# Ridesharing Price Prediction: Exploring the strategies of Dynamic Pricing

Haolun Huang *

Boston College, Chestnut Hill, MA, US

* Corresponding Author Email: haolun.huang@bc.edu

**Abstract.** This paper explores the application of dynamic pricing algorithms in rideshare industries and examines the key variables that influence trip prices by analyzing Uber and Lyft Dataset of Boston in 2018. This dataset contains Uber and Lyft data mainly from November and December 2018. It contains essential variables such like distance and price for rideshare price prediction, together with many other variables which could be included in regression model such as weather and hours in the day. The Ordinary Least Square regression method is used in this paper, and results shows that the most influential and statistically significant variables on price are the distance of trip and surge multiplier which indicates traffic surge. The regression result is in accordance with the intuition that price is positively correlated with distance and price goes up during surge time. The result implies that the dynamic pricing algorithm that adjusts live prices heavily relies on the current traffic condition and distance of the trip, while the price is not determined by hour of the trip during the day and temperature.

**Keywords:** Dynamic pricing, Rideshare, dynamically-priced labor markets.

## 1. Introduction

Imagine a tourist finished the Thanksgiving break and arrived at the airport. When the tourist tried to request a ride home using Uber or Lyft, the price is higher than the cost of the same trip a few weeks ago. This is because of the dynamic pricing algorithm of ridesharing companies that adjust the fare rate based on a number of variables such as time, traffic, the demand of riders, and the supply of drivers. In the example, because of the Thanksgiving break, there might be more riders and fewer drivers. Therefore the dynamic pricing algorithm raises the price of the tourist's trip home.

The core idea of the dynamic pricing algorithm lies behind the idea of supply and demand, for example, when this tourist tried to request a ride home, the ride is valued more than the amount of money paid for the trip fare. However, if the dynamic pricing algorithm was not designed well and raised the price too high, the tourist might not be willing to request such an overpriced ride anymore and might deviate to public transportation as an alternative. Therefore, in this case, the tourist values the trip fare more than the ride and stops requesting a ride, thus the ridesharing company loses a potential customer.

This problem raises further questions: is the strategy of dynamic pricing guaranteed to bring more profit, and how the dynamic price can be determined such that the rideshare company could make more profit, while the customers are still willing to pay for the raised fare? Whether or not industries make more profit through dynamic pricing is not a trivial problem. Although the possibility that dynamic pricing does not bring more profit might be unintuitive at first thought, industries might be worse off if they employ a dynamic price instead of a fixed price, if their dynamic pricing is poorly designed. Motivated by the questions above, this paper explores how ridesharing companies dynamically adjust the price of trips.

## 2. Literature Reviews

For works of literature on dynamic pricing, dynamic pricing policies has been studied for ridesharing giants: Uber and Lyft. Siddhartha et al. build a two-sided model on ridesharing platforms in order to capture the strategic decisions of drivers, passengers, platform, and the marketplace stochastic dynamics. They reached the conclusion that dynamic pricing has an advantage over static pricing, not just because it performs better, but because it is more resilient to uncertainty inside the system characteristics [1]. Additionally, Keith and Michael investigated how the "gig" economy's dynamic pricing of tasks impacted the supply of labor and discussed the ramifications of their findings for salary, flexibility in the workplace, and the effectiveness of labor markets with dynamic pricing [2]. Additionally, Guo et al. accomplish dynamic price prediction using data from multiple urban sources. Passengers' anxieties are allayed by price prediction, which enables them to determine whether a lesser price might be available nearby or soon. Their findings demonstrate that using multiple urban data sources does, in fact, help to increase forecast accuracy, and that different datasets may have varied effects on dynamic prices [3]. In addition, Arnoud gives a succinct overview of the historical development of quantitative research on pricing and demand estimation. He also highlights various subfields within the field of dynamic pricing, offers a thorough analysis of the literature on the subject of dynamic pricing and learning, discusses connections with other methodologically related research areas, and identifies future research directions [4]. The last evaluation of the literature and contemporary methods for dynamic pricing is provided by Wedad and Pinar. Our focus is on dynamic (intertemporal) pricing in the presence of inventory considerations due to its relevance in the majority of markets and its growing popularity in practice [5].

Sendhil and Jann propose a way of thinking about machine learning that gives it its own home in the econometric toolkit and they examine its capacity to uncover a sophisticated structure that was not stated beforehand with regard to utilizing machine learning in econometrics. It finds functions that perform well outside of a sample and fits sophisticated and highly flexible functional forms to the data without simply overfitting [6]. A dataset created by Uber for New York City is also explained in Rishi's research, along with how the k-means clustering method was used to the dataset to categorize the various areas of the city [7]. The model is used to forecast the demand for various city locations. On the Uber dataset, the authors Nikolay, Jason, Li, and Slawek also combine univariate forecasting models like Holt-Winters and machine learning techniques like random forest. Motivated by the recent resurgence of Long Short Term Memory networks, they propose a novel end-to-end recurrent neural network architecture that outperforms the most cutting-edge event forecasting techniques on Uber data and generalizes well to a public M3 dataset used for time-series forecasting competitions [8].

Additionally, there is some study on the results and tactics of Uber's surge pricing. In this case study, Jonathan et al. describe how Uber's "surge pricing" algorithm assigned a straightforward "multiplier" that multiplied the usual fare in the event of surging [9]. In addition, Wang et al.'s paper uses an aggregate and static technique to analyze the taxi market in the presence of a single taxi-hailing app [10]. They also addressed the characteristics of appropriate price perturbations that enhance societal welfare and/or the platform's profitability. In Brishen’s paper, the first section explores how Uber is fundamentally changing the automobile rental industry. Part II analyzes how Uber affects labor standards, safety, privacy, and discrimination, and it describes how lawmakers should modify current legislation to apply to Uber and other ride-sharing companies. Uber and the future of low-wage jobs are discussed as the third part comes to a close [11].

## 3. Methodology

### 3.1. Data description

This dataset contains Uber and Lyft data mainly from November and December 2018. It contains essential variables such like distance and price for rideshare price prediction, together with many other variables which could be included in regression model such as weather and hours in the day. It could almost be concluded that the dataset is uniformly distributed by hours. There might be some reasons behind this uniform distribution of rides, as it is supposed that there would be more rides during the day compared with the early morning. Notice that Lyft and Uber have a similar market share and there is no significant difference between Uber and Lyft, although the market share of Uber is a bit larger (56% over 44%) (Figure 1). It is also worth noting that there are more counts in December than November. The dataset Uber and Lyft Dataset Boston, MA is available on Kaggle [12].

**Table 1**. Summary statistics

| | hour | distance | surge_multiplier | temperature | price |
|---|---|---|---|---|---|
| count | 693071 | 693071 | 693071 | 693071 | 693071 |
| mean | 11.6191 | 2.1894 | 1.0138 | 39.5843 | 16.5451 |
| std | 6.9481 | 1.1389 | 0.0916 | 6.7260 | 9.3243 |
| Min | 0.00 | 0.02 | 1.00 | 18.91 | 2.50 |
| Median | 12.00 | 2.16 | 1.00 | 40.49 | 13.50 |
| Max | 23.00 | 7.86 | 3.00 | 57.22 | 97.5 |

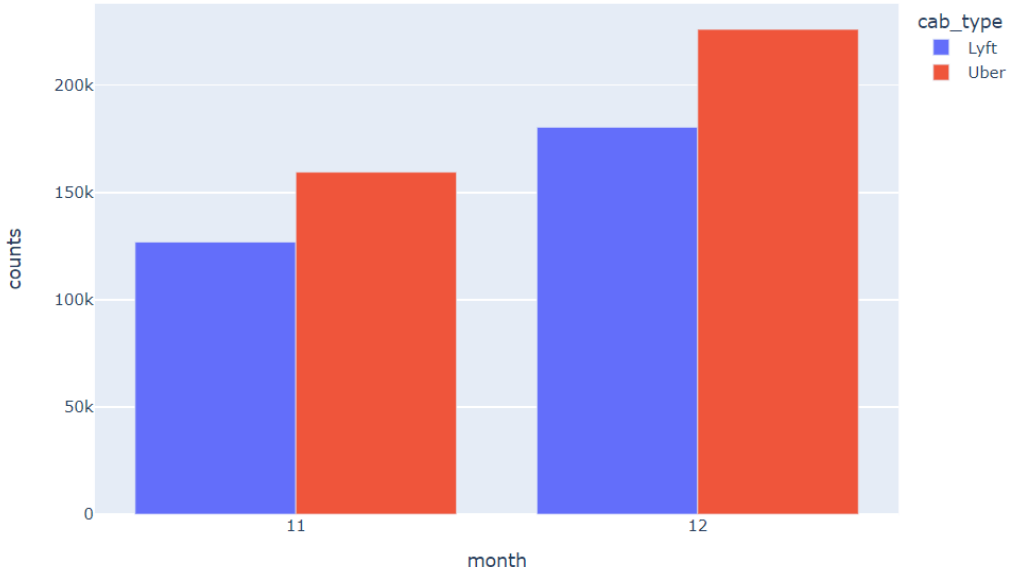


**Fig 1.** Market share of Uber and Lyft

### 3.2. OLS regression result

Run an Ordinary Least Square (OLS) regression on price with distance, surge multiplier, hour, and temperature which indicates if the ride is during a surge. The regression result is in accordance with the intuition that price is positively correlated with distance and price goes up during surge time. The regression result also suggests that hour and temperature are not statistically significant, while distance and surge multiplier is statistically significant. The coefficient of distance is around 2.78, and it tells that on average, one unit increase in distance results in a 2.7844 unit increase in the price of the trip, holding other variables the same. Similarly, the variable surge multiplier measures if it is during a period of a surge. It tells that on average, during the surge period, the trip fare is increased by 22.6375, ceteris paribus. The R-squared value of the model is 0.173 which is not very high. This suggests that there might be more variables to be taken into consideration (Table 2).

Then, run two regressions separately on Uber and Lyft. Notice that the adjusted R-squared value of regression on Lyft data is much larger than Uber, which tells that the model fits better for the Lyft dataset.

For Uber, the regression result still suggests that hour and temperature are not statistically significant, while distance and surge multiplier is statistically significant. The coefficient of distance is around 2.44, and the coefficient of surge multiplier is 10.33 (Table 3). In comparison, for Lyft data, the coefficient of distance is around 3.22, and the coefficient of surge multiplier is 21.73 (Table 4). Comparing the regression result of Uber and Lyft, it tells that the Lyft pricing algorithm charges less base fare, but customers need to pay more for the distance of their trip, and more during the traffic surge. Therefore, customers might prefer to choose Lyft for short-distance trips during non-surge hours, and choose Uber during surge hours or for long-distance trips. The R-squared value of the model for the Uber dataset is 0.113 (Table 3), while it is 0.217 for the Lyft dataset (Table 4). This suggests that the model works better for the Lyft dataset over Uber.

**Table 2**. OLS regression

| Model | OLS | R-squared | 0.173 |
|---|---|---|---|
| Dep. Variable | Price | Adj. R-squared | 0.173 |
| | coef | Std err | P > \|t\| |
| Distance | 2.7844 | 0.009 | 0.000 |
| Hour | -0.0007 | 0.002 | 0.677 |
| Temperature | 0.0018 | 0.002 | 0.258 |
| Surge-multiplier | 22.6375 | 0.111 | 0.000 |

**Table 3**. Uber OLS regression

| Model | OLS | R-squared | 0.113 |
|---|---|---|---|
| Dep. Variable | Price | Adj. R-squared | 0.113 |
| | coef | Std err | P > \|t\| |
| Distance | 2.4412 | 0.012 | 0.000 |
| Hour | -0.0009 | 0.002 | 0.675 |
| Temperature | 0.0029 | 0.002 | 0.168 |
| Surge-multiplier | 10.3393 | 0.088 | 0.000 |

**Table 4**. Lyft OLS regression

| Model | OLS | R-squared | 0.217 |
|---|---|---|---|
| Dep. Variable | Price | Adj. R-squared | 0.217 |
| | coef | Std err | P > \|t\| |
| Distance | 3.225 | 0.015 | 0.000 |
| Hour | -0.0005 | 0.002 | 0.839 |
| Temperature | 0.0005 | 0.002 | 0.822 |
| Surge-multiplier | 21.7377 | 0.118 | 0.000 |

## 4. Conclusion

In conclusion, this paper intends to explore the dynamic pricing strategies of ridesharing giants Uber and Lyft by analyzing the Uber and Lyft datasets, mainly from November and December 2018 which contain essential variables such as distance and price for rideshare price prediction. The regression results show that distance and the indicator for traffic surge are statistically significant variables for predicting the price. Notice that the R-squared value of the model is 0.173 which suggests that there might be more variables to be taken into consideration (Table 2).

One of the limitations is that the ridesharing companies Uber and Lyft did not make their data available to the public, therefore the current dataset is one of the most adequate datasets available on the internet. The research results could be improved with better datasets, for example, one dataset that contains year-long data instead of the current one clustered in November and December. In addition, some variables such as the value of the car, i.e., a Uber driver would charge more if he is driving a brand new luxury car compared with an outdated regular car. Other variables such as precipitation or snowfall quantity that would influence the traffic condition and difficulty of driving, would also be helpful to the current research. The current OLS regression is limited by its primitive nature, while some variables might not be linearly correlated with price. The research could be improved in the future by applying other methods such as random forest, k-Means, Lasso, or Ridge regression.

## References

[1] Banerjee, Siddhartha, Ramesh Johari, and Carlos Riquelme. 2016. "Dynamic Pricing in Ridesharing Platforms." ACM SIGecom Exchanges 15 (1): 65–70. Fangfang. Research on power load forecasting based on Improved BP neural network. Harbin Institute of Technology, 2011.

[2] Chen, M Keith, and Michael Sheldon. n.d. "Dynamic Pricing in a Labor Market: Surge Pricing and Flexible Work on the Uber Platform," 19. Ma Kunlong. Short term distributed load forecasting method based on big data. Changsha: Hunan University, 2014.

[3] Guo, Suiming, Chao Chen, Jingyuan Wang, Yaxiao Liu, Ke Xu, and Dah Ming Chiu. 2018. "Dynamic Price Prediction in Ride-on-Demand Service with Multi-Source Urban Data." In Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, 412–21. MobiQuitous '18. New York, NY, USA: Association for Computing Machinery

[4] Boer, Arnoud V. den. 2015. "Dynamic Pricing and Learning: Historical Origins, Current Research, and New Directions." Surveys in Operations Research and Management Science 20 (1): 1–18. Ma Kunlong. Short term distributed load forecasting method based on big data. Changsha: Hunan University, 2014.

[5] Elmaghraby, Wedad, and Pınar Keskinocak. 2003. "Dynamic Pricing in the Presence of Inventory Considerations: Research Overview, Current Practices, and Future Directions." Management Science 49 (10): 1287–1309. Fangfang. Research on power load forecasting based on Improved BP neural network. Harbin Institute of Technology, 2011.

[6] Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." Journal of Economic Perspectives 31 (2): 87–106. Ma Kunlong. Short term distributed load forecasting method based on big data. Changsha: Hunan University, 2014.

[7] Srinivas, Rishi, B. Ankayarkanni, and R. Sathya Bama Krishna. 2021. "Uber Related Data Analysis Using Machine Learning." In 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 1148–53.

[8] Laptev, Nikolay, Jason Yosinski, Li Erran Li, and Slawek Smyl. n.d. "Time-Series Extreme Event Forecasting with Neural Networks at Uber," 5.

[9] Hall, Jonathan, Cory Kendrick, and Chris Nosko. n.d. "The Effects of Uber's Surge Pricing: A Case Study," 8.

[10] Wang, Xiaolei, Fang He, Hai Yang, and H. Oliver Gao. 2016. "Pricing Strategies for a Taxi-Hailing Platform." Transportation Research Part E: Logistics and Transportation Review 93 (September): 212–31.

[11] Rogers, Brishen. 2015. "The Social Costs of Uber." University of Chicago Law Review Dialogue 82: 85–102.

[12] Uber and Lyft Dataset Boston, MA, available on Kaggle, https://www.kaggle.com/datasets/brllrb/uber-and-lyft-dataset-boston-ma