Learning from disruption: the taxicab market case

Estefania Vergara-Cobos*
PRELIMINARY VERSION
Theoretical Part
April 11, 2018

Abstract
Uber joined the NYC taxicab market with 3 advantages: a dynamic pricing scheme, a rating system, and no regulation on entry. In 2015, the Taxi and Limousine Commission approved the use of a matching technology on yellow taxicabs similar to Uber’s but with no surge pricing and no rating system. Despite the incorporation of a similar technology, Uber is still growing, indicating that dynamic pricing and rating are key in its success. This paper presents an structural dynamic model of a market where some drivers are regulated on entry and price, while the rest are not regulated on entry and face dynamic pricing. Specifically, the model captures the dynamic learning process of drivers about the market conditions and the changes on labor supply decisions. Using the data of 1.1 billion yellow trips and 19 million Uber rides, I estimate the model and study policy implications for this partially-regulated market.

Keywords: disruption, regulation, entry.

1 Introduction
The yellow taxicab market is presenting fissures with the the entrance of “e-hailing” companies such as Uber and Lyft. These start-ups are turning anyone with a car and interested in providing taxi services into a cab driver, but most important of all is that they are creating a business model where the investment in the medallion permit, the cost of leasing a medallion, and taxi fare regulation by the public administration are eliminated. The difference in the cost structures and, therefore, Uber’s comparative advantage is such that some countries have forbidden Uber from operating in their territory.

Heavily regulated industries such as transportation, healthcare and finance, are becoming targets for market disruption driven by entrants with higher technology and different cost structures. According to the Silicon Valley Bank, the expansion of technology is addressing frustrating issues of historically regulated industries while changing how consumers access and how they differ types of services.

A marketplace is a platform that uses technology to match transactions between independent providers and immediate consumers of goods and services. The success of companies like Uber and AirBnB illustrates the growing potential of marketplaces and of disruption over established companies. The industries perceived with the highest threats of marketplace disruption are: Finance, Healthcare, Small Business, Education, and Transportation.

Among the comparative advantages of marketplaces are: low transaction costs, highly efficient matching platforms, increased quality supplied, and network creation. On the other hand, accord-
ing to a survey conducted to marketplace entrepreneurs, the greatest barriers to a marketplace’s success ranked supply and demand balance first, followed by establishing trust with consumers. Only 7% of the entrepreneurs chose regulatory hurdles as a threat.  

In relation to my case of study, the literature considers regulatory hurdles such as licensing (the cost of a medallion plate) and other high start-up costs as the strongest barriers to entry. However, the new marketplaces present different cost functions allowing them to compete without incurring in initial high costs. It is often argued that early entrants to a market have comparative advantages over later entrants. In this paper, we consider the case in which the regulatory policies implemented by the incumbent become a disadvantage when an entrant with a better technology joins the market. In these cases, classic strategies such as predatory pricing or raising the rivals’ costs cannot be adopted immediately by the incumbent due to possible high adverse effect of deregulatory policies.

1.1 Evidence

The data suggest that Uber is acting as a direct competitor to yellow taxicabs. Uber has grown dramatically in Manhattan, notching a 275% increase in pickups from June 2014 to June 2015, while taxi pickups declined by 9% over the same period of time. Uber made 1.4 million more Manhattan pickups in June 2015 than it did in June 2014, while taxis made 1.1 million fewer pickups. Figure 1-a shows the evolution of the taxi stock price of the only firm trading medallions in the stock market. It is clear that after the entrance of Uber in late 2012, the value of the stocks starts declining. This figure shows the correlation of the medallions’ market value and the entrance of Uber, and it suggest that if the trend continues, medallions will lose market value completely. The same can be seen from Figure 2-b, which shows the market price of a medallion. It is worth noting that such changes only take place a few months after the entrance of Uber because there was a lot uncertainty about whether Uber would be allowed to stay and compete in this market, or not.

Figure 1: Effect of the entrance Uber in the stock market and the price of medallions

---

1Worldwide investment in start-ups doing on-demand hiring, raised from USD 0.5 billion in 2010 to almost USD 1.6 billions in 2013. The number of companies was 117 in 2013, 39 more than in the previous year. Given its significant grow, the on-demand economy represents new challenges for policy makers.
Learning from disruption: the taxicab market case

Figure 2: Effect of the entrance of Uber on the labor supply

(a) Hours per driver

(b) Trips per day

(c) Unique number of drivers

(d) Farebox

2 Literature Review

Taxicab models had been studied from different focus points such as: elasticity of demand, supply and demand equilibrium, negative externalities and returns to scale [Salanova et al., 2011]. Specifically, such studies had focused on consumers’ theory and regulatory policies. The specific topics of interest in these areas address aggregation of supply and demand for taxis, congestion caused by taxicabs, economies of scale produced by externalities, social optimum, multiple equilibria, subsidizing as a policy of achieving a first best, optimum fares, consumer’s waiting time estimation, and search and match algorithms. Optimal pricing and different market conditions are modeled in studies, such as De Vany [1975], that analyzes both competition and monopolistic structure and conclude that monopolistic taxi markets restrict supply to the high-income demanders. Studies in other fields such as computer science and civil engineering address the estimation of the taxi industry focusing mainly in the elasticity of the demand side and in matching algorithms [Wong et al., 2001].

Despite the broadness of these studies, little has been done in terms of the price elasticity of labor supply and the study of the effects of the new “e-hailing” dispatching method. Theoretically, the closer approach there is to an “e-hailing” dispatching model is the work done by Massow and Canbolat [2010] which designs a double queue model in which drivers are assigned to queues in zones of high demand. Among the few studies that focus on flexible labor supply schedules (independent workers) there are two main groups: those who describe a positive elasticity of labor supply and those who do the opposite. The seminal work of Camerer et al. [1997] addresses the economic particularities of modeling the labor supply decision of New York City cab drivers. Using “trip-sheets” of 1988, 1990, and Spring 1994; the authors concluded that drivers work until a targeted wage is reached, which differs from the traditional economic theory that predicts that increasing wages would increase the number of hours devoted to work. They found negative wage elasticities even when controlling for different variables like experience. The work of Koszegi and Rabin [2006] supports the findings of Camerer et al. [1997] by using a utility-based model with...
drivers’ endogenous targets. This is, they assume the existence of individuals’ reference bound-
aries for income and effort based on which the driver will decide to keep working or not. Based
on [Camerer et al. 1997], [Crawford and Meng 2011] introduced the theory of reference-dependent
preferences to the analysis of independent labor force. Following [K˝ oszegi and Rabin 2006], they
state that a driver’s utility function reflects the standard income and leisure preferences but also
reference-dependent gain-loss notions. The authors try to capture evidence supporting targeting
preferences. Like [Camerer et al. 1997], they found that when early earnings are high, hours of
work have a strong effect on the stopping probability. To this [Farber 2003] adds that the number
of hours worked is significant to stopping however not the income level gained so far.

On the other hand, a relevant study is done by [Oettinger 1999], which states that the find-
ings in [Camerer et al. 1997] lack of data of the demand shifters which causes a downward labor
supply elasticity. [Oettinger 1999] studied the daily labor decision of a group of stadium vendors
with panel and aggregate data of demand and supply. This paper makes an aggregate analysis of
the participation decisions and it addresses the importance of also modeling effort choices. The
theoretical model is based on the assumption that the workers will provide services as long as the
opportunity cost is less than the expected earnings. The author found a positive elasticity in the
participation with respect to hourly earnings.

Using the new data released by the NYC-TLC, in a more recent study, [Farber 2014] also stud-
ies this topic and concludes that there is no reference dependence in the taxi market. Therefore,
there is a positive wage elasticity of daily hours of work, and an upward sloped labor supply curve.
Given the complexity of estimating independent worker’s decisions, he studies different scenarios
depending on the experience of the driver. He concludes that the traditional labor model fits
best the whole population of taxicab drivers’ behavior. [Farber 2014] also claims that the reason
why one cannot find taxis in the rain is mainly because of driving conditions rather than drivers
achieving faster their targeted goals.

3 Overview of the market

3.1 Description of the yellow taxicabs market

Up until late 2012, yellow taxicabs were the only vehicles certified to pick up passengers in New
York City. The medallion, a plate attached to a yellow taxicab, certifies its operation. It is sold by
the administration (TLC) and its price is set by them depending on the year’s fare rates, demand
for rides, costs, etc. In the 2004 - 2012 period, the average price of an independent medallion
increased by 214% and the return on investment was 19.5%. In 2014, the approximate value of a
medallion permit was USD 1’000,000. However, after the entrance of Uber, the medallions’ price
decreased significantly (See Figure 1).

There are 13,437 yellow taxicab medallions, 18,000 boro taxis, and approximately 50,000 drivers
in New York City. Yellow taxis are allowed to pick up only in Manhattan while boro taxis,
implemented in 2012, can pick up passengers anywhere but Manhattan below E96th or W110th
street. Other industries include liveries (25,000), black cars (10,000), limousines (7,000), paratran-
sit (2,000), and commuter vans (500).

The average time of a driver’s daily shift is 9.5 hours. Yellow taxicabs provide approximately
485,000 trips per day, this number is higher in holidays and dangerous weather conditions such as
hurricanes. The fare structure is

- $2.50 upon entry. $0.50 for each additional unit,
- if average speed is less than 12 miles add one more unit,
• night surcharge of $.50 after 8:00 PM and before 6:00 AM,
• peak hour weekday surcharge of $1,
• the passenger is responsible for paying all bridges and tunnel tolls,
• tips are optional.

The entry barriers require that
• each taxi should have a medallion permit, sold by the TLC,
• drivers should have special licenses,
• fares be regulated by the Taxi and Limousine Commission.

3.2 Description of the “e-hailing” market structure

Uber matches demanders with drivers using a mobile application based on the passenger and driver’s GPS location, enabling consumers to hail rides electronically from their smartphones. Uber is growing quickly, its current value is $18.2 billion (five times its value a year ago). It provides five types of services depending on the type of cars: X, XL, Black, SUV, T; the first being the most popular among consumers. Uber’s base fare is $3, $0.40 per minute or $2.15 per mile; with a minimum fare requirement of $8. Tips are set to 20% of the bill and are charged automatically to the passenger. Among the regulations, drivers should have special licenses, fares are established by the Uber company, and the drivers keep 80% of the profit and the rest is paid to the company.

4 The data

The dataset was collected by the New York City Taxi & Limousine Commission (TLC) and obtained through a Freedom of Information Law request. It consists of three datasets corresponding to all the yellow taxicab trips, Uber trips, and the boro or green-cabs trips. It contains the medallion number, driver’s license number, exact date and GPS trajectory, as well as all the information of the bill. It corresponds to a total of over 900 million taxi trips from 2010 to 2016. This data was merged with the National Weather Service hourly database.

Figure 3: Descriptive statistics of the yellow taxicab sub-market

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of trips (million)</td>
<td>698</td>
<td>169</td>
<td>177</td>
<td>179</td>
<td>173</td>
</tr>
<tr>
<td>No. of medallion licenses</td>
<td>13,437</td>
<td>13,237</td>
<td>13,237</td>
<td>13,237</td>
<td>13,437</td>
</tr>
<tr>
<td>No. of taxi drivers</td>
<td>43,135</td>
<td>39,603</td>
<td>39,399</td>
<td>40,310</td>
<td>43,063</td>
</tr>
<tr>
<td>Average trip distance</td>
<td>3.42</td>
<td>3.39</td>
<td>3.44</td>
<td>3.36</td>
<td>3.51</td>
</tr>
<tr>
<td>(miles)</td>
<td>(3.48)</td>
<td>(3.40)</td>
<td>(3.46)</td>
<td>(3.48)</td>
<td>(3.59)</td>
</tr>
<tr>
<td>Average trip duration</td>
<td>12.41</td>
<td>12.03</td>
<td>12.41</td>
<td>12.39</td>
<td>12.64</td>
</tr>
<tr>
<td>(minutes)</td>
<td>(9.09)</td>
<td>(8.77)</td>
<td>(9.02)</td>
<td>(9.06)</td>
<td>(9.35)</td>
</tr>
<tr>
<td>Average trip fare</td>
<td>10.89</td>
<td>9.99</td>
<td>10.25</td>
<td>10.96</td>
<td>12.35</td>
</tr>
<tr>
<td>(dollars)</td>
<td>(8.83)</td>
<td>(7.86)</td>
<td>(8.11)</td>
<td>(8.89)</td>
<td>(10.09)</td>
</tr>
<tr>
<td>Share of credit card payment</td>
<td>0.46</td>
<td>0.36</td>
<td>0.43</td>
<td>0.48</td>
<td>0.54</td>
</tr>
<tr>
<td>Average trip tip</td>
<td>2.39</td>
<td>2.20</td>
<td>2.30</td>
<td>2.39</td>
<td>2.61</td>
</tr>
<tr>
<td>(dollars)</td>
<td>(2.20)</td>
<td>(1.99)</td>
<td>(2.06)</td>
<td>(2.20)</td>
<td>(2.42)</td>
</tr>
</tbody>
</table>
5 Static model of a semi-regulated taxicab industry

Assume that there is a matching function that determines the number of matches $q^j_t$ in very period $t$, for all $j \in (0...J)$ submarkets, given the number of passengers waiting for a cab, $I^j_t$, and the number of vacant cabs, $V^j_t$. Let $I^j_t = \{ i : E[U^j_{it}] > E[U^j_{ikt}] \forall k \}$, where $i$ indexes passengers and $E[U^j_{it}]$ is the expected utility of consuming $j$ at time $t$. To reduce notation, we omit for now the submarket superscripts. Call the matching function $q_t = \epsilon (I^\alpha V_1^{1-\alpha})$ where $\epsilon \in (0, 1)$ represents the efficiency of the matching technology and $\alpha$ is a constant between 0 and 1. Let $m^j_t = q_t/I_t$ and $m^j_t = q_t/V_t$ be the probabilities that a waiting passenger finds a vacant cab and that a vacant cab finds a waiting passenger, respectively.

Assume that each consumer’s expected utility consists of the utility of consumption, the monetary disutility of paying for the services, and the expected disutility of the time spent waiting and traveling. Let the disutility of time spent for consumer $i$ be

$$ y_i = \frac{1}{\phi_i} \int (w + h_i)^2 f(w) dw $$

where $w \sim \text{Exp}(1/w_0)$ is the random waiting time variable for $w_0 > 0$ let the pdf be $f(w) = \frac{1}{w_0} e^{-w/w_0}$, $h_i$ is the length of a ride, and $\phi_i \sim \text{Normal}$ is a parameter representing how much $i$ values the supplier’s features in comparison to the average valuation among consumers. If $\phi_i$ is high then $i$ is willing to spend more time waiting and traveling than what the average consumer does.

The decision of each traveler can be analyzed in two ways. First, we can assume that the consumer decides on what mode of transportation to use given a randomly assigned travel time. On the other hand, we can assume that each passenger knows in advance what mode of transportation to use and decides on the length of the ride. In this paper, I follow the second thought and later on will discuss the implications of using an aggregate demand without outside option instead of one with it. Given this assumption, at time $t$ consumer $i$’s expected utility of consumption of services of type $j$ is

$$ E[U^j_{it}] = (\beta_{1i} x^j + \beta_2 \mu^j) h^j_i + \alpha (y_i - p^j_i h^j_i) - \frac{1}{\phi^j_i} \int (w^j + h^j_i)^2 f(w^j) dw^j $$

where $p^j_i$ is the price of a unit of time (or distance) of $j$ at time $t$; and $x^j$ and $\mu^j$ represent the observable and unobservable characteristics of $j$, respectively. The variable $x^j$ captures features like the availability of a rating system, the type of car, and other observables. Whereas, $\mu^j$ includes safety conditions and other features unobserved by the econometrician. Additionally, it is assumed that $\beta_{1i}$ is normally distributed with unknown mean and variance. Preferences for the unobserved characteristic are constant across individuals, and preferences for the outside good are constant across individuals conditional on income (Bajari). Preferences for the unobserved characteristic are constant across individuals because such a variable includes things like safety conditions.

Given the above conditions and assuming that the range of waiting time is $[0, 24]$ hours, the optimal choice is

$$ h^j_{it} = \frac{2 - \lambda \phi_i (\beta_{1i} x + \beta_2 \mu - \alpha p) - e^{-24\lambda}(2 + 48\lambda)}{2(e^{-24\lambda} - 1)\lambda} $$

where $\lambda = 1/w_0$, $\partial h^j_{it}/\partial p < 0$, and $\partial h^j_{it}/\partial \lambda > 0$. Recall that the matching function follows a Poisson process with parameter $m^j_t$, which represents the probability that a passenger meets an
empty taxicab, thus, \( \lambda = 1/w_0 = m_t^{l*} \).

In each period of time the passenger consumes only from the option that gives him the highest expected utility,

\[
E[U_i(h_{it}^k)] > E[U_i(h_{it}^k)] \forall k.
\]

Thus, the aggregate demand \( H_t^j = \sum_i h_{it}^j \). \( \beta_{1i} \) is normally distributed with unknown mean and variance. Since both \( \phi_i \) and \( \beta_{1i} \) are independent and normal distributed, we have to integrate over a bivariate normal distribution \( f(\phi_i, \beta_{1i}) \). Assume that \( \phi_i, \beta_{1i} \) are uncorrelated and let \( f(\phi_i, \beta_{1i}) \) be the pdf of a bivariate normal distribution of uncorrelated random variables. Call \( \phi_i \) the level of patience. Thus,

\[
H_t^j = \int \int 2 - \lambda \phi_i (\beta_{1i}x + \beta_{2i}u - \alpha p) - e^{-24\lambda}(2 + 48\lambda) \frac{\beta}{\gamma} f(\phi_i, \beta_{1i})d\phi_i d\beta_{1i}.
\]

Let \( \nu \sim \text{Exp}(\kappa) \) be a variable describing the time until a driver finds a new passenger. Given the means \( \kappa \) and \( h^* \), the proportion of time that a taxicab is occupied on average is \( \frac{k^*}{\kappa + h^*} \). Recall that \( m_t^e \) is the probability that an empty taxicab meets a passenger at time \( t \), thus asymptotically I let the expected occupancy time of a driver \( e \) of type \( j \) at time \( t \) be

\[
\frac{h_t^j}{\kappa_t + h_t^j} = m_t^e.
\]

In each period, each driver observes the period’s \( \kappa_t \) and \( h_t^j \) with which he forms a signal \( \eta \) of the real \( m_t^e \). The demand is Poisson distributed and the signal \( \eta \) at time \( t \) is also Poisson distributed. The signal is based on the time occupied and the time vacant. The driver observes each period’s average occupancy time and vacancy time. Therefore, he knows what was the average time been occupied and vacant, and so he knows the probability of finding a passenger.

Following Farber (2015), assume each drivers’ utility depends on income and leisure. Explicitly, let the expected utility of driver \( e \) be a function of the expected earnings and the disutility of work,

\[
U_{el}^j = \int p^j \cdot (L_{el}^j \cdot \eta_{el}) f(\eta) d\eta - \frac{\gamma}{1 + \iota} (L_{el})^{1+i} \int \eta_{el} \lambda_t \eta e^{-\lambda_t \eta} d\eta - \frac{\gamma}{1 + \iota} (L_{el})^{1+i}.
\]

with integration boundaries \((0,1)\) and where \( L_{el} \) is the total supply of hours in the taxicab market, \( \gamma \) is a parameter representing the disutility of work, and \( \iota \) is the price elasticity of labor supply. Also, assume that the disutility of work and the price elasticity of labor are common across drivers. Thus, the optimal labor supply is

\[
L_{el}^{j*} = \left( \frac{p(1 - e^{-\lambda_t \eta \gamma})}{\lambda_t \gamma} \right) ^{\frac{1}{\iota}}
\]

as long as \( 0 < E(U_{el}^{j*}) > E(U_{el}^{l*}) \) \( \forall k \). Heterogeneity in labor supply elasticity arises endogenously from differences in reservation wages (Karabarbounis, 2012).

The aggregate labor supply is

\[
L_t^{l*} = e \cdot \left( \frac{p(1 - e^{-\lambda_t \eta \gamma})}{\lambda_t \gamma} \right) ^{\frac{1}{\iota}}
\]

\(^2\)Let \( f(q) = m_t^e e^{-m_t^e q} \) be the Poisson probability distribution of the matching function \( q \) with rate parameter (or average number of events per interval of time \( n \)) equal to \( m_t^e \). The mathematical proof that \( m_t^e = 1/w_0 \) is given by using the maximum likelihood estimate of \( m_t^e \).

\( L(m_t^e) = \prod_{i=1}^n m_t^e e^{-m_t^e q_i} = (m_t^e)^n e^{-m_t^e \sum_{i=1}^n q_i} \)

\(^3\)Let \( h_t^{j*} \) be the vector of optimal choices of all consumers, and \( h_t^{j*} = \sum h_t^{j*} / I \) be the average length of a ride.
6 Estimation of the static model (preliminaries)

The following are the equations needed for Maximum Likelihood Estimation. Optimal individual demand:

\[ h_{jt}^* = 2 - \lambda \phi_i(\beta_1 x^j + \beta_2 u^j - \alpha p) - e^{-24\lambda}(2 + 48\lambda) \]

where \( p_j^t \) is the price of a unit of time (or distance) of \( j \) at time \( t \); and \( x^j \) and \( \mu^j \) represent the observable and unobservable characteristics of \( j \), and \( \lambda = 1/w_0 = m_i^j \) is the inverse of the average waiting time or the probability of finding a cab.

Optimal individual supply:

\[ L_{jt}^* = \left( \frac{p_j^t(1 - e^{-\lambda\eta(\lambda\eta+1)})}{\lambda\eta\gamma} \right)^{\frac{1}{\epsilon}} \]

\( L_{jt}^* \) is the labor supply of the average driver; which is a function of the price elasticity of supply \( \epsilon \), the price \( p_j^t \), the average occupancy time \( \lambda\eta \), and the disutility of work \( \gamma \). Assume that \( \epsilon > 0 \) therefore, \( \partial L/\partial p > 0, \partial L/\partial \lambda\eta < 0, \partial L/\partial \gamma < 0 \). From \( L_{jt}^* \) we get the following equation of the price.

6.1 Estimation from the optimal supply equation (preliminaries)

Given the assumption that \( \epsilon > 0 \), \( L_{jt}^* \) is a continuous and monotonic function of \( \epsilon \). Note that it is possible to get the distribution of \( \epsilon \) non-parametrically by inversion,

\[ \epsilon = \frac{\log \left( \frac{p_j^t(1 - e^{-\lambda\eta(\lambda\eta+1)})}{\lambda\eta\gamma} \right) + \log \left( \frac{1}{\gamma} \right)}{\log(L_{jt}^*)} = \frac{C}{\log(L_{jt}^*)}. \]

Let \( \hat{F}_{jklm} \) be the empirical distribution of \( L_{jt}^* \) over a particular day \( j \) of week \( k \) of month \( l \) in year \( m \). We can see that (refer to Figure 4 and Figure 5):

- there are two main effects that determine the amount of labor supply: the calendar date effect and the day of the week effect,
- in most cases, the day of the week effect is stronger that the date effect. However, this is not true for the first and last days of each month.

Among the day of the week effects, we can see that:

- Sundays are more normally distributed than the rest of the days,
- on Fridays and Saturdays most drivers work a same amount of time, making the empirical distribution of these days less “normal”,
- Tuesday, Wednesday, and Thursday follow very similar distributions,
- some Holidays, such as Christmas, increases the normality of the distributions; which means that in these days the date effect is higher than the day of the week effect,
- there’s a small effect associated to the week of the month and month of the year. In general, the normality increases as the week/month increases.
6.1.1 Generalized normal

The distributions $\hat{F}_{jklm}$ deviate from normality by kurtosis and skewness (see Figure 4 and Figure 5). Therefore, I model the density as a generalized normal with changing kurtosis and skewness parameters. This can be done using the $\text{sinh} - \text{arcsinh}$ transformation from Jones, M.C. and Pewsey A. (2009) which finds a transformation of a normal sequence that changes in both skewness and kurtosis. Let $x$ be a random variable, the transformation is defined as

$$H(x; \epsilon, \delta) = \sinh[\delta \sinh^{-1}(x) - \epsilon]$$

where $\epsilon$ is the skewness and $\delta$ the kurtosis. The density of the transformed variables is

$$f_{\epsilon,\delta} = \frac{1}{\sqrt{2\pi(1 + x^2)}} e^{-\frac{S^2_{\epsilon,\delta}(x)}{2}} \delta C_{\epsilon,\delta}(x)$$

where $C_{\epsilon,\delta}(x) = \cosh\{\delta \sinh^{-1}(x) - \epsilon\} = \{1 + S^2_{\epsilon,\delta}(x)\}^{1/2}$, and $S_{\epsilon,\delta}(x) = \Phi[H(x; \epsilon, \delta)]$ is the transformation applied to the normal CDF which produces a unimodal distribution with parameters $\epsilon$ and $\delta$. Jones and Pewsey (2009) normalized the kurtosis of the normal distribution to $\phi = f_{0,1}$. Therefore, the tail-weight decreases with increases in $\delta$. Specifically, if $\delta > 1$ the distribution is more light-tailed and if $\delta < 1$ it is more heavy-tailed. It is proven that $f_{\epsilon,\delta}$ has a better maximum log-likelihood fit than competing distributions (see Figure 6).

The task is now to choose an appropriate sequence of parameters $(\epsilon_n, \delta_n)$ for each empirical distribution $\hat{F}_{jklm}$.

**Figure 6: Fit: normal vs. sinh-arcsinh distribution**

Note: The blue curve refers to the normal distribution and the red curve is the sinh-arcsinh distribution.
6.2 Estimation from the optimal demand equation

Given the aggregate demand

\[ H_t^l = \int \int \frac{2 - \lambda \phi_i (\beta_{1i} x + \beta_2 u - \alpha_p) - e^{-24\lambda} (2 + 48\lambda)}{2(e^{-24\lambda} - 1)\lambda} f(\phi_i, \beta_{1i}) d\phi_i d\beta_{1i}, \]

assume that \( \phi_i, \beta_{1i} \) are uncorrelated and let \( f(\phi_i, \beta_{1i}) \) be the pdf of a bivariate normal distribution of uncorrelated random variables.

6.3 Estimation for the matching function and market equilibrium (preliminaries)

Given \( q_t = \epsilon (I_t^o V_t^{1-\alpha}) \) where \( I_t^o = \{ i : E[U_{it}^r] > E[U_{it}^k] \forall k \} \) is the number of passengers waiting for a cab, and \( V_t^j \) is the number of vacant cabs; the market equilibrium is defined by

\[ H_t^l = \int \int \frac{2 - \lambda \phi_i (\beta_{1i} x + \beta_2 u - \alpha_p) - e^{-24\lambda} (2 + 48\lambda)}{2(e^{-24\lambda} - 1)\lambda} f(\phi_i, \beta_{1i}) d\phi_i d\beta_{1i} \]

and

\[ p = \frac{\lambda_{\eta, \gamma} (L_{el}^{j*})^e}{(1 - e^{-\lambda_{\eta, \gamma}}) (\lambda_{\eta, \gamma} + 1)}. \]

Given that \( L_t^j = O_t^j + V_t^j \), the occupancy and vacancy time respectively; and by the clearing market equilibrium condition, \( L_t^j \cdot m_o = H_t^l \) where \( m_o \) is the average occupancy time. Thus, the market equilibrium can be rewritten as

\[ \sum_e L_{el}^{j*} \cdot m_o = \int \int \frac{2 - \lambda \phi_i (\beta_{1i} x + \beta_2 u - \alpha_p) - e^{-24\lambda} (2 + 48\lambda)}{2(e^{-24\lambda} - 1)\lambda} f(\phi_i, \beta_{1i}) d\phi_i d\beta_{1i} \]

and

\[ p = \frac{\lambda_{\eta, \gamma} \sum_e (L_{el}^{j*})^e / e}{(1 - e^{-\lambda_{\eta, \gamma}}) (\lambda_{\eta, \gamma} + 1)}. \]

A joint estimation of the market equilibrium equations involves estimating a vector of structural parameters \( \theta = (m_o, \beta_2, \alpha, \lambda, \gamma, \lambda_\eta) \) and the sufficient statistics of \( f(\phi_i, \beta_{1i}) \).

7 Dynamic model of sequential learning

Let \( w \) be the real state of the world distributed based on the search time random variable \( s \sim Exp(\kappa) \). Let \( \rho^1 \) be the common prior belief that state of the world is \( w_i \) at the beginning of time. In every period each driver decides on how much to work \( l_t \). Assume that the payoff depends on the driver’s ability to act according to the state of the world. Explicitly, the payoff is \( \pi_i > 0 \) if the action is \( l_t \) when the state is \( w_i \) and \( \pi_i = 0 \) otherwise [Bikhchandani et al., 1992].

At the end of each period a driver observes his payoff, action, vacancy time, occupied time, and mean and variance of each of this variables. Let \( \mu_t = \sum_{i=1}^T \frac{\pi_i}{\pi} \) and \( \sigma_t = Var(\pi) \); and assume that each driver uses the mean and variance of their own profits as private signals of the state of the world. These signals are the state variables, together with the driver’s history of actions \( H_t = \{ l_0, ..., l_t \} \). The realization of the private signal depends on the state of the world. Explicitly,

\[ Pr(\sigma_i | w_j) = \begin{cases} \delta & \text{if } i = j, \\ 1 - \delta / 2 & \text{otherwise}. \end{cases} \]
Let the public belief, conditional on $H^t$, that the state is $w_1$ be

$$
\rho^n = \Pr(w_1|H^n) \quad \text{and let the private belief that the state is } w_1 \text{ given } H^n \text{ and the realization of } \mu_t, \sigma_t \text{ be } f_i(\rho') = \Pr(w_1|H^t, \mu_t, \sigma_t).
$$

By Bayes’ rule

$$
f_i(\rho') = \frac{\Pr(w_1 \cap \mu_t \cap \sigma_t|H^n)}{\Pr(\sigma_t|H^n)} = \frac{\Pr(\sigma_t|H^n) \cdot \Pr(w_1|H^n) \cdot \Pr(H^n)}{\sum_{j=1}^{\infty} \Pr(\sigma_t|H^n) \cdot \Pr(w_j|H^n) \cdot \Pr(H^n)} \quad \forall i = \{1, 2, 3\} \text{ or }
$$

$$
f_1(\rho^n) = \frac{\delta \rho^n}{\delta(\rho^n) + \frac{(1-\delta)\rho^n}{2}} \quad \text{and}
$$

$$
f_{2,3}(\rho^n) = \frac{(1-\delta)\rho^n}{\delta(1-\delta)\rho^n + \frac{(1-\delta)\rho^n}{2}} = \frac{(1-\delta)\rho^n}{\delta(1-\delta)\rho^n + \frac{(1-\delta)\rho^n}{2}}.
$$

These posterior beliefs define the expected profit

$$
E[\pi(w_1, \sigma_j)] = \begin{cases} 
  f_1(\rho^n) \ast \pi & \text{if } j = 1 \\
  f_{2,3}(\rho^n) \ast \pi & \text{otherwise}.
\end{cases}
$$

Thus, the optimization process is given by

$$
V(H^t, \sigma_{t-1}) = \max_{\{H^t\}} \left[ \sum_{i=1}^{\infty} f_i(\rho') \ast \sum_{t=1}^{n} \beta^{t-1} \pi(w_{it}, \sigma_{it}, l_t) \right]
$$

where $w_{t+1} \sim G(w'|w_t)$.

8 Preliminary reduced form results

See Table 1.

9 Results of the dynamic Bayesian estimation (not ready yet)
### Table 1: All shifts

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>Probit</th>
<th>Logit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent variable:</strong></td>
<td>Survival status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time of the day</td>
<td>0.002***</td>
<td>0.019***</td>
<td>0.043***</td>
</tr>
<tr>
<td></td>
<td>(0.00003)</td>
<td>(0.0003)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Weekday dummy</td>
<td>−0.013***</td>
<td>−0.227***</td>
<td>−0.554***</td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
<td>(0.006)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Total fare so far</td>
<td>−0.001***</td>
<td>−0.005***</td>
<td>−0.010***</td>
</tr>
<tr>
<td></td>
<td>(0.00001)</td>
<td>(0.0001)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>Total time so far</td>
<td>0.00003***</td>
<td>−0.001***</td>
<td>−0.001***</td>
</tr>
<tr>
<td></td>
<td>(0.00001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
</tr>
<tr>
<td>Weather</td>
<td>0.001**</td>
<td>0.015**</td>
<td>0.024*</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.007)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.007***</td>
<td>2.638***</td>
<td>4.940***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.011)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Observations</td>
<td>999,387</td>
<td>999,387</td>
<td>999,387</td>
</tr>
<tr>
<td>R²</td>
<td>0.069</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.069</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td></td>
<td>−142,150,600</td>
<td>−144,140,300</td>
</tr>
<tr>
<td>Akaike Inf. Crit.</td>
<td></td>
<td>284,313,100</td>
<td>288,292,500</td>
</tr>
<tr>
<td>Residual Std. Error</td>
<td>0.192 (df = 999381)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F Statistic</td>
<td>14,889.790*** (df = 5; 999381)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** *p<0.1; **p<0.05; ***p<0.01
Reference


B Petrongolo, and Christopher Pissarides. Looking into the black box: A survey of the matching function. *Journal of Economic literature*

Frechette, Guillaume R and Lizzeri, Alessandro and Salz, Tobias Frictions in a Competitive, Regulated Market Evidence from Taxis.


Figure 4: Daily labor supply distribution
Example: 03-2010