Surge Pricing Solves the Wild Goose Chase

JUAN CAMILO CASTILLO, Stanford University  
DAN KNOEPFLE, Uber Technologies  
GLEN WEYL, Microsoft Research and Yale University

ACM Reference format:
DOI: http://dx.doi.org/10.1145/3033274.3085098

Ride-hailing applications (apps) like Uber and Lyft introduced a matching technology and market design that recent research has found is more efficient than traditional taxi systems [2]. However, unlike traditional street-hailing taxi systems, they are prone to a failure mode first anticipated by [1]. In this paper we model and empirically establish the existence of these dynamics. We then show how surge pricing and, to a lesser extent, other market design interventions can prevent this problem from crippling a ride-hailing market.

An over-burdened dispatch system results in available idle drivers being too thinly spread throughout a city, forcing matches between drivers and passengers that are far away from each other. Cars are thus sent on a wild goose chase (WGC) to pick up distant customers, wasting drivers’ time and reducing earnings. This effectively removes cars from the road both directly (as the cars are busy making pick-ups) and indirectly (as cars exit in the face of reduced earnings), exacerbating the problem. This harmful feedback cycle results in a dramatic fall in welfare, hurting both drivers and passengers. A ride-hailing market that falls into WGCs frequently might therefore perform worse than traditional street-hailing taxi systems, so it is essential to understand WGCs in order to design markets in a way that avoids WGCs and exploits the potential welfare gains from the new technology.

[1] dismissed WGCs as Pareto-dominated equilibria and thus just a theoretical curiosity. However, we show that when prices are too low relative to demand all equilibria of the market are WGCs when using a first-dispatch protocol, in which an idle driver is immediately dispatched every time a rider requests a trip (as many ride-hailing services have committed to).

This suggests two ways in which pricing can avoid WGCs. First, one might set a single high price all the time, sufficiently high to avoid WGCs even at peak-demand periods. Of course this design has the drawback that prices will
be unnecessarily high, and thus demand inefficiently suppressed, at times of low demand. A more elaborate mechanism is to use dynamic “surge pricing” that responds to market conditions. Such a system was introduced by Uber early in its development. Prices are set high during peak-loads, but can fall when demand is more normal. Thus, against the common perception, surge pricing allows ride-hailing apps to reduce prices from the static baseline instead of increasing them.

We build a theoretical model composed of demand, supply and matching in a homogeneous spatial region in order to highlight the phenomenon of WGCs. After calibrating physical features of the transportation process based on Uber data, this theory yields a sharp and simple prediction about conditions causing WGCs. This allows us to empirically investigate the fit of the theory by seeing whether the system does enter a failure mode when these conditions are satisfied. We use data from Uber in Manhattan, and while surge appears sufficiently responsive to avoid WGCs during most of the day, during thin hours and in thin areas, our theoretical conditions for WGCs occur with some regularity. We show that during these periods a variety of negative performance metrics spike: many riders and drivers cancel rides and waiting times jump.

We then use our calibrated model to analyze the welfare effects of surge pricing. The effects of WGCs dominate more traditional price theoretic considerations and imply that welfare and profits fall dramatically as price falls below a certain threshold and only gradually move in price above this point. Thus, the main concern for a ride-hailing platform when deciding how to price is to avoid WGCs and bans on surge pricing would lead to prices that are always close to the surge prices charged at peak times.

Pricing is not the only tool ride-hailing apps can use to avoid WGCs. We discuss two alternative approaches. First, rationing rides when demand is high avoids over-burdening the market and WGCs. However, this makes the service unreliable, eliminating one key advantage of ride-hailing over traditional taxis. Second, setting a small maximum dispatch radius also avoids WGCs, but it creates passenger queues. Passengers then have to wait without being matched to a driver and without knowing how long they will have to wait to be picked up. A maximum dispatch radius is thus in tension with a user interface feature of current ride-hailing apps—that riders know immediately upon request the location and trajectory of a car driving towards them. This feature is considered very appealing to riders and our internal interviews suggest product leaders at Uber would be loath to compromise that element of the rider experience. Hence, although surge pricing is not the only way to avoid WGCs, alternative approaches have drawbacks that limit their appeal to ride-hailing platforms.

In the last few years, ride-hailing apps have promoted the use of pooling services like UberPool and Lyft Line. We find that these services can complicate and exacerbate WGCs.

REFERENCES