Observing Algorithmic Marketplaces In-the-Wild

LE CHEN
and
CHRISTO WILSON
Northeastern University

In this letter, we briefly summarize two recent works from our group that use observational data to study the mechanisms used by two large markets. First, we examine Uber's surge price algorithm, and observe that its incentive model may not be effective at changing driver behavior. Second, we study the adoption of dynamic pricing strategies by sellers on Amazon Marketplace, and investigate how these strategies interact with Amazon's "Buy Box" matching algorithm. We make our data available to the research community.

Categories and Subject Descriptors: K.4.4 [**Computing Milieux**]: Computers and Society—*Electronic Commerce*; J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Economics*

General Terms: Observational Study, Market Design, Pricing

Additional Key Words and Phrases: Empirical, Amazon, Uber, Ridesharing, Dynamic Pricing

1. INTRODUCTION

Much of the classic literature in economics deals with mechanism design, i.e., the construction of markets that maximize some useful quantity like revenue or welfare. As commerce has moved online, it has become easier to directly apply these ideas from economic theory in practice. One obvious example of this are online advertising auctions, but more broadly, many companies are now experimenting with differential [Mikians et al. 2012; 2013; Hannak et al. 2014] and dynamic pricing [Chen 2016] strategies in contexts ranging from retail to ridesharing.

Although academics are beginning to propose models for modern e-commerce platforms [Banerjee et al. 2015; Fang et al. 2016], we lack a comprehensive empirical understanding of the actual mechanisms adopted by companies in their marketplaces. The opacity surrounding widely used platforms raises fundamental questions for researchers and consumers: what objectives are these systems optimized for, and are they achieving these objectives? What features do they consider? Are the markets fair, and if so, for what definition of fairness?

In our recent work, we attempt to answer these questions through empirical measurements of major online markets. Using observed data, we quantify the basic properties of markets over time, such as number of participants and prices. In some cases, we must develop novel data gathering methodologies to acquire this information. We then leverage this raw data to infer implementation details of the markets themselves, e.g., the weights of key features, or how algorithms discretizes prices across time and physical space. Finally, we examine the implications of deployed mechanisms on market participants.

Authors' addresses: `leonchen@ccs.neu.edu`, `cbw@ccs.neu.edu`

In this letter, we briefly overview two of our recent observational studies of major online marketplaces:

(1) In [Chen et al. 2015], we examined Uber's dynamic pricing algorithm, which is known as "surge pricing". Using 43 emulated copies of the Uber smartphone app, we blanketed midtown Manhattan and downtown San Francisco for about one month in order to collect data about available supply of rides, fulfilled demand, prices, and estimated wait times. This data enabled us to quantify the sensitivity of surge prices to fluctuations in supply and demand, as well as analyze the impact of surges on rider and driver behavior.
(2) In [Chen et al. 2016], we study two types of algorithms on Amazon Marketplace. First, we investigate Amazon's "Buy Box" algorithm, which determines the default seller that will fulfill orders for each product in the market. It is estimated that ∼80% of purchases on Amazon go through the Buy Box [Taft 2014], so understanding this algorithm is key to being competitive on Amazon Marketplace. Second, we examine the dynamic pricing strategies adopted by individual sellers. Although we find that only a small fraction of sellers have adopted dynamic pricing (and that their strategies are relatively unsophisticated), we also observe that *algo sellers* have a significant competitive advantage versus non-algo sellers, especially with respect to winning the Buy Box.

Overall, we view our empirical work as being complementary to theory. Our data can be used to refine existing models, bound their parameters, or evaluate their behavior under realistic conditions. More broadly, our observations about the strategies adopted by businesses in practice can potentially motivate the design of new models. We make the data and code from many of our studies publicly available (additional data is available by request) at: `http://personalization.ccs.neu.edu`.

2. PEEKING BENEATH THE HOOD OF UBER

In this work [Chen et al. 2015], we examined Uber's surge pricing system, which aims to balance the demand for rides with the available supply by varying price dynamically. In the literature these ridesharing systems are conceptualized as traditional two-sided platforms [Banerjee et al. 2015; Fang et al. 2016] serving passengers and drivers. However, in terms of *implementation*, these systems are untraditional: rather than having an open marketplace and allowing the two parties to converge towards equilibrium, ridesharing companies have adopted centralized algorithms that attempt to balance supply and demand. The closed nature of these ridesharing platforms raises questions about whether the dynamic pricing mechanisms are efficient and fair.

2.1 Methods and Data Collection

To collect data for our study, we emulated 43 copies of the Uber smartphone application. By default, Uber's app requests fresh data from Uber's servers every 5 seconds, including: 1) the eight closest available cars to the user (based on GPS coordinates), 2) the current surge multipliers, and 3) the Estimated Wait Times (EWTs) for cars. By carefully spoofing the GPS coordinates for our 43 emulated

Fig. 1: Uber measurement points in downtown San Francisco.

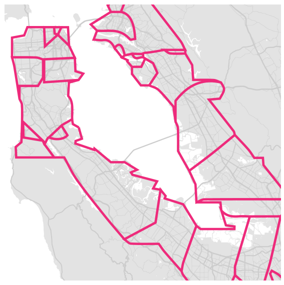

Fig. 2: Uber surge area map for the Bay Area.

users, we were able to place them in a grid throughout a target city, thus enabling us to passively collect data that covered the entire area. Figure 1 shows the measurement grid we used to collect data from downtown San Francisco. Furthermore, we were able to observe when each car became unavailable, which implies that either the driver had logged-off from Uber, or that they accepted a ride request. This enabled us to place an upper-bound on *fulfilled* demand.

Using our methodology, we collected data from midtown Manhattan between April 3–17, 2015, and from downtown San Francisco between April 18–May 2, 2015. Additionally, before collecting data at scale, we performed trials to make sure that our emulated users did not *induce* surge pricing. In these tests, we placed all 43 emulated users at a single, remote GPS coordinate late at night, and did not observe any surge pricing for one hour. We repeated this trial many times at many locations, and never observed surge prices.

2.2 Surge Pricing Algorithm

Based on our observed data, we are able to draw several conclusions about Uber's surge pricing mechanism. The system divides each city into areas (that we suspect are statically defined by human operators), and updates the surge multiplier for each area at five minute intervals. Figure 2 shows the surge areas for a subset of the San Francisco Bay Area.

If we treat surge prices as a time series and calculate the cross-correlation versus other variables, we observe statistically significant correlations between surge prices, available supply, and fulfilled demand at a time delta of -5 minutes. This suggests that Uber's algorithm calculates the surge price $s_t(a)$ at time t in area a using the supply and demand from area a measured over the previous five minute interval. While this demonstrates that Uber's dynamic pricing algorithm is highly responsive, it also means that surge prices are quite noisy (60% of surges last $\leq$10 minutes in our data).

2.3 Incentives

One of our most interesting findings concerns the incentives of the surge pricing system. Recall that Uber's goal is to equalize supply and demand by increasing the former (by incentivizing drivers with high prices) and decreasing the latter. To understand if drivers and customers are responding to these incentives in the expected way, we model the behavior of each Uber driver as a discreet-time Markov

Fig. 3: Example Buy Box from Amazon Marketplace.

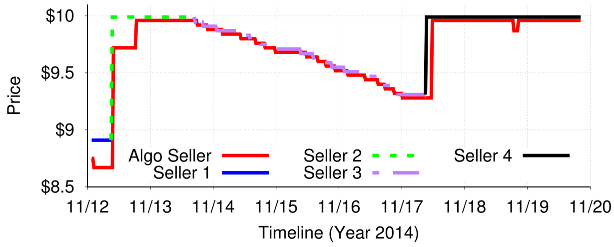

Fig. 4: Observed algorithmic seller that consistently underprices other sellers by several cents.

chain, where the state of a driver at time $t + 5$ minutes is determined by its state at t. In this model, "state" encompasses the physical location of the car and the surge prices in all areas containing and surrounding the driver. Possible state transitions include remaining stationary, accepting a ride request, or driving into an adjacent area. Using this model, we calculate the probability of state transitions at times when all areas have equal surge prices and when exactly one area of the city has surge multiplier $\geq$0.2 higher than all other areas (i.e., drivers have a strong monetary incentive to travel to the surging area).

Our results paint a complicated picture of Uber's dynamic pricing system. As expected, we find that high surge multipliers reduce demand. Bookings decrease by 7% on average in the surging area (compared to when it is not surging), while drivers that do not get booked increases by 14%. However, we also find that drivers are 13% less likely to drive into the area that is surging (compared to times when all surge multipliers are equal); in fact, the number of drivers who *leave* the surging area increases by 14%! These results suggest that the surge mechanism is not effective at incentivizing drivers. Our findings echo qualitative findings by Lee et al. who interviewed Uber drivers, and found that veteran drivers find it futile to "chase the surge" [Lee et al. 2015]. Our results also stand in contrast to Uber's own with respect to the benefits of the surge system [Hall and Krueger 2015].

3. ALGORITHMIC PRICING ON AMAZON MARKETPLACE

In this study [Chen et al. 2016], we examine two separate types of algorithms on Amazon Marketplace. *First*, we examine Amazon's Buy Box algorithm. This algorithm determines, for each product in the Marketplace, which seller's offer price will be shown to customers (and consequently, which seller makes the sale when the product is purchased). Figure 3 shows an example of the Buy Box for a product. In essence, the Buy Box algorithm functions as a matching mechanism between buyers and sellers in the Marketplace, and therefor it must balance the interests of customers (low prices, good service), third party sellers (revenue), and Amazon itself (revenue, overall health of the platform). However, despite the critical importance of the Buy Box algorithm, little is known about it beyond online folk wisdom.

Second, we investigate dynamic pricing strategies adopted by sellers on Amazon Marketplace. Amazon offers APIs that allow sellers to track competitors' prices in real-time and respond with their own price changes. Theoretically, sellers may also adopt strategies that increase their chances of winning the Buy Box. Although subscription-based tools like Feedvisor and Sellery have made dynamic pricing tools widely available to third party sellers on Amazon, it is unclear how widely dynamic pricing has been adopted, or what strategies are used by sellers.

3.1 Data Collection

To bootstrap our study, we crawled roughly four months of data from Amazon Marketplace. We chose 1000 best-selling products that were each offered by >1 seller, and crawled them every 25 minutes. Each time the crawler visited a product, it recorded the seller and price in the Buy Box, as well as up-to two additional pages of sellers offering that product (each page contains up-to 10 sellers, sorted roughly from low-to-high price). We chose to wait 25 minutes between crawls and limit the number of seller pages visited as a tradeoff between recency and completeness: Amazon implements strict rate limits, so more frequent visits (or more pages per visit) would have forced us to crawl fewer products overall. Furthermore, we could not use Amazon's APIs to collect data since the only way to get price updates for products is to list them for sale. We crawled data in two phases, between September 15–December 8, 2014, and between August 11–September 21, 2015.

3.2 The Buy Box

To investigate the features behind the Buy Box algorithm, we trained a Random Forest (RF) classifier to predict Buy Box winners (given a list of offers for a specific product). The intuition behind this process is that if we can train an accurate predictor, then it is likely that the feature weights in our model correspond closely to those used by the actual Buy Box algorithm. We input seven features into RF classifier, including each sellers': offer price relative to the lowest offer for the given product, average customer rating, positive feedback percentage, and enrollment in the Fulfilled By Amazon (FBA) program.

After performing 10-fold cross validation, we found that our RF classifier was able to predict Buy Box winners with 75–85% accuracy (depending on the total number of offers for a given product). In contrast, a simple predictor that always chooses the seller with the lowest offer only achieves 50-60% accuracy. This demonstrates that our RF classifier does a reasonable job of approximating the Buy Box algorithm, and that price alone is not the sole feature used by the true algorithm. Indeed, by examining the Gini coefficients associated with features in our RF model, we find that while the price feature has the highest weight, customer feedback also has significant weight.

3.3 Dynamic Pricing

To identify sellers on Amazon that use dynamic pricing, we look for sellers whose offer prices have high correlation over time with a specific *target* price time series. Example targets include the lowest overall price for a given product, or Amazon's offer price for that product. Intuitively, this methodology attempts to identify sellers whose offer price is pegged to an observable benchmark over time.

Using this methodology, we identify XXX sellers who we are confident have adopted dynamic pricing. Although this only represents 2.4% of the sellers in our dataset, their listings cover 51% of the products we crawled. 70% these *algo* *sellers* choose to peg their offer price within $1 above the lowest available price for each product. However, despite setting higher prices than competitors, we observe that algo sellers much more likely to win the Buy Box than non-algo sellers. This clearly demonstrates that sellers who adopt automation are at a competitive advantage versus sellers who do not. Finally, we note that algo sellers are responsible for the vast majority of price and Buy Box changes on Amazon Marketplace; we even observe a small number of products with thousands of price changes over the course of a month.

4. CONCLUSION

As commerce moves online, the opportunities to construct fluid, dynamic marketplaces increase. In some cases, like online display advertising, the structure and mechanisms in these new markets are relatively well understood. However, in other cases, like ridesharing and e-commerce, the algorithms being adopted by industry are opaque.

The overarching goal of our work is to increase the transparency of algorithms in online markets. This can help consumers and producers make more informed choices about how to best optimize their behavior on these platforms. We also hope that our work is beneficial to the theory community, as a starting point for evaluating existing models, or even motivating new designs.

REFERENCES

Banerjee, S., Johari, R., and Riquelme, C. 2015. Pricing in Ride-Sharing Platforms: A Queueing-Theoretic Approach. In *Proc. of EC*.

Chen, L., Mislove, A., and Wilson, C. 2015. Peeking Beneath the Hood of Uber. In *Proc. of IMC*.

Chen, L., Mislove, A., and Wilson, C. 2016. An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace. In *Proc. of WWW*.

Chen, M. K. 2016. Dynamic pricing in a labor market: Surge pricing and flexible work on the uber platform. In *Proc. of EC*.

Fang, Z., Huang, L., and Wierman, A. 2016. Boosting sharing economy: Social welfare or revenue driven? *CoRR abs/1604.01627*.

Hall, J. V. and Krueger, A. B. 2015. An analysis of the labor market for uber's driver-partners in the united states.

Hannak, A., Soeller, G., Lazer, D., Mislove, A., and Wilson, C. 2014. Measuring Price Discrimination and Steering on E-commerce Web Sites. In *Proc. of IMC*.

Lee, M. K., Kusbit, D., Metsky, E., and Dabbish, L. 2015. Working with machines: The impact of algorithmic and data-driven management on human workers.

Mikians, J., Gyarmati, L., Erramilli, V., and Laoutaris, N. 2012. Detecting price and search discrimination on the internet. In *Proc. of HotNets*.

Mikians, J., Gyarmati, L., Erramilli, V., and Laoutaris, N. 2013. Crowd-assisted search for price discrimination in e-commerce: First results. In *Proc. of ACM CoNEXT*.

Taft, D. K. 2014. Amazon buy box: The internet's $80 billion sales button. eWeek. `http://www.eweek.com/enterprise-apps/slideshows/amazon-buy-box-the-internets-80-billion-sales-button.html`.