Impact of Rideshare on the NYC Taxi Industry

Economics Honors Thesis

By Rohan Kumar*

Monday, April 6, 2020

Abstract: During the last decade, rideshare companies such as Uber and Lyft entered New York City. At the same time, taxi medallion sales prices significantly declined. This paper studies the effect of rideshare on various aspects of the taxi industry and also looks at how weather interacts with taxi revenue. By analyzing data from taxi rides, and using multiple linear regressions, it becomes evident that the introduction of rideshare reduced profitability in the taxi industry. While cab revenues have decreased, they have also become more unevenly distributed across the day. Via the Gordon Growth Model, it is apparent that the changes in profitability significantly influenced medallion values. Overall, the price-to-earnings ratio (PER) for medallions declined substantially in recent years, thus indicating a negative future outlook for the taxi industry. Nonetheless, the city's ride market as a whole — taxi plus rideshare — has expanded tremendously.

* Rohan Kumar: kumarrp@umich.edu

Good advice goes a long way. I am forever grateful to Matthew Shapiro, my thesis advisor, whose suggestions and guidance helped bring this project to fruition. I was fortunate to get additional structure and recommendations from Kathryn Dominguez, the Director of the Honors Program. Finally, I would like to thank the University of Michigan Department of Economics for its endless support, resources and enthusiasm.

Introduction

In the 1930s, the New York City taxi industry experienced a significant change. In years prior, the industry had been relatively unregulated. There was a licensing system but no limit to the number of taxis. As a result, there was a high supply of taxis that created competition among drivers to get passengers. Additionally, there were unlicensed taxis — known as wildcat taxis — that often lowered fares to capture the demand. This situation led to unrest among drivers. With the Great Depression, further decreases in the demand for taxis heightened the tension. To ease the tension, New York City Mayor Fiorello H. La Guardia passed the Haas Act in 1937. This law created the medallion licensing system that exists today. Basically, in order to operate a taxi one was required to own a government-issued medallion, and these medallions were tradable. The number of medallions was capped at 16,900 to reduce competition for passengers, thus improving economic conditions for drivers who had medallions. The city also increasingly used the police to enforce regulations and industry standards. More than 70 years after the creation of the medallion system, the taxi industry faced another monumental change — the introduction of rideshare companies, predominantly Uber and Lyft. This added competition for taxis, tapping into their exclusivity and potentially crumbling medallion investments.

This paper serves a few purposes that stem from two fundamental questions. First, to what extent do medallion sales prices respond to profitability in the taxi industry? And second, how severely has growth in the rideshare industry impacted taxi profits? I incorporate medallion prices from 2011 to 2018 into the Gordon Growth Model to quantify the link between the value of a medallion and taxi profitability. To study the holistic effect from rideshare, I use rideshare data starting in 2015. After delving into the first two questions, though, I strive to further decompose taxi earnings. It is likely that rideshare transforms earnings in the taxi industry through multiple paths. I pose two potential channels — the daily distribution of taxi revenue and also linkage between temperature and precipitation — and I analyze the evolution of these channels from 2009 to 2018 to identify changes stemming from rideshare. Essentially, I use Gini coefficients and also quantify hourly effects to investigate how revenue is distributed across hours of the day, and whether that distribution becomes more or less even after rideshare enters the city. To inspect how the role of weather has changed, I incorporate quadratic and quartic approximations. This paper also looks at interesting correlations that help pinpoint different ways rideshare has impacted the taxi industry, such as to what extent rideshare acts as a substitute for medallion taxis. Finally, this study serves the broader goal of articulating who benefits and who is harmed with rideshare.

Literature Review

My paper investigates medallion prices, revenue distribution, weather's impact on hourly earnings, and how the growth of rideshare affects all of those topics. Given the recency to the boom in rideshare, closely related studies are still forthcoming. There's vast related economic literature on the taxi industry, with much attention to the supply side. These studies focus on much narrower time periods than mine, which looks at a 10 year span. Studying a larger time interval adds validity to results and also sheds light on long-term trends. Regardless, analyzing the established literature helps in understanding the inner-workings of the market and decisions drivers make each day. Camerer et al. (1997) dive into taxi drivers' allocative planning of when to drive versus when to choose leisure. Per neoclassical theory, in maximizing income and leisure drivers should display positive wage elasticity; they should work longer hours on good, busy days and fewer on slow days. The taxi industry is a suitable environment for testing neoclassical theory, as there is wage correlation across the hours of a particular day but not for multiple days. But the results indicate negative wage elasticities. Drivers appear to set a target income for each day and then stop working once they reach said target. On busy days, drivers reach their income targets quicker. As a result drivers end up working fewer hours on busy days and more on slow days, showing irrationality in behavior. Farber (2015) follows that line of thinking but uses data from 2009 to 2013, compared to older more limited data used in the earlier research. He coins the income targeting behavior with the behavioral economic concept of reference dependence, then once again tests how it fares against neoclassical theory. By contrast, the new data yields positive labor supply elasticities and favors neoclassical optimization over reference dependence. A further extension also showed that a driver's experience is positively correlated with elasticity, meaning that drivers more optimally manage income and leisure as time goes on.

On that topic of driver experience, Haggag et al. (2017) dive further in. A driver's revenue depends on his or her skill of finding customers, so their research studies the magnitude of learning by doing for NYC cab drivers. Their analysis shows a seven percent productivity uptick between a driver's first and 100th shift. They also show evidence that neighborhood-specific driving experience — as opposed to just general city experience — has a statistically significant impact on a driver's efficiency in finding passengers following a trip completion. Frechette et al. (2015) offer additional prominent supply-side research of the taxi industry. They note the industry differs from the efficient market hypothesis because of government restrictions on entry and also since demand and supply must look for each other. They analyze the impact of these distinctions by running simulations allowing for market entry and improved driver-passenger matching techniques. The simulations incorporate 2011-2012 taxi ride data, and the study concludes entry and improved matching could bring substantial gains. Looking back now shows a level of foresight in those conclusions; rideshare surpassed entry limits, bringing superior matching technology compared to taxis that helped capture those gains — and perhaps more.

Supply-side research stretches into the rideshare market, too, but is more constrained due to limitations in accessing proprietary rideshare data. Hall and Krueger (2016) try to better understand the on-demand economy via Uber. They look at inner-company surveys for a demographic snapshot of drivers. In the surveys, drivers predominantly note schedule flexibility as a reason for driving Uber. The study then confirms through Uber ride data that drivers do indeed use the flexibility. The key finding, though, is that earnings per hour and hours worked are independent for Uber drivers. This suggests a sort of equilibrium in the Uber market, perhaps created by pricing mechanisms. Brodeur and Nield (2017) follow a related path, showing such mechanisms — in particular surge pricing — help churn an 18 percent increase in volume of Uber rides when it is raining. They identify an uptick in volume for taxis during rain, too, but it is not as substantial. They offer an explanation based on the following principles: Demand for ride services is greater during rain. Surge pricing offers Uber's labor pool an incentive to start driving, in turn shifting the supply curve. The taxi industry lacks such a mechanism. Cab drivers may be able to pick up more customers while it's raining, but fare rates don't increase. Thus, the incentive is often lesser in magnitude than with surge pricing, so taxi drivers must determine whether it's worth it to stay on the road during harsh conditions.

Zhang et al. (2017) take a closer look at Uber surge pricing. They mine spring 2017 ride request data — such as location of request, distance of request, time of day and Uber vehicles nearby — from the Uber app to distinguish patterns in pricing. They look at how surge pricing varies with different parts of the day and week, and find the multiplier is highest (2X) on Tuesday and Wednesday mornings as well as late Sunday nights. For different initial conditions, the analysis also calculates probabilities of an Uber ride costing more than a cab ride. Finally, Cramer and Krueger (2016) compare capacity utilization rates for Uber and taxis in New York, LA, Boston, San Francisco and Seattle. Utilization refers to the proportion of time a passenger is in the vehicle out of the total time a driver spends working, and I incorporate related utilization metrics for my study. They find Uber has significantly higher utilization rates in all the cities except for New York, where it's roughly the same. However, the study uses 2013 cab data with 2014 and 2015 Uber data. The authors purport New York City could likely have the same utilization disparity as other cities after accounting for the lag in the data. The paper tries to explain the utilization difference between Uber and cabs. Cab drivers work longer hours on average, so one theory is these excess hours naturally lead to lower utilization rates. The study rejects this hypothesis, though, since utilization rates for Uber drivers prove to be independent of how many hours a driver works. This indicates a natural utilization balance in the Uber market. These results tie into other studies mentioned above. The matching technology helps raise utilization, thus creating a competitive edge; Uber can charge lower fares and still churn the same hourly revenue. For instance, the paper mentions it could often charge 28% lower fares than taxis and still match earnings. And Uber's advanced pricing mechanism helps it change prices exactly so, thus luring in customers. With such an advantage, rideshare has grown tremendously. I will look at how this advancement has impacted the New York City taxi market.

Institutional Background

New York City made some adjustments to the medallion system after creating it. Early on, the number of medallions shrunk down to 11,787 because some owners didn't pay renewal fees due to the turbulent economy. The count stayed steady until 1996. Since then, it has increased in small increments and is now at 13,587. There are two key types of medallions — independent and corporate. The former are meant for individuals while the latter are for companies looking to own big fleets of taxis. Within these two main categories there are further distinctions; some medallions require the owner to operate a vehicle that is wheelchair accessible, while others entail alternative-fuel vehicles. Medallions that are free from these specifications are called "unrestricted." When the city creates new medallions it auctions them off, but since the number of medallions has been relatively fixed most transactions happen among owners at mutually-agreed-upon prices. These licenses are essentially assets. Owners have the right to make money through driving the vehicle, but can also lease the vehicle out and collect rents. These assets were once seen as sound investments. Prices skyrocketed from $10 in 1937 to an average of $5,000 in 1950. They kept growing, and both categories of medallions have crossed the million-dollar mark in the last decade.

Aside from traditional medallion taxis, other means of transportation — called for-hire vehicles — entered the city later. The traditional yellow medallion taxis still get preferential treatment. They can pick up passengers anywhere in New York City, through both street-hails and prearranged rides. There are now also green, non-medallion cabs. They can pick up both types of rides, but are geographically restricted to outer boroughs and the northern region of Manhattan. They were created to help meet demand in less central areas where customers had trouble catching street-hails. Black Cars, luxury limousines and liveries also play a role but can only take prearranged rides. Uber launched in the city in 2011 while Lyft entered in 2014. These rideshare companies make up another category of for-hire vehicles, called High Volume, since they deal with more than 10,000 rides a day. These companies can only take prearranged rides, yet have changed the landscape. They've enabled a multitude of individuals to offer rides with their own personal vehicles, whenever desired.

Method/Model

Since medallions enable one to operate a taxi and make profits, they essentially offer a stream of future payments. Therefore, the price of a medallion should gauge the present discounted value of that flow. To capture this aspect, this paper relies on the Gordon Growth Model, which states the following:

$$P_t = \frac{E_t}{r_t - g_t}$$

In this equation, P_t represents the price of an asset and E_t stands for earnings or dividends associated with said asset. In the denominator, r_t is the cost of equity — the minimum rate of return an investor would need to invest in the asset. It is linked to the underlying risk, because investors require a higher rate of return when investing in riskier assets. For a risk-free investment, r_t would be the real interest rate. Finally, g_t is the expected growth rate of earnings. Together, r_t and g_t act to discount future earnings and thus appropriately value the asset. Using this framework, I will take a close look at how taxi profitability influences medallion prices by estimating the following model:

(1) $$P_t = \beta_0 + \beta_1 E_t + u_t$$

Here, P_t represents the average medallion price and E_t is the average profitability of a taxi, both in month t. The slope estimate is essentially a measure of the inverse of $r_t — g_t$. The Gordon rule is also linked to the price-to-earnings ratio (PER). Treating the cost of equity as a constant and moving earnings to the left hand side derives the following:

$$\frac{P_t}{E_t} = \frac{1}{\bar{r} - g_t}$$

The ratio on the left side represents the PER. This ratio indicates perceived future growth in the value of an asset. For instance in the stock market, the belief that a company will excel in the future coincides with a higher expected earnings growth rate (g_t). In the above equation, as the expected growth rate (g_t) rises, the PER will also increase. Put differently, if investors believe a company will grow, they are willing to pay a higher multiple of earnings for the stock in the present. Similarly, the PER in the medallion market illustrates the market's future outlook on the taxi industry. I will analyze changes in the PER of medallions over time.

Model (1) does not look directly at the effect of rideshare, so I will use the following model as well to see how rideshare changes the taxi industry:

(2) $$E_t = \beta_0 + \beta_1 HighVolume_t + u_t$$

where $HighVolume_t$ represents the average number of daily rides, in 1000s, conducted through High Volume for-hire vehicle services in New York City, for a given month. Note, rideshare companies fall into this category. Here, estimating the β_1 coefficient shows how the volume of rideshare trips impacts the profitability of medallion taxis, on average.

I will also estimate a third model that decomposes a taxi's profitability as follows:

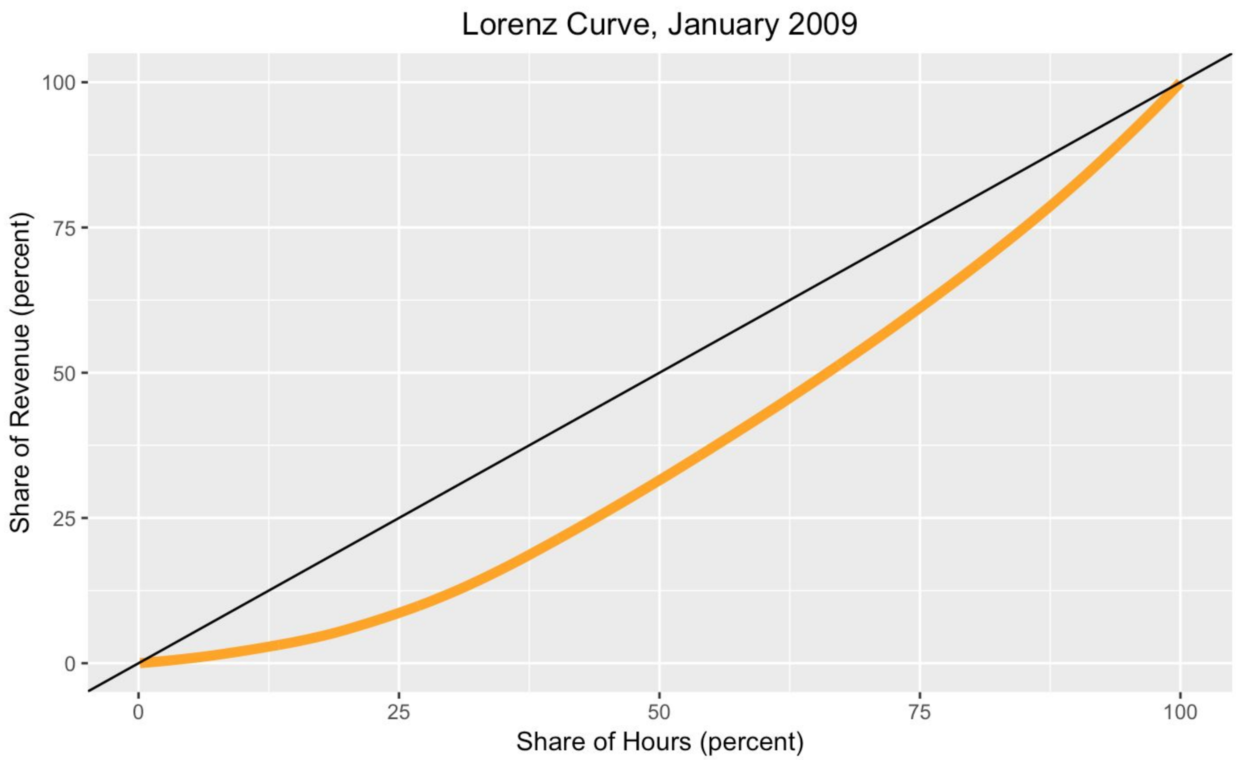
(3) $$log(E_t) = \beta_0 + \beta_1 util_t + \beta_2 faredensity_t + \beta_3 mpg_t + \beta_4 fuelprice_t + u_t$$

The $util_t$ term represents the average daily utilization hours per cab for a given month. The $faredensity_t$ term is a measure of a cab's hourly efficiency for a given month; essentially, it is how much fare the average cab generates in an hour. These variables are included because the revenue a cab makes will depend both on how many hours it is driven and how much it can make in an hour. The mpg_t term shows the average fuel economy of taxis, while $fuelprice_t$ is the average monthly fuel price. These two variables help represent the cost side. Estimating model (3) essentially gives a breakdown of how these different variables determine profitability.

A key part of this paper will be to look at correlation coefficients between various variables. This will allow me to analyze industry trends and see how the volume of rideshare potentially impacts the taxi industry. I will look at changes in the volume of total rides and taxi rides, as well as hourly efficiency, total daily industry fare, the number of drivers and vehicles, driver shifts, monthly and daily vehicle utilization, trip durations and payment types that correspond with the introduction of rideshare.

Model (1) estimates how medallion prices depend on taxi profits, and model (2) quantifies how those profits change with growth in rideshare. As an extension, this paper will look at specific channels that fold into taxi earnings and the evolution of those channels before and after the introduction of rideshare. The first channel I study is time of day. The data indicates significant reductions in the taxi industry, but this paper also serves to identify where the losses are occuring. The cab industry is one of peaks and troughs. Certain hours of the day tend to be significantly busier than others, and drivers generate most of their revenue in these peak hours. I analyse whether shrinkage in the market occurs consistently across all hours of the day, or whether the curtailment is predominantly during peak periods. The principle behind the latter scenario is that perhaps with rideshare, taxi drivers make considerably less money in the hours they used to rely on, while the other slower hours remain relatively untouched. If the latter turns out to be true, the revenue should become more evenly distributed across hours of the day following the introduction of rideshare. To test this idea I use Lorenz curves, because they are practical for representing a distribution. I construct 120 curves — one for every month — by sorting all the hours of a month based on revenue. Then I link the hours with what share of that month's total revenue they generate, as shown in Figure 1 using January 2009 as an example. I find Gini coefficients by calculating the area between the black and orange curves and then dividing by the total area under the black curve. The Gini coefficient is a measure of inequality, so by studying coefficient changes I can better understand how revenue is spread across the hours and how that spread changes over time.

Figure 1



Notes: This Lorenz Curve example shows how all the hours in the month of January 2009 cumulatively contribute to the total revenue of that month. Basically, this helps display the degree of evenness in that month's distribution of earnings and the curve is later used to calculate Gini coefficients.

I also run the following regression to further analyze distribution changes:

(4)

$$log(totalfare_t) = \beta_0 + \beta_1 month_t + \beta_2 weekend_t + \beta_3 temp1_t + ... + \beta_{26} temp24_t$$
$$+\beta_{27} precip1_t + ... + \beta_{31} precip5_t + \beta_{32} hour1_t + ... + \beta_{54} hour23_t$$
$$+(temp * precip) + u_t$$

I use hourly observations, so $totalfare_t$ refers to the sum of industry revenue generated in a particular hour. The specification controls for month, weekend, temperature and precipitation effects. I create 25 different temperature ranges, with narrower ranges at the low and high end of the temperature spectrum. The $precip1_t$ variable is a dummy for non-zero precipitation that is under .125 inches. The other subsequent precipitation dummies refer to increments that grow in size, with $precip5_t$ referring to at least one inch of precipitation. The *hour* dummies categorize the data by time of day, with *hour1* referring to 1-2am. Finally, I interact the first 11 temperature dummies with the precipitation dummies. The first 11 reach up to the freezing point of water, and I am assuming the interaction between temperature and precipitation is more relevant for snow. The purpose of all these controls is to account for some natural variation in the data, before fully focusing on the hourly effects. I run this regression individually for all 10 years. Next, I plot estimates of the hourly coefficients (β_{32} — β_{54}) from each year to see how hourly effects vary across time. Additionally, this will allow me to understand shift changes for cab drivers. It's often said that finding a taxi in New York City during the late afternoon is more difficult. Drivers are known to be switching off shifts, so during these hours there seem to be less cabs on the road. I will take a look at the shift change hours to see if the magnitude of the dead period varies across the years.

The second channel I analyze is how weather impacts cab revenue, through using the following quadratic and quartic regression models:

(5)

$$log(totalfare_t) = \beta_0 + \beta_1 month_t + \beta_2 weekend_t + \beta_3 hour1_t + ... + \beta_{25} hour23_t$$
$$+\beta_{26} post_t + \beta_{27} precip_t + \beta_{28} temp_t + \beta_{29} temp_t^2 + \beta_{30} temp_t * precip_t$$
$$+\beta_{31} temp_t^2 * precip_t + \beta_{32} precip_t * post_t + \beta_{33} temp_t * post_t$$
$$+\beta_{34} temp_t^2 * post_t + \beta_{35} temp_t * precip_t * post_t$$
$$+\beta_{36} temp_t^2 * precip_t * post_t + u_t$$

(6)

$$log(totalfare_t) = \beta_0 + [\beta_1 month_t + \beta_2 weekend_t + \beta_3 hour1_t + ... + \beta_{25} hour23_t]$$
$$+\beta_{26} post_t + \beta_{27} precip_t + \beta_{28} temp_t + \beta_{29} temp_t^2 + \beta_{30} temp_t^3 + \beta_{31} temp_t^4$$
$$+\beta_{32} temp_t * precip_t + \beta_{33} temp_t^2 * precip_t$$
$$+\beta_{34} temp_t^3 * precip_t + \beta_{35} temp_t^4 * precip_t$$
$$+\beta_{36} precip_t * post_t + \beta_{37} temp_t * post_t$$
$$+\beta_{38} temp_t^2 * post_t + \beta_{39} temp_t^3 * post_t + \beta_{40} temp_t^4 * post_t$$
$$+\beta_{41} temp_t * precip_t * post_t + \beta_{42} temp_t^2 * precip_t * post_t$$
$$+\beta_{43} temp_t^3 * precip_t * post_t + \beta_{44} temp_t^4 * precip_t * post_t + u_t$$

Specification (5) is the quadratic model. It uses the same month, weekend and hourly controls as model (4) but introduces some new variables. $Post_t$ characterizes each observation as either before or after the rise of rideshare, defining 2015-2018 as the post period as this is when rideshare grew substantially in the city. I add a dummy for precipitation and use a continuous temperature variable rather than temperature bins. I add a quadratic temperature variable, too. The specification includes multiple interactions to see how temperature and precipitation together impact a particular hour's total fare, under a quadratic relationship, and how the effect changes with rideshare on the rise. Model (6) follows suit, adding cubic and quartic temperature terms plus related interactions to approximate a quartic relationship. To interpret regression results from the quadratic and quartic models, I plot partial fitted value functions. For each model, I plot the four combinations of pre versus post and precipitation versus no precipitation functions. The eight total functions are all solely dependent on temperature. The following is an example of the quartic relationship plotted in the post period with precipitation, and the simple steps to derive it:

Starting with (6), assume post = 1 and precip = 1 to yield:

$$log(totalfare_t) = \beta_0 + \beta_{26} + \beta_{27} + \beta_{28} temp_t + \beta_{29} temp_t^2 + \beta_{30} temp_t^3$$
$$+\beta_{31} temp_t^4 + \beta_{32} temp_t + \beta_{33} temp_t^2 + \beta_{34} temp_t^3 + \beta_{35} temp_t^4 + \beta_{36}$$
$$+\beta_{37} temp_t + \beta_{38} temp_t^2 + \beta_{39} temp_t^3 + \beta_{40} temp_t^4 + \beta_{41} temp_t + \beta_{42} temp_t^2$$
$$+\beta_{43} temp_t^3 + \beta_{44} temp_t^4 + u_t$$

Then simplify by grouping:

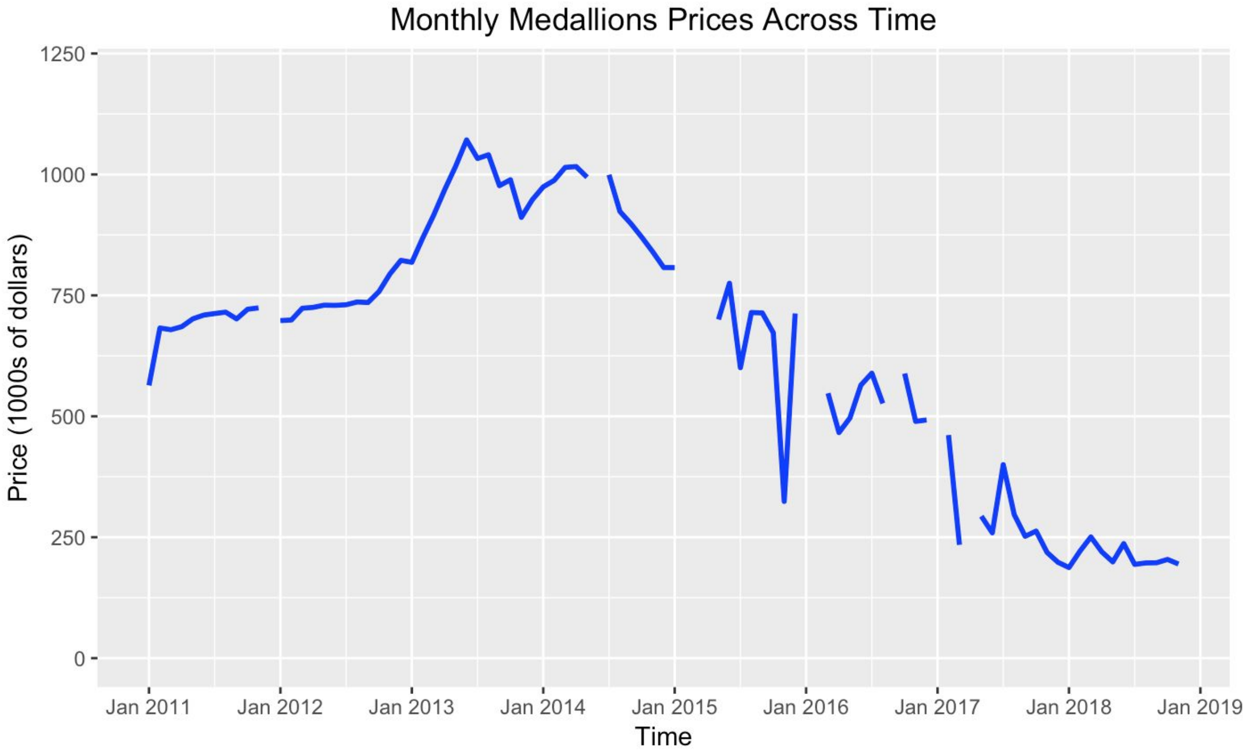
$$log(totalfare_t) = \beta_0 + \beta_{26} + \beta_{27} + \beta_{36} + (\beta_{28} + \beta_{32} + \beta_{37} + \beta_{41})temp_t$$
$$+(\beta_{29} + \beta_{33} + \beta_{38} + \beta_{42})temp_t^2 + (\beta_{30} + \beta_{34} + \beta_{39} + \beta_{43})temp_t^3$$
$$+(\beta_{31} + \beta_{35} + \beta_{40} + \beta_{44})temp_t^4 + u_t$$

Data

Most of the data comes from the New York City Taxi and Limousine Commission (TLC). The TLC regulates the various types of vehicles discussed in this paper, as well as medallion sales. Additionally, CPI data from the Federal Reserve Bank of St. Louis is used to adjust medallion prices, taxi fares and fuel prices for inflation.

As mentioned before, there are different types of medallions. For medallion prices this paper focuses solely on the independent unrestricted category, as opposed to corporate and or restricted medallions, because records for this type are recorded most consistently and are thus most easily interpretable. Also medallions are sometimes sold in bundles or fractions; however, this paper sticks to sales of single medallions. This avoids the effect of discounting or markups and helps yield the true price of an individual medallion. The records contain notes depicting some sales as “Estate” or “Family,” and the corresponding prices are often $0.00 or significantly lower than the bulk of other sales prices. These may inaccurately reflect the true price, so they are ignored. Some other price outliers are dropped. As an illustration, eight independent unrestricted medallions were transacted in October 2013, none of which contained notes. Four of them had prices listed as $0.00, thus were likely just transfers, so I ignore them. Three were sold around $1,000,000, and that had been close to the trend in prior and latter months. One of them had a price of $500,000. This was dropped, because it is considerably less than the trend; it was likely a sale of a fractional medallion but was not recorded properly. After following this process for each record, I calculated average sales prices for each month. For medallions, this paper studies the period from January 2011 through November 2018. Reports are missing for December 2011 and June 2014. Also, seven months didn’t have relevant medallion sales. Thus overall, I calculate average medallion prices for 86 months. Figure 2 tracks individual unrestricted medallion prices. Prices peaked in mid 2013, at over $1,000,000, and have significantly declined since.

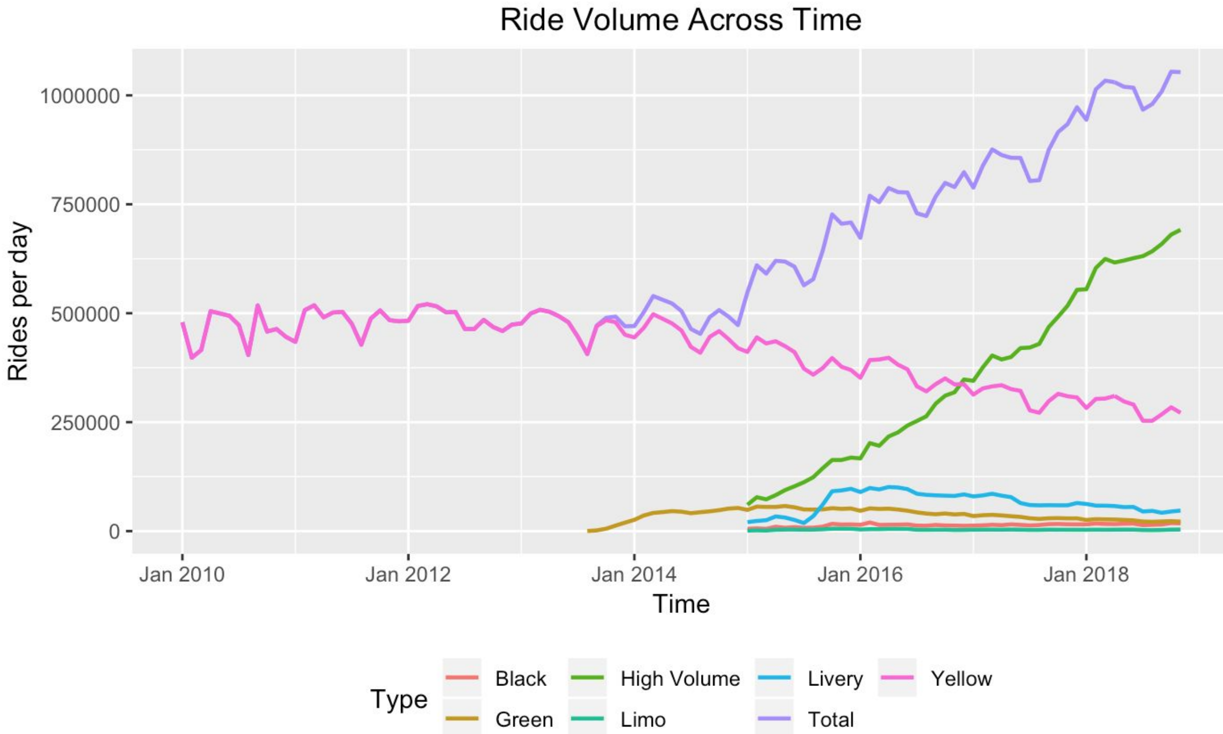
Figure 2



Notes: The figure shows the recent decline in sales prices of individual unrestricted taxi medallions. Gaps represent missing data.

Much of my analysis of medallion taxis dates from January 2009 through the end of 2018. Green cab data is recorded beginning in August 2013, while Black Cars, Luxury Limousines, Liveries and High Volume services join the dataset in January 2015 — accounting for 47 observations. It should be noted that these different categories existed for some time before being added to the dataset. For instance, as mentioned earlier, Uber and Lyft entered the city before 2015. They were fairly small though, and weren't accounted for until later. It is important to specify that since the High Volume category applies to all companies that offer more than 10,000 rides a day, it is not certain that all of them are rideshare firms. However, this group is predominantly rideshare companies — such as Uber, Lyft, Gett and Via — thus studying this category provides a fair representation of the rideshare industry. Figure 3 shows ride volumes for the different services in New York City. For a while, the market size stayed quite constant. Over the last four years the High Volume sector took off, spurring a significant expansion in the market as a whole. Simultaneously, the yellow cab industry has declined from around 500,000 rides a day to just over 250,000.

Figure 3

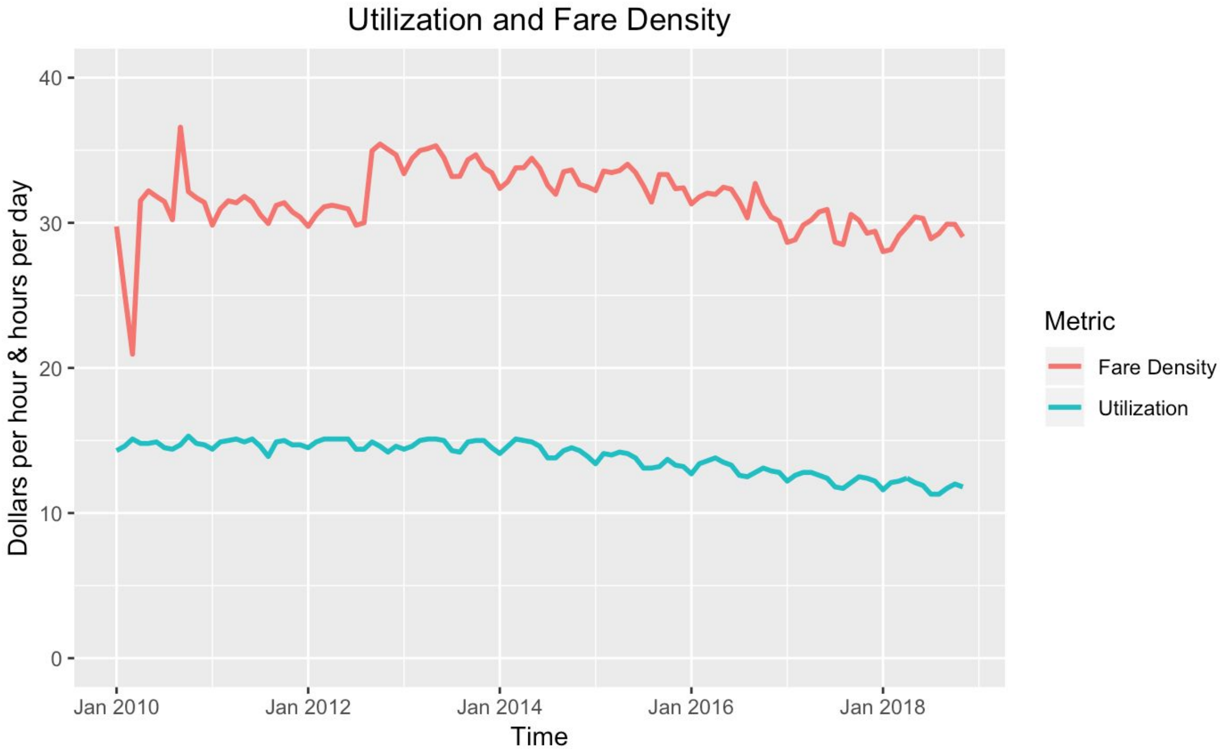


Notes: The figure shows the evolution of different ride services in New York City and the composition of the total industry.

The TLC dataset contains a monthly measure of how many hours a taxi operates in a day, on average. This data is used for the $util_t$ variable. It also records the average daily total fare amount the yellow cab industry makes each month, as well as the average amount of unique vehicles on the road each day. Dividing the fare variable by both hours of operation and amount of unique vehicles yields the total hourly fare each medallion cab generates, on average, for each month. This is used for the $faredensity_t$ term. Figure 4 plots utilization and fare density across time; both have been declining in recent years.

The monthly profit numbers, used for E_t, are composed of data from a few different sources. To calculate these figures, I first find the average amount of revenue a taxi generates, for each month. I start with the TLC monthly data on how much the yellow cab industry makes on the average day. Multiplying by the amount of days in the respective month and dividing by the number of cabs yields the revenue a cab makes, on average, for a particular month.

Figure 4

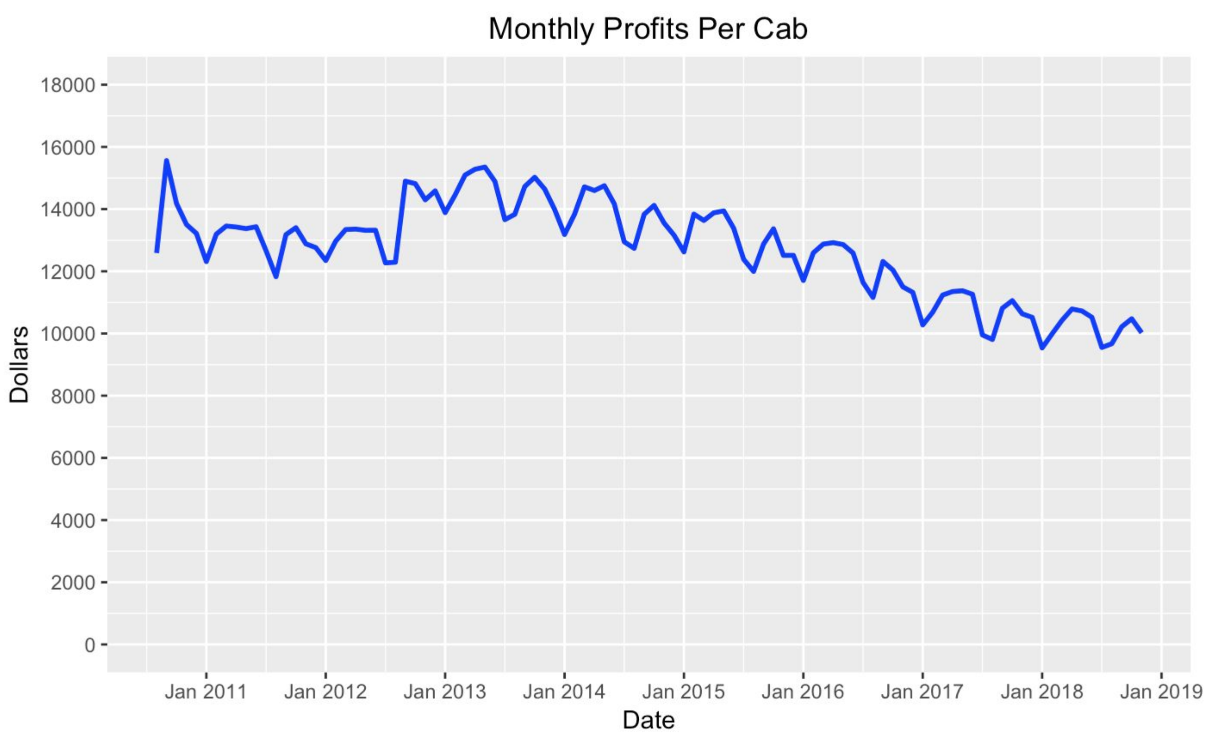


Notes: The blue line indicates how many hours a cab drives in a day, on average, for a specific month. The red line shows how much money a cab makes in an hour of driving, on average, for a given month.

Owning a taxi presents a variety of costs — such as labor and depreciation costs. However, this paper solely incorporates fuel costs, as data for this is most accessible. Ignoring these other costs entails that my total costs are understated and profit figures are overstated. But this is fine; an estimate for the magnitude of the total costs and profits is sufficient, because it is the change in these figures that really matters. Also, fuel costs are likely much more volatile than labor and depreciation. Thus focusing on fuel captures adequate variability in costs and profits.

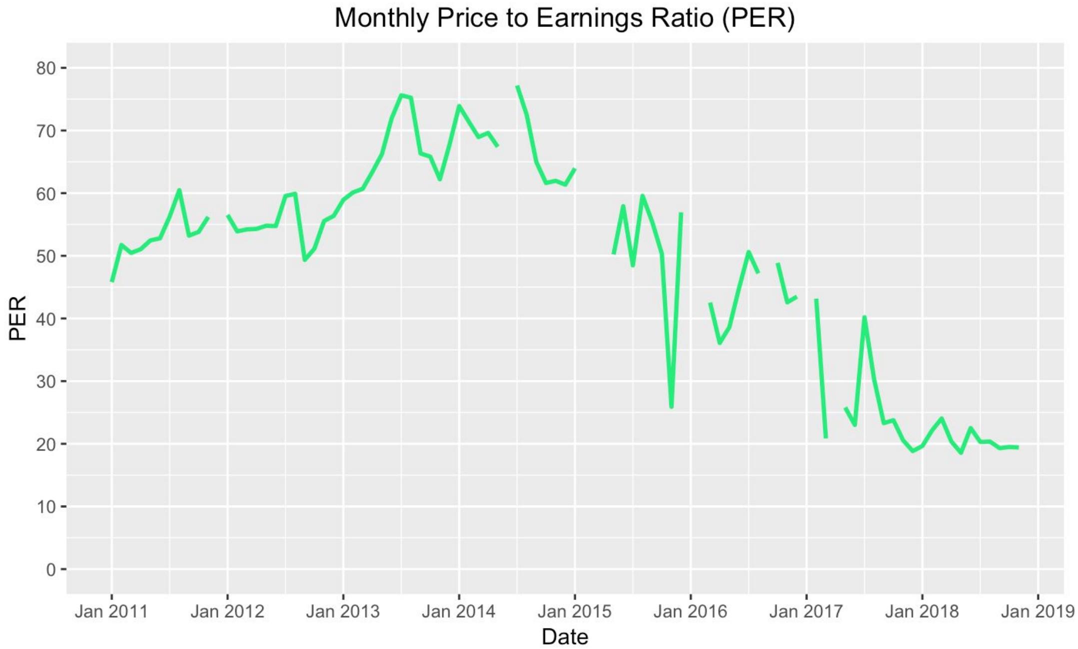
To calculate monthly fuel costs, I first find how many miles the average cab drives in a given month. The TLC provides average taxi trip duration figures for each month, in minutes. Multiplying these by the amount of trips a taxi takes yields total driving time. To get mileage, I assume taxis drive 30 mph, on average. The Bureau of Transportation Statistics (BTS) provides the average new vehicle fuel efficiency for each year. My analysis is monthly, though. Also, it is more realistic that these fuel economy improvements spill into the taxi industry in smaller increments, rather than just big steps once every year. Thus I use linear regression on the yearly data to get fuel economy projections for each month. TLC documents indicate the average taxi is around four years old, so I use fuel economy data with a four year lag. For instance, for September 2013 I use the fuel economy fitted value from September 2009. The Energy Information Administration offers average monthly regular fuel prices for New York City. Combining these with fuel economy and miles driven generates a taxi's monthly fuel costs. The earnings data for E_t is simply monthly revenue minus fuel costs. Figure 5 shows these profitability numbers across time, which have been decreasing recently. I also calculate PER from January 2011 to November 2018. Data is missing for nine months, so overall I have PER for 86 months. Figure 6 shows that PER peaked in 2014 and has been on a downward trend since. This indicates the perceptions about the industry's future are worsening.

Figure 5



Notes: This figure shows the monthly profitability of owning a cab.

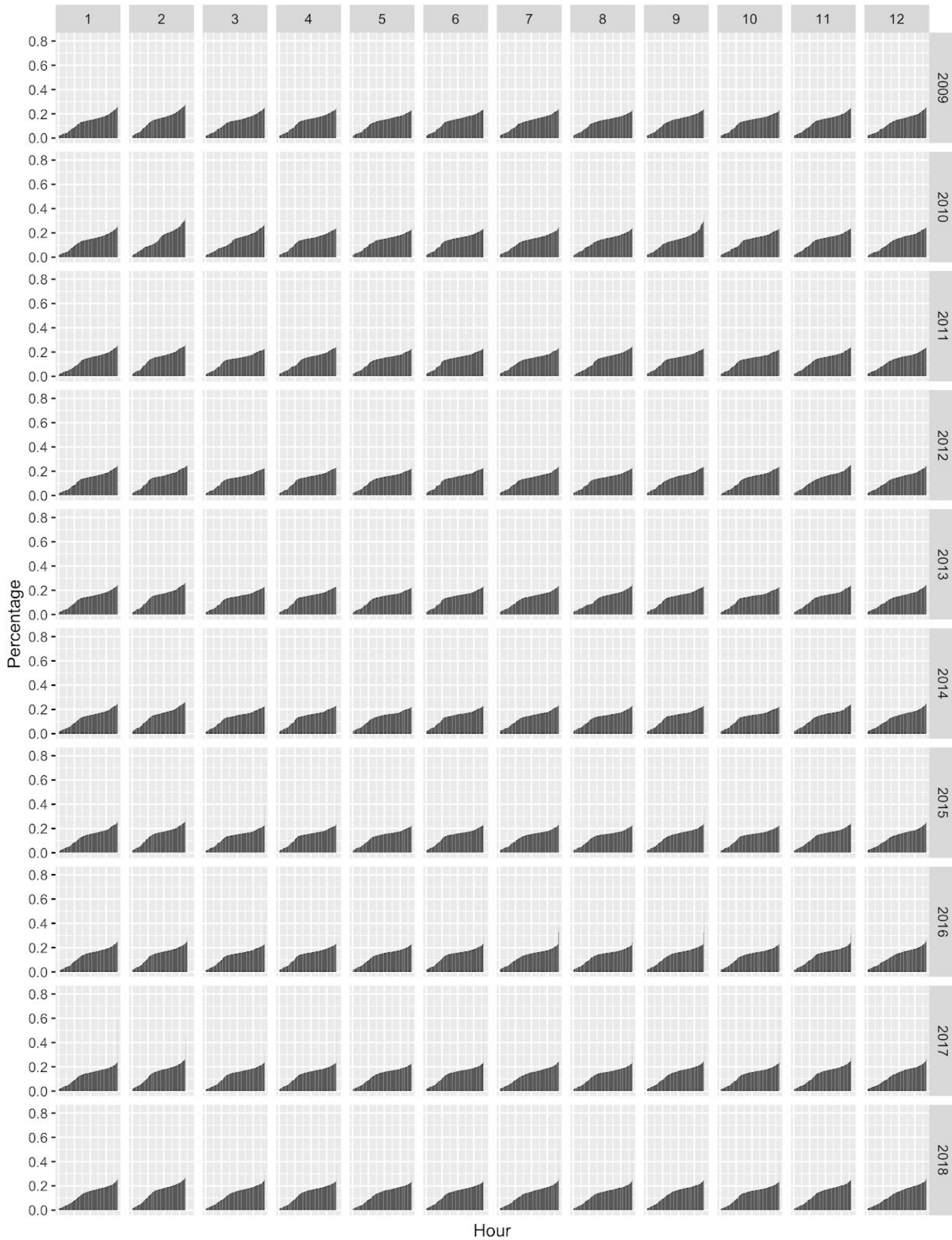
Figure 6



Notes: This figure shows the monthly price-to-earnings ratio (PER) in the market for taxi medallions. Gaps indicate missing data.

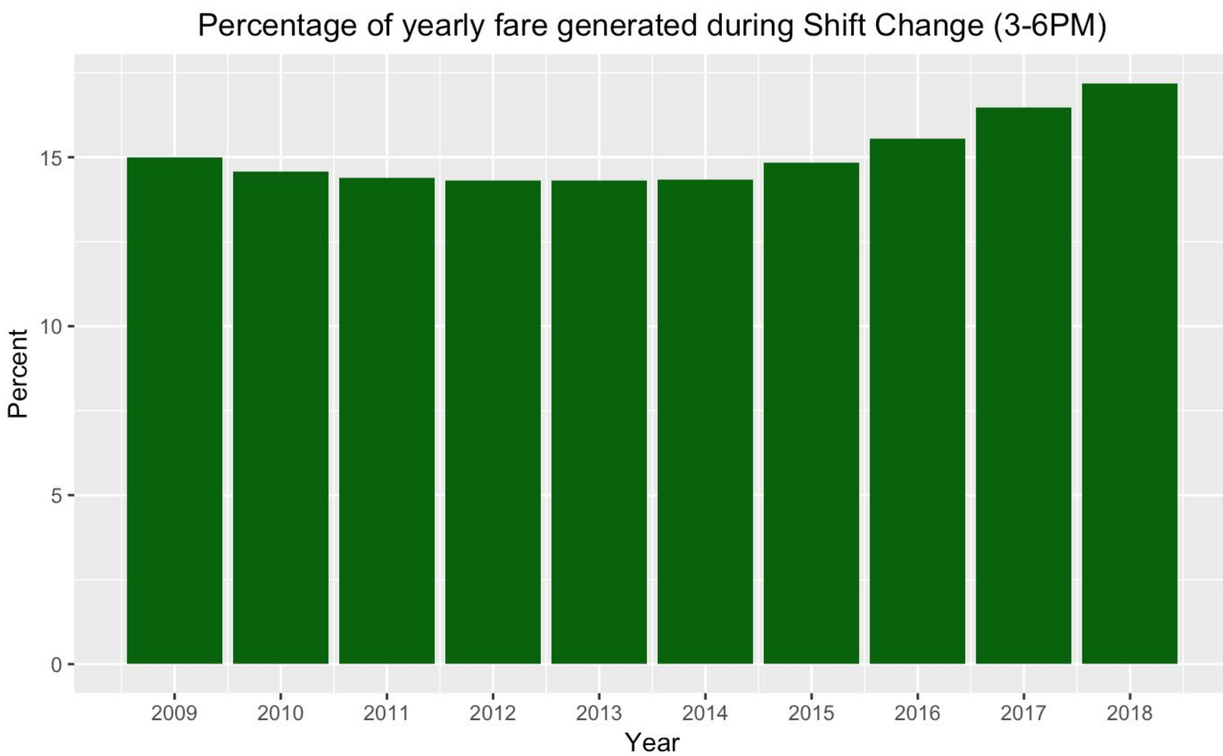
The TLC publishes monthly aggregate ride data. This works well for the medallion analysis, as I calculated the average price of medallion transactions each month. But since much of my analysis focuses on hourly effects, I composed hourly data too. Through the TLC, ride statistics of every individual cab ride from 2009 to 2018 are publicly available. These data sets are massive, though, as each monthly file may contain more than 15 million ride observations. And since my research uses a vast range of months, conducting analysis at the individual ride level becomes impractical from a data management standpoint. So, I summarized the data at the hourly level by calculating the number of rides and total fare collected for each particular hour of the 10 year span. Figure 7 arranges each month's hours by revenue, and then plots what percentage of that month's total revenue each hour generated. Merely looking at the figure, it's hard to distinguish the distribution changes across time. The purpose of the figure is to illustrate that the cab industry is indeed one of peaks and troughs, and my look at the Gini coefficients will help quantify movement. A preliminary look at the hourly ride data shows an interesting trend related to shift changes, as displayed in Figure 8. The percentage of yearly fare generated in shift change hours decreases for the first few years of my analysis. In the latter half, though, it increases substantially and reaches levels higher than the starting point. Finally, regressions four through six incorporate weather, and I obtained weather data from the Network for Environment and Weather Applications via the Cornell University division. I use hourly weather observations taken at Central Park.

Figure 7 — Revenue Distribution by Month



Notes: For each month of the 10 years, this chart arranges hours based on revenue generated then shows what percent of the month's total revenue each hour generated. It's evident that there are peaks and troughs; revenue is not distributed evenly.

Figure 8



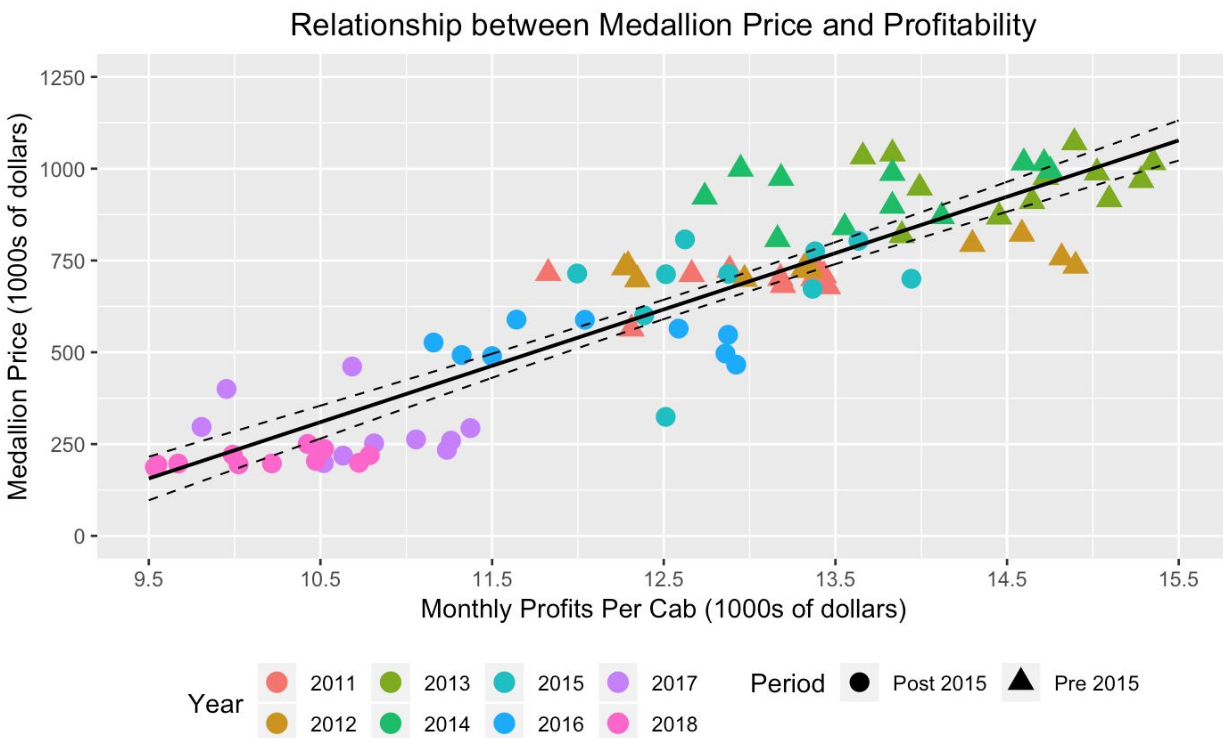
Notes: This figure shows how much of each year's taxi revenue is generated during the afternoon shift change.

Results

A. Gordon Growth Model & Baseline Findings

The regression results from model (1), which regressed average monthly medallion prices on monthly taxi profitability, are woven into Figure 9. Shapes categorize the data by pre and post 2015, synonymous with before and after vast growth in rideshare. Color further classifies the data by individual year. Evaluating the first model yields a slope estimate of 153.57 that is statistically significant at the one percent level, and I accordingly construct a regression line plus a 95 percent confidence interval. The interpretation of the slope estimate is that, on average, every $1000 increase in taxi profitability spurs a $153,570 increase in medallion value, holding all else constant. The chart indicates that the model fits the data quite well in both the pre and post periods. Because of the close fit, it appears the Gordon Growth Model holds up well in this setting. A medallion's market value responds heavily to taxi profitability, which is a consequential result considering the substantial variation in profits over time.

Figure 9



Notes: This chart incorporates regression results from model (1) for a sample size of 86 months. Shape and color organize the data by a pre and post period as well as by individual year. The regression line plus a 95% confidence interval are shown in black, and the slope estimate of 153.57 was significant at the one percent level. Finally, the model yielded an R^2 of 0.7939

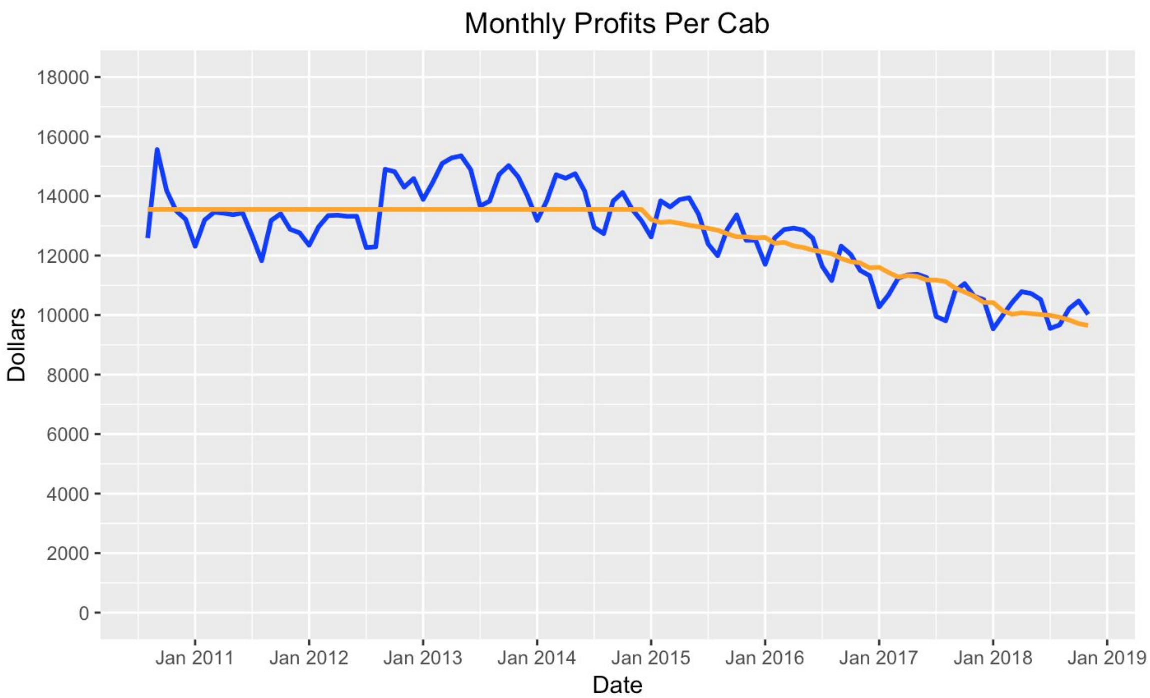
Table 1 shows the regression results for model (2), which helps explain the profitability variation by regressing profitability on the amount of High Volume rides (in 1000s). The results are statistically significant at the one percent level, and can be interpreted as follows: if the number of daily trips through High Volume services increases by 100,000, the monthly profitability of a taxi falls by $563.86, on average. This appears economically significant, because the number of High Volume daily trips increased from just over 60,000 in January 2015 to nearly 700,000 in November 2018.

Table 1: Profitability regressed on amount of High Volume rides

	Model (2)
constant	13548.89*** (183.9)
$HighVolume_t$	-5.6386*** (0.0004287)
R^2	0.7790
Sample size	47

Notes: This table shows estimation results from model (2). *** Significant at the one percent level. Heteroskedasticity robust standard errors reported in parentheses.

Figure 10 — Plotting profitability fitted values from model (2)



Notes: The blue line shows monthly cab profits while the orange line shows predictions based on estimating model (2).

Figure 10 plots actual monthly profitability figures as well as profitability fitted values derived from estimating model (2). The orange regression line is constant for many years because there were very few High Volume rides and thus this sector was not even part of the data set yet. The orange line accurately depicts the changes in profitability — including the downward trend beginning in 2015 and the slight leveling off in profitability after 2018. So, it appears model (2) holds up very well.

Table 2 shows the results from model (3), which regressed the log of profitability on utilization, fare density, fuel efficiency and fuel price. The slope estimates are statistically significant at the one percent level for all explanatory variables except for fuel efficiency (mpg_t). The results indicate that if daily utilization falls by one hour, profits decline by 6.8% on average, holding other factors fixed. If the fare density falls by $1, profits fall by 3.6%. Finally, a $1 increase in the price of fuel results in a 1.2% fall in profitability. The purpose of this is to illustrate how these various factors impact profitability, because they have changed considerably over the years.

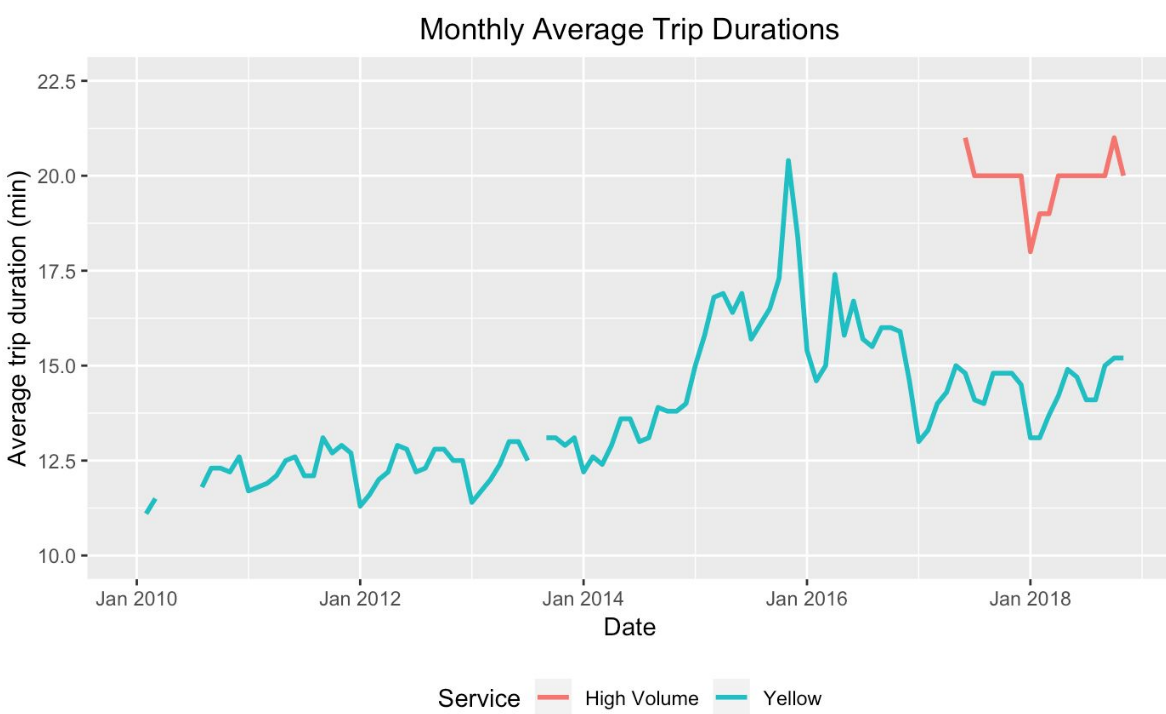
Table 2: Log(profitability) regressed on utilization, fare density, fuel efficiency & fuel price

	Model (3)
constant	7.386*** (.00756)
$util_t$	.068011*** (.0047817)
$faredensity_t$	.035664*** (.0017502)
mpg_t	.0007042 (.0017477)
$fuelprice_t$	-0.0124709*** (.0024404)
R^2	0.9963
Sample size	101

Notes: This table shows the results from estimating model (3). *** Significant at the one percent level. Heteroskedasticity robust standard errors reported in parentheses.

I calculated correlation coefficients between the number of High Volume rides and many other variables. The results are shown in Table 3. It's possible that some of the correlation is grasping a common trend, but the results are still pertinent. As one might expect, the amount of High Volume rides is negatively correlated with many variables, thus showing an overall contraction in the yellow cab industry. For instance, there is strong negative correlation between High Volume rides and the number of taxi rides. There is negative correlation between High Volume rides and the three variables relating to cab drivers; the increase in High Volume coincides with less taxi drivers, less working days each month for the drivers and also shorter shifts. There is strong positive correlation between High Volume and the total rides per day, indicating that rideshare spurred a market expansion. This is crucial to note. Rideshare does not solely steal rides from the taxi industry, but it also creates new rides.

Figure 11



Notes: This figure shows that rideshare trips are longer than taxi trips, on average.

There is also a strong positive correlation between High Volume and the percentage of yellow cab rides paid with credit cards, and this is quite interesting. One explanation is that general credit card use likely increased in this time. One could also make the following argument: there were barriers to using credit cards in medallion cabs in the past, and part of the reason customers like rideshare is that it is hassle-free since it does not use cash. The growth of rideshare, in turn, could have created pressure on the cab industry to be more open to credit cards. At the start of 2010, credit cards accounted for roughly one third of the transactions. By the end of 2018 though, it was up to 70%. Finally, the negative correlation between High Volume rides and the duration of yellow cab trips is also peculiar. It shows that rideshare services may be stealing some of the longer rides. This coincides with Figure 11, which shows that High Volume rides are significantly longer, on average, than taxi rides. Yes, rideshare services take away rides from the taxi industry; there is some substitution, and this harms medallion owners. But they also induce people to take longer rides that they simply weren't taking before. In part, they are offering a new product (longer rides) and this benefits consumers.

Table 3: Correlation between number of High Volume rides and other variables

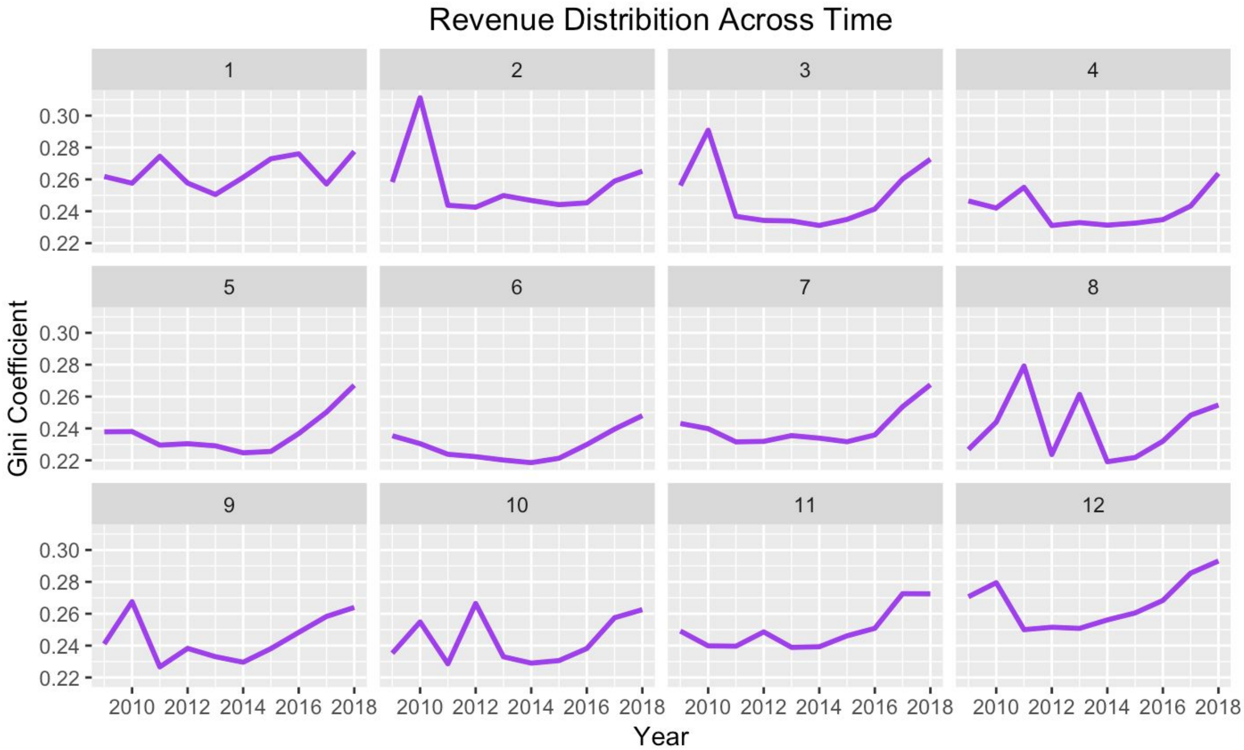
	High Volume Rides Per Day
Yellow Cab Rides Per Day	-0.9018
Fare Density	-0.8257
Avg Hours Per Day Per Yellow Cab	-0.8768
Total Rides Per Day	0.9726
Total Yellow Cab Industry Fare Per Day	-0.9122
Avg Days Yellow Cabs On Road Per Month	-0.4460
Yellow Cab Driver Count	-0.9674
Avg Days Driver On Road Per Month	-0.1505
Avg Hours Per Day Per Driver	-0.4468
Avg Minutes Per Trip	-0.6164
Percent Of Yellow Cab Trips Paid Through Credit Cards	0.8890

Notes: This table displays correlation coefficients between the average amount of High Volume rides per day, for a given month, and various variables from the taxi industry. The numerous negative correlations indicate a contraction in the taxi industry

B. Distributional Changes and Hourly Effects

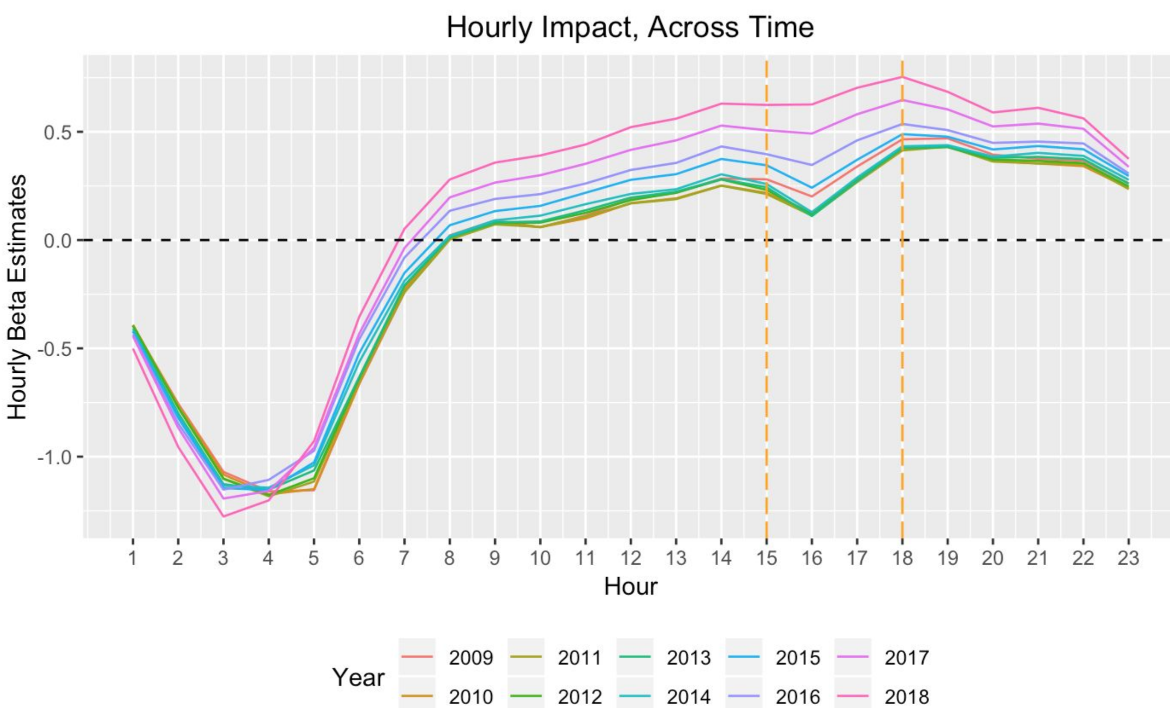
The main takeaway from the results discussed above is that rideshare impacts taxi profitability, which heavily influences medallion prices. I now shift focus to the specific channels through which profitability has changed, starting with changes in the distribution of profits. Figure 12 displays Gini coefficients for each month and how those coefficients changed over time. In all 12 of the subplots, there appear to be 'U' shapes. The underlying trend is that the Gini coefficients decrease for the first few years and then begin to increase. This indicates that revenue had been becoming more evenly distributed, but then as rideshare started growing the distribution became more unequal. The most recent Gini coefficients are relatively high, hinting that peak periods are becoming relatively more extreme. Results from model (4) further support this finding. Figure 13 plots estimates of that model's hourly coefficients (β_{32} — β_{54}). The chart can be understood as follows: looking at a specific year's trend line, each data point shows hourly effects on total fare relative to the first hour of the day (midnight to 1 am). For example, the first six observations on the pink line (2018) are below zero, which means that each of the hours between 1 am and 7 am generate less revenue than the first hour of the day — on average, for 2018. The pink line is above zero for the rest of the observations, showing that in 2018 hours between 7 am and midnight churn more total revenue than the first hour of the day, on average. A key pattern is that the most recent trend lines are the lowest for the first three observations — the slow period. And then in the busier period from 7 am to midnight, the recent trend lines are the highest. This confirms that the peakedness in earnings becomes more prominent while rideshare grows. Looking at the trend lines between 3-6 pm (hours 15 through 18 on the figure) reveals the evolution of the afternoon shift change. Earlier years show substantial drops during those hours, while recently the drop has become less and less prominent. This is in agreement with the initial takeaway from Figure 8. There are plenty of potential explanations for the lessening magnitude of the shift change. The taxi market faced shrinkage starting in 2014 around the advent of rideshare. Perhaps as revenues shrunk, cab drivers were forced to adapt and became more efficient when changing shifts. Maybe now, drivers need to capitalize on those hours more than they've needed to in the past, thus leading to a less eminent dead period.

Figure 12 — Gini Coefficients, by Month



Notes: These charts track Gini coefficients for monthly distribution of taxi revenue. Numbers 1 through 12 indicate the month.

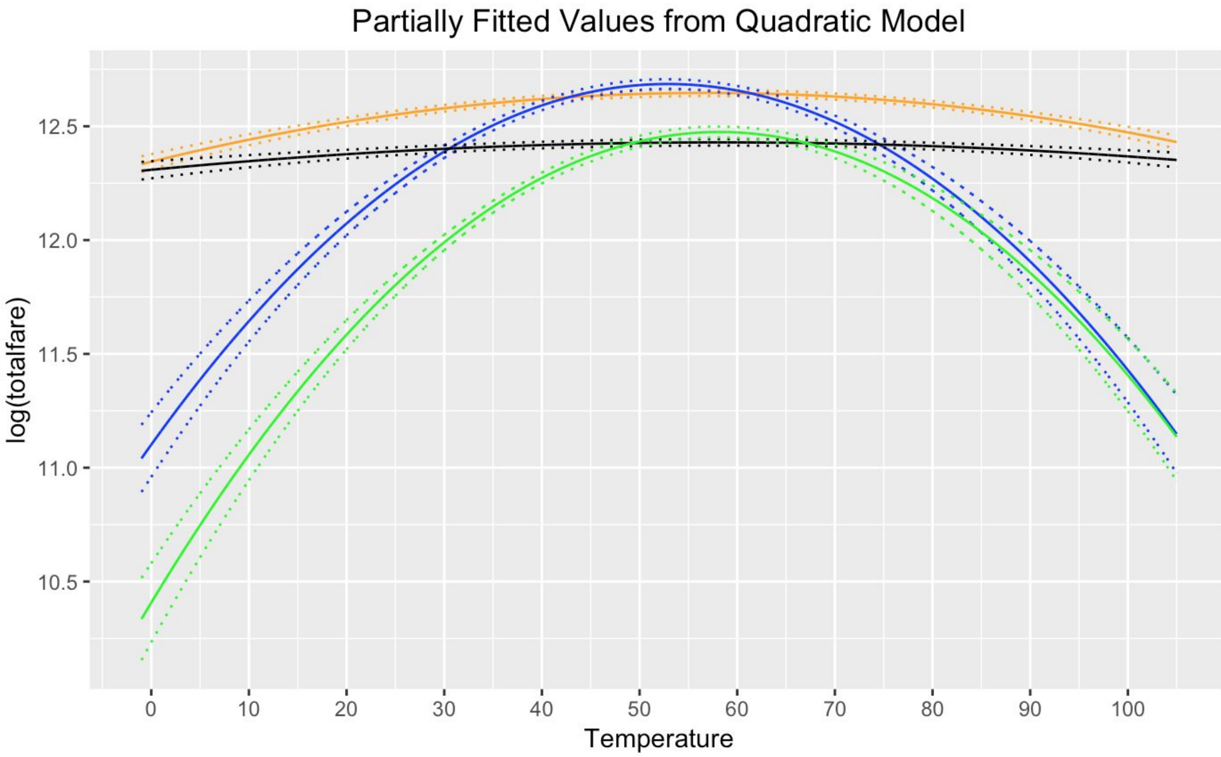
Figure 13 — Hourly Coefficient Estimates from Model (4)



Notes: This chart displays estimated hourly effects on log(total fare) for each of the 10 years. Increased peakedness is evident in recent years, and there's also less of a slow period during the shift change (the period between the two vertical lines).

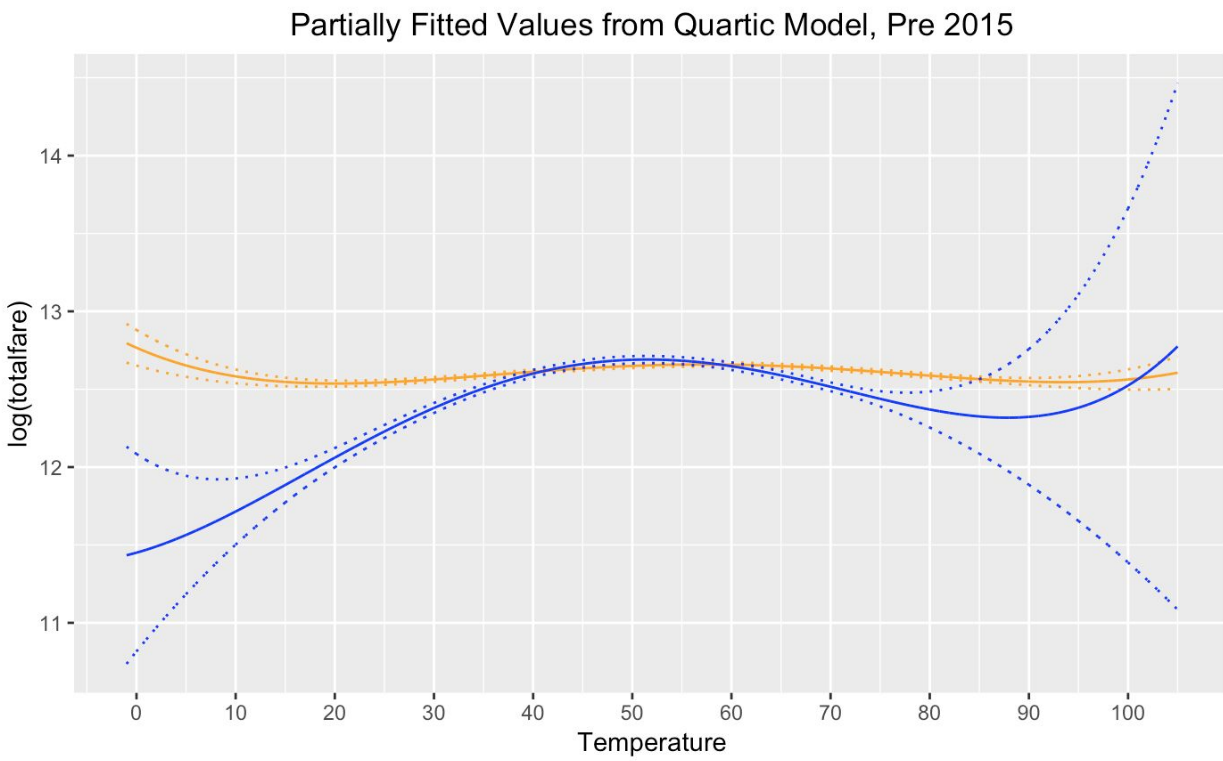
C. Weather

Figure 14



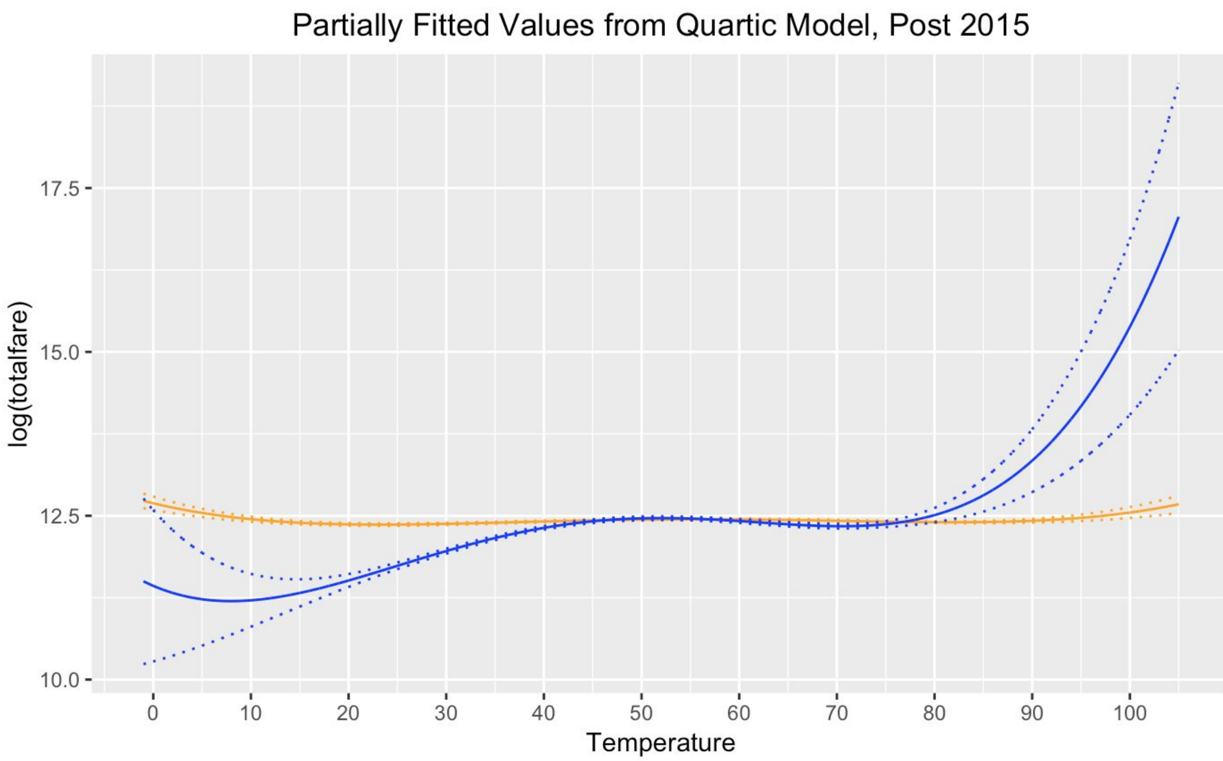
Notes: This chart displays partially fitted values based on the quadratic specification from model (5). The orange and black curves are the non-precipitation functions from the pre and post periods, respectively. The blue and green curves are the functions with precipitation from the pre and post periods, respectively. Dotted lines indicate 95% confidence intervals.

Figure 15



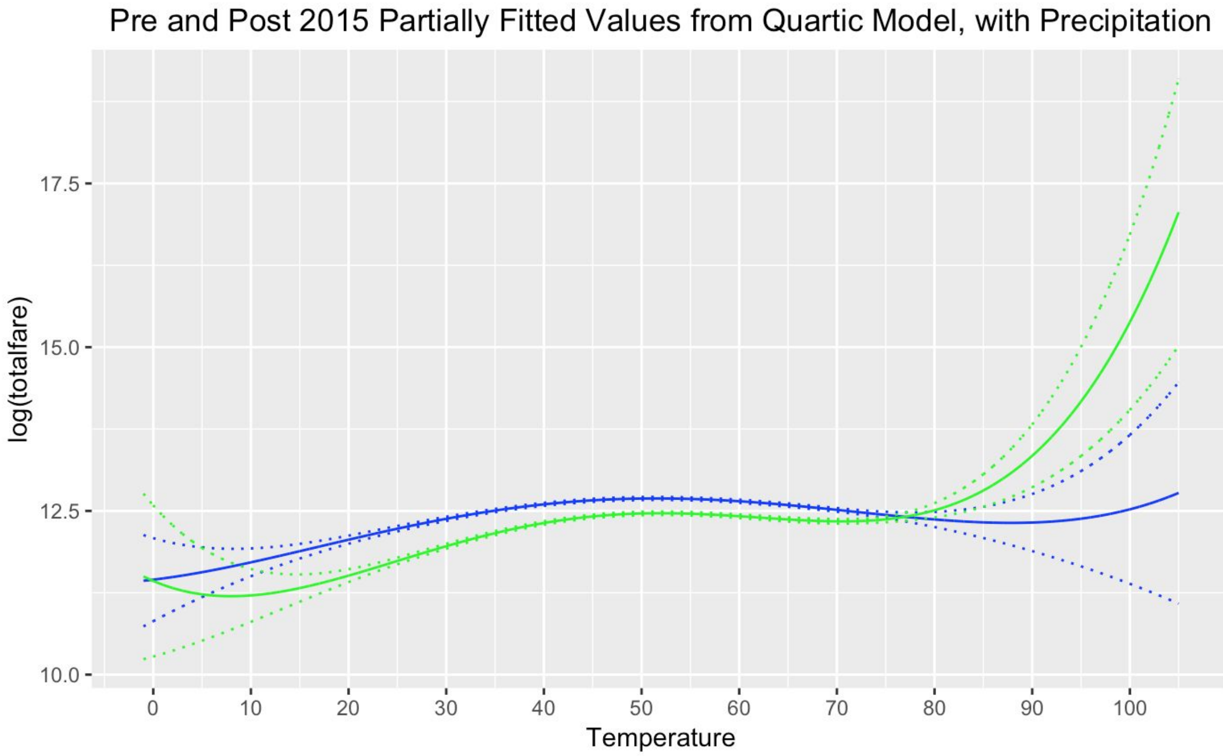
Notes: This chart displays partially fitted values based on the quartic specification from model (6). The orange and blue curves are the non-precipitation and precipitation functions from the pre 2015 period, respectively. Dotted lines indicate 95% confidence intervals.

Figure 16



Notes: This chart displays partially fitted values based on the quartic specification from model (6). The orange and blue curves are the non-precipitation and precipitation functions from the post 2015 period, respectively. Dotted lines indicate 95% confidence intervals.

Figure 17



Notes: This chart displays partially fitted values based on the quartic specification from model (6). The blue and green curves are the precipitation functions from the pre and post 2015 periods, respectively. Dotted lines indicate 95% confidence intervals.

Results from the quadratic (5) and quartic (6) weather approximations are exhibited on pages 27 through 30 via partial fitted value functions. There is a link between weather and taxi profitability, and through examining the charts it's more clear how rideshare has impacted this channel. The dotted lines represent 95 percent confidence intervals. Due to some of the significant overlapping of confidence intervals in the figures at high temperatures — especially evident in Figure 15 — I focus more on the model results for the low to mid temperature range. Figure 14 shows that under the quadratic model, hourly fare is substantially more sensitive to temperature changes when there is precipitation. This is evident through the more limited variation in the orange and black curves compared to the blue and green. This increased temperature sensitivity with precipitation is apparent under the quartic model, too, in Figures 15 and 16. A key finding from all three charts is that at the low end of the temperature spectrum, precipitation correlates with a decrease in hourly taxi fare. At more moderate temperatures, though, precipitation can actually slightly increase revenue. The following is a possible justification: with precipitation there is more demand for rides, regardless of the temperature. But the combination of precipitation and low temperatures creates harsh driving conditions, so the taxi drivers may not have enough of an incentive to look for passengers. This is especially prevalent since the taxi industry lacks the surge pricing necessary to promote driving during rough conditions. As temperature increases, though, conditions improve and drivers seize the excess demand.

One final observation, from Figure 14, is that the gap between pre and post precipitation curves (blue and green) is wider under low temperatures compared to the middle range. Figure 17 compares the quartic functions with precipitation for the pre and post periods. The blue and green curves represent pre and post 2015, respectively. The confidence intervals overlap at the far left side, so one shouldn't focus on the extreme low temperatures. But comparing the 15 to 45 degree range with the 45 to 75 degree range, the gap between the blue and green curves is larger with the colder temperatures. This coincides with takeaways from Figure 14. One potential explanation is as follows: rideshare services have surge pricing, so they can encourage driving under bad weather. Perhaps the cab drivers who are still willing to drive in tough conditions thus face more competition. This leads to an even larger reduction in fare under precipitation and cold weather, compared to just precipitation and moderate temperatures, and this is consistent with the changing size of the gaps between curves.

Conclusion

All of this analysis is pertinent because New York City created the medallion system to reduce tensions in the industry. Many invested life savings into medallions because they seemed like sound investments. There was truth to that for a while, as the taxi industry was successful for generations. This study shows that with rideshare though, taxi revenues and medallion values have plummeted. The industry outlook has become bleak. Thus, the tensions are front and center again. Broader implications of this paper pertain to government regulations on the supply dynamics of an industry. The New York City taxi industry is a prime example of the delicate balance between market regulation and innovation. In a sense, the medallion system warped the supply-demand dynamics in the city's ride service industry, leading to a somewhat artificial increase in medallion value. Rideshare companies demonstrated that such long-standing market regulations often create untapped opportunities. They used technological change to circumvent supply regulations and the implications were tremendous. Consumers appear to have benefitted, due to overall increased availability of rides via market expansion. But medallion owners were drastically harmed, and multiple taxi drivers reportedly committed suicide due to decimated investments. One can only wonder how events may have unfolded differently if the medallion system never existed. Change was necessary to calm the tides in the 1930s, but perhaps less stringent regulations would've been more ideal.

Overall, this paper showed a range of things. I presented weather and time of day as critical components in the taxi industry. Then, I examined changes in the relationship between those channels and taxi revenue while rideshare grew. My analysis showed growth in rideshare has led to significant reductions in taxi profits, which is pivotal since I revealed — via the Gordon Growth Model — that medallion prices are very sensitive to taxi profitability. Further research could focus on specific ride locations to see if medallion cabs are losing rides in a particular part of the city. Subject to availability of Uber and Lyft data, one might also extend my revenue distribution and weather analyses to the Uber and Lyft markets. Finally, how New York City responds to the changing industry is a stimulating topic for future study. For instance, the city tried placing a cap on the amount of rideshare vehicles allowed. Time will tell how this — and potential other actions — will play out. One thing is certain: the history of the medallion system provides a special lens for designing better regulations.

References

"Average Fuel Efficiency of U.S. Light Duty Vehicles." Bureau of Transportation Statistics. United States Department of Transportation .

Brodeur, Abel, and Kerry Nield. "Has Uber Made It Easier to Get a Ride in the Rain?" *University of Ottawa*, May 2017.

Camerer, C., L. Babcock, G. Loewenstein, and R. Thaler. "Labor Supply of New York City Cabdrivers: One Day at a Time." *The Quarterly Journal of Economics* 112, no. 2 (1997): 407–41

"Consumer Price Index." Federal Reserve Economic Data. Federal Reserve Bank of St. Louis.

Cramer, Judd, and Alan B. Krueger. "Disruptive Change in the Taxi Business: The Case of Uber." *American Economic Review* 106, no. 5 (May 2016): 177–82.

"Data." Data and Research. NYC Taxi and Limousine Commission. Accessed April 6, 2020.

Falsetta, Gabe, and Gary Bono. "The Decline and Fall of the NYC Taxi Industry." People's World, August 29, 2017.

Farber, Henry S. "Why You Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers*." *The Quarterly Journal of Economics*, vol. 130, no. 4, (2015): 1975–2026.

Fitzsimmons, Emma G. "Why Are Taxi Drivers in New York Killing Themselves?" *The New York Times*, December 2, 2018.

Frechette, Guillaume R., Alessandro Lizzeri, and Tobias Salz. "Frictions in a Competitive, Regulated Market Evidence from Taxis." *SSRN Electronic Journal*, 2015.

Gelder, Lawrence V. "Medallion Limits Stem From the 30's." *The New York Times*. May 11, 1996.

Gordon, Myron J., and Eli Shapiro. "Capital Equipment Analysis: The Required Rate of Profit." *Management Science* 3, no. 1 (1956): 102–10.

Haggag, Kareem, Brian Mcmanus, and Giovanni Paci. "Learning by Driving: Productivity Improvements by New York City Taxi Drivers." *American Economic Journal: Applied Economics* 9, no. 1 (2017): 70–95.

Hall, Jonathan, and Alan Krueger. "An Analysis of the Labor Market for Uber's Driver-Partners in the United States." *NBER*, November 2016.

"NEWA Hourly Weather Data Page." Network for Environment and Weather Applications. Cornell University.

"New York City Regular All Formulations Retail Gasoline Prices." Petroleum & Other Liquids. U.S. Energy Information Administration.

Prial, Frank J. "When the Taxi Customer Complains, the Hack Bureau Listens." *The New York Times*. January 13, 1971.

"Regulation and Prosperity: 1935 - 1960." History. NYC Taxi and Limousine Commission.

Zhang, Wenbo, Dheeraj Kumar, and Satish V. Ukkusuri. "Exploring the Dynamics of Surge Pricing in Mobility-on-Demand Taxi Services." *2017 IEEE International Conference on Big Data*, 2017.