Yellow Cabs as Red Corpuscles

Timothy H. Savage
Berkeley Research Group
Visiting Scientist
Center for Urban Science & Progress
Email: timothyhsavagephd@gmail.com

Huy T. Vo
Center for Urban Science & Progress
New York Univeristy
Email: huy.vo@nyu.edu

***Abstract*—Data obtained from New York City's Taxi & Limousine Commission highlights the importance of yellow cabs in the city's transportation infrastructure. In 2011, for example, there were nearly 180 million cab rides. In this paper, we provide detailed visualizations of these numerous journeys. Using a random sample, we also develop several relevant economic metrics. For example, we find that nearly half of all rides were taken by a single passenger traveling short distances. In contrast, out-of-pocket costs faced by passengers averaged less than $6 per mile, with 90% paying less than $8 per mile, even when the social value of the city's roads is highest. From a city planner's perspective, the current fare structure for yellow cabs may not be optimal. Finally, we propose a statistical method to analyze dynamic flow patterns of yellow-cab rides that can ultimately be extended to a multivariate model.**

I. Introduction

In Manhattan, yellow taxicabs (cabs) seem ubiquitous. They appear in almost all of the numerous movies filmed in New York City (NYC), and an award-winning situation comedy was devoted to the cabbie profession. In NYC, the Taxi & Limousine Commission (TLC) licenses all yellow cabs through a medallion system, which allows it to regulate fares and standards. TLC reports that there were 13,237 licensed yellow cabs on NYC's roads in 2011.[1] Each yellow cab typically operates on a shift schedule with two or more different drivers per day, which permits a single cab to be in near-continuous daily operation.

Data were obtained from the TLC on all yellow cab rides in 2009, 2011, and 2012. The focus of this study is 2011, during which there were nearly 180 million yellow cab rides. When a cab picks up passengers curb side, its meter records detailed information about the journey that is ultimately relayed to the TLC. This information includes the time, longitude, and latitude of the pick-up and drop-off location, as well as the number of passengers, the fare (including tolls and taxes), the tip (if a credit card is used), as well as a trip's (piecewise linear) distance and duration.

We have developed animated visualizations and summary statistics of these journeys to demonstrate the important role that yellow cabs play in NYC's transportation infrastructure. We examine the spatial and temporal characteristics of these journeys. We also develop useful economic metrics to examine whether, from a planning perspective, the current fare schedule is optimal. Finally, we propose a statistical approach to examine the dynamic patterns of yellow cab journeys.

[1]Ted Mann, "New Cab Plan Curbs Hybrids," The Wall Street Journal, September 19, 2012.

II. Related Work

There is a considerable academic literature on transportation economics, and a comprehensive literature review is outside of the scope of this workshop paper. In brief, as with other areas of economics, transportation economics deals with the allocation of scarce resources in a transportation network. People and goods flow through transportation networks at certain speeds along defined paths, and a single trip may involve the use of multiple modes of transportation. Modeling the demand for transportation amounts to the thought experiment of an individual making a choice among a set of discrete and finite alternatives. An individual will choose the option that yields the highest benefit. Not surprisingly, this modeling approach is called discrete choice theory, the econometrics of which was developed and advanced by economist Daniel McFadden. More information can be found in [1].

Regarding the analysis of taxi data, several efforts have been done in recommendation systems [2]–[4], modeling [5]–[8], and visual analytics [9], [10]. Only [10], however, examined yellow cab data for NYC, focusing on visual exploratory tasks. In this paper, we look at the data from a hypothetical policy planner's perspective using various metrics to begin to understand better the role yellow cabs play in NYC's transportation infrastructure. The visualizations here supplement those in [10].

III. Millions of Journeys

A. One Day in the Life of a Yellow Cab

There were nearly 180 million yellow cab journeys throughout NYC in 2011. With only about 13,000 yellow cabs on the road, this implies that each cab is making nearly 14,000 journeys each year. Figure 1 maps out a day in the life of a single yellow cab. On April 30th, 2011, this yellow cab made curb-side pickups of 58 total passengers who paid $650 in total fares. In Figure 1(a), hired trips from the yellow cab are highlighted in red. Because our data does not include GPS tracking for each trip, we constructed plausible routes by utilizing fastest paths based on speed limits. Figure 1(b) shows a similar visualization but also includes journeys when the taxi is roaming free without a passenger. Hired trips were marked with blue lines, while empty trips were marked with red lines. This combined visualization expresses how much time and distance a taxi spent looking for passengers compared to fare-generating time and distance. Animations are available

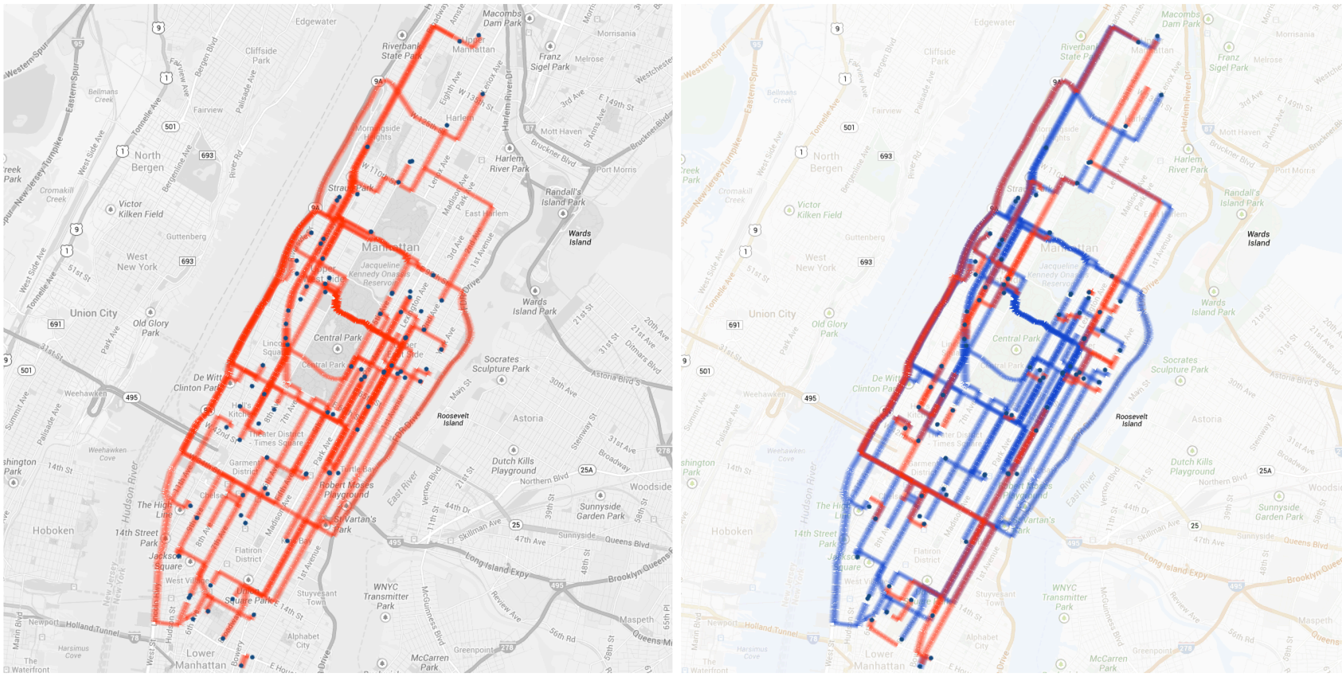

Fig. 1. Road coverage by a single cab on 04/30/2011: (a) when hired; and (b) when both empty (red) and full (blue).

online for Figure 1(a) [2] and Figure 1(b) [3].

B. Spatial and Temporal Characteristics

For each curb-side pick up and drop off, a cab's meter records the time in seconds elapsed since January 1, 1970 and the location in longitude and latitude. Longitudes and latitudes can be mapped to any level of geographic aggregation, and we have geocoded this information to the neighborhood map used by the TLC.[4] This neighborhood map would likely be recognized by any New Yorker.

For our analysis, we generated a simple random one-in-24 sample of the all of the 2011 cab rides. While future extensions will examine all of the data obtained from the TLC, the results presented here are unbiased summary statistics. We have eliminated the limited number of journeys that begin or end in the Bronx, Staten Island, or any of the three regional airports (JFK, LaGuardia, and Newark). We also eliminated journeys of exceedingly long duration (60 minutes) or distance (60 miles), as well as journeys that record no passengers. Finally, we aggregated all locations in Brooklyn and Queens to those respective boroughs, as well as a limited number of locations in Manhattan into a category we call "Other in Manhattan" or "OIM".

The first observation we note is that almost all yellow cab rides begin and end in Manhattan, making it an obvious focus for this workshop paper. Within Manhattan, a large number of cab rides begin and end in Midtown (MID), the Upper East Side (UES), or the Upper West Side (UWS). This concentration of journeys is hardly surprising. First, population densities in the UES and UWS exceed 40,000 persons per square kilometer, more than double the highest population densities in either Brooklyn or Queens.[5] Only the East Village has a population density that approaches this magnitude. Second, there is a considerable concentration of business establishments in MID.[6]

Figure 2 displays a histogram of pick up locations (based on the geocoding discussed above). For those neighborhoods not already discussed, the legend is: BR is Brooklyn; CH is Chelsea; EV is East Village; FD is Financial District; GD is Garment District; GR is Gramercy; GV is Greenwich Village; HK is Hell's Kitchen; LES is Lower East Side; QN is Queens; SO is Soho; and WV is West Village. Fewer than 3% of rides begin in either Brooklyn or Queens. Nearly 40% of rides begin in MID, UES, or UWS. Note that MID contains Grand Central Terminal, but does not include Pennsylvania Station, which is located in CH (based on the TLC map). Nearly 7.5% of rides begin in CH.

Figure 3 displays a histogram of drop off locations using the same naming convention. The results are largely consistent with those seen in Figure 2 with two main exceptions. The share of cab rides that end in either Brooklyn or Queens (regardless of their origin) rises markedly. The share of rides that end in Queens rises to 2.24%, while that in Brooklyn rises to 3.31%. Again, we will address this point in greater detail in the section on dynamic modeling, but a bit of New York fokelore appears wrong. Cabs do go to Brooklyn (and Queens)–they just do so infrequently.

We now examine the temporal characteristics of yellow cab rides. Figure 4 displays a histogram of cab rides by month,

[2] http://vgc.poly.edu/files/hvo/cab_hired.mp4

[3] http://vgc.poly.edu/files/hvo/cab_hired_empty.mp4

[4] http://www.nyc.gov/html/tlc/downloads/pdf/passenger_info_map.pdf

[5] 2010 U.S. Census.

[6] 2011 County Business Patterns.

while Figure 5 and 6 display them by day of week and by time of day (24-hour clock), respectively.

Yellow cab rides are fairly uniform throughout the year. (We have confirmed that 2011 is consistent with other years.) They reach a slight peak in March and a slight trough in August. In January 2011, the average air temperature at JFK Airport was about 30°F.[7] Peak temperatures were not reached until July and August, with averages of 79°F and 75°F respectively. Those months also recorded maximum temperatures that exceed 90°F. As a result, it appears that yellow cab rides are largely invariant to temperature. We continue to explore the role of rainfall, but we do not believe that general weather conditions are a strong driver of cab rides (so to speak). In 2011, nearly 51 million tourists visited NYC.[8] We are exploring whether seasonal variation in tourist patterns help to explain the monthly variation seen in Figure 4. We think this implausible largely because tourism in NYC typically peaks in July and August.

Turning to daily characteristics, Figure 5, in contrast with Figure 4, shows a clear pattern. The distribution of yellow cab rides rises steadily throughout the work week. About 12.5% of rides occurs on Sunday, which is lower than one would expect if journeys were uniform over the week (indicated by the horizontal line). This share rises by 2.5 points to its Friday peak. In terms of the total daily trips taken, this difference amounts to approximately 12,000 yellow cab rides. Again, the temporal pattern here is clear: yellow cab rides are concentrated late in the week, with nearly one-third of all cab rides taken on Friday and Saturday.

Figure 6 displays a histogram of journeys by hour, which demonstrates the most temporal variability. The share of journeys declines steadily from midnight to 5AM. As the morning rush hour begins at 6AM, however, the volume of journeys rises markedly. This level is largely sustained throughout the day, albeit with a small dip at 4PM, which may be associated with the shift change in yellow cab drivers. Journey volume again ramps up with the evening rush hour, which TLC defines to be 4 to 6PM. Peak volume, however, is reached at 7PM, declining steadily thereafter, only to resume the cycle. The sustained volume throughout the day, together with a 7PM peak, suggests that the current TLC definition of rush hours may not be optimal.

C. Journey Characteristics

We now turn to importance characteristics of yellow cab journeys in NYC, particularly in Manhattan. There is a striking but simple characterization: a single passenger, slowly traveling a short distance. (This bit of New York fokelore is correct.) Figure 7 displays a histogram of the number of passengers for a given ride. Figure 8 displays a kernel density plot (or probability density function) of the distance in miles of a given journey.

From Figure 7, we find that nearly 70% of yellow cab journeys are occupied by a single passenger. This share spikes between 6 to 8AM, hours the TLC designates to be morning rush hour. During these hours, nearly 80% of journeys are

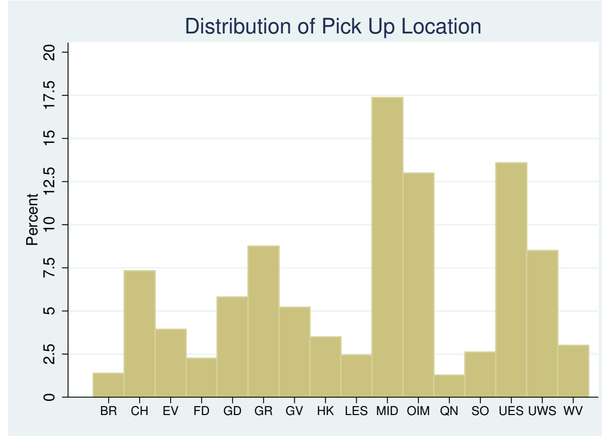

Fig. 2. Distribution of pickup location

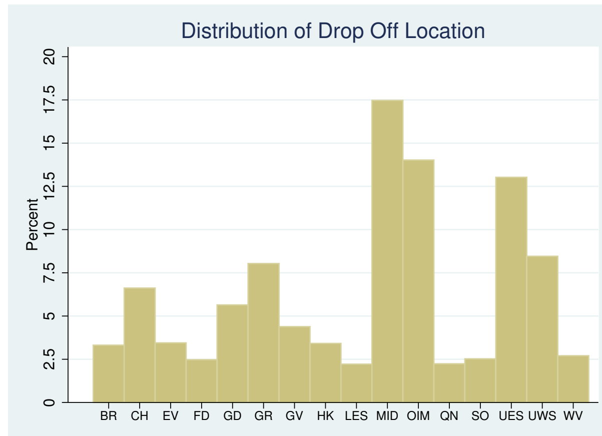

Fig. 3. Distribution of drop off location

occupied by only a single passenger (another fact that suggests variability is not driven by tourism). The high share of single passengers implies that, on a daily basis, cabs carry far fewer than one million passengers. In contrast, New York's Metropolitan Transit Authority (MTA) reports that, in 2011, NYC's subway system carried nearly 4.5 million passengers per day, while MTA buses running throughout NYC carried over 1.8 million passengers.[9]

Figure 8 displays a density plot of the distance of cab rides, omitting the density beyond 10 miles. (Note that this is a probability density function, not a cumulative density function from which one could readily read percentiles.) The mode of the distribution falls almost exactly at one mile. The average distance is 2.1 miles, and slightly over 70% of journeys are 2.5 miles or less. Simply put, this is a highly skewed but important characterization. Yellow cab journeys are overwhelmingly short distance.

Figure 9 displays a density plot of the average speed in miles per hours (MPH) for a cab journey. Recall that a yellow cab meter measures piecewise linear distance, which is not necessarily the same thing as the distance between two locations. (For example, a yellow cab traveling one mile east along 42nd Street and turning north on 8th Avenue to

[7] National Oceanic and Atmospheric Administration (NOAA).
[8] http://www.nycandcompany.org/research/nyc-statistics-page
[9] http://www.mta.info/nyct/facts/ridership/

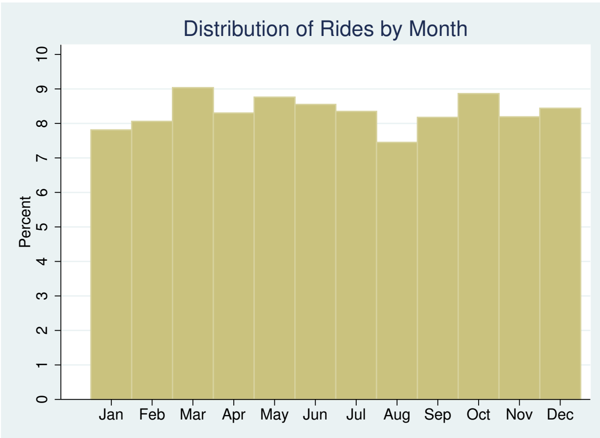

Fig. 4. Taxi rides by month

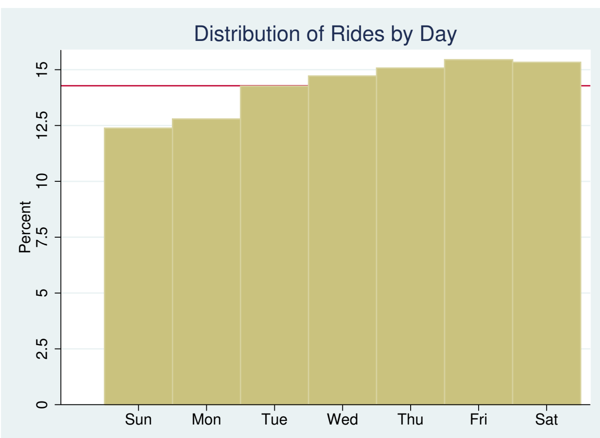

Fig. 5. Taxi rides by day of week

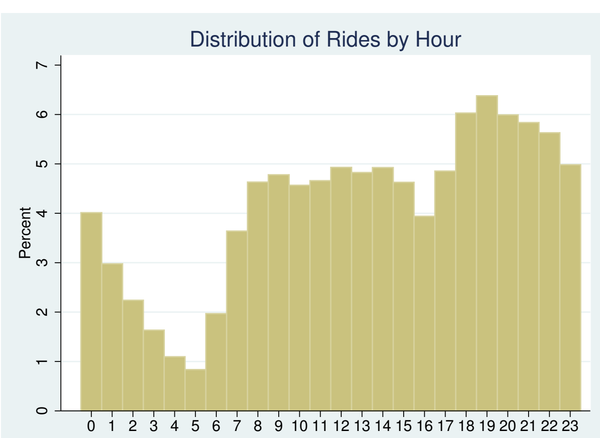

Fig. 6. Taxi rides by time of day

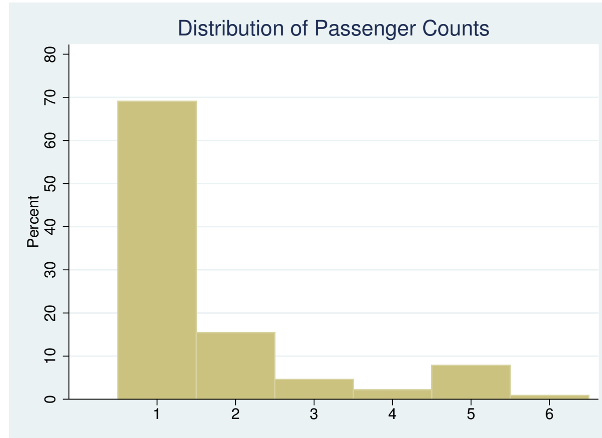

Fig. 7. Number of passengers for a given ride

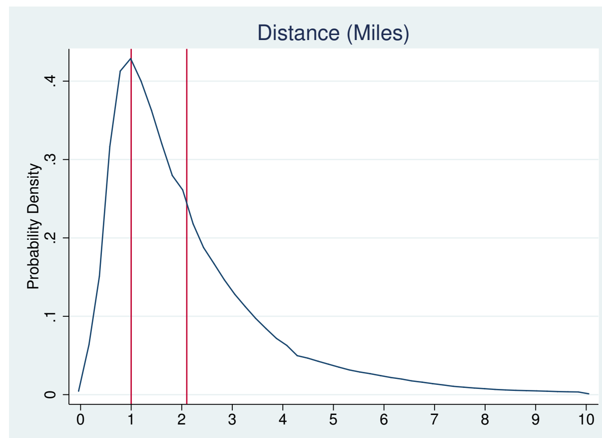

Fig. 8. Distance in miles for a given ride

travel one mile would record a journey of two miles. In linear distance, these two locations are only 1.4 miles apart.) The speed limit throughout NYC is 30MPH, unless otherwise posted.[10] Figure 9 shows a modal speed of about 10MPH. On average, yellow cabs move at about 12MPH, while about 76% of journeys move at 15MPH or slower, half the posted speed limit throughout NYC.

There is another way to examine this characteristic, which is the time in minutes necessary to travel one mile or time per mile. Figure 10 displays a density plot of time per mile. The modal value is about four minutes to travel one mile. The average is 5.9 minutes with about a quarter of the journeys requiring at least seven minutes to travel one mile. A pedestrian walking “briskly” would take about 15 minutes to walk one mile. In other words, the typical single passenger who is traveling a short distance is doing so slowly.

D. A City Planner’s Perspective

An extended discussion of public economics and externalities is outside of the scope of this workshop paper. Any introductory textbook on public economics addresses the topic at length. For example, Rosen and Gayer define an externality as

[10] http://www.nyc.gov/html/dot/html/motorist/knowthespeedlimit.shtml

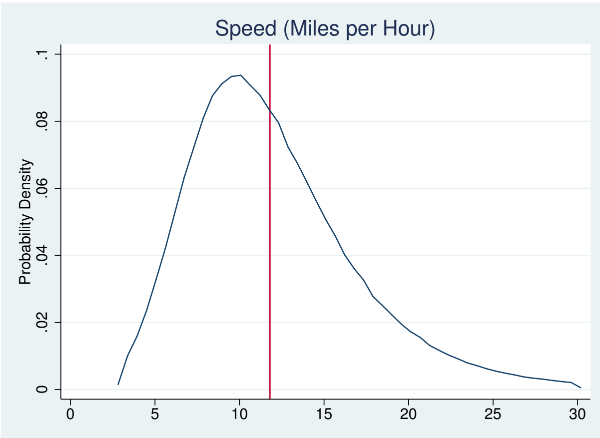

Fig. 9. Density plot of speed in miles per hour for taxi rides

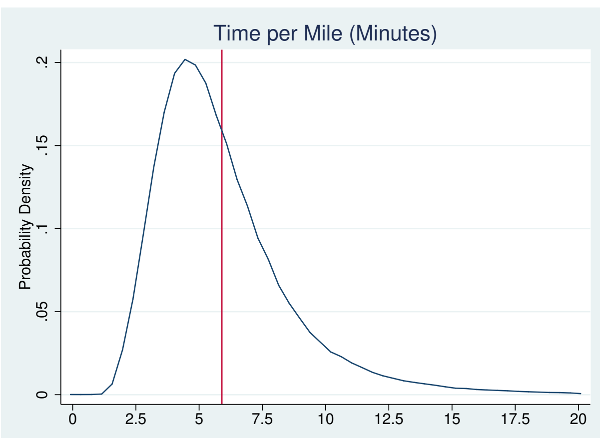

Fig. 10. Density plot of time per mile

occuring "when the activity of one entity affects the welfare of another in a way that is outside the market. Unlike effects that are transmitted through market prices, externalities adversely affect economic efficiency." [11] Conceptually, an externality drives a wedge between the value or cost of goods to an individual and the value or cost of goods to a group or society, which gives rise to the economic inefficiency. Externalities often arise with public goods, such as roads. For example, in transport economics, road congestion is a negative externality.

We see such congestion when we examine time per mile at 5PM. Figure 11 displays the same density plot as Figure 10, but limited to those journeys that commence between 5 and 6PM. The entire density shifts to the right, meaning it takes longer to travel one mile at 5PM than it does for the day as a whole. The average time to travel one mile rises by nearly 10%. Based the density plots, one should note that this increase is occuring for all trips at this time as the density has shifted right in its entirety. Some famous New Yorkers have complained about yellow cab drivers changing their shifts around 5PM, when "the number of active taxicabs on the streets falls by nearly 20 percent compared with an hour before."[11] Having fewer yellow cabs on the road at this time, while an inconvenience

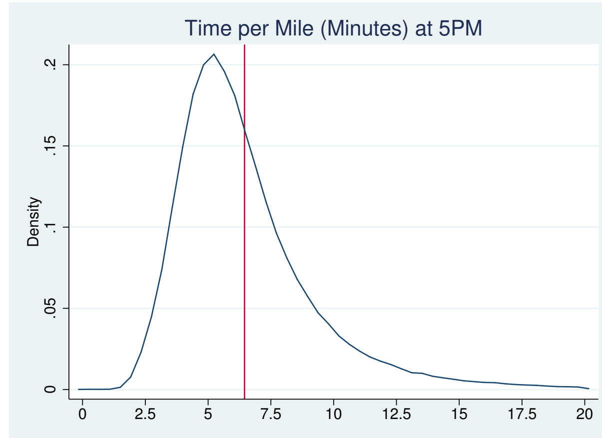

Fig. 11. Density plot of time per mile at 5PM

for a former NYC mayor and his law partner, likely reduces traffic congestion when the social value of NYC's roadways is its highest. This is particularly true when one considers the array of transportation alternatives available in NYC.

An obvious question arises at this point: what are (largely-single) passengers paying to go short distances slowly? Yellow cab pricing is a two part tariff: a fixed (or flat) fee of $3.00 ($2.50 entry and $0.50 MTA tax) and a variable fee of $0.50 for each additional "unit," where a unit is determined by a combination of speed and distance. An additional dollar is added to the fixed fee on weekdays between 4 and 8PM, as are any bridge or tunnel tolls.[12] We calculate the total cost of a journey faced by passengers: the sum of fare, taxes, and tolls. We exclude the tip from this calculation as this amount is chosen by the passenger rather than being a fee schedule faced by the passenger.[13] We then normalize this amount by the total distance of the journey to obtain the passenger cost per mile.

Figure 12 displays the density plot of the cost per mile overall and between 5 and 6PM. Both density plots have pronounced modes at about $5, which is more pronounced at 5PM. Both plots have several "spikes" that are associated with mass points in total costs induced by tolls. Passengers are paying $5.26 on average to travel one mile, which rises by about 5% to $5.51 per mile for journeys between 5 and 6PM. While both densities are somewhat skewed, 90% of journeys cost a passenger $7.77 or less per mile overall and $8.08 or less per mile between 5 and 6PM.

To date, we have found no solid statistics on the value to New Yorkers of a mile of their roadways. Nevertheless, it is certainly true that this value rises, perhaps markedly, during morning and evening rush hours. On the other hand, yellow cabs cannot carry the passenger volume for a given amount of roadway space relative to, say, MTA buses. In 2011, the base fare for a single ride on local MTA buses was $2.25. The base fare for interborough express buses, which cover distances well in excess of one mile, was $5.50. We note that the average

[11] Michael Grynbaum, "Where Do All the Cabs Go in the Late Afternoon?" The New York Times, January 11, 2011

[12] http://www.nyc.gov/html/tlc/html/passenger/taxicab_rate.shtml

[13] Because tips are electronically recorded when they are charged to a credit cards, we do not observe the entire distribution of tips. For those that are recorded, the median tip value was about 20% of the fare.

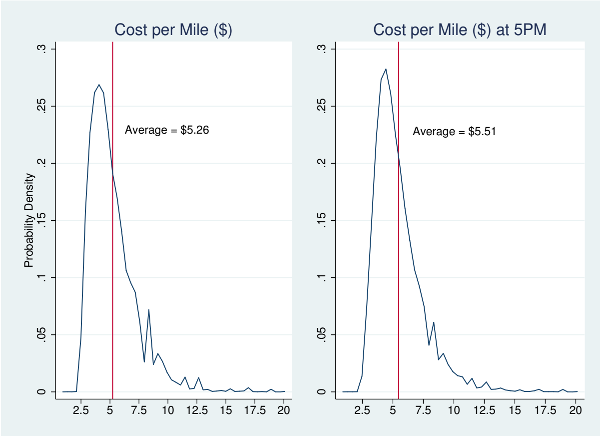

Fig. 12. Cost per mile for taxi rides in overall and between 5pm and 6pm

passenger cost per mile during evening rush hours is $5.51.

IV. Dynamic Transitions

The visualizations in Figure 1 and discussion above present an obvious method of analysis: given a cab ride begins in a particular location, what is the likelihood that it ends in the same or some other location? If such dynamic transitions do not depend the next to last journey, we call such likelihoods a first-order Markov process. For example, if the likelihood a cab goes from Midtown to the Upper East Side does not depend on the fact that it previously picked up and dropped off in the Upper West Side, then the process can be called a first-order Markov. (Given the structure of these data, it is trivial to use higher-order Markov processes. For our purposes here, the first-order approach is satisfactory.) The use of Markov processes has become widespread in dynamic stochastic modeling. For example, its use is ubiquitous in macroeconomics (dynamic stochastic general equilibrium), finance (dynamic asset pricing), and areas of microeconomics (dynamic programming).

Formally, a Markov process $\{X_t\}$ is a stochastic process with the property that, given the outcome of X_t, the outcomes of X_s for $s > t$ are unaffected by the outcomes of X_u for $u < t$. (Our discussion of Markov processes is deliberately brief. For a thorough treatment, see [12].) The probability that outcome j is observed this period, given outcome i was observed last period, is called a one-step transition probability, customarily denoted P_{ij}.

$$\begin{aligned} P\{X_t = j | X_{t-1} = i_{t-1}, X_{t-2} = i_{t-2}, \ldots\} \\ = P\{X_t = j | X_{t-1} = i\} \forall t \end{aligned} \quad (1)$$

If the one-step transition probabilities are stationary, they can be arrayed into a square matrix with a dimension that is determined by the number of possible (and finite) outcomes. For example, if there are J possible outcomes, the transition probability matrix has dimension $J \times J$ with elements that determine the probability of transitioning from the row outcome to the column outcome. Naturally, a row sum is equal to 1. A Markov process is completely defined by its transition probability matrix and some initial starting condition, X_0. For a complete application of this statistical approach to dynamic land use, see [13].

Using the neighborhood location discussed above, our Markov transition probability matrix would be 16×16 with on-diagonal elements capturing the likelihood that a yellow cab ride began and ended in the same neighborhood. We refer to these probabilities as "unconditional" as they are simply the observed outcomes. Figure 13 displays the Markov transition matrix.

We find several salient features of these unconditional probabilities. For example, the probability a cab ride begins and ends in Brooklyn is 52%, while that for Queens is 60%. (Recall we have excluded journeys to and from airports.) Consider that if these Markov states were represented with a 16-sided die, the probability of any outcome on a roll would be 6.25%. Therefore, these two intra-location transition probabilities are extraordinarily large. None of the other 14 on-diagonal transition probabilities are nearly as large. This suggests that particular yellow cab drivers may "specialize" in intra-neighborhood journeys, a subject that we are currently exploring. We also find that the on-diagonal elements are 13% for Chelsea, 12% for Gramercy, 22% for Midtown, 32% for the Upper East Side, and 31% for the Upper West Side. These findings reinforce the point that yellow cab journeys are short distance. In contrast, a journey starting in Soho is likely to end anywhere else, including Soho, almost uniformly.

For future iterations of this work, we are developing a multivariate Markov model that will allow us condition the transition probabilities on a number of independent covariates from other data sources. For example, information from the U.S. Census allows us to compute population density (which normalizes the differences in land area), as well as concentrations of business establishments. We will also be able to include information on weather conditions by hour in NYC, which will provide a time-varying determinant that is common to all yellow cab journeys. Ultimately, we intend to develop a simulation model, based on the transition probabilities, to examine whether and how the distribution of journeys would be affected by alternative fare structures and tolls.

V. Conclusion

The visualizations in this paper highlight the important role that yellow cabs play in NYC's transportation infrastructure. As shown, however, this role typically appears to be limited to picking up a single passenger and transporting her slowly over a short distance. The out-of-pocket cost she should expect to pay is less $6 per mile, and almost certainly less than $8 per mile, even when the social value of the city's roads is highest. Given the richness of these data, we propose a statistical method to analyze dynamic flow patterns of taxi rides that can ultimately be extended to a multivariate model.

Acknowledgment

The authors thank the Taxi & Limousine Commission of New York City for providing the data used in this paper. They also thank Michael Holland, Claudio Silva, and Juliana Freire for ideas and discussions.

Unconditional Markov Transition Probabilities

	BR	CH	EV	FD	GD	GR	GV	HK	LES	MID	OIM	QN	SO	UES	UWS	WV
BR	52.16%	3.05%	4.13%	3.01%	1.74%	3.63%	3.23%	0.96%	4.93%	4.59%	7.73%	2.85%	2.17%	2.68%	1.45%	1.67%
CH	2.79%	12.97%	3.88%	2.12%	7.35%	9.84%	6.99%	5.54%	2.00%	15.32%	11.52%	1.13%	3.65%	5.20%	4.98%	4.72%
EV	7.03%	7.06%	7.57%	3.18%	4.84%	12.55%	8.38%	2.13%	6.18%	9.60%	12.37%	2.04%	3.60%	6.62%	3.02%	3.82%
FD	5.62%	6.16%	4.50%	7.46%	5.38%	7.18%	4.90%	2.80%	5.64%	14.66%	17.74%	1.33%	4.99%	5.50%	2.44%	3.69%
GD	2.02%	7.20%	2.55%	2.18%	5.09%	10.19%	3.67%	5.28%	1.30%	28.10%	12.52%	1.60%	2.04%	8.88%	5.45%	1.92%
GR	2.56%	7.66%	5.75%	2.60%	7.28%	12.05%	6.34%	2.39%	2.95%	17.20%	13.58%	1.52%	3.11%	8.87%	3.26%	2.89%
GV	4.58%	10.11%	7.26%	3.11%	5.59%	9.95%	7.95%	2.43%	4.02%	11.70%	12.96%	1.30%	5.24%	5.36%	3.37%	5.07%
HK	1.89%	11.99%	2.32%	2.06%	7.21%	6.31%	3.66%	8.08%	1.35%	22.09%	11.38%	1.53%	1.99%	5.80%	8.90%	3.43%
LES	11.04%	5.78%	9.52%	5.29%	3.77%	9.45%	7.16%	1.92%	6.82%	8.01%	12.63%	2.10%	4.61%	5.32%	2.73%	3.85%
MID	1.50%	5.65%	2.14%	2.13%	8.08%	8.64%	3.16%	4.58%	1.13%	21.85%	12.05%	1.99%	1.82%	14.60%	8.73%	1.94%
OIM	2.68%	5.46%	3.16%	3.39%	5.22%	7.85%	4.09%	2.54%	2.20%	16.25%	18.21%	1.33%	2.49%	13.99%	8.44%	2.71%
QN	2.86%	1.73%	1.19%	0.94%	2.02%	3.08%	0.98%	1.14%	0.86%	8.29%	6.24%	60.35%	0.62%	7.14%	2.08%	0.49%
SO	5.52%	10.71%	5.18%	4.91%	5.38%	7.44%	7.88%	2.59%	4.66%	12.10%	13.35%	0.98%	4.40%	4.43%	3.20%	7.28%
UES	0.81%	2.51%	1.46%	1.22%	3.96%	5.39%	1.70%	1.52%	0.89%	19.33%	16.26%	1.42%	0.92%	32.32%	9.39%	0.88%
UWS	0.73%	3.65%	0.94%	0.66%	2.96%	3.15%	1.54%	4.11%	0.55%	17.71%	16.37%	0.77%	0.77%	14.29%	30.71%	1.10%
WV	3.51%	12.82%	5.02%	3.13%	5.37%	7.84%	8.00%	3.78%	3.00%	13.15%	13.13%	0.98%	5.72%	4.89%	4.39%	5.27%

Fig. 13. The unconditional Markov transitional matrix by neighborhoods with on-diagonal elements capturing the likelihood that a yellow cab ride began and ended in the same neighborhood. Elements with more than 20% are highlighted in green.

References

[1] M. Ben-Akiva and S. Lerman, *Discrete Choice Analysis: Theory and Application to Travel Demand*, ser. MIT Press series in transportation studies. MIT Press, 1985.

[2] Y. Ge, H. Xiong, A. Tuzhilin, K. Xiao, M. Gruteser, and M. Pazzani, "An energy-efficient mobile recommender system," in *SIGKDD*. ACM, 2010, pp. 899–908. [Online]. Available: http://doi.acm.org/10.1145/1835804.1835918

[3] J. Yuan, Y. Zheng, L. Zhang, X. Xie, and G. Sun, "Where to find my next passenger?" in *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 2011, pp. 109–118. [Online]. Available: http://doi.acm.org/10.1145/2030112.2030128

[4] G. Pan, G. Qi, Z. Wu, D. Zhang, and S. Li, "Land-use classification using taxi gps traces," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 113–123, 2013.

[5] M. Veloso, S. Phithakkitnukoon, C. Bento, P. Olivier, and N. Fonseca, "Exploratory study of urban flow using taxi traces," in *First Workshop on Pervasive Urban Applications (PURBA)*, 2011.

[6] M. Veloso, S. Phithakkitnukoon, and C. Bento, "Urban mobility study using taxi traces," in *Proceedings of the 2011 international workshop on Trajectory data mining and analysis*, ser. TDMA '11. ACM, 2011, pp. 23–30.

[7] X. Liang, X. Zheng, W. Lv, T. Zhu, and K. Xu, "The scaling of human mobility by taxis is exponential," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 5, pp. 2135–2144, 2012.

[8] C. Peng, X. Jin, K.-C. Wong, M. Shi, and P. Liò, "Collective human mobility pattern from taxi trips in urban area," *PloS one*, vol. 7, no. 4, p. e34487, 2012.

[9] Z. Liao, Y. Yu, and B. Chen, "Anomaly detection in gps data based on visual analytics," in *IEEE Symposium on Visual Analytics Science and Technology (VAST)*, 2010, pp. 51–58.

[10] N. Ferreira, J. Poco, H. Vo, J. Freire, and C. Silva, "Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips," *IEEE Transactions on Visualization and Computer Graphics*, 2013, to appear.

[11] H. Rosen and T. Gayer, *Public finance*. McGraw-Hill Irwin, 2008.

[12] H. Taylor and S. Karlin, *An Introduction to Stochastic Modeling*. Academic Press, 1998.

[13] T. Savage. (2011, Jun.) Forecasting land use from estimated markov transitions. [Online]. Available: http://dx.doi.org/10.2139/ssrn.1866003