were given in miles. Fare information was available including total fare amount, tips, surcharge (e.g. extra fee for night trips) and toll amount if applied. Payment type was recorded with the two options as credit/debit card and cash payments, both types were kept for the analysis. Since cash payments did not include tip information, only trip fare amounts including tolls and...
surcharges were used in this study. Other data fields, namely, GPS coordinates of pick-up and drop-off locations, and number of passengers were not used for the analysis in this paper.

In this paper we refer the term “wage” as the gross earnings of a driver in a day. It is a fact that drivers’ net earnings can be different depending on several factors such as being an owner or a renter of taxicab and other finance costs. However the taxi dataset did not provide such detailed information about the drivers, therefore this study only considered gross earnings.

The taxi data used for the analysis in this study covers four complete months (i.e. January, April, July and October) of 2013, each of which accounts for a different season. The rest of the months were left out mainly to avoid unusually high computational overhead. A number of filtering steps were applied to remove noisy data points. First, erroneous data such as trip duration or trip distance of zero values, and zero total fare amounts were deleted. The records with empty fields and trip shifts that took more than 12 hours were screened out. Finally, trip shifts of a driver that start within the next 12 hours of the previous shift end were omitted. The number of observations (i.e. trips) in the raw data and filtered data for the time periods that were used in empirical test are given in Table 2. From each month approximately 8% of the original raw data were eliminated.

The statistics reported in 2014 NYC Taxi Fact Book (TFB) were used to validate the accuracy of filtered data. Among the four months, highest number of trips (i.e. 15,067,357), highest mean hours worked and highest mean total revenues were observed in April. This was consistent with TFB which showed the highest daily average usage in spring months. Lowest number of hours worked and average wage were observed in July, similarly, TFB showed the lowest taxicab usage in summer months. In addition, TFB reported an average daily taxi trip number of 485,000, which compared accurately with the raw data (20).

Median daily work hours in the data set were ranging in a narrow band between 8.65-8.78 hours. Average hourly earnings did not significantly differ among the months that represent four different seasons. When compared with Camerer et al. (9)’s data (which was also from NYC), average work hours were higher (i.e. 9.67 hours for their TLC1 data set) however average hourly earnings were almost half of what the data shows in this current study. Average earnings can be easily explained by the taxi rate hikes\(^1\). The lower average hours worked in 2013 data could be a result of larger sample size, considering the existence of drivers that work very low hours that decrease the sample mean. Another possible reason can be the different definition of the shifts. For the data at hand, we did not have information about the exact shift start time for drivers. Therefore in calculating hours worked, we did not have the chance to include the time searching for the first customer or the empty driving time after last customer. This information was usually provided in classical trip sheets. Work hours in this study was defined as the time period from the pickup of

\(^1\) Since 1996 taxicab rates in NYC increased three times. The rates during Camerer et al. (1) were $2.00 initial rate, $0.30 per mile and $0.30 per 90 seconds waiting time. In 2013, initial rate was $2.50, $0.50 per mile and $0.50 per one minute waiting time. For peak weekday hours $0.50 extra fee was charged. For more information please see Brent Cox, [http://www.theawl.com/2012/07/how-much-more-do-taxi-fares-cost-today](http://www.theawl.com/2012/07/how-much-more-do-taxi-fares-cost-today) (Accessed on 7/11/14)
the first customer to the drop off time of the last customer. It should be also noted that hourly
average wages could be slightly lower if those excess times were included in the total hours
worked. However, we did not have any validation data about the measurement error that could be
generalized to the entire dataset, therefore no correction methodologies were implemented.

TABLE 2: Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Number of Drivers</strong></td>
<td>32,224</td>
<td>33,111</td>
<td>32,788</td>
<td>33,456</td>
</tr>
<tr>
<td><strong>Total number of trips:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw Data:</td>
<td>14,744,391</td>
<td>15,067,357</td>
<td>13,791,052</td>
<td>14,971,100</td>
</tr>
<tr>
<td>Filtered Data:</td>
<td>13,423,993</td>
<td>13,820,121</td>
<td>12,767,771</td>
<td>13,784,076</td>
</tr>
<tr>
<td>Perc. Removed:</td>
<td>8.95%</td>
<td>8.27%</td>
<td>7.42%</td>
<td>7.93%</td>
</tr>
<tr>
<td><strong>Hours Worked</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>8.34</td>
<td>8.58</td>
<td>8.40</td>
<td>8.52</td>
</tr>
<tr>
<td>Median</td>
<td>8.55</td>
<td>8.78</td>
<td>8.6</td>
<td>8.7</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>2.44</td>
<td>2.38</td>
<td>2.46</td>
<td>2.40</td>
</tr>
<tr>
<td><strong>Average Wage ($/hour)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>36.50</td>
<td>38.12</td>
<td>35.77</td>
<td>38.00</td>
</tr>
<tr>
<td>Median</td>
<td>36.27</td>
<td>38.01</td>
<td>35.63</td>
<td>37.86</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>8.62</td>
<td>8.31</td>
<td>8.41</td>
<td>8.17</td>
</tr>
<tr>
<td><strong>Total Revenue ($)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>299.28</td>
<td>323.20</td>
<td>295.81</td>
<td>319.94</td>
</tr>
<tr>
<td>Median</td>
<td>296.72</td>
<td>320.81</td>
<td>296.3</td>
<td>317.44</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>100.27</td>
<td>104.07</td>
<td>100.04</td>
<td>103.76</td>
</tr>
<tr>
<td><strong>Number of Trips</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>22.03</td>
<td>22.63</td>
<td>20.76</td>
<td>21.71</td>
</tr>
<tr>
<td>Median</td>
<td>22.00</td>
<td>22.00</td>
<td>20.00</td>
<td>21.00</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>9.27</td>
<td>9.17</td>
<td>8.59</td>
<td>8.95</td>
</tr>
<tr>
<td><strong>Number of Customer Service Hours</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>4.08</td>
<td>4.56</td>
<td>4.18</td>
<td>4.59</td>
</tr>
<tr>
<td>Median</td>
<td>4.06</td>
<td>4.57</td>
<td>4.22</td>
<td>4.61</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>1.45</td>
<td>1.54</td>
<td>1.49</td>
<td>1.53</td>
</tr>
</tbody>
</table>
LABOR SUPPLY MODELS

The main purpose of this study was to test the alternative labor supply models using large-scale, comprehensive data sets of observed taxicab behavior. In doing so, two main empirical analysis were carried out. The first analysis was the “log hours function” that follows from Camerer et al. (9) and the second analysis uses the “discrete choice (i.e. probit) stopping model” that was explained in Farber (14).

Log Hours Function

The analysis starts with assessing simple correlation between log hours and log wages (i.e. hourly average wages). Table 2 gives the correlation statistics as negative values for all four months. Scatter plots of the log hours and log wages that were given in Figure 2 also visually support the negative correlation. The graphs indicate that there were no obvious differences in the relation of hourly average wage and work hour among different months of four seasons. One interesting indicator for income-targeting can be seen in the peak points in log hours. The points representing hours worked show that up to a certain wage, which can be the reference point, the log hours gradually increase and after the peak there exist a sharp decrease in hours worked.

An ordinary least square (OLS) regression with log hours as the dependent variable and log wages as the independent variable was estimated. The main assumption in this approach was that the driver makes the decision to stop considering the average hourly wages. Thus, hour to hour fluctuations in average hourly wage were not taken into account. To test this assumption hourly wage autocorrelations were calculated. If these autocorrelations were negative, that could mean a rational driver is more likely to quit early on a day with early high earnings, since upcoming hours are more likely to be low-wage hours. That would be an intentional violation of the labor supply behavior what neoclassical model claims (9). Median hourly wages for all drivers were calculated and regression analysis was conducted for median hours on the previous hour’s median wage. For all the data sets, positive autocorrelations were found, similar to Camerer et al. (9). Therefore, the alternative hypothesis of intentional violation was rejected.

Under the assumption of wages being uncorrelated across days and not fluctuating within day, labor supply response to temporary wage changes can be estimated using the following regression form (14):

Estimated autocorrelations: January data, 0.022; April data, 0.069; July data, 0.057; October data: 0.076.
\[ \ln H_{it} = \eta \ln W_{it} + X_{it} \beta + \varepsilon_{it}, \quad (1) \]

in which \( H_{it} \) is the hours worked by driver \( i \) on day \( t \), \( W_{it} \) is the average hourly wage of driver \( i \) on day \( t \), and \( X_{it} \) are the additional parameters (e.g. weather conditions, night shift) that are considered to affect labor supply decision. Parameter \( \eta \) represents the wage elasticity that shows the labor supply response to transitory wage changes. The term \( \varepsilon_{it} \) is a random error component.

FIGURE 2: Hours Worked – Average Wage Relationships

Discrete Choice Stopping Model
Discrete choice stopping model used in this study for labor supply is in the form of survival time model. The decision to stop working for a taxicab driver was assumed to take place at discrete points in time. Average hourly wage was considered in a cumulative way, which was a major difference compared to the “log hour model” which only calculated the ratio of total wages over total hours worked. Similarly, hours worked were handled in a cumulative way. For a time point \( \tau \), and an optimal time point \( \tau^* \), a driver was assumed to decide to stop if \( \tau \geq \tau^* \). A latent variable, 

\[
R_{idc}(\tau) = \gamma_1 h_{\tau} + \gamma_2 y_{\tau} + X_{idc}'\beta + \epsilon_{idc},
\]

(2)

where \( h_{\tau} \) is the cumulative hours worked in the shift until time point \( \tau \), and \( y_{\tau} \) is the cumulative earnings until time point \( \tau \). \( X_{idc} \) accounts for the other factors related to the stopping decision. The term \( \epsilon_{idc} \) is a normally distributed random error term. Therefore, a driver was assumed to stop at \( \tau \) if \( R_{idc} > 0 \). A probit model was used to estimate Eqn.2. A similar specification was used by Crawford and Meng (17), assuming decisions were made based on earnings in the first hour. In this study we followed the original specification proposed by Farber (14) in which cumulative work hours were considered and not split within the shift.

MODEL ESTIMATIONS

OLS model for “log wage function” was estimated using three other explanatory variables. Night shift, weather conditions, -i.e. rainy (including snow) and high temperature (above 85° Fahrenheit) days, were included to account for the labor supply decisions under different driving conditions. The estimation results are presented in Table 3. Goodness of fit measures used in Camerer et al. (9)’s models were adjusted \( R^2 \)'s and they were ranging between 0.146 and 0.484 (see Table II in their paper) which were not wildly different than the model fits shown in Table 3. All of the shown parameters were significantly different from zero. For the months of January and April, there were no observations for high temperature variable therefore that term was omitted in the estimations. As consistent with Camerer et al. (9)’s findings, the estimated wage elasticities (i.e. the coefficient of “Log hourly wage”) were negative for all of the months. In addition, the findings in this current analysis showed even stronger negative values compared to Camerer et al. (9), who found a negative elasticity estimation of -.618 in the most extreme case when they included driver fixed effects in their estimation. Night shift variable had a higher influence on the dependent variable (i.e. log hours) compared to weather related variables for all cases.

Possible measurement errors resulting from the simplistic average hourly wage computation, i.e. dividing total wage to reported hours, were addressed by introduction of instrumental variables into the model. Instruments should be selected that they are not correlated with the measurement error in hours. In particular, these errors could lead to high hours-low wage or low hours-high wage observations that both exacerbate the negativity of the estimated wage elasticity (9; 10). Camerer et al. (9) used the statistical measures from the distribution of hourly wages of other drivers that worked on the same day assuming average wage of other drivers should
be uncorrelated with a driver’s measurement error. Thus, downward bias in wage elasticity estimation because of over/understated hours could be reduced with the help of summary statistics (25th, 75th and median) of the wage distribution of other drivers.

### TABLE 3: OLS Estimation Results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Log hourly wage</td>
<td>-.6564 (.002)</td>
<td>-.6614 (.002)</td>
<td>-.6474 (.002)</td>
<td>-.6495 (.002)</td>
</tr>
<tr>
<td>Night</td>
<td>.1313 (.001)</td>
<td>.1274 (.001)</td>
<td>.1352 (.001)</td>
<td>.1280 (.001)</td>
</tr>
<tr>
<td>Rain^3</td>
<td>.0087 (.001)</td>
<td>.0011 (.001)</td>
<td>.0453 (.001)</td>
<td>.0127 (.001)</td>
</tr>
<tr>
<td>High Temp. ^*</td>
<td>-</td>
<td>-</td>
<td>.0141 (.001)</td>
<td>.107 (.003)</td>
</tr>
<tr>
<td>Sample Size (# of shifts)</td>
<td>640,084</td>
<td>641,088</td>
<td>641,534</td>
<td>665,725</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>.143</td>
<td>.126</td>
<td>.131</td>
<td>.124</td>
</tr>
</tbody>
</table>

Notes:
Values in parenthesis are the standard errors.
* High temperature variable was omitted for January and April due to collinearity.
- All coefficients were significant at 1 percent level.

Table 4 gives the instrumental variable regression results along with the first stage regression of log hourly wage variable on the instrumental variables. Adjusted $R^2$ values are comparable with the model fits found in Camerer et al. (9) in a sense that when they used larger sample sizes they found model fit measures such as 0.019. However when the sample size was low, their model fits improved dramatically, for example in a model with only eight drivers they found an adjusted $R^2$ value of 0.642 (see table III in their paper). Thus, considering the gigantic data size used in the models shown in Table 4, the model fits can be considered reasonable. The model results, surprisingly, did not reconcile with the findings Camerer et al. (9) since for three out of four data sets, the sign of wage elasticity turned to be positive. All instrumental variable coefficients were significantly different than zero. Using the same instrumental variables with the same methodology, Camerer et al. (9) found negative wage elasticities in five of his six models.

---

^3 Historical weather reports were downloaded in comma separated text file format from [http://www.wunderground.com/](http://www.wunderground.com/). Accessed on 7/16/14
That single model with positive wage elasticity (i.e. using TRIP data set with driver fixed effects) showed negative elasticities when instrumental variables were not used.

The only negative wage elasticity was found for the month of January 2013, indicating a possible income-targeting behavior among the drivers for this month. One interesting feature of this month January is that it has the second lowest number of total trips among the four months (see Table 2). Considering the negative impact of rainy (or snowy) days on work hours, drivers’ setting daily targets and quitting after reaching their expected earnings could make more sense, especially when compared to working conditions in other months. Moreover, median average hourly wage is the second lowest and standard deviation of average hourly wage is the highest in January, among the four months. The higher variation in wages could be an indication of driver preference heterogeneity, which can be modeled using more advanced modeling approaches such as Mixed Logit (21). Driver fixed effects were generally employed in the literature, however for large-scale data sets including thousands of drivers, introduction of driver-specific dummy variables into the model is not computationally feasible.

### TABLE 4: Instrumental Variables Regression Results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Log hourly wage</td>
<td>-.2309 (.014)</td>
<td>1.031 (.026)</td>
<td>.3052 (.011)</td>
<td>.599 (.019)</td>
</tr>
<tr>
<td>Night</td>
<td>0.1147 (.001)</td>
<td>.0265 (.002)</td>
<td>.1122 (.001)</td>
<td>.0756 (.002)</td>
</tr>
<tr>
<td>Rain</td>
<td>-.0136 (.001)</td>
<td>.0243 (.002)</td>
<td>.0137 (.001)</td>
<td>.0306 (.002)</td>
</tr>
<tr>
<td>High Temp.*</td>
<td>-.</td>
<td>-.</td>
<td>.005 (.001)</td>
<td>.114 (.068)</td>
</tr>
</tbody>
</table>

**First-Stage Regressions**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>-.0294 (.001)</td>
<td>-.0067 (.001)</td>
<td>-.012 (.002)</td>
<td>-.008 (.001)</td>
</tr>
<tr>
<td>25th percentile</td>
<td>.0302 (.001)</td>
<td>.0193 (.001)</td>
<td>.025 (.001)</td>
<td>.0229 (.001)</td>
</tr>
<tr>
<td>75th percentile</td>
<td>.0262 (.001)</td>
<td>.0119 (.001)</td>
<td>.0163 (.001)</td>
<td>.0114 (.001)</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>.087</td>
<td>.035</td>
<td>.068</td>
<td>.124</td>
</tr>
<tr>
<td>P-value for instruments</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
</tbody>
</table>

Notes:
Values in parenthesis are the standard errors.
* High temperature variable was omitted for January and April due to collinearity.
- All coefficients were significant at 1 percent level.
Strong positive elasticity observed in April, on the other hand, again, could be compromised with the demand for taxicabs, which is highest during spring months throughout the year. Drivers might be more likely to drive longer hours while the demand is at its yearly peak, and behave more consistent with neoclassical labor supply model to avoid the opportunity cost. One interesting extension to current analysis could be including experience levels of drivers in the models. Camerer et al. (9) conducted an analysis based on the hack license numbers, which were issued in order (i.e. smaller numbers meaning high experience), and found that high-experienced drivers’ wage elasticities were higher than low-experienced drivers. That is, income targeting were more common within experienced drivers. In our data, however, the hack license numbers were anonymized therefore it was not possible to validate this finding. For July and October data, again positive wage elasticities found but with much smaller magnitudes compared to April data.

The “log hours function” analysis yielded in three main observations. First, the average hourly wage calculation by dividing the entire daily income to total work hours, did not give consistent results when instrumental variables were introduced. This econometric shortcoming was questioned by Farber (14) earlier but was not tested empirically mainly because of the lack of sufficient amount of data to calculate the instrumental variables in the form of statistical summary of other drivers. It was empirically shown in this current study that using the same instrumental variables as in Camerer et al. (9) could change the sign of the wage elasticity, as it happened for the three out of four data sets in this study. Therefore similar model specifications using a broad average hourly wage must be implemented with care. Second, it was still not possible to reject income targeting behavior, since January data indicated negative wage elasticities with and without instrumental variables. Third, varying demand levels for taxicabs could have an effect on income targeting, as April data showed a significantly positive wage elasticity when the demand was the highest in the study year. The latter two findings could be an implication of the seasonal variations in income-targeting behavior of drivers.

Table 5 shows marginal probability effects using “discrete choice stopping model”. Cumulative work hours (i.e. total work hours including breaks, cruising and with passenger), cumulative on-duty hours with passengers, cumulative wage and control variables (i.e. night shift, weather conditions) were considered as independent variables. The effects of the variables on stopping probability was evaluated for two different key points. The first point was the same values used in Farber (14)’s analysis (8.0 hours total work, $150 shift income) and the second point was with an updated daily income level based on Table 2 (8.0 hours total work, $300 shift income). The estimated coefficients of cumulative hours and cumulative wage had positive signs for all cases which accorded with intuition. The sensitivity to the selected evaluation point was also reasonably estimated with increasing marginal effect on probability of stopping with higher cumulative shift income in the second evaluation point.

The probability of stopping was significantly related to wage and hours worked for all of the study months. The effects of cumulative wage were always higher than cumulative hours, both of which had significant coefficients at one percent level. This was a consistent observation with
the results seen in the literature when only two variables, cumulative hours and cumulative wages, were used (14; 17). However when the dummy variables were included in the models, those literature reported insignificant effect of cumulative wage, while cumulative hours usually remained significant. The results in Table 5 did not agree with these findings, in which cumulative wage was still a significant factor on stopping probability when the models were estimated with control variables. Therefore, this finding suggested that daily target earnings was an important factor in taxicab drivers’ labor supply. The effect of a $10 increase in income on the probability of stopping work was found 0.24 to 0.48 per cent for a driver that has already earned $300 for that day.

**TABLE 5: Marginal Effects Based on Stopping Probability: Discrete Choice Stopping Model**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Cumulative Hours</td>
<td>.006 (.001)</td>
<td>.013 (.001)</td>
<td>.004 (.001)</td>
<td>.010 (.001)</td>
</tr>
<tr>
<td>Cumulative Hours with Passengers</td>
<td>-.010 (.001)</td>
<td>-.023 (.001)</td>
<td>-.006 (.001)</td>
<td>-.013 (.001)</td>
</tr>
<tr>
<td>Cumulative Wage/100</td>
<td>.032 (.001)</td>
<td>.076 (.001)</td>
<td>.025 (.001)</td>
<td>.056 (.001)</td>
</tr>
<tr>
<td>Night</td>
<td>.009 (.001)</td>
<td>.021 (.001)</td>
<td>.013 (.001)</td>
<td>.030 (.001)</td>
</tr>
<tr>
<td>Rain</td>
<td>-.004 (.001)</td>
<td>-.011 (.001)</td>
<td>.001 (.001)</td>
<td>.001 (.001)</td>
</tr>
<tr>
<td>High Temp.*</td>
<td>-</td>
<td>-</td>
<td>-.001 (.001)</td>
<td>-.016 (.001)</td>
</tr>
<tr>
<td>Sample Size (# of taxi trips)</td>
<td>13,671,075</td>
<td>14,074,120</td>
<td>12,806,892</td>
<td>14,000,203</td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>.140</td>
<td>.161</td>
<td>.151</td>
<td>.156</td>
</tr>
</tbody>
</table>

Notes: (1): Evaluation point for marginal effects: Cumulative Total Hours= 8.0, Cumulative Income/100=1.5
(2): Evaluation point for marginal effects: Cumulative Total Hours= 8.0, Cumulative Income/100=3.0
* High temperature variable was omitted for January and April due to collinearity.
-Values in parenthesis are the standard errors.
-All coefficients were significant at 1 percent level. The coefficient for the “Rain” variable was insignificant for January, 2013 estimate, therefore excluded from the table.
-Trip shifts that last less than 0.5 hours were excluded from the analysis.

Moreover, the variation in the marginal effect of daily income among the months in the “discrete choice stopping model” interestingly matched the findings from “log hour function” model. For example, January data showed the highest marginal effect of cumulative income (.076) on the stopping decision, as shown in Table 5. In other words, drivers were more likely to stop...
after reaching $300 evaluation point for cumulative income in January. Similarly, as shown in Table 3 and Table 4, significant negative wage elasticities were found for the January data. Therefore two completely different methodologies were found to support income-targeting behavior of drivers for the same data set. April data in Table 4 showed strong positive wage elasticities in “log hour model”, -i.e. opposite of income-targeting behavior, whereas the results in Table 5 for “discrete choice stopping model” showed the second lowest magnitude of the marginal effect of income (.056) on the decision to stop driving. Thus, even if there exists an income targeting behavior for the month of April, the impact of this behavior was not as significant as two other months, namely January and July.

CONCLUSION

This paper deals with the taxicab drivers’ labor supply behavior. In particular, the goal of this paper is to take advantage of the new Big Data to shed additional light to the well-established theories of labor supply in response to transitory wage changes and also demonstrate usefulness of it in the context of well-studied economic theories, especially the ones with major implications on urban transportation. One of the potential uses of these models is for the understanding of the impact of demand changes for traditional taxi services on the supply as result of the introduction of new technologies.

This study adds to literature by presenting an empirical investigation for income-targeting approach. The data used in this study were collected under TPEP and included almost every taxi trip in NYC during January, April, July and October months of 2013. The main objective of this study is to investigate the relationship between income and worked hours of taxicab drivers based on the large size data and to seek evidence for income-targeting hypothesis which has been a topic of debate in labor economics literature. The methodologies used in this study were purposely borrowed from earlier contributions of Camerer et al. (9) and Farber (14), and the results based on the new data were compared to their findings.

The findings were interesting in that they indicated several differences in estimations that could lead to different interpretations as seen in previous studies. For the “log-hour function” approach, simple OLS estimation showed the evidence for income-targeting behavior as proposed in Camerer et al. (9). However, when instrumental variables were included in the model to reduce possible biases due to measurement errors in average hourly wage, positive wage elasticities were found for three of the four data sets (9). Income-targeting assumption still could not be rejected since for January data, negative wage elasticities were estimated. It is hypothesized that seasonal taxicab demand could be a determinant factor in labor supply response of drivers. For “discrete choice stopping model”, both cumulative income and cumulative hours worked were found to affect the probability of stopping. Farber (14) and Crawford and Meng (17) found income as an insignificant factor whereas total hours worked as a major factor for similar models. On the other hand, two different methodologies predicted somewhat similar variation among the study months, January is found to be the month when income-targeting behavior was most supported.
Both models used in this study were well established and had a wide usage that have led to development of various extensions. Empirical testing, however, needs more work to better understand the real-world implications of these models. The results presented in this study showed the potential use of Big Data for conducting unique empirical analysis of economic theories where limited data can significantly bias analysis results. There are many implications of the provided analysis regarding the reliability of taxicab service. For example, a forthcoming study by Farber (22) discusses the reasons for undersupply of taxicabs during rainy weather. Using a large-size taxicab data Farber (22) suggests a rain surcharge can be used to increase the taxicab supply.

Future similar studies should include detailed spatial characteristics of taxi trips. Although Farber (14) found a weak relation between location and key labor supply model parameters, i.e. income and hours worked, an estimation using increased sample size could be worth investigating. One limitation of the provided analysis was the nonexistence of the information about the driver experience which was shown to have an impact on labor supply (9). However large size taxi data sets were usually provided with anonymized driver information due to privacy concerns, therefore further cooperation with city agencies or taxi medallion companies is needed to receive data about driver experience and similar characteristics. It is important to note that, new taxi data continues to become available and it will be possible to search for any changes in traditional taxi service demand and supply as a result of on-line taxi hailing applications. This remains as a topic of future research.

ACKNOWLEDGMENTS AND DISCLAIMER

This study was partially funded by Polytechnic School of Engineering and Center for Urban Science and Progress (CUSP) of New York University. The authors are solely responsible for the findings presented in this research work.

REFERENCES


