Clipboard Health's product team case study #1: Lyft Toledo case solutions

Danis Kiziridis[1]

November 28, 2023

Abstract
In this scenario of launching Lyft's ride-scheduling services in Toledo, riders call for rides, and drivers might answer, earning a ride by paying a commission fee to the company. The problem is that the higher this fee is—and consequently Lyft's income—the lower is the number of rides the drivers offer, leading to more dissatisfied riders (and drivers) who quit more often, hence increasing the costs of acquiring new ones. Thus, the aim was to find the optimum driver cost that maximises the total net revenue in the first 12 months of operation.

Two settings were applied. First, a fixed ratio of 166 riders per driver was assumed, to comply with the provided empirical data, with the optimum driver cost estimated computationally and mathematically at $\$2.33$. Second, a fluctuating ratio around $166{:}1$ was assumed, with simulations and mathematical derivations showing that dynamic pricing is more profitable than fixed at $\$2.33$. In specific, by regulating the driver cost linearly with the ratio, extra profit mainly arose when increasing the driver cost by $\$1$ for every 18 fewer riders per driver.

It was also found that this optimum driver cost is the one leading to exact answering of all calls. Any further decrease would reduce costs from the fewer quitting and replenished drivers, but this would not compensate for the relatively higher decrease of income from the nevertheless all-answered calls (drivers offering more rides than requested). Finally, it is shown that the results are robust when drivers quit even less (saving further replenishment costs), and if the numbers of customers are larger or smaller (still at the $166{:}1$ ratio).

[1] Personal website

Contents

1. Problem overview

As pricing product managers of Lyft in Toledo, we need to investigate the launching of the company in one route, i.e. airport–downtown in either direction. The conventional cost for the rider has been $\$25$, from which the driver cost has been $\$6$ (paid to Lyft; the rest to the driver).

Under this $\{\$25, \$6\}$ scheme, a driver has been found to serve around 100 rides per month, while there is 20% monthly chance to quit. A new driver's acquisition cost for Lyft is $\$500$.

A rider needs a lift around once per month, and there is a 10% monthly base rate of quitting, unless a call was not answered at least once, thereafter changing the feeling of the rider, quitting irrevocably with 33% chance on any subsequent month. A new rider's acquisition cost is $\$10$–$\$20$ for the company.

The problem is that under the prevalent $\{\$25, \$6\}$ scheme, only 60% of the riders are answered to their monthly call. After experimentally changing the scheme to $\{\$25, \$3\}$, though, 93% of the riders were given a lift.

The aim is to estimate the optimum driver cost per ride that maximises the company's total net revenue for the first 12 months of the company's operation in Toledo.

2. Thoughts, assumptions, relations

For the openness and reproducibility of the present analyses, it is good to share among us the code developed when trying out the Toledo problem (downloadable here).

We know that the company will not charge the riders with more than $\$25$ per ride. We will assume that this rider cost will not be less than $\$25$ either, since this would limit our potential for profit. Additionally, we do not have information relating rider cost to the acquisition rate of new riders, or the number of rider calls per month. Thus, we will simply assume a fixed rider cost of $\$25$, allowing us to focus only on the driver cost.

The answering rate must have increased from 60% to 93% because drivers were more motivated to give a lift when experimentally charged $\$3$ per ride, presumably leading them to offer more than 100 rides per month. An issue here is that this experiment returned data for only about the rate of answered calls; that is, we do not have data about the change in the driver rides per month, neither how this lower driver cost affected the 20% quit rate of the drivers. We will then assume an equivalent percent change in the driver quit rate, i.e. the 55% increase in the answering rate would be accompanied by a 55% decrease in the driver quit rate, decreasing from 20% to 9% per month when charged from $\$6$ to $\$3$ per ride, respectively.

For simplicity, we will furthermore assume that the answering rate is equivalent to the percent of riders being given a lift in one month, since a rider calls once per month on average.

The problem and the available empirical data are in monthly rates, and hence we will work with months as our unit of time, i.e. in monthly time steps.

Related to time, a question arises about the time of quitting. With the rationale that drivers and riders are sensible enough to try out the service for at least one month before quitting, we can assume that they can resign by the end of the month. That is, quitting occurs after all riders made their monthly call, and after all drivers attempted to make all of their expected rides. By the end of a month, Lyft can know the number of quitted customers, and can invest in attracting more of them.

Related again to time, but also to customer acquisition and replenishment, there is another question about the time of paying the customer acquisition cost (CAC), and the time of acquiring a driver or rider who will be ready for service. Although paying CAC and acquiring new customers is supposed to happen in a continuous-in-time fashion, let's assume for modelling simplicity and for monitoring purposes that the payment of CAC occurs by the end of each month, after customer quitting, and that new drivers or riders are acquired and ready by the start of the next month.

A follow-up question, then, is what strategy does the company adopt to acquire new customers. Does the company acquire a fixed number of drivers and riders independently of how many quit? Alternatively, does the company have a pre-specified strategy of increasing or keeping steady the numbers of customers, or perhaps using a more adaptive strategy? The description of the problem provides us with a clue which will encourage a simplifying assumption: we were given with the percentage of riders with answered calls each month, and actually for two different driver costs. Consequently, we have to assume that the company adopts the strategy of keeping steady the numbers of drivers and riders, or at least the riders:drivers ratio. If only the riders or the drivers change, the 60% answer rate would not be valid. Any provided percentage of answered calls would also not be usable if the ratio of riders:drivers could change between months, or between prevalent and experimental cost for the driver.

In specific, if according to the problem's description a driver gives 100 lifts per month, which results to 60% answering rate under the $\$6$ charging rate, then there must be approximately 166 riders calling per month for each driver, with the 66 of the riders not being answered (40%). For the 100 monthly lifts per driver, the 60% would have been lower (higher) if there had been more (fewer) riders calling. We hence have to assume a fixed riders:drivers ratio of $166{:}1$.

Thus, to leave out another dimension of the problem, for better focusing on our aim of total net revenue maximisation, we will finally assume that the company's strategy is to initially invest to a certain number of drivers and riders before the start of the first month, and will thenceforth replenish any quitting customers at the end of each month, so that their absolute numbers and their $166{:}1$ riders:drivers ratio are preserved at the start of each month. By solving any model with smaller or larger absolute numbers of riders and riders, while keeping the same ratio, we can test whether the results are robust to the numbers of customers (Section 8).

3. Selecting the type of model

Taking into account the thoughts, assumptions and relations of Section 2, an easy and fast starting point in modelling the Toledo process would be a simulation model. A simulation model forces us to explicitly state and implement all the inner workings of the process, yet without formal mathematical language. In this exploratory way, we can straightaway emulate the system, enquire and validate its behaviour, hence testing our understanding and plausibility of the rules implemented in the modelled process.

A second step after working with the simulation model would be to build its equivalent, mathematical model. While a simulation run will be a stochastic realisation of the Toledo process, since drivers and riders have a random chance of quitting, a mathematical model can neatly describe its average behaviour. Thus, we need to analytically solve the mathematical model only once, in comparison to the simulation model that we have to numerically solve enough times for estimating its average behaviour.

Additionally, the mathematical model provides values continuously for any parameter value, whereas the simulation model will return results for discrete-valued input. Finally, the mathematical model can provide explicit and concise expressions about specific quantities or relations of interest, deepening our understanding of the system.

A comparison between the average behaviour of the simulation model and the solution of the mathematical model is commonly an additional test of our understanding and implementation of the modelled system, both in rule-based and in mathematical terms, respectively.

4. Simulation model for fixed ratio

Before describing the steps of the model during each time step of a simulated month, we have to introduce the two relations which were assumed on the basis of the empirical data provided in the problem's description (Section 4.1). After the description of the simulation model (Section 4.2), we can inspect its behaviour both in stochastic runs of the net revenue during the 12-month period (Section 4.3), and in their 12-month total net revenue for different values of driver cost (Section 4.4).

4.1 Implemented relations

First, we have to establish a relation between the driver cost per ride and the number of rides per month that a driver is willing to give (left y-axis of Fig. 1). A simple, linear relation was assumed, passing exactly through two pairs of values. Note that only the riders' 93% answer rate was given in the description of the experimentally reduced driver cost of $\$3$, i.e. the monthly number of provided rides by the drivers was not given. Thus, we had to derive the monthly number of rides per driver, since the simulation model required it as input, i.e. the answer rate was just a consequence. Given the fixed numbers of $D = 100$ drivers and $R = 16,600$ riders that were assumed for each month, each driver had to provide 154 rides per month, to obtain an answer rate of 93% of the riders. Note that the 54% increase of the monthly rides, i.e. from 100 to 154 rides, is close to the 55% increase in the answer rate, i.e. from 60% to 93% of the riders when reducing the driver cost from $\$6$ to $\$3$.

Second, the driver quit rate was also a linear function of driver cost per ride (right y-axis of Fig. 1). Again, here, the driver quit rate was not provided for the experimentally reduced driver cost of $\$3$. As already explained in one of the assumptions of Section 2, an equivalent to the answer rate's percent change in the driver quit rate was assumed, i.e. the 55% increase in the answering rate would be accompanied by a 55% decrease in the driver quit rate, decreasing from 20% to 9% per month when charged from $\$6$ to $\$3$ per ride, respectively.

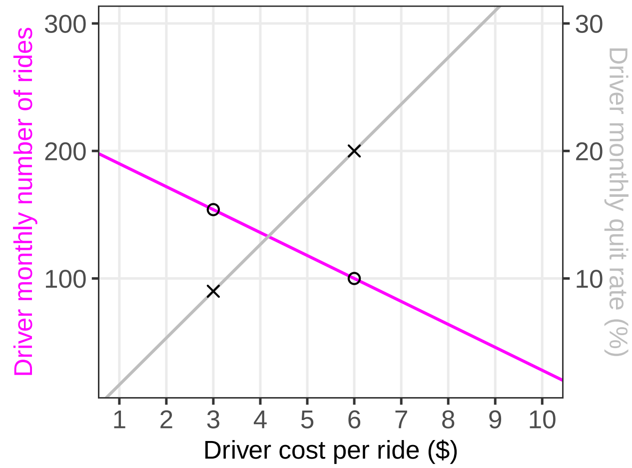

Figure 1. The two assumed linear relations between the company's commission fee to a driver per ride and: (left y-axis) the driver's monthly number of rides willing to give; and (right y-axis) the driver's monthly quit rate. The two pairs of points through which the two lines pass are real or assumed empirical data from the description of the problem.

4.2 Model steps

A simulation was run for a given driver cost as the focal parameter of input. In each simulation, time was simulated in months in a loop, while on each month the same processes were taking place.

Before the start of the first month, two tasks were necessary. First, given the driver cost, we had to predict the number of rides a driver would be able to give in each month, together with the driver's quit rate, from the two linear relations (Fig. 1). Second, we had to record the company's CAC of acquiring all the initial drivers and riders, and register all riders as initially feeling happy about the service.

After completing these two initialisation tasks, the simulation model executed for each month the following steps:

1. Calculate the total number of rides offered by all drivers, and select randomly that number of riders whose calls will be answered this month.

2. If a rider's call was not answered, and the feeling of the rider was *happy*, turn the rider's feeling to *sad*.

3. Record the company's income from the rides and the driver cost per ride.

4. For each driver and rider, make a binomial test with probability equal to their quit rate, to find out who quit.

5. Record the company's CAC of replenishing all quitted drivers and riders.

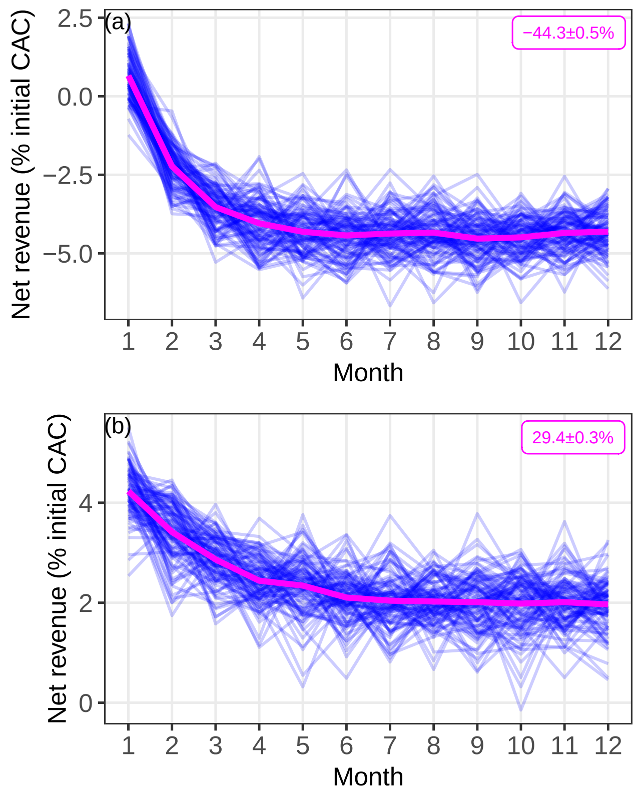

Figure 2. Simulated net revenue when drivers pay $\$6$ (a) or $\$3$ (b) per ride during the 12-month period. (a) Each driver did 100 rides per month, resulting in 60% monthly answer rate. (b) Each driver did 154 rides per month, resulting in 93% monthly answer rate. Net revenue is expressed as a percentage of the company's cost of acquiring the fixed numbers of drivers and riders before the start of the first month. The thinner and darker trajectories are from 100 stochastic runs of the simulation model, whereas the thicker and brighter curves are the monthly average. The labels indicate the 100 runs' mean total (sum of) net revenue across the 12 months ($\pm$ the 95% confidence interval).

6. Remove the quitted riders from the *feelings* list, and register the new riders as feeling happy about the company.

7. Record the month's net revenue, i.e. the income from the answered calls minus the CAC from the replenishments.

After finishing the iteration of these steps for all months sequentially, the total net revenue was calculated as the sum of net revenue from all months.

4.3 Sample runs

As a first test, the 12-month period was simulated 100 times with the prevalent driver cost of $\$6$ per ride (Fig. 2a). For this driver cost, a driver provided 100 rides per month, according to the linear relation (Fig. 1). As a result, 60% of the riders got answered to their monthly ride, in agreement with the description of the problem. Note that net revenue is expressed as percentage of the CAC for acquiring the initial drivers and riders before launching the service. In that way, we can assess how much of the initial investment is paid off during the 12-month period. For the $\$6$ driver cost per ride, the mean total net revenue was $-44.3 \pm 0.5\%$ of the initial CAC.

As a second test based on the experimentally decreased driver cost of $\$3$, the simulation model predicted a positive total net revenue of $29.4 \pm 0.3\%$ the initial CAC (Fig. 2b). The documented 93% monthly answer rate was confirmed from the simulations, with drivers now offering 154 rides per month.

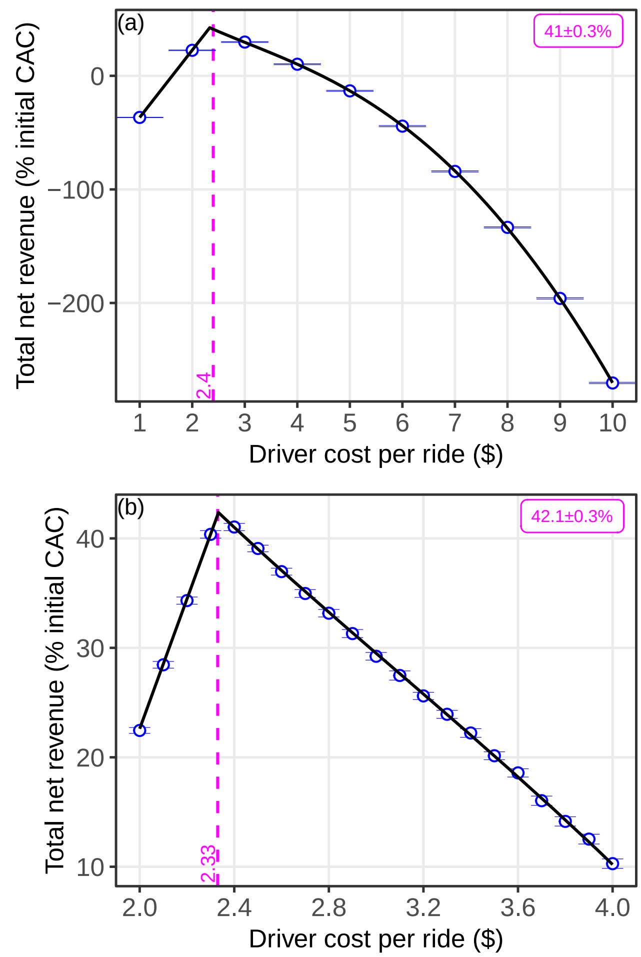

Figure 3. Simulated 12-month total net revenue for different driver costs. (a) The dashed vertical line indicates the optimum driver cost for a higher, ¢10 accuracy according to panel (b). (b) The dashed vertical line indicates the driver cost for the highest, ¢1 accuracy (not shown). The circled points are the average from 100 simulations, and the error bars are 95% confidence intervals. The labels indicate the mean maximum total net revenue at the driver cost given at the bottom of the vertical line. The solid curve is the mathematically derived function for the average behaviour of the simulated process (Eq. 1).

4.4 Maximum total net revenue

The simulations showed that the relation between the 12-month total net revenue and the driver cost per ride is hump-shaped (Fig. 3). This shape hence confirmed the existence of a maximum revenue for a specific driver cost.

By running the simulation model 100 times for different values of driver cost, the maximum total net revenue of the 12-month launching of the company in Toledo was initially located in the $\$2$–$\$4$ range (Fig. 3a). Additional simulations at the higher accuracy of ¢10 further identified a maximum total net revenue in the $\$2.3$–$\$2.5$ range (Fig. 3b). A last set of simulations in the latter range and at ¢1 accuracy, finally located the maximum revenue when the driver cost per ride was $\$2.33$. This maximum total net revenue was on average $42.1 \pm 0.3\%$ of the initial CAC.

5. Mathematical model for fixed ratio

The quantity we aimed to maximise was the 12-month total net revenue $\dot{N}(c)$ as a function of driver cost c per ride. As a 12-month total, $\dot{N}(c)$ was the sum of net revenue $N_t(c)$ from each month t:

$$\dot{N}(c) = \sum_{t=1}^{12} N_t(c). \tag{1}$$

A monthly net revenue was equal to the $E_t(c)$ money earned minus the $L_t(c)$ money lost during a month t:

$$N_t(c) = E_t(c) - L_t(c). \tag{2}$$

We will now analyse the different components of earning and losing (Section 5.1), and especially the more demanding components of losing which regard the quitting riders (Section 5.2), and their replenishment (Section 5.3). Putting back to Eq. 1 all the calculated components, we will identify the driver cost that leads to the maximum 12-month total net revenue (Section 5.4).

5.1 Earning and losing

For the money $E_t(c)$ earned monthly (Eq. 2), the company charged the drivers a cost c per ride, for each of the $A(c)$ answered calls during a month t:

$$E_t(c) = cA(c). \tag{3}$$

Based on the empirical data given by the problem's description, the number of answered calls $A(c)$ was equal to the number of drivers D times the monthly number of $a(c)$ rides each driver was willing to give, which was assumed to be a linear function of the driver cost c (Fig. 1, left y-axis):

$$A(c) = a(c)D = (\delta c + \varepsilon)D, \tag{4}$$

with the intercept $\varepsilon = 208$ monthly rides available when driver cost is zero, and the slope $\delta = -18$ monthly rides per extra dollar charged.

Note that as the driver cost c was decreased, more and more of the R riders' calls were answered, until all of them were answered. Thus, for any further decrease in c, the answered calls remain at the upper boundary of $A(c) = R$. Given the fixed numbers of D drivers and R riders in a month that were assumed, the maximum driver cost c_A that leads exactly all riders to be answered can be calculated from Eq. 4 by solving for c if $A(c) = R$:

$$c_A = (\frac{R}{D} - \varepsilon)/\delta. \tag{5}$$

For the fixed $R = 16,600$ riders and $D = 100$ drivers assumed monthly, $c_A = \$2.33$ if $A(c) = R$.

For the money $L_t(c)$ lost monthly in CAC (Eq. 2), the driver acquisition cost DAC and the rider acquisition cost RAC were applied to the $Q_t(c)$ quitting drivers and $U_t(c)$ app-uninstalling riders at the end of each month t, since it was assumed that the company's strategy was to merely replace any quitting customers, in order to keep their numbers fixed by the start of the next month:

$$L_t(c) = Q_t(c)DAC + U_t(c)RAC. \tag{6}$$

The $Q_t(c)$ drivers quitting at the end of month t were the product of the D drivers' quit rate $q(c)$. The latter was again based on the empirical data given by the problem's description, assuming to be a linear function of the driver cost c (Fig. 1, right y-axis):

$$Q_t(c) = q(c)D = (\zeta c + \eta)D, \tag{7}$$

with the intercept $\eta = -2\%$ of riders quitting when driver cost is zero, and the slope $\zeta = 3.6\%$ additional chance of quitting for every extra dollar charged.

Despite the implausibility of having a negative quit rate as an intercept, the intended $\$1$–$\$10$ range of studied driver cost was using only non-negative quit rates, since the driver cost for a zero quit rate was:

$$\zeta c + \eta \le 0 \Rightarrow c \le \frac{-\eta}{\zeta} = \frac{2\%}{3.6\%\$^{-1}} = \$0.56, \tag{8}$$

below any reasonable charging price under consideration. Nevertheless, any negative quit rate for the drivers could be better forced to be equal to zero, by re-formulating Eq. 7 as a piecewise function with the help of Ineq. 8:

$$Q_t(c) = \begin{cases} 0, & \text{if } c \le -\eta/\zeta, \\ (\zeta c + \eta)D, & \text{if } c > -\eta/\zeta. \end{cases} \tag{9}$$

5.2 Riders quitting

For the $U_t(c)$ uninstalling riders by the end of month t, it is convenient to distinguish the two sub-phases which were implied in the simulation model: (1) the $\{t, 0\}$ start of the month (pre-calling), with the replenished from the end of the previous month drivers and riders; and (2) the $\{t, 1\}$ end of the month (post-calling), with the riders'

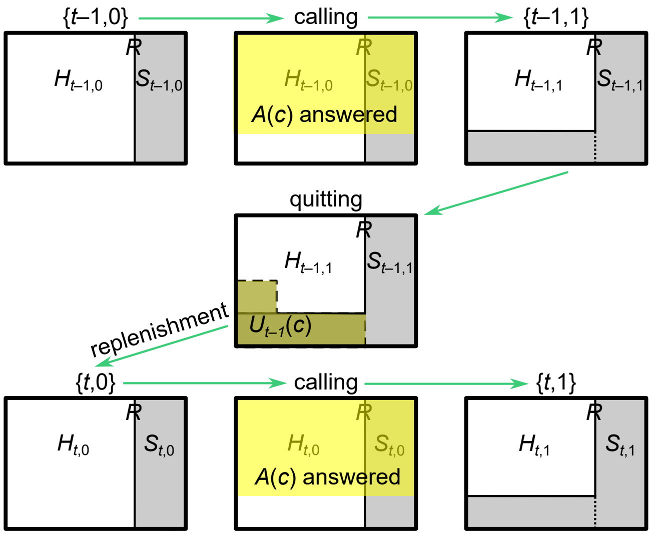

Figure 4. Schematic representation of the fate of the R riders from month $t-1$ (first and second row) to month t (third row). Happy riders are in white background, sad ones in grey. In the first row, the $H_{t-1,0}$ happy and $S_{t-1,0}$ sad riders at the $\{t-1,0\}$ start of month $t-1$ call for a ride, $A(c)$ of the R riders are answered, leading to a change of attitude in the $H_{t-1,0}$ who did not get a lift by the $\{t-1,1\}$ end of month. In the second row, $U_{t-1}(c)$ happy and sad riders quit before the start of the next month, and are replenished with happy riders at the start $\{t,0\}$ of the next month t. In the third row, the process described in the first row for the previous month $t-1$ is repeated for this month t.

feeling about the company's service possibly changing from *happy* to *sad*, depending on their current feeling and whether their calls were answered.

Any rider quitting on month t was the product from the latter $\{t,1\}$ sub-phase, i.e. of the $H_{t,1}(c)$ happy and $S_{t,1}(c)$ sad riders after calling occurred, given their respective quit rates u_h and u_s:

$$U_t(c) = u_h H_{t,1}(c) + u_s S_{t,1}(c), \tag{10}$$

with $u_h = 10\% < u_s = 33\%$ of the respective riders.

The numbers of happy and sad riders post-calling depended on the $H_{t,0}(c)$ happy and $S_{t,0}(c)$ sad riders pre-calling, from the $\{t,0\}$ sub-phase of the month t (Fig. 4, third row). In specific, if we reasonably assume that a driver answers independently of whether a calling rider feels happy or sad, the same proportion $A(c)/R$ of answers must be distributed to the happy and sad riders. Thus, the $H_{t,1}(c)$ happy riders post-calling are the happy ones that their calls were answered:

$$H_{t,1}(c) = \frac{A(c)}{R} H_{t,0}(c). \tag{11}$$

The $S_{t,1}(c)$ sad riders post-calling are the sum of the already sad riders who anyway do not change their feeling—answered or not—and of the happy ones that their calls were not answered:

$$\begin{aligned} S_{t,1}(c) &= S_{t,0}(c) + (1 - \frac{A(c)}{R}) H_{t,0}(c) \\ &= S_{t,0}(c) + H_{t,0}(c) - \frac{A(c)}{R} H_{t,0}(c) \\ &= R - \frac{A(c)}{R} H_{t,0}(c), \end{aligned} \tag{12}$$

i.e. the sad riders at the end of the month are equal to all the riders minus the happy riders with answered calls (Fig. 4, third row). Utilising the fact that the happy and sad riders at any month phase constitute all the riders R, i.e. $H_{t,0}(c) + S_{t,0}(c) = R$, Eq. 12 hence has the helpful characteristic of depending only on the happy riders at the start of the month, just like Eq. 11 for the happy riders.

Plugging the Eqs. 11 and 12, for respectively the happy and sad riders at the end of a month, back into Eq. 10 for the number of uninstalling riders, we have:

$$\begin{aligned} U_t(c) &= u_h \frac{A(c)}{R} H_{t,0}(c) + u_s R - u_s \frac{A(c)}{R} H_{t,0}(c) \\ &= u_s R - (u_s - u_h) \frac{A(c)}{R} H_{t,0}(c). \end{aligned} \tag{13}$$

In other words, the $U_t(c)$ uninstalling riders by the end of month t are expressed as all quitting with the higher rate u_s of the sad riders, but in the second term this number is reduced by taking into account the lower quit rate u_h of the happy riders with answered calls.

5.3 Riders replenished

Having expressed all the $U_t(c)$ uninstalling riders of month t in terms of only the $H_{t,0}(c)$ happy riders at the start of the month, it is now necessary to express $H_{t,0}(c)$ in terms of the $H_{t-1,0}(c)$ happy riders of the previous month $t-1$. This is important because of the transfer of riders from month to month, as well as of their feeling and their quitting–replenishment dynamics, all of which must be implemented in the mathematical model to better mimic the simulation model.

By inspecting the schematic representation of quitting and replenishment in the second and third rows of Fig. 4, we can mathematically express the $H_{t,0}(c)$ happy riders at the start of the month as equal to the $H_{t-1,1}(c)$ happy ones from the end of the previous month, plus the quitting ones from the $S_{t-1,1}(c)$ sad which were replaced by happy:

$$H_{t,0}(c) = H_{t-1,1}(c) + u_s S_{t-1,1}(c). \tag{14}$$

The replacement of quitting happy riders by new happy riders cancels out, and hence does not appear in Eq. 14 which is expressed in terms of riders from the end of the previous month.

We can then express the $H_{t,0}(c)$ riders in terms of riders from the start of the previous month by plugging the already derived Eqs. 11 and 12 into Eq. 14:

$$H_{t,0}(c) = \frac{A(c)}{R} H_{t-1,0}(c) + u_s \Big(R - \frac{A(c)}{R} H_{t-1,0}(c) \Big)$$
$$= u_s R + (1-u_s) \frac{A(c)}{R} H_{t-1,0}(c). \quad (15)$$

Thus, we have the desired $H_{t,0}(c)$ happy riders to feed into Eq. 13 for the $U_t(c)$ uninstalling riders, which will feed Eq. 6 for all the CAC-related losses $L_t(c)$, which will then feed Eq. 2 for the net revenue of month t, in order to calculate the total net revenue as the sum across the 12 months in our final target Eq. 1.

The problem here is that we cannot use Eq. 15 in the total net revenue's sum of the months, because each month depends on the previous month in this recursion Eq. 15. Nevertheless, we can obtain a solution of the recursion relation because we know its value at the first month $t = 1$. In specific, we know that at the start of the first month, all riders are happy, i.e. $H_{1,0}(c) = R$. With this initial condition, we can calculate the solution of Eq. 15, which will provide us with the $H_{t,0}(c)$ happy riders of each month without any dependence on $H_{t-1,0}(c)$:

$$H_{t,0} = \frac{R^2 \Big(u_s A(c) + M^t R - M^t A(c) \Big)}{A(c) \Big(R + u_s A(c) - A(c) \Big)}, \quad (16)$$

with $M = (1-u_s)A(c)/R$.

5.4 Maximum total net revenue

We can finally calculate the 12-month total net revenue with Eq. 1. Note, though, that the number of answered calls can be at most equal to the number of riders, since each rider calls once per month, i.e. $A(c) \leq R$. We have already derived the maximum driver cost c_A for exactly $A(c) = R$ with Eq. 5. Thus, the final result is a piecewise Eq. 1 of 12-month total net revenue as a function of driver cost c, which uses $A(c)$ normally according to Eq. 4 when $c \geq c_A$, but uses $A(c) = R$ when $c < c_A$.

The average total net revenue from the simulation model was approximated remarkably well by this function of the mathematical model (solid line in Fig. 3a,b). Within the 95% confidence interval of the simulation model (Fig. 3b), the mathematical model estimated a maximum total net revenue of 42.3% the initial CAC under the optimum driver cost of $\$2.33$ per ride (in ¢1 accuracy).

6. Combining the two models' results

We estimated with both the simulation and the mathematical model that the company will be paid off around the 42% of its initial investment in CAC after the first 12 months of operation by charging the optimum driver cost of $\$2.33$ per ride (Fig. 3b). A subsequent question would consider after how many months all the initial CAC would be paid off. The numerical and analytical solution of respectively the simulation and the mathematical model for 36 months agreed that the cumulative total net revenue will first exceed the initial investment to CAC on the 29^{th} month of operation (Fig. 5). The mathematical model's estimation of the average monthly net revenue was equal to 3.51% of the initial CAC (solid, black, horizontal line in Fig. 5), within the 95% confidence interval of the simulation model's temporal average of $3.5 \pm 0.01\%$. By the end of the 36^{th} month, the total net revenue was estimated to be around $126 \pm 0.5\%$ of the initial CAC.

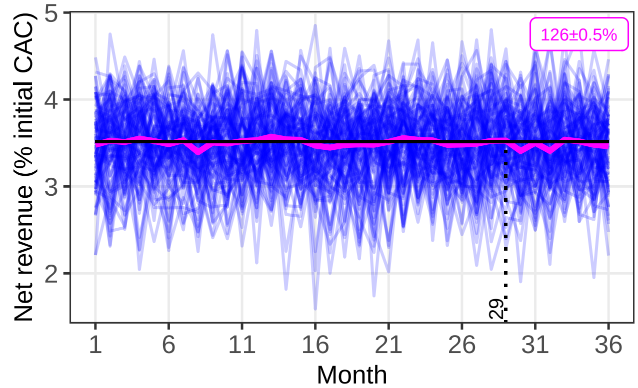

Figure 5. Simulated net revenue when drivers pay the optimum $\$2.33$ per ride during a 36-month period. This driver cost was estimated by the simulation and the mathematical model as the one maximising total net revenue. The thinner and darker trajectories are from 100 stochastic runs of the simulation model, whereas the thicker and brighter curve is the monthly average. The label indicates the 100 runs' mean total (sum of) net revenue across the 36 months ($\pm$ the 95% confidence interval). The black, solid, horizontal line is the estimation of the average monthly net revenue from the mathematical model. The dotted vertical line indicates the month when the cumulative net revenue first exceeded the initial investment in CAC.

The simulations showed that the maximum total net revenue is obtained for the optimum driver cost of $\$2.33$ per ride. The same commission fee was found during the derivation of the mathematical model when searching for the maximum driver cost that leads exactly to 100% of riders being answered to their monthly call (Eq. 5).

Combining these results from the two models, it seems that if the driver cost drops even further after attaining a 100% answer rate, all riders are still happy, quitting with the lower quit rate, but the company has lower income from the rides. At the same time, though, there is a further drop in the quit rate of drivers, which could save the company from driver CAC (right y-axis of Fig. 1).

Since the maximum total net revenue decreases for any further decrease of the optimum driver cost (Fig. 3), the loss of income from the drop of driver cost must have been greater than the saving of CAC from the

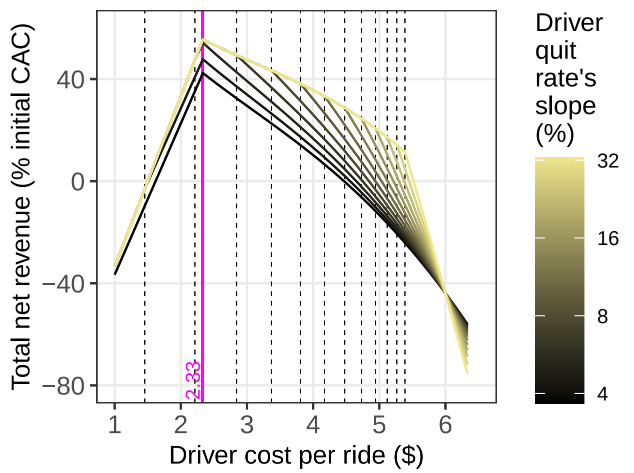

Figure 6. The mathematical model of Eq. 1 for the 12-month total net revenue as a function of the driver cost c, and of the slope of driver quit rate ζ. The different curves are for higher slopes ζ of the Eq. 9 between driver quit rate and c, when the line is forced to pass through the only empirical point of $q(\$6) = 20\%$ (Fig. 1). For steeper slopes of the relation, the dashed vertical lines indicate the higher driver costs below of which the driver quit rate was zero. The solid vertical line indicates the driver cost which resulted to the maximum total revenue in the mathematical model for any slope ζ. The minimum $\zeta = 3.6\%\$^{-1}$ of the darkest curve is the one originally used in the simulation and the mathematical model (right y-axis in Fig. 1).

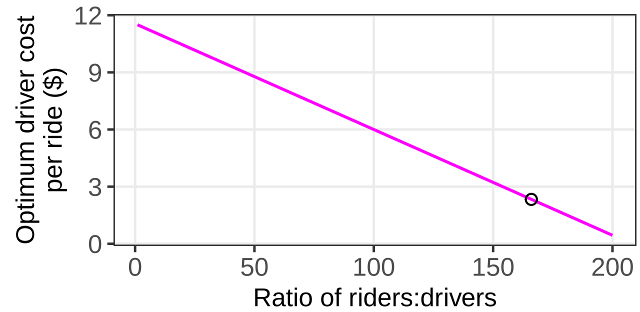

Figure 7. Optimum driver cost per ride for any given ratio of riders:drivers. This linear relation was taken from Eq. 5 for the maximum driver cost that leads to exactly all riders' calls to be answered, but now with a variable R/D ratio, i.e. after relaxing the assumption of a fixed ratio. The circled point is located at the $\$2.33$ driver cost for the 166 riders per driver ratio used in the simulation and the mathematical model so far.

fewer quitting drivers who all need to be replenished. Indeed, if the price drops, for example, from $\$2.33$ to $\$2$, the income from answering all $R = 16,600$ riders would decrease:

- from $\$2.33 \times 16,600$ answered riders $= \$38,678$,
- to $\$2 \times 16,600$ answered riders $= \$33,200$.

At the same time, the cost from the quitting drivers' $\$500$ CAC would only decrease:

- from $\$500 \times 100$ drivers $\times\ 0.07$ quit rate $= \$3,500$,
- to $\$500 \times 100$ drivers $\times\ 0.05$ quit rate $= \$2,500$.

This is a $\$5,478$ decrease in income while only saving $\$1,000$ from the fewer quitting drivers' replenishment.

However, someone could hypothesise that these savings from the decreasing quit rate of the drivers could have been higher if a steeper slope had been assumed between the quit rate and the driver cost (Fig. 1). Perhaps, under a steeper slope, the savings would have been greater from fewer drivers quitting, surpassing the loss of income from the reduced driver cost. By forcing this relation to pass through the only empirical point of $q(\$6) = 20\%$ (Eqs. 7 and 9), we can plot the mathematical model of Eq. 1 for different values of steeper slopes of this linear function (Fig. 1).

For steeper slopes of the relation between driver quit rate and driver cost per ride, the maximum total revenue was still located at the driver cost of $\$2.33$, although the maximum total net revenue was higher, as expected due to the lower and even zero driver CAC (Fig. 6). As the slope of this relation becomes steeper, the driver cost below of which the quit rate is zero becomes higher (dashed vertical lines in Fig. 6). Nevertheless, the increase in the total revenue is uniform for the $\leq \$6$ region, such that there is no relocation of the $\$2.33$ driver cost associated with the revenue's maximum.

7. Relaxing the fixed ratio assumption

Returning to the evidently important Eq. 5 for the maximum driver cost that leads to exactly all riders' calls to be answered, we can now relax the assumption of a fixed riders:drivers ratio, by considering the R/D term as a variable. Thus, this linear relation essentially provides a simple kind of dynamic pricing scheme with the optimum driver cost that the company should charge to the drivers so that exactly all riders' calls are answered for any ratio of riders:drivers (Fig. 7). Given the parameter values of Eq. 5 from Eq. 4, the relation of Fig. 7 tells us that the driver cost must decrease by $\$1$ for every 18 more riders per driver, so that each driver offers 18 more rides monthly (slope δ in Eq. 4). This dynamic pricing scheme can be implemented in the simulation model (Section 7.1), and extra mathematical derivations can shed light on the reasons of its efficiency in comparison to a fixed pricing scheme (Section 7.2).

As long as we charge drivers with the price of $\$2.33$, each driver will be willing to offer 166 rides monthly, and all riders will be answered if the ratio is $166{:}1$. If the ratio becomes larger than 166 riders per driver, then the company should reduce the driver cost according to the linear function of Fig. 7, to motivate the drivers to offer more rides, for serving exactly the proportionally more numerous riders. Conversely, if the ratio becomes lower than $166{:}1$, then the company should increase the driver cost to decrease the offered rides, leading to the

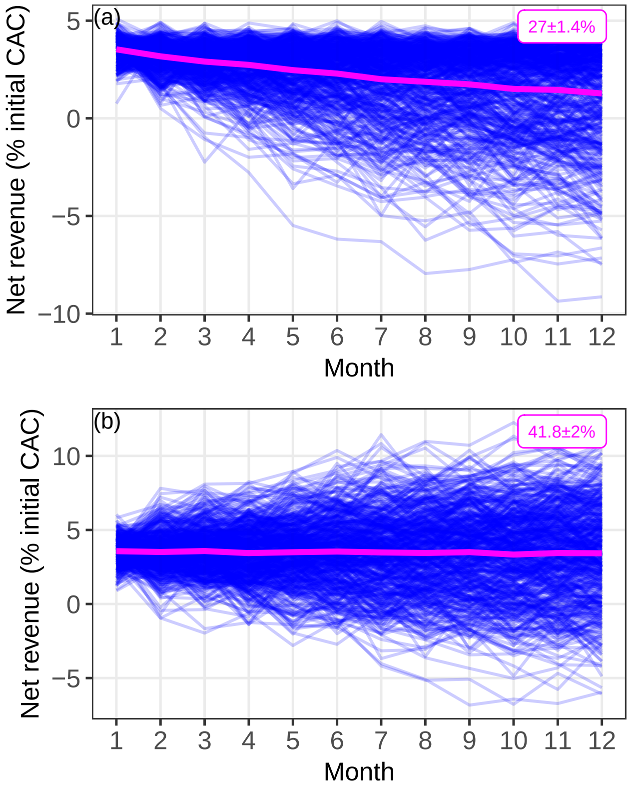

Figure 8. Running the simulation model when the riders:drivers ratio is variable, under a fixed driver cost (a), versus a dynamic pricing scheme (b). The initial CAC were equal to the ones used with the simulation model so far (Figs. 2 and 3). The driver cost per ride for the dynamic pricing scheme of panel (b) was obtained from the variable ratio via the relation of Fig. 7. The thinner and darker trajectories are from 500 stochastic runs of the simulation model, whereas the thicker and brighter curves are their monthly average. The labels indicate the 500 runs' mean total (sum of) net revenue across the 12 months ($\pm$ the 95% confidence interval).

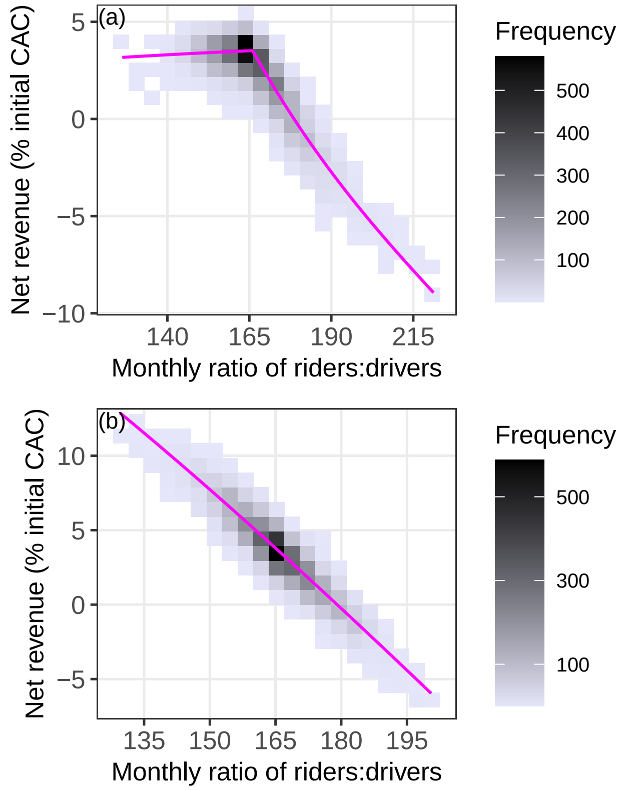

Figure 9. Monthly ratio of riders:drivers and respective net revenue in the 500 simulation runs corresponding to the two schemes of Fig. 8a,b: (a) fixed driver cost; and (b) variable driver cost. Tile darkness denotes the frequency of months with the specific pair of ratio–revenue values. The bright curves are mathematically derived expectations (Appendix A).

maximisation of net revenue by not allowing drivers to wait idle for long at the route's two endpoint stands. The basic idea of this dynamic pricing scheme resembles the one used by Lyft (Crapis and Sholley, 2020).

7.1 Simulation model

The only change to the simulation model that was needed for implementing the variable riders:drivers ratio was in the replenishment of the quitting customers. Instead of replenishing exactly their numbers for preserving a fixed ratio, replenishment became a Poisson distributed random variable with a mean equal to the number of quitting riders or drivers.

To test the efficiency of this dynamic pricing scheme of the linear relation in Fig. 7, the simulation model was run under two pricing schemes while the riders:drivers ratio was now fluctuating.

Under the first, fixed pricing scheme, the simulations started with the usual, initial CAC of introducing $R = 16,600$ riders and $D = 100$ riders, and the driver cost was fixed at $\$2.33$ which was the optimum for the initial $166{:}1$ ratio (Fig. 8a).

The second, dynamic pricing scheme started similarly to the first one, but the optimum driver cost was changing according to the relation of Fig. 7, following any change in the ratio at the start of each month (Fig. 8b).

The dynamics of net revenue from 500 runs of each scheme showed that dynamic pricing led to significantly higher on average total net revenue than the fixed pricing scheme (labels in Fig. 8a,b). This average total net revenue from the dynamic pricing scheme was similar to the one obtained for the fixed $\$2.33$ driver cost but back when also the ratio was fixed (Fig. 3b), since the mean of the random replenishments was around the number of quitting customers, leading to a variable ratio which fluctuated around the same ratio of 166 riders per driver.

The lower average of the total net revenue from the fixed pricing scheme was basically due to its monthly revenue trajectories which were bounded from above at around 5% of the initial CAC (Fig. 8a), corresponding to ratios lower than $166{:}1$ (Fig. 9a). Ratios higher than 166 riders per driver corresponded to lower monthly net revenues. On the contrary, dynamic pricing with higher driver cost for these lower ratios resulted in higher net revenues than with fixed pricing, although net revenues for ratios higher than $166{:}1$ were similar (Fig. 9b).

7.2 Mathematical derivations

The underlying reasons for the shape of these two relations from the fixed and dynamic pricing schemes could be revealed by deriving mathematically their expectations shown with the bright curves in Fig. 9a,b (described mathematically in Appendix A). For both pricing schemes under a variable ratio, the fluctuations of the ratio were mainly driven by the fluctuating number of drivers (Eq. 17), and not due to the riders whose number was assumed steady on average and equal to the initially acquired R (details in Appendix A.1).

Under the fixed pricing scheme (bright curve in Fig. 9a), and for ratios lower than $166{:}1$, the number of answered calls from the proportionally more drivers was bounded by the relatively stable number of riders, since the number of offered rides was fixed at 166 per driver due to the fixed driver cost (Eq. 21). Additionally, all the riders were happy for any lower ratio, since they were all answered, hence quitting with the lower rate, resulting in lower CAC for their replenishment (Eq. 20). Thus, the slight decrease in the mathematically derived net revenue for ratio lower than 166 riders per driver was due to the increasing number of drivers, which led to more quits and hence replenishments (Eq. 18), although driver quitting occurred at a fixed rate due to the fixed driver cost of $\$2.33$. For ratios greater than 166 riders per driver, the decreasing relationship of net revenue with the ratio in Fig. 9a was due to the decreasing number of drivers resulting in lower income due to fewer provided rides (Eq. 21), but also to higher rider CAC due to more unanswered and hence sad riders who quitted more often (Eq. 20), although driver-related CAC was lower due to the fewer (quitting) drivers.

Under the dynamic pricing scheme (bright curve in Fig. 9b), there were no different regimes below and above the $166{:}1$ ratio for a couple of reasons. First, since the driver cost was taking the maximum value such that drivers were offering exactly as many rides as needed, the number of answered calls was on average equal to the number R of riders (Eq. 25), and the CAC from the quitting riders was on average the same across the different ratios because all riders were answered and happy (Eq. 24). Second, the income from all the answered calls was lower for a higher ratio, since the driver cost was lower for the same number R of answered calls (Fig. 7). We could expect that the dynamic pricing performs better than fixed pricing also in this higher ratio region, since higher ratio was associated with fewer drivers and with lower quit rate due to the associated lower driver cost, leading to lower CAC. Nevertheless, it appeared that with the given CAC values per customer, the lower CAC from driver quitting and replenishment was surpassed by the lower income from the lower driver cost, as illustrated in the following.

For example, by focusing on two ratios which are 20 riders per driver below and above the $166{:}1$ ratio, we can compare the changes in income and driver CAC. For a ratio of 146 riders per driver, the driver cost was $\$3.4$, leading to a monthly income of $\$3.4 \times 16,600$ riders $= \$57,178$, and to driver CAC from on average 13 quitting drivers $\times$ $\$500 = \$6,500$. For a ratio of 186 riders per driver, the driver cost was $\$1.2$, leading to an income of $\$1.2 \times 16,600$ riders $= \$20,289$, and to a driver CAC of 3 quitting drivers $\times$ $\$500 = \$1,500$. Thus, by an increase in the ratio from $146{:}1$ to $186{:}1$, and consequently by decreasing the driver cost from $\$3.4$ to $\$1.2$, we lose $\$36,889$ in income from the rides, while we only save $\$5,000$ in driver CAC.

We could, of course, implement a more adaptive dynamic pricing scheme, to increase the net revenue, especially for the $> 166{:}1$ ratios, but this would be out of scope in this already lengthy report.

8. Final considerations

A final test of the models would consider the effect of having different absolute numbers of drivers and riders, while keeping the (average) rider:driver ratio equal to the desired 166 riders per driver. After solving the models with lower and larger absolute numbers of customers, it was found that the average behaviour of the system was not significantly different, and only the variation around the average behaviour would be larger for lower numbers of drivers and riders (results not shown).

References

Crapis, D. and Sholley, C. Dynamic pricing to sustain marketplace balance, 2020. URL. Accessed: February 02, 2023.

A. Derivations for variable ratio

Any mathematical calculation of monthly net revenue for our problem had three components: the profit from the answered calls for rides, the CAC for the replenished drivers, and the CAC for the replenished riders. Some preliminary analyses about the characteristics of the variable ratio will be useful (Appendix A.1), before proceeding to the derivation of the three components of monthly net revenue for the fixed pricing scheme (Appendix A.2), and for the dynamic pricing scheme (Appendix A.3).

A.1 Preliminaries

There are three elements to consider about the fluctuating ratio, which will help in the derivations.

First, note that since the numbers of replenished customers were Poisson distributed with a mean equal to the numbers of quitted customers, the variances of these Poisson distributed numbers were equal to their respective means, and the customers are expected to fluctuate around their initial numbers of $R = 16,600$ riders and $D = 100$ drivers. For the fixed driver cost of $\$2.33$ per ride, each driver offered 166 rides monthly, and all riders were initially happy since 100 drivers $\times$ 166 rides $= 16,600$ answered calls, equal to the average number of riders. This means that 10% of the riders were quitting on average at the end of the first month ($1,660$ riders on average), while 6.5% of the drivers (≈ 7) were quitting on average, given the fixed driver cost and the associated linear relation in Eq. 7 (right y-axis in Fig. 1).

Second, returning to the Poisson distributed replenishments, their main implication on the ratio was that its fluctuations were mainly influenced by the fluctuations in the numbers of drivers. This can be illustrated by looking at how the fluctuating denominator or numerator of the ratio can affect the ratio. On the one hand, any fluctuation of the replenished drivers within three standard deviations from its mean, i.e. from the number of quitted drivers, would change the 166:1 ratio in the limits of $16,600$ riders / [100 drivers $\pm 3\sqrt{(}7$ quitted drivers)], i.e. from 154 to 180 riders per driver. On the other hand, any equivalent fluctuation of the riders would change the 166:1 ratio in the limits of [$16,600$ riders $\pm 3\sqrt{(}1,660$ quitted riders)] / 100 drivers, i.e. from 165 to 167 riders per driver only. Thus, we can assume that the fluctuating ratio is mainly driven by the fluctuating number of drivers, while the number of riders remains on average equal to the initial $R = 16,600$ riders.

Third, in that way, we can estimate the average number of drivers at the extreme ratios that appeared in the simulations. For the lowest ratio $r_{min} \approx 125$ riders per driver, the drivers D_{min} must have been $R/D_{min} = r_{min} \Rightarrow D_{min} = R/r_{min} \approx 133$ drivers on average; equivalently, for the highest ratio $r_{max} \approx 225$ riders per driver, the drivers $D_{max} \approx 74$ drivers. Taking these two pairs of ratio–drivers values, together with the known pair of the 166:1 ratio corresponding to the number of initial drivers $D = 100$, we can build a quadratic equation that passes exactly through these three points, to predict the variable number of drivers $D_v(r_v)$ for any ratio r_v in the range produced by the simulations:

$$D_v(r_v) \approx 0.004r_v^2 - 1.8r_v + 306.6. \quad (17)$$

This relation, together with the assumption of steady R riders and mainly variable $D_v(r_v)$ drivers, were confirmed by plotting the actual numbers and dynamics from the simulations (not shown).

A.2 Net revenue for variable ratio and fixed cost

Let's now derive the three components of net revenue for the fixed pricing scheme: the quitting drivers first, then the quitting riders, and finally the answered calls.

Given the variable number $D_v(r_v)$ of drivers as a function of the variable r_v ratio (Eq. 17), we can calculate the expected monthly number $Q_{v,f}(r_v, c = \$2.33)$ of quitted drivers for any observed value of the variable ratio, given the fixed driver cost c leading to 6.5% driver quit rate:

$$Q_{v,f}(r_v) = 0.065D_v(r_v). \quad (18)$$

For the number $U_{v,f}(r_v)$ of quitting riders, we can assume that from the relatively steady number R of riders all are happy by the end of any month when the ratio $r_v \leq 166$, because the call of everyone is answered, and hence they quit with 10% rate, leading to $U_{v,f}(r_v \leq 166) = 0.1R = 1,660$ quitting riders monthly. As the ratio r_v becomes larger than 166 riders per driver, with each driver offering 166 rides monthly due to the $\$2.33$ fixed driver cost, not all riders' calls are answered. In specific, we can start with the base rate of all R riders quitting like happy ones, but the ones whose calls are not answered and were happy since the start of the month t will quit with the additional chance of sad riders (33% sad quit rate $-$ 10% happy quit rate $= 23\%$). The number $S_{v,f,t}$ of these newly sad riders can be calculated from the variable ratio's excess of riders per driver, times the number of drivers (given the fixed driver cost and hence the fixed 166 monthly rides per driver):

$$S_{v,f,t} = (r_v - 166)D_v(r_v), \quad (19)$$

with $r_v > 166$:1. The contribution of these newly sad drivers to the quit rate of the so far happy riders will then be $0.23S_{v,f,t}$.

Additionally, we have to take into account the extra riders who were already sad from previous months. Since their monthly quit rate is 33%, half of the newly sad riders of a month will quit and be replenished in approximately two months, while all of them will on average quit and be replenished in three months. In specific, since the quit probability for the sad riders is $0.33 \approx 1/3$, the probability of not quitting can be taken as around $1 - 1/3 = 2/3$. We can then ask how many months will pass for a sad rider to have $1/2$ probability to quit (or in other words, for half of the sad riders to quit). This can be formulated as $(2/3)^n = 1/2$, and if we solve for n, we get $n = 1.71$ months for half of the sad riders to quit, counting the month they got sad. This is the median, and the average can be taken from the expected value of the Geometric distribution with $p = 1/3$, i.e. $\sum_{n=1}^{\inf} np(1-p)^{n-1} = 1/p = 3$ months for all sad drivers to quit on average. Thus, half of the already sad riders became sad in the previous month (median), and the rest became sad on average in the month before (mean). Given the already derived number $S_{v,f,t}$ of happy riders who became sad at any month t (Eq. 19), the already sad riders originate on average from the two previous months, and are equal to $S_{v,f,t-1} = 0.5S_{v,f,t}$ for the previous month, and $S_{v,f,t-2} = 0.5S_{v,f,t}$ for the month before. In total, then, the sad quitting riders of a month are $S_{v,f} = S_{v,f,t} + S_{v,f,t-1} + S_{v,f,t-2} = 2S_{v,f,t}$. The contribution of all $S_{v,f}$ sad riders can then be added to the so far happy quitted riders of the month with the use of Eq. 19, and the number of quitting riders for both ratio regimes is:

$$U_{v,f}(r_v) = \begin{cases} 0.1R, & \text{if } r_v \leq 166, \\ 0.1R + 0.23[2(r_v - 166)D_v(r_v)], & \text{if } r_v > 166. \end{cases} \tag{20}$$

Finally, for the number $A_{v,f}(r_v)$ of answered calls, we expect all the R riders' calls to be answered when $r_v \leq 166$, while fewer riders will be answered as r_v becomes higher than 166 riders per driver, times the number of drivers, each offering the fixed number of 166 monthly rides:

$$A_{v,f}(r_v) = \begin{cases} R, & \text{if } r_v \leq 166, \\ R - (r_v - 166)D_v(r_v), & \text{if } r_v > 166. \end{cases} \tag{21}$$

Taking all three components together (Eqs. 18,20,21), the expected monthly net revenue under the fixed pricing scheme for a variable ratio r_v with a driver cost c is:

$$N_{v,f}(r_v, c = \$2.33) = A_{v,f}(r_v)c - Q_{v,f}(r_v)DAC - U_{v,f}(r_v)RAC, \tag{22}$$

for the fixed driver cost $c = \$2.33$, and the CAC for each driver ($DAC = \$500$) and each rider ($RAC = \$15$).

A.3 Net revenue for variable ratio and variable cost

Let's derive the three components of net revenue for the dynamic pricing scheme in the same order as in Appendix A.2. In the dynamic pricing scheme, the driver cost c was not fixed, but it was taking the maximum value c_A such that all riders' calls were answered under a variable ratio $R/D = R/D_v(r_v) = r_v$, according to Eqs. 5 and 17 (Fig. 7).

Under this scheme, the driver quit rate was additionally changing as a linear function of c_A (Eq. 7; right y-axis of Fig. 1). Thus, the number of quitting drivers under a variable ratio and driver cost was:

$$Q_{v,v}(r_v, c_A) = (\zeta c_A + \eta)D_v(r_v). \tag{23}$$

For the number of quitting riders, since at the start of each month the driver cost c_A was adjusted according to the riders:drivers ratio, such that all riders' calls will be answered, all riders were quitting with the happy quit rate of 10%:

$$U_{v,v}(r_v) = 0.1R. \tag{24}$$

Finally, the number $A_{v,v}(r_v)$ of answered calls was on average equal to the riders R, again because all were answered due to the appropriate driver cost c_A:

$$A_{v,v}(r_v) = R. \tag{25}$$

Taking all three components together (Eqs. 23,24,25), the expected monthly net revenue under the dynamic pricing scheme for a variable ratio r_v with a shifting optimum driver cost c_A is:

$$N_{v,v}(r_v, c_A) = A_{v,v}(r_v)c_A - Q_{v,v}(r_v, c_A)DAC - U_{v,v}(r_v)RAC, \tag{26}$$

with the same CAC for each driver ($DAC = \$500$) and each rider ($RAC = \$15$).