Willis, G., & Tranos, E. (2020). Using 'Big Data' to understand the impacts of Uber on taxis in New York City. *Travel Behaviour and Society*, 22, 94-107. <https://doi.org/10.1016/j.tbs.2020.08.003>

Peer reviewed version

License (if available):
CC BY-NC-ND

Link to published version (if available):
[10.1016/j.tbs.2020.08.003](https://doi.org/10.1016/j.tbs.2020.08.003)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Elsevier at <https://doi.org/10.1016/j.tbs.2020.08.003> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Using ‘Big Data’ to understand the impacts of Uber on taxis in New York City

George Willis¹, Emmanouil Tranos¹,

Abstract

Uber has coupled their ability to dodge regulation with their ride-hailing app, to grow from a Silicon Valley start-up, to a global powerhouse challenging established taxis in over 700 Metropolitan areas. Through the lens of Schumpeter’s ‘Creative Destruction’ and Roger’s ‘Diffusion of Innovation’, this paper aims to investigate how Uber challenges the established taxis, focusing on yellow taxis in New York City. To capture Uber’s popularity we employ innovative new data from the New York Times API. We couple these data with data about taxi usage and quality of service from the New York City Taxi & Limousine Commission. We then employ well established time-series modelling techniques in order to test if and how Uber affected the well established taxi market in New York City. Our results illustrate the disruptive nature of Uber in terms of leading to a decrease in traditional taxi rides, but also to a behavioural change of customers raising more complaints about taxi services.

Keywords: Uber, big data, time series, Granger causality, New York City, yellow taxis

Introduction

Founded in 2009, Uber is a ride-hailing app that aims to connect drivers with customers through their technology platform. Based in San Francisco, the self-proclaimed technology company is at the heart of major court cases and mass-demonstrations across the world (Bowcott 2017). Despite this, Uber has continued to grow, and it has now penetrated over 600 cities in 82 countries across the world. In many cases, it has been argued that Uber revolutionises a taxi market that was stuck in the past and in desperate need of an update (Cordray 2015). Using a novel source of big data, this article illustrates the impacts that Uber has had on the traditional yellow taxi in New York City.

When creating the medallions system, regulators did not anticipate the rapid rise of ‘ride-hailing’ businesses such as Uber or Lyft. While it has been accepted with open arms by consumers in some cities, it has been disputed, and in some

Email addresses: gcw519@student.bham.ac.uk (George Willis), e.tranos@bristol.ac.uk (Emmanouil Tranos)

cases, banned. Uber uses its technology platform to link ‘self-employed’ drivers with consumers looking for a ride, taking a small percentage of the fare for itself. A major problem that traditional taxis face, is that Uber and Lyft can avoid regulation and therefore compete unfairly in the market (Malhotra and Alstyne 2014). Governments do not have strong regulatory frameworks that suit the new technology/transport companies (Posen 2015). For example, Uber drivers can operate without a taxi medallion or taking any licensing exams. This immediately puts the established taxi firms at a disadvantage in the market and has caused disruption in many cities such as London and Vancouver (Rhodes 2017).

However, while some groups have openly criticised these new companies, other have greeted them as they aim to revolutionise a dated industry (Cordray 2015). In 2014, the price of a medallion reached a peak of around \$1.3 million, compared to 1937 where the price was \$10 (\$150 in current prices). This increase perhaps suggests there was an under-supply of taxi medallions, and thus taxis. This, therefore, suggests that before the entrance of Uber, there was a strong case for an increased number of taxis. Cervero (1985, 226) has suggested that as the availability of taxi service improves, smaller taxi companies or even individuals will be able to service “marginal markets abandoned by large fleets”. Furthermore, Hall, Palsson, and Price (2018) find that Uber can compliment, rather than hinder public services in transport. Research by FiveThirtyEight provides evidence to suggest that Uber has served the outer reaches of New York City far better than the traditional yellow taxis, rather than being focused in the centre of New York City.

Despite the controversies linked to Uber, its impact on the taxi industry is difficult to be illustrated as, among other things, Uber does not provide access to their data.¹ This paper aims to overcome this problem by employing novel and diverse sources of big data and analysing them using data science methods and statistical modelling. To do so, we build a proxy of Uber’s popularity by mining the New York Times (NYTimes) – New York’s prominent newspaper – Application Programming Interface (API). This enables us to create counts of NYTimes articles, which refer to Uber, over time and, therefore, we are able to capture the volume of interest in the public realm about Uber and how this changes over time. Researchers from other fields have employed data from public media before to capture the public interest (Damak et al. 2011; Raghavan and Anshul 2013; Lee and Kim 2008), but to our knowledge this is the first time that such data is utilised in transport studies

While Uber does not provide much data, the TLC has provided data containing every taxi trip since 2009. We also include in our data set other variables to control for external conditions which may affect taxi usage including weather and income and time. We visualise our data and we then employ time series modelling to capture the effect of Uber on taxi usage. Specifically, we are inter-

¹Uber has only recently started providing some aggregated data via <https://movement.uber.com>.

ested in illustrating whether the public interest in Uber, as reflected in NYTimes articles, has led to a decrease to taxi usage in New York City. Moreover, we are also interested in exploring whether this increased competition has led to an improvement in the quality of service provided by taxis.

Our empirical work builds upon some key theoretical developments. Uber is approached as a disruptive innovation (Christensen and Raynor 2005) and we use the work of Rogers (1995) to understand how it diffuses. We then adopt a Schumpeterian view point to explain how the creative destruction initiated by Uber can lead to service quality improvements (Schumpeter 1942).

The structure of the paper goes as follows. The next section discusses the relevant literature. Then, we present the methodological framework and the data we utilise to answer the research questions. The next section presents the analysis and the results and the paper ends with a conclusions section.

Literature Review

Regulation within the taxi market

The main focus of this paper is the taxi industry, which is well known for being heavily regulated, and entrance into the market is often controlled (Gilbert et al. 2002). In 1937, policy-makers in New York city decided to control the number of taxis in the city by using a medallion system. This meant capping the number of taxis allowed to operate in the city by controlling the number of medallions (which were required to operate), immediately removing the free market of taxis in New York City. The number of taxis to be licensed by regulators is a vital decision. Oversupply of taxis can lead to unsustainable prices, high vacancy rates (Zhang and Ukkusri 2016) and poorly-qualified drivers (Schaller 2005). However, under-supply can cause extortionately high prices, long wait times and can lead to certain areas being deprived of service at all (Benz and Sumares 2015).

The number of medallions is often controlled by political forces, rather than market conditions. While there is an abundance of literature regarding the regulation of taxis in the United States, little has been produced as to how the demand of taxis can change and what this means for the traditional yellow taxi. An exception is the work of Schaller (2005), who employs statistical modelling to determine the factors which affect the number of taxis required.

Uber's impact on the taxi market

While Uber's ability to avoid regulation has enabled it to challenge the traditional taxi firms, it has also made its operation controversial. In Chicago, medallion holders sued local regulators over their alleged failures to regulate Uber (Sidak 2016). In Taiwan, Chang (2017) found Uber reduced regular taxi drivers' service revenue by approximately 18 percent in the third year of entry of Uber, while also suggesting that in order to compete with Uber, taxis were more likely to provide a better service to their customers. In 2015, Rogers (2015) suggested that despite both the media and academic attention that Uber was

receiving, little had actually been revealed in regards to the real impact of Uber on public goods and services. Watanabe *et al* (2017) provide numerous sources contrasting both the success and failures of Uber expansion into various markets, further highlighting the need to investigate the empirical impact of Uber behind a strong theoretical background. One such method is Dudley (2017) who uses Schumpeter’s concept of ‘creative destruction’ as a means to explain Uber’s impact, describing it as a disruptive innovator. Joseph Schumpeter first coined this term in 1942 in order to depict an economic process that “revolutionizes the economic structure from within, continually destroying the old one, incessantly creating a new one” (Schumpeter 1942, 83). Uber resembles such a process as traditional taxis have suffered a loss of business and it has been suggested that unless regulated strongly, they would eventually die out (Holt 2016). While Schumpeter’s theory is a wider, macro view of innovation, Christensen and Raynor (2005) takes this further by focusing on smaller scale creative destruction, termed ‘Disruptive Innovation’. Markides and Geroski (2006) suggest that established firms should not aim to disrupt the market and instead leave it to the smaller scale firms. This is exactly what Uber did when they started their operation as their aim was to disrupt the market as a small-scale enterprise. Cramer & Krueger (2016) found that Uber were able to have higher capacity utilization rate than taxi drivers in multiple cities across the United States, perhaps caused by their novel technology platform. The wide usage of smartphones and geolocation technologies allowed Uber to create a product which the taxi market had never seen before. Their technological platform radically changed the market as they drastically undermined competition from established transport services. This paper aims to find how much Uber has disrupted the market by applying this theory to the New York City taxi market and how it has been affected by Uber’s diffusion.

A lot of effort has been spent in order to theorise how innovation processes evolve and diffuse over time. Rogers’s (1995) influential theory of ‘Diffusion of Innovation’ understands the evolution of an innovation as an orderly sequence of events (Robertson 1967). Rogers proposes that there are 5 ‘adopter categories’ for a new technology or idea (shown in Figure 1). These groups are classed by their ‘innovativeness’, meaning how quick they are to adopt a new technology. Innovators and early adopters are seen to be more social, cosmopolitan, and have a greater exposure to mass media channels (Rogers 1995). The early adopters are also seen to have the highest degree of opinion leadership and are often people working in review magazines and newspapers, where potential adopters would look for advice (Rogers 2002). Later adopters will only accept the new idea once their surrounding peers have already adopted it and are satisfied by this; for example, good reviews in the media. To reach a critical level of take-off, innovations need to reach around 15-20% adoption by the population. For instance, by launching in San Francisco, Uber was able to grab young, tech-savvy adopters (high opinion leadership) to gain their base (Leddin 2015). This technology community is always looking for ways to make life simpler with new technologies. Increasing reports and opinion articles supported Uber’s growth. Uber themselves estimated that “every 7 rides, create a new Uber user” (Kriss

Figure 1: Rogers' Diffusion of Innovation (Rogers 1995)

2014). Our proxy of the popularity and the public interest in Uber, as reflected in the counts of NYTimes articles, as well as the longitudinal nature of our research framework and data enable us to test how Uber has led to disruptive effects on traditional taxi services. Rogers' model can not only be used to see how Uber was able to expand so quickly, but also to track how Uber continues to grow and what we can expect in the future.

Benefits of Competition

The negative effects of new competition such as predatory pricing (Baloch 2018) and the plummeting prices of taxi medallions (Wallsten 2015; Gabel 2016) have been well documented both in academic literature and the media. The Mayor of New York City stated it was necessary to cap the number of ride-sharing vehicles to limit congestion in the city, however, this claim was found to be baseless by Li, Hong, and Zhang (2016). Despite these effects, there are many positive benefits for the taxi market that need to be considered. Established companies may see the new competition as an incentive to increase the quality of their own service. Kranton (2003) shows that in a situation where consumers have lots of choice and base their choices on quality, competition provides companies an incentive to increase quality of service. Hirschman's (1970) study provides a framework on how consumers express their 'voice' through complaints. He suggests that consumers will express their dissatisfaction with the service through complaints. Recent empirical studies have utilised this framework in order to investigate how increased competition may affect the quality of transport services. For instance, Beard, Macher, and Mayo (2015) and Forbes (2008) utilised data from transport providers about user complaint and both illustrate an inverse relation between transport service quality and competition. Hence, one might expect that as Uber entered the market, traditional taxis may have tried to improve their quality to keep their market share (Wallsten 2015). This may include better up-keep of their vehicles, fixing broken card machines and better customer service as Wallsten (2015) found in Chicago. As Uber was queried more on Google, he was able to find a decrease in complaints of broken credit card machines as the taxis improved their service in response to the increased

competition. Complaints may provide a proxy of the quality of traditional taxis in the city and thus provide a means to investigate the service quality differences in yellow taxis since the arrival of Uber. The next section discusses the data and the methods we utilise for this paper.

Research framework

Methodological approach

In order to answer the empirical questions stated in introduction we will estimate two empirical models, which will have the following general forms:

$$TaxiUsage_t = \beta_0 + \beta_1 UberPopularity_{t-1} + \beta_2 Controls_t + \epsilon_t \quad (1)$$

$$TaxiQuality_t = \beta_0 + \beta_1 UberPopularity_{t-1} + \beta_2 Controls_t + \epsilon_t \quad (2)$$

Equation 1 will test whether taxi usage in time period t has been disrupted because of Uber popularity in period $t - 1$ after controlling for a number of other variables, which might also affect taxi usage in New York City. As briefly discussed in the introduction section we are not able to directly observe Uber usage and, therefore, we propose a proxy of Uber' public interest and popularity. The data we use for this proxy is analysed below in this section, but it is worth saying here we use a yearly lag of this variable ($t - 1$) to partially address potentially endogeneity issues related to reserve causality. Similarly, Equation 2 will test whether the quality taxi services has been affected because of the increased competition. Again, we will employ the Uber popularity proxy in time $t - 1$.

To estimate the above models we employ well established time series techniques. After estimating the models presented in Equations 1 and 2 using Ordinary Least Square (OLS), we explore whether the residuals suffer from autocorrelation. The existence of such a structure in the residuals would violate the OLS assumptions. According to the tests presented in the analysis and results section this is indeed the case. Therefore, we also estimate these models using Generalized Least Squares (GLS), which can account for autocorrelation (Brockwell and Davis 1987).

In addition, we also employ Granger causality tests. These are widely used tests in time series analysis, which enable to test whether, for example, the popularity of Uber in time t proceeds the decreases of yellow taxis rides in time $t + 1$. The starting point of Granger causality tests is a bivariate model such as the one represented by Equation 3, where variable y is regressed against lagged values of y and lagged values of x . The Granger causality test examines whether the inclusion of lagged values of x improve the prediction of y or whether the y is better predicted by including only the lagged values of y (Granger 1969). To perform this test, Equation 3 is estimated using OLS including both lagged

values of y and x and then without the lagged values of x . A Wald test is then performed to indicate the preferred specification, which illustrates whether the lagged values of x improves the prediction of y or not. While the former result indicates that x Granger-causes y , the latter would signify the opposite (Hurlin and Venet 2001; Tranos 2012; Tranos and Mack 2016). In essence, Granger causality tests comprehend causality as temporal precedence between y and x . To test the direction of causality we follow the above process twice and interchange y and x .

$$y_t = \sum_{j=1}^k \alpha_1 y_{t-j} + \sum_{j=1}^k \alpha_2 x_{t-j} + \epsilon_t \quad (3)$$

Data

In order to estimate the above models and address the various data gaps, including capturing Uber usage, which are discussed in detail below we will employ a variety of data some of which fit under the ‘big data’ umbrella term. The growth of new sources of big data has been spurred on by the improvements of technology, allowing sensors and computers to be embedded within almost every conceivable type of object (Batty 2013). As citizens interact with these objects, individual records are captured usually at a real-time basis leading to an unprecedented volume of data being produced. These data are characterized by vast volumes, that are created in near-real time while also being highly detailed, allowing a wide range of applications (Kitchin 2013). Transport researchers were always looking for more granular and extensive data and, therefore, the interest towards such new data sources should not come as a surprise (Winter et al. 2011; Tranos and Mack 2019). However, such data do not come without biases and epistemological as well as technological challenges (see the discussion in Milne and Watling (2019)).

To begin with the taxi market and its fluctuation, the New York City Taxi and Limousine Commission (TLC) has released a vast data set under New York State’s Freedom of Information Law, providing every yellow taxi journey taken in New York from 2009 to present. The entire data set provides trip-level data on well over 1.3 billion taxi journeys, including taxi price, dates, and times of each journey. Importantly, these data are public domain and can be freely download.² These data are collected by independent technology providers authorised under the Taxicab & Livery Passenger Enhancement Programs (TPEP/LPEP) (TLC 2017). Given the size of these stream data, extensive data cleaning processes were applied in order to utilise them. Hence, we were able to collect the daily number of taxi journeys from 2009 to 2016 and create a time series variable, which is the dependent variable for the models presented in Equation 1.

²<https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

The taxi service quality data, which is the dependent variable for the models in Equation 2 come from the New York City’s Open Data Platform³. Specifically, they provide the number of complaints submitted about yellow taxi to the TLC per month. Following the competition discussion from the literature review section, these data are used to measure the quality of taxi service provision. To account for the fact that rides may be decreasing and, therefore, the absolute number of complaints may also be decreasing, we standardise this variable by dividing it by the number of yellow taxis ride in New York City (complaints per 1000 trips by yellow taxis).

As discussed in the literature review section, the diffusion of innovation concept is the foundation for understanding how popular Uber is in New York City. Given the lack of Uber usage data, we build a proxy variable to measure its popularity and the public interest in Uber. Specifically, we employ articles published by the NYTimes, which include a reference to Uber and we build a variable that captures the volume of interest in the public realm about Uber. To do so, we use the NYTimes open API, which enables users to search all their articles from 1851 to present using, among other filters, user-defined keywords. We queried the API using the terms ‘Uber’ and ‘taxis’ to both appear in the same article, from 2009 to 2016 in New York City. We specified ‘taxis’ because our preliminary analysis indicated that many articles contained the word Uber, not relating to the transport company. From this, we were able to produce a variable showing how many articles were published about Uber since 2009 and when and, consequently, build a longitudinal measure of the public interest on Uber.

Researchers have employed data from public media before to capture the public interest. For instance, Damak et al. (2011) used the NYTimes API to extract keywords of major stories to compare to microblogs on Twitter. Raghavan and Anshul (2013) employed keywords from the main text or headlines of NYTimes articles to extract emotions from news articles. Lee and Kim (2008) used the number of news articles as a proxy to track the popularity of a subject within the media, and, therefore its impact on the public. In the same vein, we use here the number of relevant NYTimes articles to track the coverage of Uber in the media and, therefore, track the popularity of and the public interest in Uber over time. Interestingly, such data have been utilised in various research fields, but not in transport studies. A potential limitation is that the NYTimes API only provides access to the main headline or paragraph, meaning some relevant articles may be missed and, therefore, our measure may underestimate Uber popularity. This also prevents us from conducting in-depth analysis of the opinions expressed in these articles, which might be a subject for further research. However, given that we interpret the counts of Uber in NYTimes articles as a proxy for the public interest in Uber, the inclusion of articles with potentially negative sentiments in the counts would have not contaminated our sample.

Transport systems are not independent and, therefore, there is a need to also

³<https://opendata.cityofnewyork.us/>

control for the quality of the broader transport service provision in New York City. For instance, if there is a closure of a subway station, commuters may decide to use taxis instead leading to an increase usage for taxi or Uber usage for a reason different than the substitution effect. Although the Metropolitan Transportation Authority collects stream transit data about New York subway, they do not provide archived data⁴. To overcome this data gap, we use Google Trends, which collects all queries of a search term globally, and from this provides an index by geographic location and category. The index begins with the total query volume for a search term in a geographic region divided by the total number of queries in that region at a point in time (Choi and Varian 2012). This number is then normalised to begin at 1 at the beginning of the period and any deviations in later dates are given in percentages for this origin. For our data we queried ‘Subway down’ in New York City at a weekly time interval from 2009. Hence, we use this variable as a proxy for instances when the subway was not operating.

We also need to control for fluctuations in disposable income, which might also affect the use of taxi and Uber. Given that the granularity of the above data is weekly, we are ideally looking for a measure that could depict income fluctuations at a similar level of temporal resolution. Because official income data is only available on a yearly basis, at least for metropolitan areas, we employ data from the New York Stock Exchange (NYSE) and use it as a proxy of the economic climate of New York on a weekly basis. Specifically, we utilise the stock market index of the New York Stock Exchange, which is a daily composite index covering all common stocks. For our analysis we utilise the weekly median, but we also create yearly aggregates in order to calculate correlation between the yearly median of the NYSE index with the yearly income of New York city. Even though we only have 8 data points, Figure 2 indicates a rather strong relationship (0.8 Pearson coefficient).

Weather is also a well established determinant of transport demand (e.g. Koetse and Rietveld 2009; Zhou et al. 2017). When looking at the daily taxi trip data the extreme lows on some days correspond with severe weather conditions. The lowest number of trips (29,028) occurred on the 28th August 2011, the same day Hurricane Irene hit. A similar pattern occurred for Hurricane Sandy (113,808 trips) and the North American Blizzard (71,833 trips) showing that weather clearly affects the number of taxi journeys and, therefore, we needed to control for this. To do so we employ data from National Climatic Data Centre⁵. Specifically, we use data about daily precipitation and snowfall from the Central Park weather station in New York City. These data enable us to control for daily weather conditions as well as seasonality.

Table 1 shows the descriptive statistics for the independent and dependent variables. The Yellow taxi and weather data were initially provided at a daily

⁴<http://web.mta.info/developers>

⁵<https://www.ncdc.noaa.gov/cdo-web/datasets/GHCND/stations/GHCND:USW00094728/detail>

Figure 2: Correlation of NYSE with income

basis. However, Subway data were only available at a weekly temporal resolution. Furthermore, while the NYTimes data were available at a daily resolution, there were not enough articles published daily about Uber to give a meaningful value. Hence, we aggregated the daily variables to weekly bins by using the median for each week, meaning all our data was in the same resolution for analysis. We then divided the yellow taxi trips into peak taxi times (06:00-10:00 & 16:00-20:00), evening peak taxi times (16:00-20:00), and non-peak taxi times. These three subsets were chosen for two reasons. The evening peak fare is based on the traditional taxi peak fare surcharge which is 4pm to 8pm, suggesting this is when they feel this is the taxi peak.⁶ The second subset was based previous research by Hochmair (2016) who shows that for selected districts in New York, the peak taxi usage begins at 06:00. Kamga, Yazici, and Singhal (2015) also show that taxi trips tend to see a rise at around 05:00, unlike subway usage which is often later in the morning. At 17:00 the taxis begin to increase again as people finish work, causing a second peak in taxi journeys. This morning peak is then added to the evening peak to form our second subset. Therefore we have a peak, based on the traditional taxi surcharge, a peak based on actual taxi trips, and a non-peak dataset. This allows us to compare the daily fluctuations,

⁶<https://www1.nyc.gov/site/tlc/passengers/taxi-fare.page>

which may have an impact on our results. Only the number of complaints of yellow taxis were not in a weekly resolution as they were collected monthly from 2010-2016. Therefore for the complaints analysis, all data were aggregated to a monthly temporal resolution, using the same method as above.

Table 1: Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Trips	416	449,799.30	57,800.45	281,208.10	406,661.50	495,875.10	609,561.30
Precipitation	416	0.95	1.06	0.00	0.19	1.41	7.36
Snow(mm)	416	0.75	2.88	0.00	0.00	0.00	27.50
Min Tem	416	48.99	15.96	11.43	35.96	63.61	76.86
Max Tem	416	63.32	17.25	26.14	48.68	79.29	92.43
Subway	416	4.70	7.32	0.00	2.00	6.25	100.00
NYSE	416	8,783.93	1,696.94	4,334.70	7,459.72	10,432.10	11,213.15
NYTimes articles	416	1.27	2.24	0.00	0.00	2.00	20.00
Peak Taxi Trips	416	162,885.60	22,028.03	82,550.14	148,985.50	179,925.30	242,809.40
Evening Taxi Trips	416	96,047.92	12,335.54	48,252.00	88,311.93	105,221.50	139,015.40
Complaints	84	1,520.32	271.93	922.00	1,316.25	1,693.25	2,328.00

Analysis and results

After discussing the relevant literature and the data we use, this section presents the analysis and discusses the results. We first utilise our methodological framework to test whether Uber popularity, as reflected in Uber’s mentions in NYTimes has affected taxi rides in New York City. We then turn the focus on the quality of taxi services and whether this has been affected by Uber.

Taxi usage and Uber

Figure 3 plots the number of yellow taxi trips between 2009 and 2016. This graph clearly indicates that prior to Uber’s entrance in the market, the number of trips taken by yellow taxis was increasing. However, after that point the numbers begin to fall sharply. This perhaps indicate that Uber may well be a disruptive innovator directly affecting previous technologies such as the traditional taxi market (Christensen and Raynor 2005). Furthermore, Figures 4 and 5 present taxi usage on a daily basis for the first and last year of the study period. These figures clearly indicate that the volume of taxi trips follows a seasonal pattern, with a drop in December and January, which may be attributed to bad weather and, therefore, decreased demand for mobility. The lowest value in 2016 is 23rd January, which is the date of the 2016 United States blizzard, which resulted in a travel ban being imposed across New York City (Yuhua and Glenza 2016). On top of illustrating the temporal variability of taxi usage, the differences between 2009 and 2016 clearly show the drop in the number of yellow taxi trips across the period of study. The above exploratory and suggestive results illustrate a

Figure 3: Trend of daily taxi journeys in New York City, before and after Uber enters New York City

negative Uber effect on yellow taxi rides, something which is also supported by the literature (Dudley, Banister, and Schwanen 2017). However, this research provided no quantitative evidence, and is therefore a gap that our research aims to exploit. The next step is to approach this research question using our modelling tools.

We checked for stationarity of our data using the Augmented Dickey-Fuller Test and found our data to be stationary (Cheung and Lai 1995). In order to meet the normality assumption for the estimations, we normalise the continuous variables using the logarithmic transformation (Schroeder, Sjoquist, and Stephan 2016)⁷. Moreover, we add a small constant (0.5) in order to utilise the observations which are equal to 0 for some of the variables presented in Table 1. This is a common strategy in the empirical literature, which does not influence the results (Ebdon 1991).

The next step is to estimate the model presented in Equation 1. The results of the estimation are presented in Table 2. The main finding is that the number of NYTimes articles (lagged) with a reference to Uber has a significant negative effect on the number of yellow taxi rides. Importantly, the coefficient of the main variable of interest – NYTimes articles – remains significant even after the

⁷We performed various skewness and kurtosis tests, which indicate that our variables do not follow the normal distribution. The results can be provided upon request.

Figure 4: Calendar plot of taxi journeys in 2009

Figure 5: Calendar plot of taxi journeys in 2016

inclusion of various control variables that are related with taxi usage. Specifically, we control for a time trend in order to capture long term changes in the taxi industry. We also control for weather conditions following the above discussion. Hence, we control for snowfall, temperature and precipitation and all but the last variable significantly affect taxi trips. In total, bad weather is correlated with less taxi rides, which can be justified with decreased demand for mobility. Furthermore, we control for subway closures, which also have a significant effect on taxi usage. In agreement with our expectations, as the number of subway failures increased, the number of yellow taxis also increased. This is due to commuters attempting to find alternative transport once their subway line has closed. The only control variable which has an effect opposite to the initial expectations is the stock market index of the New York Stock Exchange, which is negatively correlated with yellow taxis. However, previous literature has been inconclusive in determining the relationship between income and taxi demand. Yang and Gonzales (2014) find that during the day wealthy people may be more likely to take taxis, but find no significant relationship during the night, while Qian and Ukkusuri (2015) suggest that that in some areas of NYC higher income decreases the number of taxi trips. Although not in agreement with our initial expectations, our results are supported by Qian and Ukkusuri (2015). Our interpretation, although speculative, is that increases in NYSE might lead to increased demand of more expensive transport options such as private limousine

services. Interestingly, the coefficient for this variable is highly significant and its inclusion leads to a small increase in the model R-squared. We also performed the same analysis on our peak and non-peak subsets of our dataset. In all three subsets, NYTimes had a significant negative effect on yellow taxis, with snowfall, minimum temperature and NYSE also showing significant relationships. The strongest relationship between yellow taxi trips and NYTimes articles is at off-peak times with -0.048^{***} , while the weakest was evening peak trips with a value of -0.035^{***} . The similarity between the subsets demonstrates the robustness in our methodology. All the results can be found in tables A1-3 in Appendix A.

Table 2: OLS estimations for taxi trips

	<i>Dependent variable:</i>			
	log(Yellow Taxi Trips)			
	(1)	(2)	(3)	(4)
log(lagged NYTimes articles)	-0.057^{***} (0.008)	-0.055^{***} (0.008)	-0.056^{***} (0.007)	-0.045^{***} (0.008)
Time Trend	-0.018^{**} (0.007)	-0.018^{**} (0.007)	-0.021^{***} (0.007)	0.016 (0.014)
log(Precipitation)		-0.002 (0.010)	-0.002 (0.010)	-0.001 (0.010)
log(Snow)		-0.031^{***} (0.009)	-0.032^{***} (0.009)	-0.032^{***} (0.009)
log(Minimum Temperature)		-0.187^{**} (0.083)	-0.195^{**} (0.083)	-0.228^{***} (0.083)
log(Maximum Temperature)		0.162 (0.103)	0.175* (0.103)	0.203** (0.102)
Subway Failures			0.013** (0.006)	0.012** (0.006)
NYSE				-0.220^{***} (0.073)
Constant	13.103*** (0.037)	13.142*** (0.138)	13.116*** (0.138)	14.931*** (0.617)
Observations	416	416	416	416
R ²	0.226	0.256	0.266	0.282
Adjusted R ²	0.223	0.246	0.253	0.268

Note:

*p<0.1; **p<0.05; ***p<0.01

However, this model assumes that the elements of the error vector are uncorrelated, meaning they do not suffer by autocorrelation. This assumption states that the previous value of y has no impact on proceeding ones (Keele and Kelly 2006). However, with time series data, autocorrelation is often a problem and can cause the model to underestimate the standard error of coefficients. To test for this suspected autocorrelation, we employed the Durbin-Watson statistic which determines whether data is autocorrelated by using the predicted y values

to return a value between 0 and 4, where 2 signals no autocorrelation (Savin and White 1977). For the model presented in Table 2 the Durbin-Watson statistic is equal to 0.843 and highly significant. This, therefore, suggests that the model suffers from autocorrelation. We address this issue by estimating the model presented in Equation 1 with GLS. This estimator fits a linear model, while accounting for auto-correlated errors in the residuals. Deciding which order of lag to use is vital in such models. To do so we created a plot of partial autocorrelation coefficients (PACF), which indicated a large spike at lag 1. This is an indication that an autoregressive structure of one order (AR(1)) can capture the structure of autocorrelation in our model (Nelson 1998). Table 3 presents the GLS results, which mirror the OLS estimation and demonstrate that the popularity of and the public interest in Uber as reflected in NYTimes articles is negatively correlated with the use of taxis the next week.

Table 3: Autoregressive estimations for taxi trips

	<i>Dependent variable:</i>			
	log(Yellow Taxi Trips)			
	(1)	(2)	(3)	(4)
log(lagged NYTimes articles)	−0.061*** (0.008)	−0.059*** (0.007)	−0.059*** (0.007)	−0.050*** (0.008)
Time Trend	−0.016** (0.007)	−0.016** (0.007)	−0.019*** (0.007)	0.016 (0.014)
log(Precipitation)		0.003 (0.010)	0.003 (0.010)	0.003 (0.009)
log(Snow)		−0.030*** (0.009)	−0.030*** (0.009)	−0.031*** (0.009)
log(Minimum Temperature)		−0.231*** (0.083)	−0.238*** (0.082)	−0.262*** (0.082)
log(Maximum Temperature)		0.221** (0.102)	0.234** (0.102)	0.251** (0.101)
log(Subway Failures)			0.013** (0.006)	0.012** (0.005)
log(NYSE)				−0.207*** (0.072)
Constant	13.093*** (0.037)	13.056*** (0.139)	13.030*** (0.138)	14.747*** (0.614)
Observations	416	416	416	416
Log Likelihood	283.382	279.691	277.989	280.358
Akaike Inf. Crit.	−556.765	−541.383	−535.978	−538.716
Bayesian Inf. Crit.	−536.647	−505.259	−495.865	−494.619

Note:

*p<0.1; **p<0.05; ***p<0.01

Although the negative coefficient and the lagged structure of our main regressor illustrates a negative relation between Uber and taxi usage, it is difficult to depict a causal relation and the disruptive capacity of Uber (Christensen and

Raynor 2005). To do so, we employ Granger causality tests. As stated in the methodological framework and data section, Granger causality tests whether a lagged value of x precedes variable y . Specifically, we test here whether the lag of the number of NYTimes articles Granger-causes a decrease in yellow taxi trips or vice versa. We employ the Granger causality at a lag of one time period (week) using the “Vars” package in R (Pfaff 2008). The results are presented in Table 4. We estimate Equation 3 using both lagged values of y (Taxi Trips) and x (NYTimes) and then without the lagged values of x . The Wald Test results are statistically significant as shown in Table 4 meaning there is a causal relationship between the two variables. As Uber has grown as a disruptive innovator in the media, the Granger Causality results enforce our hypothesis that this has led to a decrease in yellow taxi trips, and is therefore negatively impacting traditional taxis. We then tested for a bidirectional relationship, which we also found to be significant, as shown. This bidirectional relationship between Uber’s popularity and decrease in Yellow Taxi trips should not come as surprise. In essence, it reinforces the complexity of human behaviour regarding transport choices. Popularity of Uber in the media leads to a decrease in more traditional taxi services. At the same time, the decrease in the use of Yellow Taxis and the consequent increase of Uber usage leads to higher popularity and, therefore, increased interest from the media.

Once, again the analysis for our peak and non-peak subsets, showed similar results. They were all found to have highly significant negative relationships with Uber in the media. Furthermore, snow, minimum temperature and NYSe were once again significant, varying very little to our overall model. We also employed the Granger-Causality tests across all the subsets to ensure our model was consistent across all time periods. For all subsets, the data showed that a rise in NYTimes articles was highly significant to a decrease in yellow taxi trips. Furthermore, like the main model, they also showed a lower, but still significant, bi-directional relationship. Full results can be found in the Appendix. Overall, the most significant relationship within the subsets was the non-peak taxi trips, where there was a stronger negative impact. This suggests that Uber has been less effective at disrupting the taxi markets during peak times, perhaps caused by their own peak price system, causing higher prices during these times. However, further analysis would be needed in order to confirm this hypothesis.

In total, our estimations showed that after controlling for other factors which might affect taxi usage, an increase of NYTimes articles related to Uber, results in fewer yellow taxi trips in the following week. This is a complex relationship though and as the Granger causality tests indicated it is a bi-directional one. Our results are in agreement with with Dudley, Banister, and Schwanen (2017), which state that Uber has been a disruptive innovator for the New York City taxi market, due to its fast growth, lack of regulation and has a strong relationship with the decrease of yellow taxi rides. The bidirectional nature of the relationship between Uber popularity and taxi usage is also in agreement with Leddin (2015); as the use of yellow taxis increase, the buzz and excitement around Uber grows, and subsequently more is written in the media.

Table 4: Granger causality results

	F	Pr(>F)	Signfi.	Res.Df
Taxis~NYTimes	7.121	0.001	***	411
NYTimes~Taxis	6.061	0.003	**	411

Note:

*p<0.1; **p<0.05; ***p<0.01

Taxi service quality and Uber

To investigate the effect of Uber on the quality of taxis, we will estimate the model presented in Equation 2. The dependent variable is the number of taxi complaints per 1000 taxi rides. Because the complaint data are only available on a monthly basis we re-aggregate our time series data set to match the temporal resolution of the complaint data. Using both Forbes', (2008) and Beard *et al's*, (2015) framework we will use complaints to estimate the quality of service of yellow taxis. As Uber grows, we expect to see a decrease in the number of taxi complaints filed reflecting an improvement in taxi service as a response to increased competition by Uber. Instead of using the raw number of complaints we divide it by the number of trips (000s) to account for the decreasing number of taxi trips above. We also employ the set of control variables we previously utilised.

The results are presented in Table 5. We initially estimated these regressions using OLS, but as previously these results suffer from temporal autocorrelation. Hence, we present here the GLS estimations. Overall the model shows that the public interest in Uber as reflected in the number of relevant NYTimes articles is positively related with the number of complaints. This result is statistically significant and consistent throughout the different specifications. However, it does not agree with previous literature on this subject. For instance, Wallsten (2015) indicated a strong negative relationship between complaints and Uber Google searches in Chicago. Specifically, he found a significant negative relationship between Uber Google searches and broken credit card machines, while any other types of complaints were found to be statistically insignificant. Our complaint variable, which combines all types of taxi complaints, has a significant positive relationship with Uber's popularity. A potential interpretation might be that Uber increased the expectations also for traditional taxi customers and, therefore, the higher the penetration of Uber and the exposure to the Uber quality of service, the higher the expectations of taxi users as well. Similarly, our results might signify a behavioural change of taxi customers, following Uber users who can assess the quality of service immediately after an Uber ride. As Figure 6 illustrates, the introduction of Uber in the market in 2011 is accompanied by an increase in the number of complaints against traditional taxis two years later. Nevertheless, we cannot claim a causal relationship. Following the research framework utilised in the previous section we also estimated a Granger causality test, which is insignificant indicating the lack of a causal relation with the

Granger causality framework⁸.

Regarding the other control variables, snow and NYSE have consistent and significant negative effects. We need to highlight though that the temporal resolution of these models is monthly and, therefore, the median values of stream-like data are not able to fully capture their variation.

Table 5: GLS estimations for taxi quality

	<i>Dependent variable:</i>			
	log(Complaints per 1000 taxi rides)			
	(1)	(2)	(3)	(4)
Lagged NYTimes articles	0.090** (0.036)	0.086*** (0.032)	0.078** (0.031)	0.079** (0.031)
Time Trend	-0.006 (0.004)	-0.005 (0.003)	0.009 (0.006)	0.009 (0.006)
PRCP		0.182 (0.122)	0.175 (0.117)	0.176 (0.119)
Snow		-0.209*** (0.045)	-0.192*** (0.044)	-0.191*** (0.045)
TMIN		1.070 (1.090)	0.945 (1.048)	0.946 (1.055)
TMAX		-1.351 (1.352)	-1.309 (1.299)	-1.313 (1.308)
NYSE			-2.245*** (0.824)	-2.260*** (0.847)
Subway				-0.010 (0.114)
Constant	3.573*** (0.171)	4.755*** (1.569)	24.799*** (7.507)	24.963*** (7.805)
Observations	84	84	84	84
Akaike Inf. Crit.	176.947	167.187	160.563	165.065
Bayesian Inf. Crit.	188.920	188.282	183.870	190.558

Note:

*p<0.1; **p<0.05; ***p<0.01

Conclusions

The paper aimed to understand how Uber has affected the taxi market in New York City. To do so we built a theoretical framework which explains why Uber has the capacity to act as a disruptive innovation especially as it relates to traditional taxi market (Christensen and Raynor 2005). We also employed the widely used work of Rogers (1995) to explain how technological innovations diffuse. Finally, we adopted a Schumpeterian perspective to explain how such

⁸The results can be provided upon request

Figure 6: Taxi complaints over time

an innovation may also affect the broader quality of the taxi market through competition (Schumpeter 1942). In order to answer these questions we had to overcome specific data gaps and especially these related with Uber usage given that Uber does not reveal detailed data. To do so we creatively employed new sources of big data and built a proxy of Uber’s popularity and public interest in New York City by counting how many NYTimes articles include the term Uber.

Our statistical analysis indicated that Uber has indeed led to a decrease in trips undertaken by traditional yellow taxis in New York City. Importantly, this is a complex bivariate relationship as decreases in taxi trips also led to increases in Uber’s popularity. Hence, our analysis confirms Uber’s nature as a disruptive innovator in the New York City taxi market. As Malhotra & Alstytne (2014) suggest, Uber’s ability to circumvent barriers to entry, such as no licensing exams or the need for medallions, has allowed them to quickly challenge the taxi market in New York City. Given the lengthy time series data set we constructed, we were able to capture these effects during the early adoption face of Uber and, therefore, observe how technological developments outside the traditional transport industry have led to disruptive effects to the taxi market (Leddin 2015). Following Rogers (2002) ideas and the curve of the diffusion of innovation it would be interesting to see the long term effect of Uber in the traditional taxi market when, for example, the majority and the laggards engage with Uber. Our models all showed snowfall and minimum temperature to have a significant

negative effect on yellow taxi trips, justified by decreased demand for mobility in poor conditions. Regarding the competition effect on quality, our results have shown that complaints are significantly related to an increase in the popularity of Uber. As Uber has gained popularity, the number of taxi complaints per 1000 trips has also increased. One might say that in this case competition has not led to a rise in the standard of yellow taxis as expected. A different interpretation though might be linked to behavioural change. The introduction of Uber and the related completion in the taxi market, instead of just affecting taxi quality might have also affected taxi user behaviour. On the one hand, taxi users might have higher expectations because they are used to or read about the services that Uber users enjoy. On the other hand, they might adopt an Uber inspired behaviour regarding assessing the taxi services they enjoy. Both of these behavioural mechanisms can potentially explain the increase in the number of complaints about taxi service.

In total, our approach to employ novel data from a news media outlet to approximate the public interest and Uber's popularity enabled us to illustrate how Uber disrupted the New York City well established taxi market. Because of the lack of granular data about Uber usage researchers had not been able to model how the growth of Uber's popularity can affect the traditional market at this level of temporal granularity. Following the diffusion of innovation arguments, it would be interesting for future research to estimate the long term disruptive effects of Uber and other similar services to established and well regulated markets as well as potential spatial effects. It is also important for transport studies to find new ways to overcome data gaps driven by corporate policies in order to better inform transport policies.

Acknowledgements

The authors acknowledge the constructive feedback from the editor, the anonymous reviewers and the participants of the 2019 NECTAR Conference.

References

- Baloch, Fareed. 2018. "A Predatory Pricing Model Is Not Competition." *Financial Times*. Financial Times. <https://www.ft.com/content/45c949da-94e4-11e8-b747-fb1e803ee64e>.
- Batty, Michael. 2013. "Big data, smart cities and city planning." *Dialogues in Human Geography* 3 (3): 274–79. <https://doi.org/10.1177/2043820613513390>.
- Beard, T Randolph, Jeffrey T Macher, and John W Mayo. 2015. "Can You Hear Me Now? Exit, Voice, and Loyalty under Increasing Competition." *The Journal of Law and Economics* 58 (3): 717–45. <https://doi.org/10.1086/684232>.
- Benz, A, and E Sumares. 2015. *The Uber Controversy: a Public Choice Perspective*.
- Bowcott, Owen. 2017. "Uber to face stricter EU regulation after ECJ rules it is transport firm." <https://www.theguardian.com/technology/2017/dec/20/uber-european-court-of-justice-ruling-barcelona-taxi-drivers-ecj-eu>.

- Brockwell, Peter J, and Richard A Davis. 1987. "Stationary Time Series." In *Springer Series in Statistics*, 1–41. Springer New York. https://doi.org/10.1007/978-1-4899-0004-3_1.
- Cervero, R. 1985. "Deregulating Urban Transportation" 5 (1): 219–37.
- Chang, Hung-Hao. 2017. "THE ECONOMIC EFFECTS OF UBER ON TAXI DRIVERS IN TAIWAN." *Journal of Competition Law & Economics* 13 (3): 475–500. <https://doi.org/10.1093/joclec/nhx017>.
- Cheung, Yin-Wong, and Kon S. Lai. 1995. "Lag Order and Critical Values of the Augmented Dickey-Fuller Test." *Journal of Business & Economic Statistics* 13 (3): 277–80. <https://doi.org/10.1080/07350015.1995.10524601>.
- Choi, Hyunyoung, and Hal Varian. 2012. "Predicting the Present with Google Trends." *Economic Record* 88: 2–9. <https://doi.org/10.1111/j.1475-4932.2012.00809.x>.
- Christensen, Clayton M, and Michael E Raynor. 2005. "The Innovators Dilemma: When New Technologies Cause Great Firms to Fail." *Academe* 91 (1): 80. <https://doi.org/10.2307/40252749>.
- Cordray, R. 2015. "Uber's Big Data Effect On The Taxi And Transportation Industry." <http://www.digitalistmag.com/industries/travel-and-transportation/2015/09/10/ubers-big-data-effect-taxi-transportation-industry-03396123>.
- Cramer, Judd, and Alan Krueger. 2016. "Disruptive Change in the Taxi Business: The Case of Uber." *The American Economic Review* 106 (5): 177–82. <https://doi.org/10.1257%7B/%7D2Faer.p20161002>.
- Damak, Firas, Ben Jabeur, Guillaume Cabanac, Karen Pinel-Sauvagnat, Lynda Tamine, and Mohand Boughanem. 2011. "IRIT at TREC Microblog." <http://lucene.apache.org/java/docs>.
- Dudley, Geoffrey, David Banister, and Tim Schwanen. 2017. "The Rise of Uber and Regulating the Disruptive Innovator." *The Political Quarterly* 88 (3): 492–99. <https://doi.org/10.1111/1467-923x.12373>.
- Ebdon, D. 1991. *Statistics in Geography: A Practical Approach-Revised*. London: Wiley-Blackwell.
- Forbes, Silke J. 2008. "The Effect of Service Quality and Expectations on Customer Complaints." *Journal of Industrial Economics* 56 (1): 190–213. <https://doi.org/10.1111/j.1467-6451.2008.00338.x>.
- Gabel, David. 2016. "Uber and the Persistence of Market Power." *Journal of Economic Issues* 50 (2): 527–34. <https://doi.org/10.1080/00213624.2016.1179060>.
- Gilbert, G, T Cook, A Nalevanko, and L Everett-Lee. 2002. "No Title." *Transit Cooperative Research Program*, 63–76.
- Granger, Clive WJ. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica: Journal of the Econometric Society*, 424–38.
- Hall, Jonathan D., Craig Palsson, and Joseph Price. 2018. "Is Uber a substitute or complement for public transit?" *Journal of Urban Economics* 108 (November): 36–50. <https://doi.org/10.1016/J.JUE.2018.09.003>.
- Hirschman, Albert O. 1970. *Exit, voice, and loyalty : responses to decline in firms, organizations, and states*.

- Hochmair, Hartwig H. 2016. "Spatiotemporal Pattern Analysis of Taxi Trips in New York City." *Transportation Research Record* 2542 (1): 45–56.
- Holt, Richard. 2016. "Black cabs must face up to reality - it's all Uber now." <https://www.telegraph.co.uk/news/2016/09/19/black-cabs-must-face-up-to-reality---its-all-uber-now/>.
- Hurlin, Christophe, and Baptiste Venet. 2001. "Granger Causality Tests in Panel Data Models with Fixed Coefficients." *Cahier de Recherche EURISCO, September, Université Paris IX Dauphine*.
- Kamga, Camille, M Anil Yazici, and Abhishek Singhal. 2015. "Analysis of Taxi Demand and Supply in New York City: Implications of Recent Taxi Regulations." *Transportation Planning and Technology* 38 (6): 601–25.
- Keele, Luke, and Nathan J Kelly. 2006. "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables." *Political Analysis* 14 (02): 186–205. <https://doi.org/10.1093/pan/mpj006>.
- Kitchin, Rob. 2013. "Big data and human geography." *Dialogues in Human Geography* 3 (3): 262–67. <https://doi.org/10.1177/2043820613513388>.
- Koetse, Mark J, and Piet Rietveld. 2009. "The Impact of Climate Change and Weather on Transport: An Overview of Empirical Findings." *Transportation Research Part D: Transport and Environment* 14 (3): 205–21.
- Kranton, Rachel E. 2003. "Competition and the Incentive to Produce High Quality." *Economica* 70 (279): 385–404. <https://doi.org/10.1111/1468-0335.t01-1-00289>.
- Kriss, Peter. 2014. "People power: turning customer experience into a marketing tool." <https://www.theguardian.com/media-network/2014/nov/06/people-power-customer-experience-marketing>.
- Leddin, Taylor. 2015. "Why Uber is gaining in popularity among Millennials." <https://theamericangenius.com/tech-news/why-uber-is-gaining-in-popularity-among-millennials/>.
- Lee, Sungjick, and Han-Joon Kim. 2008. "News Keyword Extraction for Topic Tracking." *2008 Fourth International Conference on Networked Computing and Advanced Information Management*. <https://doi.org/10.1109/ncm.2008.199>.
- Li, Ziru, Yili Hong, and Zhongju Zhang. 2016. "Do Ride-Sharing Services Affect Traffic Congestion? An Empirical Study of Uber Entry." *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2838043>.
- Malhotra, Arvind, and Marshall Van Alstyne. 2014. "The dark side of the sharing economy ... and how to lighten it." *Communications of the {ACM}* 57 (11): 24–27. <https://doi.org/10.1145/2668893>.
- Markides, Constantinos. 2006. "Disruptive Innovation: In Need of Better Theory{\ast}." *Journal of Product Innovation Management* 23 (1): 19–25. <https://doi.org/10.1111/j.1540-5885.2005.00177.x>.
- Milne, Dave, and David Watling. 2019. "Big Data and Understanding Change in the Context of Planning Transport Systems." *Journal of Transport Geography* 76: 235–44.
- Nelson, Brian K. 1998. "Time Series Analysis Using Autoregressive Integrated Moving Average ({ARIMA}) Models." *Academic Emergency Medicine* 5 (7): 739–44. <https://doi.org/10.1111/j.1553-2712.1998.tb02493.x>.

- Pfaff, Bernhard. 2008. "VAR, Svar and Svec Models: Implementation Withinrpackagevars." *Journal of Statistical Software* 27 (4). <https://doi.org/10.18637/jss.v027.i04>.
- Posen, Hannah A. 2015. "Ridesharing in the Sharing Economy: Should Regulators Impose Uber Regulations on Uber." *Iowa Law Review* 101.
- Qian, Xinwu, and Satish V. Ukkusuri. 2015. "Spatial Variation of the Urban Taxi Ridership Using Gps Data." *Applied Geography* 59: 31–42. <https://doi.org/10.1016/j.apgeog.2015.02.011>.
- Raghavan, Bharad, and Samar Anshul. 2013. "Extracting Emotions from News Headlines." <http://www.cse.unt.edu/~rada/affectivetext/2http://www.cse.unt.edu/~rada/affectivetext/1>.
- Rhodes, Anna. 2017. "Uber: Which countries have banned the controversial taxi app." <https://www.independent.co.uk/travel/news-and-advice/uber-ban-countries-where-world-taxi-app-europe-taxi-us-states-china-asia-legal-a7707436.html>.
- Robertson, Thomas S. 1967. "The Process of Innovation and the Diffusion of Innovation." *Journal of Marketing* 31 (1): 14. <https://doi.org/10.2307/1249295>.
- Rogers, Brishen. 2015. "The Social Costs of Uber." *University of Chicago Law Review Dialogue* 82. <https://heinonline.org/HOL/Page?handle=hein.journals/uchidial82%7B/&%7Ddid=85%7B/&%7Ddiv=7%7B/&%7Dcollection=journals>.
- Rogers, Everett M. 1995. *Diffusion of Innovations, 4th Edition*. Free Press.
- . 2002. "Diffusion of preventive innovations." *Addictive Behaviors* 27 (6): 989–93. [https://doi.org/10.1016/s0306-4603\(02\)00300-3](https://doi.org/10.1016/s0306-4603(02)00300-3).
- Savin, N E, and Kenneth J White. 1977. "The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors." *Econometrica* 45 (8): 1989. <https://doi.org/10.2307/1914122>.
- Schaller, B. 2005. "A Regression Model of the Number of Taxicabs in U.S. Cities." *Journal of Public Transportation* 8 (5): 65–78.
- Schroeder, LD, DL Sjoquist, and PE Stephan. 2016. *Understanding regression analysis: An introductory guide*. Vol. 57. Sage Publications.
- Schumpeter, J. 1942. *Capitalism, socialism and democracy*. London: Routledge.
- Sidak, J. Gregory. 2016. "Is Uber Unconstitutional?"
- TLC. 2017. "NYC Taxi & Limousine Commission - Trip Record Data." http://www.nyc.gov/html/tlc/html/about/trip%7B/_%7Drecord%7B/_%7Ddata.shtml.
- Tranos, Emmanouil. 2012. "The Causal Effect of the Internet Infrastructure on the Economic Development of European City Regions." *Spatial Economic Analysis* 7 (3): 319–37.
- Tranos, Emmanouil, and Elizabeth Mack. 2019. "Big Data: A New Opportunity for Transport Geography?" *Journal of Transport Geography* 76: 232–34.
- Tranos, Emmanouil, and Elizabeth A Mack. 2016. "Broadband Provision and Knowledge-Intensive Firms: A Causal Relationship?" *Regional Studies* 50 (7): 1113–26.
- Wallsten, Scott. 2015. "Is Uber Making Taxi Service Better?" <https://www.theatlantic.com/business/archive/2015/07/uber-taxi-drivers-complaints-chicago-newyork/397931/>.

- Watanabe, Chihiro, Kashif Naveed, Pekka Neittaanmäki, and Brenda Fox. 2017. “Consolidated challenge to social demand for resilient platforms - Lessons from Uber’s global expansion.” *Technology in Society* 48 (February): 33–53. <https://doi.org/10.1016/j.techsoc.2016.10.006>.
- Winter, Stephan, Monika Sester, Ouri Wolfson, and Glenn Geers. 2011. “Towards a Computational Transportation Science.” *Journal of Spatial Information Science* 2011 (2): 119–26.
- Yang, Ci, and Eric J. Gonzales. 2014. “Modeling Taxi Trip Demand by Time of Day in New York City.” *Transportation Research Board* 2429 (1): 110–20. <https://doi.org/10.3141/2429-12>.
- Yuhas, Alan, and Jessica Glenza. 2016. “Blizzard buries eastern US as New York bans travel – as it happened.” <https://www.theguardian.com/us-news/live/2016/jan/23/blizzard-snowstorm-washington-new-york-east-coast>.
- Zhang, W, and S V. Ukkusri. 2016. “Optimal Fleet Size and Fare Setting in Emerging Taxi Markets with Stochastic Demand.” *Computer-Aided Civil and Infrastructure Engineering* 21 (9): 647–60.
- Zhou, Meng, Donggen Wang, Qingquan Li, Yang Yue, Wei Tu, and Rui Cao. 2017. “Impacts of Weather on Public Transport Ridership: Results from Mining Data from Different Sources.” *Transportation Research Part C: Emerging Technologies* 75: 17–29.

Appendix

The following tables show the same analysis conducted for peak taxi times (06:00-10:00 & 16:00-20:00), evening peak taxi times (16:00-20:00), and non-peak taxi times. We followed the exact steps used in the paper, with the results below.

Table A1: OLS estimations for Peak taxi trips

	<i>Dependent variable:</i>			
	log(Peak Yellow Taxi Trips)			
	(1)	(2)	(3)	(4)
log(lagged NYTimes articles)	−0.053*** (0.008)	−0.051*** (0.008)	−0.051*** (0.008)	−0.040*** (0.009)
Time Trend	−0.020** (0.008)	−0.021** (0.008)	−0.022*** (0.008)	0.019 (0.016)
log(Precipitation)		0.006 (0.011)	0.007 (0.011)	0.008 (0.011)
log(Snow)		−0.019* (0.011)	−0.019* (0.011)	−0.019* (0.011)
log(Minimum Temperature)		−0.171* (0.094)	−0.176* (0.094)	−0.212** (0.094)
log(Maximum Temperature)		0.164 (0.116)	0.172 (0.116)	0.203* (0.116)
Subway Failures			0.007 (0.006)	0.007 (0.006)
NYSE				−0.246*** (0.082)
Constant	12.096*** (0.041)	12.069*** (0.155)	12.054*** (0.156)	14.086*** (0.697)
Observations	416	416	416	416
R ²	0.181	0.193	0.196	0.213
Adjusted R ²	0.177	0.181	0.182	0.197

Note:

*p<0.1; **p<0.05; ***p<0.01

Table A2: OLS estimations for Evening taxi trips

	<i>Dependent variable:</i>			
	log(Evening Yellow Taxi Trips)			
	(1)	(2)	(3)	(4)
log(lagged NYTimes articles)	−0.046*** (0.008)	−0.045*** (0.008)	−0.045*** (0.008)	−0.035*** (0.008)
Time Trend	−0.034*** (0.007)	−0.033*** (0.007)	−0.035*** (0.007)	0.003 (0.014)
log(Precipitation)		0.003 (0.010)	0.003 (0.010)	0.005 (0.010)
log(Snow)		−0.021** (0.010)	−0.022** (0.010)	−0.022** (0.009)
log(Minimum Temperature)		−0.184** (0.084)	−0.189** (0.084)	−0.223*** (0.084)
log(Maximum Temperature)		0.145 (0.104)	0.154 (0.104)	0.183* (0.104)
Subway Failures			0.009 (0.006)	0.008 (0.006)
NYSE				−0.227*** (0.074)
Constant	11.637*** (0.037)	11.729*** (0.139)	11.711*** (0.140)	13.584*** (0.624)
Observations	416	416	416	416
R ²	0.239	0.265	0.269	0.286
Adjusted R ²	0.235	0.254	0.256	0.272

Note:

*p<0.1; **p<0.05; ***p<0.01

Table A3: OLS estimations for Non-Peak taxi trips

	<i>Dependent variable:</i>			
	log(Non-peak Yellow Taxi Trips)			
	(1)	(2)	(3)	(4)
log(lagged NYTimes articles)	−0.059*** (0.008)	−0.057*** (0.008)	−0.057*** (0.008)	−0.048*** (0.009)
Time Trend	−0.017** (0.008)	−0.017** (0.008)	−0.021** (0.008)	0.013 (0.016)
log(Precipitation)		−0.006 (0.011)	−0.006 (0.011)	−0.005 (0.011)
log(Snow)		−0.041*** (0.010)	−0.041*** (0.010)	−0.041*** (0.010)
log(Minimum Temperature)		−0.212** (0.092)	−0.221** (0.091)	−0.251*** (0.091)
log(Maximum Temperature)		0.181 (0.114)	0.197* (0.113)	0.222** (0.113)
Subway Failures			0.015** (0.006)	0.015** (0.006)
NYSE				−0.201** (0.080)
Constant	12.645*** (0.041)	12.698*** (0.152)	12.667*** (0.152)	14.329*** (0.681)
Observations	416	416	416	416
R ²	0.195	0.236	0.247	0.259
Adjusted R ²	0.192	0.225	0.234	0.244

Note:

*p<0.1; **p<0.05; ***p<0.01

Table A4: Autoregressive estimations for Peak taxi trips

	<i>Dependent variable:</i>			
	log(Peak Yellow Taxi Trips)			
	(1)	(2)	(3)	(4)
log(lagged NYTimes articles)	−0.056*** (0.008)	−0.055*** (0.008)	−0.055*** (0.008)	−0.045*** (0.009)
Time Trend	−0.019** (0.008)	−0.019** (0.008)	−0.020** (0.008)	0.019 (0.016)
log(Precipitation)		0.011 (0.011)	0.011 (0.011)	0.011 (0.011)
log(Snow)		−0.018* (0.011)	−0.018* (0.011)	−0.018* (0.011)
log(Minimum Temperature)		−0.212** (0.093)	−0.216** (0.093)	−0.243*** (0.093)
log(Maximum Temperature)		0.219* (0.115)	0.227** (0.116)	0.245** (0.115)
log(Subway Failures)			0.007 (0.006)	0.007 (0.006)
log(NYSE)				−0.232*** (0.082)
Constant	12.088*** (0.041)	11.989*** (0.156)	11.974*** (0.157)	13.906*** (0.695)
Observations	416	416	416	416
Log Likelihood	238.618	230.671	227.205	229.649
Akaike Inf. Crit.	−467.236	−443.343	−434.411	−437.298
Bayesian Inf. Crit.	−447.118	−407.219	−394.298	−393.201

Note:

*p<0.1; **p<0.05; ***p<0.01

Table A5: Autoregressive estimations for Evening taxi trips

	<i>Dependent variable:</i>			
	log(Evening Yellow Taxi Trips)			
	(1)	(2)	(3)	(4)
log(lagged NYTimes articles)	−0.053*** (0.008)	−0.050*** (0.007)	−0.050*** (0.007)	−0.043*** (0.008)
Time Trend	−0.029*** (0.007)	−0.030*** (0.007)	−0.032*** (0.007)	−0.001 (0.013)
log(Precipitation)		0.006 (0.009)	0.005 (0.009)	0.009 (0.009)
log(Snow)		−0.021** (0.009)	−0.021** (0.009)	−0.020** (0.009)
log(Minimum Temperature)		−0.190*** (0.046)	−0.177*** (0.047)	−0.283*** (0.060)
log(Maximum Temperature)		0.153*** (0.037)	0.136*** (0.039)	0.270*** (0.061)
log(Subway Failures)			0.008 (0.005)	0.009* (0.005)
log(NYSE)				−0.185*** (0.065)
Constant	11.614*** (0.036)	11.706*** (0.045)	11.724*** (0.046)	13.097*** (0.486)
Observations	416	416	416	416
Log Likelihood	285.090	279.237	275.992	278.169
Akaike Inf. Crit.	−560.180	−540.474	−531.983	−534.339
Bayesian Inf. Crit.	−540.063	−504.351	−491.871	−490.242

Note:

*p<0.1; **p<0.05; ***p<0.01

Table A6: Autoregressive estimations for non peak taxi trips

	<i>Dependent variable:</i>			
	log(non peak Yellow Taxi Trips)			
	(1)	(2)	(3)	(4)
log(lagged NYTimes articles)	−0.063*** (0.008)	−0.061*** (0.008)	−0.061*** (0.008)	−0.053*** (0.009)
Time Trend	−0.015* (0.008)	−0.015* (0.008)	−0.018** (0.008)	0.013 (0.016)
log(Precipitation)		−0.001 (0.011)	−0.001 (0.011)	−0.001 (0.010)
log(Snow)		−0.040*** (0.010)	−0.040*** (0.010)	−0.040*** (0.010)
log(Minimum Temperature)		−0.256*** (0.091)	−0.265*** (0.091)	−0.287*** (0.090)
log(Maximum Temperature)		0.242** (0.113)	0.257** (0.112)	0.272** (0.112)
log(Subway Failures)			0.015** (0.006)	0.015** (0.006)
log(NYSE)				−0.187** (0.080)
Constant	12.634*** (0.041)	12.610*** (0.153)	12.579*** (0.152)	14.136*** (0.679)
Observations	416	416	416	416
Log Likelihood	240.676	239.647	238.513	239.660
Akaike Inf. Crit.	−471.352	−461.294	−457.026	−457.321
Bayesian Inf. Crit.	−451.235	−425.170	−416.913	−413.224

Note:

*p<0.1; **p<0.05; ***p<0.01

Table A7: Granger Causality Results

	F	Pr(>F)	Signif.	Res.Df
Taxis~NYTimes	7.121038	0.0009120	***	411
NYTimes~Taxis	6.060951	0.0025469	**	411
Peak Taxis~NYTimes	10.519811	0.0000351	***	411
NYTimes~Peak Taxis	5.289206	0.0053967	**	411
Evening Taxis~NYTimes	8.154118	0.0003369	***	411
NYTimes~Evening Taxis	7.995250	0.0003925	***	411
Non-peak Taxis~NYTimes	7.002598	0.0010226	**	411
NYTimes~Non-peak Taxis	5.981937	0.0027501	**	411

Note:

*p<0.1; **p<0.05; ***p<0.01