Frictions in a Competitive, Regulated Market: Evidence from Taxis.

Guillaume R. Fréchette, Alessandro Lizzeri, and Tobias Salz

Online Appendix

A Additional Figures

Figure 15 shows the fraction of medallions of each type that stop at each hour of the day. The graph shows, especially for the day shift, that fleet medallions are more likely to stop around the typical shift transition times, with more stops at 2PM, 3PM, and 4PM.

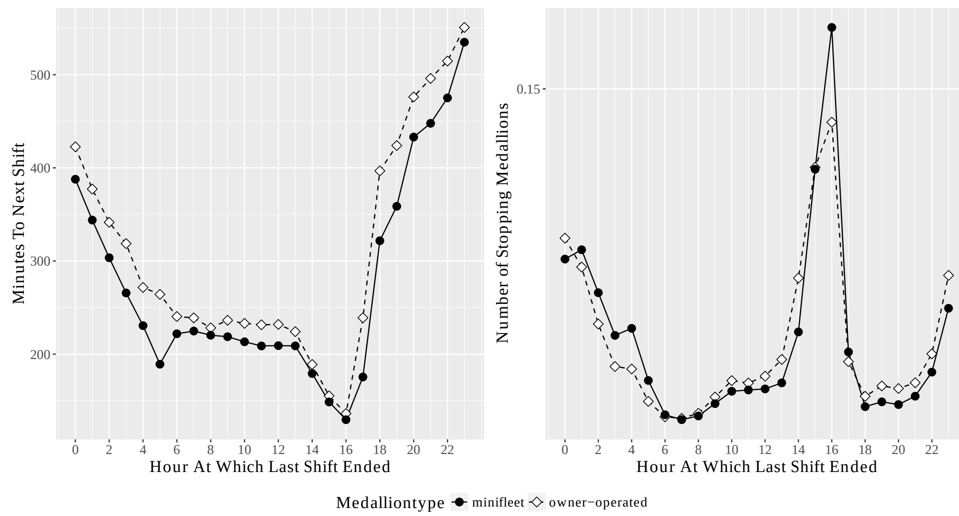


Figure 15: Time Medallion is Unutilized Conditional on Hour of Drop-Off.

Figure 16 shows the histogram of constrained transitions for each hour. We define a constrained transition as one in which the change between the last drop-off by the quitting driver and the pickup by the next starting driver lasts less than an hour. We have picked the transition hours so that the vast majority of medallions fits into one of these patterns. In the case of day-shifts, 8% of transitions occur after the hour that we call constraint and 2.5% of transitions two hours after that. In the case of night-shifts it is 5% and 1% respectively.

Figure 17 shows the fraction of rides served by the dispatch platform in the segmentation counterfactual in which half the drivers are committed to the dispatch platform and the remaining half is committed to the search platform. The dispatch platform serves more than half the rides at every hour of the day.

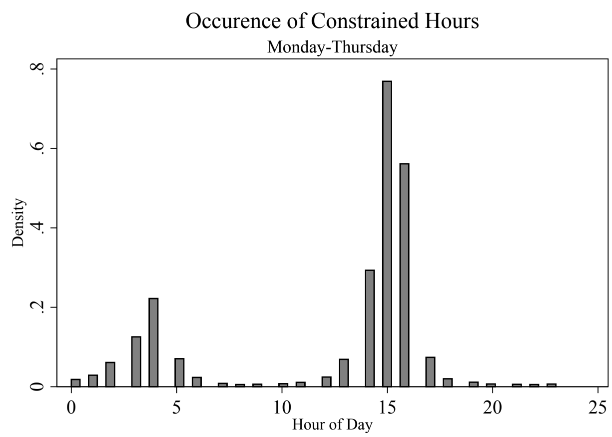


Figure 16: Prevalence of Constrained Hours Throughout the Day

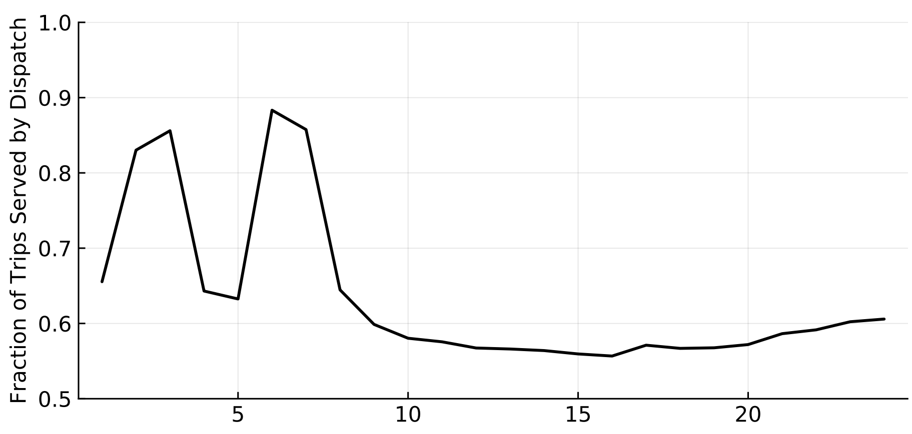


Figure 17: Fraction of Rides Served by Dispatch

B Multiplicity

A look at the inter-temporal patterns of the data shows that it follows a remarkably stable pattern. This clearly suggests that equilibrium multiplicity is not an issue in the time dimension. To illustrate this we plot the same weekdays (for those weekdays that we use in the data) on one plot (Figure 18 and Figure 19), which allows a comparison of intra-daily demand and supply patterns. We therefore believe that a single equilibrium in the data is a reasonable assumption. Even across weekdays, the numbers at different hours are remarkably similar (one of the Mondays is a holiday).

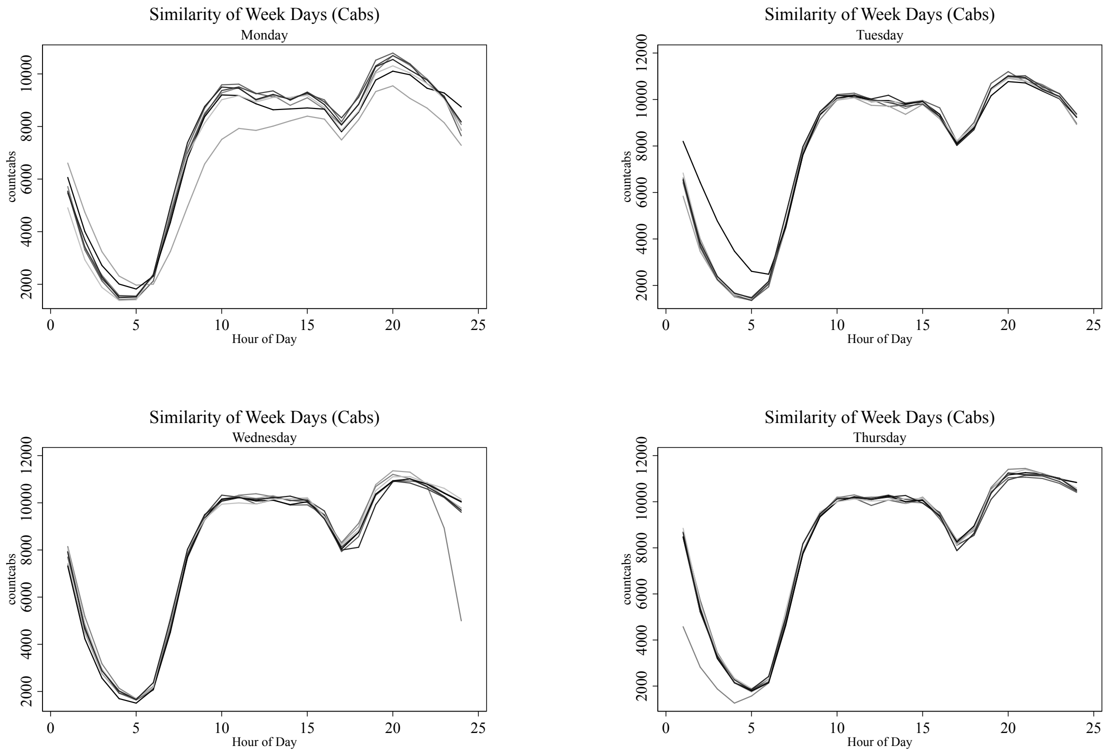


Figure 18: Activity per Hour on Different Mondays and Tuesdays in our Sample, Cabs

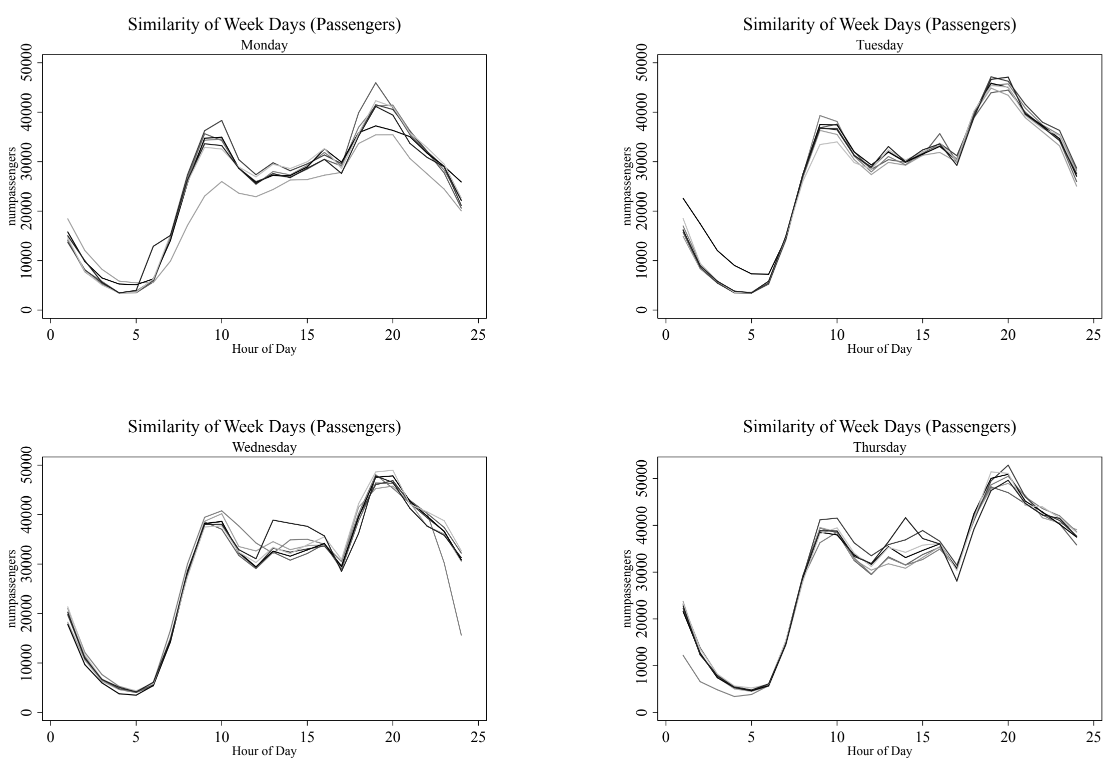


Figure 19: Activity per Hour on Different Weekdays in our Sample, Passengers

C Full Table of Summary Statistics

Table 6: Summary Statistics by Hour, Mean and (Standard Deviation)

Hours	Cabs	Passenger	Search time (minutes)	Wait time (minutes)	MPH (Manhattan)	MPH (other)
0	9188.2 (1311.6)	28849.74 (6536.2)	9.3 (1.7)	1.24 (.31)	15.13 (.907)	17.62 (1.05)
1	6997.4 (1273.6)	18201.71 (3401.0)	12.3 (1.0)	1.03 (.20)	16.6 (1.09)	18.44 (1.55)
2	4351.2 (904.2)	10481.33 (2307.3)	14.4 (1.1)	1.14 (.19)	17.58 (.77)	18.05 (1.07)
3	2744.3 (608.9)	6611.52 (1448.7)	15.6 (.86)	1.5 (.26)	18.24 (.59)	17.47 (1.03)
4	1875.7 (435.4)	4533.09 (1208.4)	16.4 (1.05)	1.94 (.24)	18.62 (.48)	17.21 (.75)
5	1652.1 (248.6)	4224.4 (831.1)	15.1 (.99)	2.00 (.14)	20.49 (.65)	18.28 (.63)
6	2208.0 (132.4)	6056.1 (1344.2)	13.2 (2.0)	1.95 (.24)	21.08 (.74)	21.19 (1.25)
7	4699.1 (348.2)	14592.1 (1015.5)	10.1 (.74)	1.18 (.17)	17.51 (.60)	19.22 (1.04)
8	7541.5 (606.6)	27263.6 (2187.9)	7.5 (.60)	1.53 (.18)	13.86 (.77)	15.05 (1.07)
9	9081.8 (618.8)	36598.5 (3289.6)	5.8 (.55)	2.29 (.22)	10.89 (.79)	12.70 (1.28)
10	9824.1 (561.8)	36818.3 (3028.7)	6.2 (.53)	2.48 (.23)	10.03 (.76)	12.65 (1.30)
11	9903.3 (528.3)	31471.9 (2649.3)	8.2 (.69)	2.00 (.24)	10.17 (.82)	13.06 (1.58)
12	9722.8 (570.9)	28986.3 (2548.5)	9.1 (.77)	1.83 (.23)	10.25 (.80)	13.14 (1.23)
13	9775.4 (579.1)	31866.6 (3153.1)	7.9 (.81)	2.08 (.30)	10.03 (.82)	12.99 (1.26)
14	9650.0 (548.6)	31281.1 (3677.4)	8.0 (1.0)	2.05 (.36)	10.27 (.88)	13.34 (1.2)
15	9745.5 (475.2)	32456.2 (2929.8)	7.4 (.79)	2.23 (.35)	10.31 (.83)	13.55 (1.16)
16	9204.7 (345.0)	33786.9 (2233.8)	6.2 (.55)	2.71 (.35)	10.16 (.85)	12.99 (1.14)
17	8072.0 (207.8)	29813.4 (1011.3)	5.9 (.29)	2.88 (.34)	10.86 (.87)	13.24 (.83)
18	8726.0 (296.2)	39035.8 (2479.2)	4.6 (.4)	2.94 (.39)	10.73 (.76)	13.01 (.81)
19	10320.8 (315.0)	45458.0 (3697.0)	4.9 (.63)	2.65 (.39)	10.28 (.73)	12.74 (.84)
20	10872.3 (403.6)	45170.5 (4402.3)	5.6 (.7)	2.12 (.40)	11.10 (.81)	13.90 (.97)
21	10767.6 (526.9)	40399.2 (4159.9)	6.7 (.92)	1.64 (.27)	12.68 (.90)	15.80 (1.04)
22	10457.3 (643.0)	37413.6 (4455.7)	7.3 (1.0)	1.49 (.25)	13.84 (.81)	16.89 (.83)
23	10027 (787.3)	34763.1 (4989.8)	7.7 (1.1)	1.44 (.26)	14.47 (.75)	17.36 (.77)

D Details on Simulation

The goal of the simulation is to obtain a mapping of the number of waiting passengers and searching cabs within an hour to the wait time and search time of those passengers and cabs. The mapping is used to infer the number of waiting passengers from observed number of active cabs and their search time. Wait and search time are also influenced by other exogenous factors, which therefore need to be arguments of the matching function. These factors are the speed, mph_t, at which the traffic flows, and the average trip length, $miles_t$, requested by passengers. Note that the average trip length determines how long a taxi is utilized for each unit of demand, and therefore shifts effective supply. The longer the trip length the higher the utilization. Table 6 provides an overview of taxi search time, the number of taxis, as well as the recovered number of passengers and their wait time.

The baseline simulation is performed under the assumption that cabs search randomly for passengers. The search is performed on an idealized map of Manhattan. subsubsection 5.1.1 provides a schematic of the grid we use for the simu-

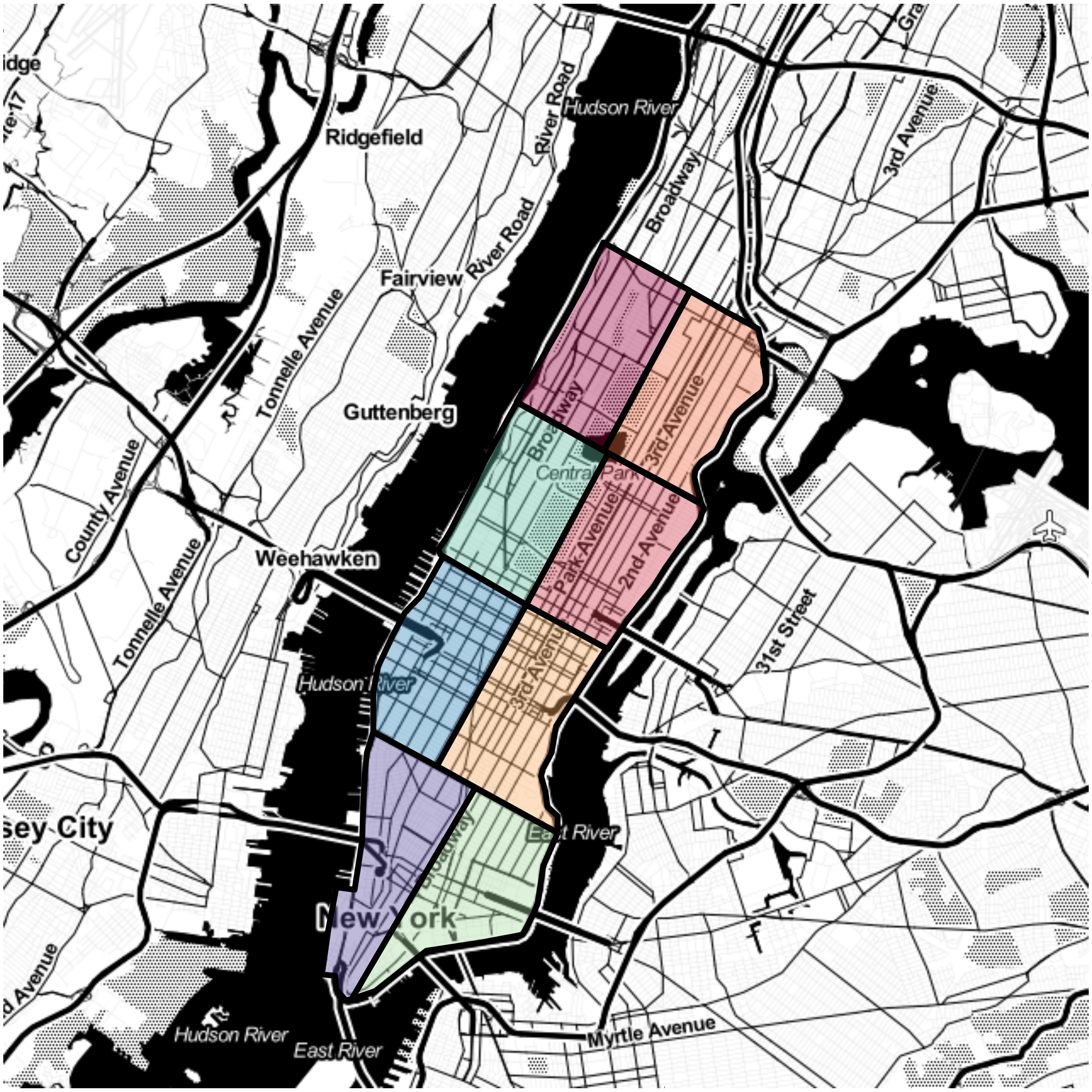


Figure 20: Division of Manhattan

lation. In line with the topography of Manhattan, we require the area to be four times longer in the north-south direction (y_t) than wide in the east-west-direction (x_t). Cabs move on nodes that are $1/20$ mile segments apart from each other, which is based on the average block length in the north-south direction. In the north-south direction they can turn at each node whereas in the east-west direction they can only turn at every fourth node. subsubsection 5.1.1 highlights nodes on which cabs can turn as gray. This map corresponds to the block structure of Manhattan, where a block is approximately $1/20$ of miles long in the north-south and $4/20$ of a mile wide in the east-west direction. Under the random-search assumption, cabs take random turns at nodes with equal probability weight on each permissible direction. However, we assume they never turn back in the direction from which they were coming (i.e., no U-turns).

Because we only model the Manhattan market (below 128th Street), our grid corresponds to an area of 16 square miles. Appendix D shows the modeled part on the map and its division into the eight equally sized different areas for which we separately compute the pickup and drop-off probabilities. Correspondingly, our grid is divided into eight equal parts, which correspond to those areas. Note however, that we map those areas onto our rectangular grid and do, therefore, not precisely model the actual street grid and shape of Manhattan.

Each node on the grid is a possible passenger location. For each hourly simulation, $\frac{d_t}{6}$ passengers are placed in 10-minute intervals randomly on the map. Those $\frac{d_t}{6}$ are divided up and placed in proportion to the corresponding (observed) pickup probabilities on the eight areas on the grid. Within those areas passengers appear with equal probability on each node.

If a cab hits a node with a passenger a match occurs, implying no additional frictions on a node, and so the number of matches is the minimum of the number of passengers and the numbers of taxis on the node, which corresponds to the assumption of a Leontieff matching function on each node. Once the match takes place the cab is taken of the grid for $60 \cdot \frac{miles_t}{mph_t}$ minutes, that is, the average measured delivery time from the data, after which is has delivered the passenger and is again placed randomly on the map with a random travel direction. Cabs reappear in locations on the grid in proportion to observed drop-off locations of the eight areas (Appendix D).

The full algorithm is described in pseudo-code below. It takes the following inputs: the number of cabs, c, the number of passengers, d, the trip length, $miles$, and the trip speed, mph. A unit of time in the algorithm is scaled so that it always represents the time it takes a cab to travel from one node to the next because a smaller time unit is unnecessary. Passengers are added to the map for one hour in 10-minute intervals. Because nodes are spaced $1/20$ of a mile apart, the last time passengers are added to the map is at $\bar{t} = 20 \cdot \text{mph}$. The set of times at which new passengers arrive is given by $\{\bar{t}/6 \cdot k | k = 1, ..., 6\}$. The following additional variables are used to describe the algorithm: $npick$ refers to the number of matches that have already taken place; $deliverytime_i$, to the remaining delivery time of taxi i; $searchtime_i$ to the time that taxi i has spent searching since the last delivery, *total_searchtime* refers to the total time taxis have spent searching for passengers; and *total_waittime* to the total wait time that passengers have been waiting.

while *npick* < *d* **do**
  $t = t + 1$ (time units represent travel time from one node to the next, scales with mph);
  if $t \in \{\bar{t}/6 \cdot k | k = 1, ..., 6\}$ **then**
    add $d/6$ passengers to random nodes on map, stratified by eight areas;
  end
  for $i = 1 : c$ **do**
    if $deliverytime_i = 0$ *(cab i is not occupied)* **then**
      update the node of cab i. Cabs only take turns on gray nodes (subsubsection 5.1.1) and do not make u-turns. All feasible travel directions are chosen with equal probability;
      if *new node of cab i has a passenger* **then**
        cab becomes occupied, set $deliverytime_i$ to $20 \cdot miles$, and add $searchtime_i$ to $total\_searchtime$;
      else
        $searchtime_i = searchtime_i + 1$
      end
    else
      $deliverytime_i$ = $deliverytime_i$ - 1;
      if $deliverytime_i$ == 0 **then**
        place cab in random area on map according to observed drop-off probabilities (all nodes within area equal probability), give cab random feasible travel direction;
      end
    end
  end
  Add one to $total\_waittime$ for each passenger that is on the map;
  for $j = 1 : p$ **do**
    Increase counter for all passengers on the map; remove passengers that have been waiting for 20 minutes.
  end
end
Result: Use $total\_waittime$ and $total\_searchtime$ to compute the average search time for taxis and average wait time for passengers.

In the dispatcher simulation, we assume that, as soon as a cab has delivered a passenger, it is matched with the closest passenger available. We also assume that neither the driver nor the passenger has an option to cancel this match for another match option. For example, a waiting passenger, who is promised to a cab, may encounter another (previously unavailable) cab earlier than the promised cab. The option to cancel might in some instances be beneficial to a market side because our search for the optimal match is only over the currently available cabs and passengers and does not take into account cabs and passengers that will soon appear somewhere close on the map.

Performing the simulations for each point in the domain of the matching function is not possible. We therefore perform them for the Cartesian product of the sets: $c \in \{500, 1000, ..., 17000\}$, $d \in \{3000, 6000, ..., 75000\}$, $miles \in \{1, 2, ..., 7\}$, $mph \in \{4, 8, ..., 24\}$. To obtain the search-time and wait time for other points we interpolate linearly between the grid points.

D.1 Details on Heterogeneity across Areas.

In the current baseline simulation we allow for heterogeneity by dividing the city map into eight different areas and we match pickup and dropoff probabilities by area. To test how sensitive our results are to this subdivision, we recompute the simulation, including inferred passenger waiting times and the demand function, for two alternative subdivisions of the map. In one case, we double the number of areas to 16; in the other, we only divide the map into four areas. We perform this simulation in the same way in which we do it for the baseline, but instead using drop-off and pick-up probabilities that are matched at different (higher and lower respectively) levels of aggregation. We find aggregate numbers that look almost identical when we move to 16 areas, whereas there are some differences when we reduce the number of areas to 4. This can be seen in table Table 7. Since the natural concern is about robustness with respect to finer subdivisions, we feel comforted by this, and also feel justified in keeping our 8 area baseline. We also re-estimate demand both for the 4-area division and for the 16-area division. The demand estimates for all specifications do not change notably. For 4 areas, the numbers are as follows. Mean wait-time: 1.88; mean number of passengers: 22234; elasticity: -1.08. For eight areas (current specification) we have the following. Mean wait-time: 2.47; mean number of passengers: 23262; elasticity: -1.225. For 16 areas the numbers are as follows. Mean wait-time: 2.61; mean number of passengers: 23551; elasticity: -1.21. Lastly, taking the eight areas that we currently use, search time is remarkably stable across locations (see Table Table 8).

Table 7: Robustness to Number of Areas (from our simulated matching function)

Passengers	Cabs	Search Time			Wait Time			Matches		
		s^4	s^8	s^{16}	w^4	w^8	w^{16}	m^4	m^8	m^{16}
15000.0	4000.0	4.83	5.58	5.77	4.05	4.8	4.93	13750.0	13201.0	13056.0
15000.0	6000.0	11.53	11.84	11.96	1.32	1.74	1.83	14747.0	14599.0	14534.0
15000.0	9000.0	23.34	23.48	23.51	0.35	0.46	0.47	14876.0	14826.0	14813.0
25000.0	4000.0	1.41	2.18	2.3	8.24	8.68	8.72	16998.0	16145.0	16013.0
25000.0	6000.0	3.51	4.42	4.59	4.09	4.9	4.99	22190.0	21014.0	20791.0
25000.0	9000.0	9.29	9.73	9.86	1.32	1.7	1.8	24204.0	23798.0	23675.0
35000.0	4000.0	0.75	1.17	1.29	10.7	10.79	10.84	17670.0	17251.0	17111.0
35000.0	6000.0	1.46	2.35	2.48	6.83	7.41	7.47	25272.0	23818.0	23606.0
35000.0	9000.0	4.2	5.05	5.19	2.89	3.6	3.68	31672.0	30234.0	29981.0

Note: superscript indicates number of areas.

Table 8: Taxi Search Time Across Areas (data)

Area	Median	Mean	Median Day	Mean Day
1	6.55	11.64	4.37	10.3
2	4.37	9.15	4.37	7.7
3	4.37	8.6	4.37	7.7
4	4.37	9.21	4.37	8.04
5	4.37	8.34	4.37	7.8
6	4.37	8.79	4.37	8.29
7	4.37	9.68	4.37	9.4
8	6.55	11.42	4.37	11.12

D.2 Robustness Check on Maximal Wait Time

Table 9: Robustness to Different Maximal Wait Time (from our simulated matching function)

		Search Time			Wait Time			Matches		
Passengers	Cabs	s^{15}	s^{20}	s^{25}	w^{15}	w^{20}	w^{25}	m^{15}	m^{20}	m^{25}
15000	4000	8.48	10.01	11.5	2.07	2.02	1.52	9809.04	10046.88	10072.36
25000	4000	10.97	13.5	15.97	1.03	1.0	0.99	10281.26	10386.78	10377.66
35000	4000	11.97	14.96	17.5	0.53	0.5	0.5	10420.64	10470.94	10453.68
15000	6000	4.99	5.12	5.68	5.25	5.25	4.95	13397.34	13579.86	13706.74
25000	6000	8.97	10.5	12.15	1.29	1.03	1.0	15174.32	15412.9	15457.64
35000	6000	10.46	12.52	15.01	1.03	0.63	0.5	15514.7	15650.06	15643.38
15000	9000	2.76	2.67	2.89	34.52	35.04	34.61	14529.56	14463.64	14666.5
25000	9000	5.47	5.55	6.29	2.9	2.97	2.65	21524.88	21686.48	22057.66
35000	9000	7.98	9.02	11.0	1.07	1.03	1.0	22894.04	23181.42	23315.9

Note: superscript indicates maximal wait time.

D.3 Robustness to Definition of Breaks

Our definition of a shift is a sequence of trips that have not been interrupted for longer than 300 minutes (we follow Farber's shift definition). However, within a shift, cabs sometimes take breaks. In the paper, define a break every gap between a drop-off and the next pickup that is longer than 45 minutes (note that we restrict the sample to trips within Manhattan). Of course, the value of mean search-time does depend to some extent on this choice, but it turns out not to be very sensitive to it. The average search time under our current break definition is 7.5 minutes. If instead we chose a 50-minute cutoff, it would be 7.73, whereas with a 40-minute cutoff, it would be 7.22. An important question is how our demand recovery is affected by the break-time cutoff. In this matter, a countervailing effect arises. Note first that a change in the break cutoff changes both the search time that we use to recover demand, but also the effective number of searching taxis. As an example, if we lower the cutoff, search time is lower, but so is effective supply since more cabs are on a break according to the new cutoff. To understand how these two changes affect the inferred number of passengers, it is helpful to consider how each change separately affects the inversion of the matching function. First, decreasing supply, ceteris paribus, leads us to infer a smaller number of passengers. Lower search time, on the other hand, implies a larger number of passengers, all else equal. These two changes counteract each other. We now report our computation of how different the inferred number of passengers is as we change the cutoff. Moving the cutoff from 45 to 50 minutes, an 11% change, leads to an inferred number of passengers that is 0.77% higher. A reduction in the cutoff to 40 minutes, leads us to infer that the number of passengers is 0.89% lower. Lastly, even if we consider a more drastic change, a cut-off of 30 minutes, a reduction of one third, we would infer 3.8% fewer passengers. Hence, the total effect on the implied number of passengers from varying the definition of a break within reasonable values is quite small. Furthermore, the changes in our counter-factuals should be even smaller because all scenarios would be similarly affected by any such change.

E Details on Supply Side Estimation.

As defined in the text, let $\boldsymbol{x}_{it}$ denote the observable part of the state (the hours on a shift, the hour of the day as well as the medallion-invariant characteristics). Let $p(\boldsymbol{x}_{it})$ be the theoretical probability that an active medallion/driver i stops at time point t, and let $q(\boldsymbol{x}_{it})$ be the probability that an inactive medallion/driver i starts at t. Correspondingly, let d^A be the indicator that is equal to one if an active driver stops and d^I be an indicator that an inactive driver starts. Using this notation, we maximize a constrained log-likelihood that we formulate as an MPEC problem. MPEC does not perform any intermediate computations, such as value function iterations, to compute the objective function. It instead treats these objects as parameters. Therefore, the solver maximizes both over the parameters of interest θ and an additional set of parameters δ. The parameter vector δ consists of all $p(\mathbf{x}_{it})$, $q(\mathbf{x}_{it})$, $\mathbb{E}_\epsilon[V(\mathbf{x}_{i(t+1)}, \epsilon_{i(t+1)})]$ for $\mathbf{x}_{i(t+1)} \in \mathbf{X}$. In other words, δ consists of expected values and choice probabilities for each point in the observable state space. With this notation in place we can express the maximization problem as follows:

$$\begin{aligned} \min_{\theta, p(\mathbf{x}_{it}), q(\mathbf{x}_{it}), \mathbb{E}_\epsilon[V(\mathbf{x}_{i(t+1)}, \epsilon_{i(t+1)})]} \sum_{j \in J} \sum_{t \in T_j} d_{it}^A \cdot log(p(\mathbf{x}_{it})) + (1 - d_{it}^A) \\ \cdot (log(1 - p(\mathbf{x}_{it}))) + d_{it}^I \cdot log(q(\mathbf{x}_{it})) + (1 - d_{it}^I) \cdot (log(1 - q(\mathbf{x}_{it})))) \end{aligned} \tag{4}$$

subject to

$$\begin{aligned} \mathbb{E}_\epsilon[V(\mathbf{x}_{i(t+1)}, \epsilon_{i(t+1)})] = \sigma_\epsilon \cdot log\Bigg(exp\left(\frac{1}{\sigma_\epsilon}\right) \\ + exp\left(\frac{\pi_{h_t} - C_{z_i}(l_{it}) - f(h_t, k_i) + \mathbb{E}_{\boldsymbol{\epsilon}_{i(t+1)}}[V(\boldsymbol{x}_{i(t+1)}, \boldsymbol{\epsilon}_{i(t+1)})]}{\sigma_\epsilon} \right) \Bigg) \\ + \gamma * \sigma_\epsilon \quad \forall \boldsymbol{x}_{it} \in \boldsymbol{X} \end{aligned} \tag{5}$$

$$p(\boldsymbol{x}_{it}) = \frac{exp\left(\frac{1}{\sigma_v}\right)}{exp\left(\frac{1}{\sigma_v}\right) + exp\left(\frac{\pi_t - C_{z_i,h_t}(l_{it}) - f(h_t,k_i) + \mathbb{E}_{\boldsymbol{\epsilon}_{i(t+1)}}[V(\boldsymbol{x}_{i(t+1)},\boldsymbol{\epsilon}_{i(t+1)})]}{\sigma_v}\right)} \forall \;\; \boldsymbol{x}_{it} \in \boldsymbol{X} \tag{6}$$

$$q(\boldsymbol{x}_{it}) = \frac{exp\left(\frac{\mathbb{E}_{\boldsymbol{\epsilon}_{i(t+1)}}[V(\boldsymbol{x}_{i(t+1)},\boldsymbol{\epsilon}_{i(t+1)})] - r_{h_t}}{\sigma_v}\right)}{exp\left(\frac{\mathbb{E}_{\boldsymbol{\epsilon}_{i(t+1)}}[V(\boldsymbol{x}_{i(t+1)},\boldsymbol{\epsilon}_{i(t+1)})] - r_{h_t}}{\sigma_v}\right) + exp\left(\frac{\mu_{h_{t+1}}}{\sigma_v}\right)} \;\forall \boldsymbol{x}_{it} \in \boldsymbol{X} \tag{7}$$

The constraint given by Equation 5 ensures the starting and stopping probabilities obey the intertemporal optimality conditions imposed by the value functions. The log-formula is the closed-form expression for the expectation of the maximum over the two choices of stopping and continuing, which integrates out the T1EV unobserved valuations. Equation 6 and Equation 7 are again the closed form expressions for the choice probabilities under extreme-value assumption.

F Impact of weather on Demand.

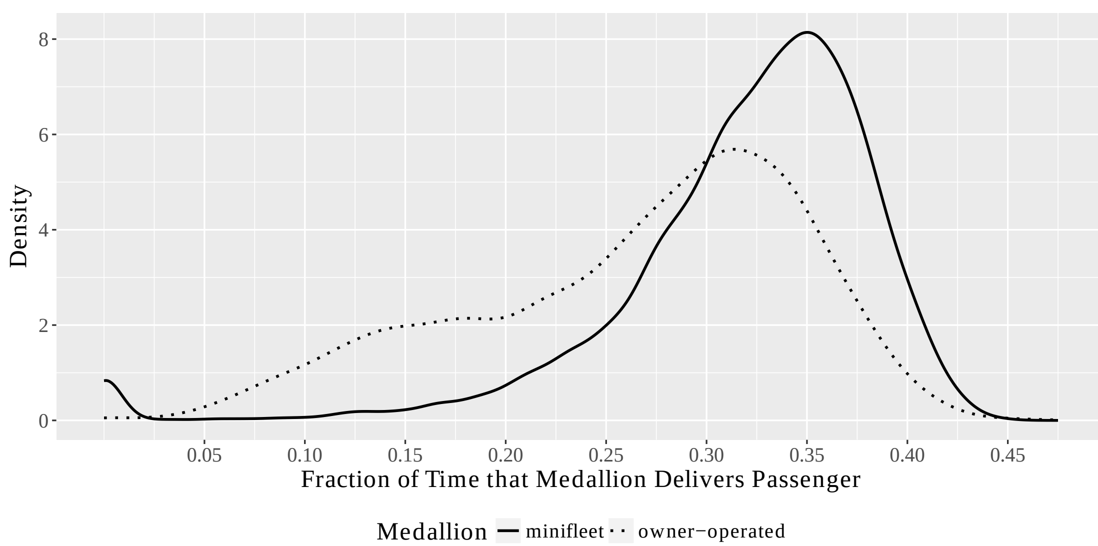
One of the important demand shocks for taxi rides is weather patterns, most notably, rainfall. To check whether our demand recovery picks up such patterns, we have merged hourly rainfall data. We then regress both log-wait-times and log-recovered-demand on a dummy for whether rainfall occurred in this hour. As the table below shows, our recovered demand and wait-times are highly correlated with rainfall. Estimates from these regressions suggest that in an hour with rainfall, our recovered demand is about 28% higher and wait times about 37% higher.

Table 10: Demand Validation with Weather Data.

	(1)	(2)	(3)	(4)	(5)	(6)
	$log(d_t)$	$log(d_t)$	$log(d_t)$	$log(w_t)$	$log(w_t)$	$log(w_t)$
Rainfall Dummy	0.292**	0.281**	0.281**	0.275**	0.371**	0.372**
	(0.0140)	(0.00839)	(0.00817)	(0.0108)	(0.00829)	(0.00800)
Hour FE	No	Yes	Yes	No	Yes	Yes
Day of Week FE	No	No	Yes	No	No	Yes

Note: $+ \ p < 0.10, * \ p < 0.05, ** \ p < 0.01$. The dependent variables are the log of wait time and demand. The rainfall dummy is an indicator whether is has rained in this hour.

G Medallion Ownership and Utilization



Notes: These densities are based on medallion-level observations, where each observation is the fraction of time this medallion spends delivering a passenger out of the total time we observe these medallions. Note that the rest of the time the medallion could either be searching for a passenger or be idle and not on a shift at all. The densities show a stark differences between owner-operated and minifleet medallions. The lower tail of low utilization is much thicker for owner-operated medallions.

Figure 21: Density of Utilization Separated by Medallions

Appendix G shows the cross-sectional distribution of the fractions of time a medallion spends delivering a passenger out of the total time that we observe a medallion. The distribution for owner-operated cabs displays a much thicker left tail of low utilization rates and is overall more dispersed.[78]

The left panel of Appendix A in Appendix C shows the length of time a medallion is *inactive* conditional on the stopping time of the last shift. Because most day shifts start around 5AM and most night shifts around 5PM, the time of non-utilization is minimized for stops that happen right around these hours, whereas a stop at any other time causes the medallion to be *stranded* for a longer time period. We see that minifleets typically return a medallion to activity faster after each drop-off. This difference is particularly large after the common night shift starting times (6PM and later). This in turn suggests that minifleets have access to a larger pool of potential drivers, making it easier for them to find a replacement for someone who does not show up at the normal transition time. In the structural model, we allow for a different set of parameters for minifleets and owner-operated medallions to capture these differences. The right panel of Appendix A in Appendix C shows the number of shifts that end conditional on the hour. We see that minifleet medallions have a more regular pattern, with most day shifts ending at 4PM. This pattern is also reflected in subsection 4.2, which shows a stronger supply decrease for minifleets before the evening shift relative to owner-operated medallions.

H Details on the Computation of Counterfactuals

Define the following six steps as **Block1(i)** for iteration i.

1. *For each hour, simulate from the observed empirical distributions of speed of traffic flow and the length of requested trips, under current guess c_h^i and d_h^i to determine the search time for taxis s_h^i, $h \in \{0, \ldots, 23\}$ under $g(.)$.*

2. *Simulate drivers earnings π_h^i, $h \in \{0, \ldots, 23\}$ from the ratios of passenger delivery time over delivery and search time (computed in step 2) and rate earned per minute of driving. Simulate new expected number of passengers d_h^i, $h \in \{0, \ldots, 23\}$ from the wait times w_h^i and the estimated demand function $d(.)$.*

3. *Compute the optimal starting and stopping probabilities $p^i(\mathbf{x}, \pi; \theta)$, $q^i(\mathbf{x}; \theta)$ under the new earnings (computed in step 3).*

4. *Use $p^i(\mathbf{x}, \pi; \theta)$ and $q^i(\mathbf{x}; \theta)$ to simulate a new law of motion for drivers c_h^i, $h \in \{0, \ldots, 23\}$. For each medallion type (z,k), we simulate 30 medallions, where each medallion starts inactive at 12PM, and iterate forward for 48 hours. Across these 30 medallions, we then compute the fraction of times the medallion has been active*

[78]The observed differences might be due to the fact that minifleets enable a more efficient utilization; another plausible argument is a selection-effect.

in this hour (using only the last 24 hours) and multiply this amount by the total number of medallions.[79]

5. *Compute* $sumsq_1 = \sum_h (c_h^i - c_h^{(i-1)})^2$

Define the following four steps as **Block2(i)** for iteration i.

1. *For each hour, simulate values from the observed empirical distributions of speed of traffic flow and the length of requested trips, under* c_h^i *,* $h \in \{0, \ldots, 23\}$*, to determine the wait times* w_h^i*,* $h \in \{0, \ldots, 23\}$ *for passengers.*

2. *Simulate the new number of passengers* d_h^i *,* $h \in \{0, \ldots, 23\}$ *from the waiting times* w_h^i *,* $h \in \{0, \ldots, 23\}$ *and the estimated demand function* $d(.)$*.*

3. *Compute* $sumsq_2 = \sum_h (d_i^h - d_h^{i-1})^2$

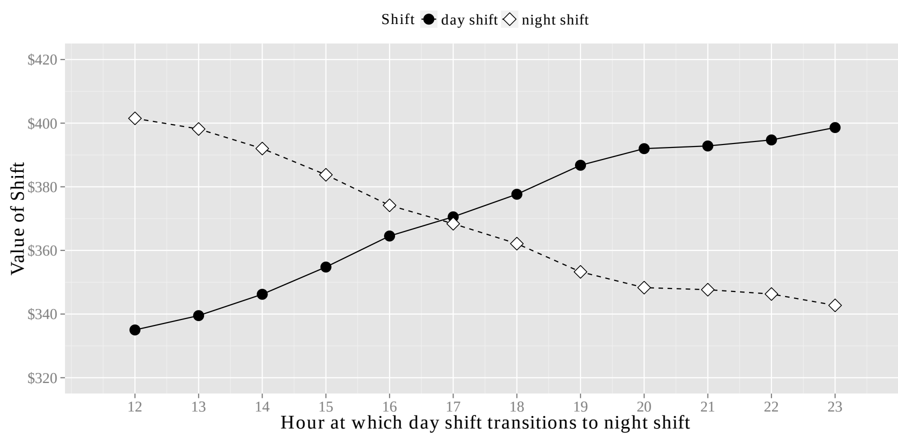
Using these definitions, the algorithm can be described as follows:

```
while STOP ≠ 1 do
    Compute sumsq1 using Block1(i).
    if sumsq1 < tol then
        STOP=1
    else
        while sumsq1 > tol do
            Block1(i) (in each iteration guess new vector of cabs with some
             non-linear solver)
        end
    end
    Compute sumsq2 using Block2(i).
    if sumsq2 < tol then
        STOP=1
    else
        while sumsq2 > tol do
            Block2(i) (in each iteration guess new vector of passengers with
             some non-linear solver)
        end
    end
    i = i + 1
end
```

[79]We have also experimented with different numbers in this step, for example simulating each medallion for more than 48 hours or increasing the number of simulations. For the final counterfactual results, these alternatives do not make a large difference.

I Shift-Transition and the Witching Hour

We now argue that the timing of the shift transition is such a supply side driven shifter. The dip in the number of active taxis in the later afternoon hours is clearly visible in subsection 4.2. The right-hand panel of Appendix A shows that this dip is mostly due to taxis ending their shift. The left-hand panel shows that the transition to the next driver is quite long. Together these suggest that this common shift transition is responsible for a prolonged reduction in the number of cabs between 4 and 6 o'clock. Multiple reasons could explain why most shifts are from 5AM to 5PM and 5PM to 5AM, but the data (and the rules) suggests some key factors.



Notes: This graph shows the average earnings that would accrue to the night-shift and day-shift driver for each possible division of the day. The x-axis shows the end-hour of the day shift and the start-hour of the night shift. Because these earnings are a function of the current equilibrium of the market, they have to be understood as the shift-earnings that one deviating medallion would give to day-shift and night-shift drivers. The graph shows that earnings are almost equal at 5PM, the prevailing division for most medallions.

Figure 22: Earnings of Day and Night Shift for Different Split Times

First, the rules are such that minifleets can only lease a medallion for exactly two shifts per day: they must operate a medallion for at least two shifts of nine hours and the lease must be on a per-day or per-shift basis.[80] Second, a cap is placed on the lease price for both day and night shifts. Anecdotal evidence from the TLC and individuals in the industry suggests that these lease caps were binding during our sample period. Given these rules, minifleets may try to equate the earning potentials for the day and night shifts, as a way to ensure they will get a similar number of drivers willing to drive each shifts. A similar argument applies for owner-drivers that want to ensure they always find a driver for the

[80] See section 58-21(c) in TLC (2011).

second shift, which they do not drive themselves. Appendix I shows the earnings for night and day-shifts under different hypothetical shift divisions. The x-axis shows each potential division-point, that is, each point at which a day shift could end and a night shift start. The y-axis reports the earnings for the day-shift (black dots) and night-shift (white diamonds).[81] As can be seen, the 5-5 division creates two shifts with similar earnings potential. Combined with the above observation, the difference in rate caps for day and night shifts may reflect different disutility from working at night. Hence, requiring two shifts and imposing a binding cap on the rates results in most medallions having shifts that start and end at the same time. Because transitions do not happen instantaneously, this correlated stopping therefore leads to a negative supply shock at a time of high demand during the evening rush hour.

J Table of Choice Probabilities

Table 11: Stopping probabilities by hour and hours on a shift.

	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$	$l=6$	$l=7$	$l=8$	$l=9$	$l=10$	$l=11$	$l=12$	$l=13$	$l=14$
$h=0$	.18	.11	.13	.14	.14	.14	.15	.21	.32	.37	.43	.43	.41	.4
$h=1$	.15	.18	.13	.15	.15	.16	.16	.19	.24	.35	.4	.42	.42	.37
$h=2$	.07	.11	.22	.15	.17	.18	.17	.19	.23	.29	.44	.46	.39	.36
$h=3$	.03	.04	.15	.32	.19	.22	.23	.24	.28	.34	.44	.58	.5	.42
$h=4$	.02	.03	.08	.23	.53	.36	.4	.45	.51	.59	.73	.81	.78	.65
$h=5$	.01	.01	.03	.07	.18	.42	.39	.46	.54	.61	.7	.79	.79	.64
$h=6$	.01	0	.01	.02	.05	.12	.27	.32	.4	.53	.62	.71	.63	.52
$h=7$	.01	0	0	.01	.01	.03	.06	.15	.21	.32	.42	.46	.41	.29
$h=8$	.01	0	0	.01	.01	.01	.03	.07	.12	.2	.29	.27	.24	.19
$h=9$	.02	.01	.01	.01	.01	.02	.03	.04	.11	.17	.21	.23	.14	.13
$h=10$	.03	.01	.01	.01	.01	.02	.02	.04	.09	.16	.25	.25	.12	.1
$h=11$	.04	.01	.02	.02	.02	.02	.02	.04	.09	.15	.22	.2	.16	.08
$h=12$	.04	.02	.02	.02	.02	.02	.03	.04	.07	.17	.29	.28	.16	.1
$h=13$	.05	.02	.03	.04	.04	.04	.04	.06	.08	.15	.29	.34	.22	.14
$h=14$	.04	.03	.05	.07	.09	.1	.1	.13	.17	.22	.34	.38	.27	.14
$h=15$	.03	.03	.06	.1	.15	.19	.24	.29	.37	.45	.5	.58	.46	.3
$h=16$	.01	.02	.05	.09	.16	.24	.33	.44	.58	.68	.74	.73	.55	.4
$h=17$	.01	.01	.02	.04	.06	.09	.12	.16	.22	.27	.3	.27	.17	.1
$h=18$	.01	0	.01	.02	.04	.05	.08	.09	.12	.14	.15	.16	.14	.1
$h=19$	.01	.01	.01	.01	.03	.05	.07	.09	.1	.14	.15	.16	.17	.15
$h=20$	.02	.01	.01	.01	.02	.04	.07	.09	.11	.13	.16	.18	.18	.17
$h=21$	.04	.02	.02	.01	.02	.03	.06	.09	.12	.14	.17	.19	.19	.19
$h=22$	.06	.03	.03	.03	.03	.03	.06	.1	.14	.18	.2	.23	.22	.23
$h=23$	.1	.05	.05	.05	.05	.05	.06	.11	.16	.2	.22	.23	.23	.23

[81] Clearly this comparison ignores any equilibrium effects of changing the shift structure. The graph can therefore be understood as the earnings that one deviating medallions could have under the current system.

Table 12: Starting probabilities by hour.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
.01	.01	.02	.05	.11	.22	.29	.26	.2	.17	.14	.12	.12	.14	.2	.43	.73	.47	.28	.15	.09	.05	.03	.03

K Other Counterfactual Numbers

Table 13: Counterfactual Results for other Segmentations

	hourly active cabs	**hourly demand**	**passenger waittime**	**matches per day**	**taxi searchtime**	**hourly taxi revenues**	**consumer surplus (minutes)**	**medallion revenue**
Segmented(2-Search)	4580.0	14163.0	3.92	12543.18	10.44	33.85	1.2	1.09
Segmented(2-Search) (perc. change)	-32.49	-39.96	52.03	-39.98	37.0	-16.14	-27.83	-29.25
Segmented(2-Dispatch)	6602.0	21316.0	2.84	20158.8	7.84	39.8	1.65	1.48
Segmented(2-Dispatch) (perc. change)	-2.7	-9.65	10.2	-3.54	2.87	-1.41	-0.86	-3.22
Fleet	7217.0	24926.0	2.46	22115.01	7.72	40.26	1.71	1.53
Fleet (perc. change)	6.37	5.65	-4.62	5.82	1.29	-0.26	2.9	-0.1

Note: The changes are the mean over all 24 hours of the day. The wait time and search time averages are weighted by the number of trips, and the hourly driver profits are weighted by the number of active drivers across hours. **PE** means partial equilibrium and holds demand fixed to give a sense of how much the demand expansion changes counterfactual results. The percentage changes $\Delta\%$ are the changes in the means over all hours compared to the baseline. Consumer surplus is computed under the assumption that the demand function is truncated above the maximal waiting time observed in the data. The reason is that, for our parameter specifications, consumer surplus would be infinite if we integrated over all wait times. This issue results from the assumption of constant elasticity, log-linear demand. A similar issues arises, for example, in Wolak (1994), who also truncates the demand distribution. Note, however, that except for the limit case, the absolute difference in consumer surplus will be well defined and the same, no matter how high we choose the truncation point to be.