This page is intentionally blank. Proper e-companion title page, with INFORMS branding and exact metadata of the main paper, will be produced by the INFORMS office when the issue is being assembled.

Appendix A: Multihoming statistics excluding low-activity users

In Section 3, we report that about 56.7% of customers ride exclusively with Uber, 26.7% exclusively with Lyft, and the remaining 16.6% ride with both Uber and Lyft. However, these statistics could be skewed by the presence of riders who only took a very small number of trips during the year. Specifically, since riders who took only one or two trips would naturally appear to use only one platform, including them inflates the observed fraction of single-homing customers.

To provide a more complete picture, we report the extent of single-homing and multihoming behavior while progressively excluding users with low total ride counts. Table EC.1 presents descriptive statistics by applying minimum ride filters, ranging from at least one trip to at least six trips in the year.

Each row corresponds to the subset of users who took at least x number of rides, with x indicated in the first column. Columns two and three report the number of users and the total number of rides taken by this filtered group, respectively. The next three columns present the percentages of users who rode exclusively with Uber, exclusively with Lyft, and those who multihomed by using both platforms. Finally, the last three columns report the distribution of total rides attributable to Uber-only users, Lyft-only users, and multihoming users.

While the fraction of multihoming users increases after excluding low-activity users, it remains relatively low at 35% among users who took at least six rides, reaffirming that most riders consistently use a single platform.

Min Rides Filter	Filtered # Users ($\times 10^5$)	Filtered # User Rides ($\times 10^6$)	Uber Only Users (%)	Lyft Only Users (%)	Multi-homing Users (%)	Uber User Rides (%)	Lyft User Rides (%)	Both User Rides (%)
1	1.56	1.39	56.72	26.71	16.57	44.42	18.27	37.31
2	1.13	1.35	52.22	24.20	23.58	43.74	17.78	38.49
3	0.88	1.30	49.50	22.62	27.87	43.04	17.31	39.65
4	0.72	1.25	47.63	21.50	30.87	42.47	16.92	40.61
5	0.61	1.21	46.38	20.60	33.02	42.03	16.57	41.40
6	0.53	1.17	45.31	19.90	34.79	41.64	16.27	42.09

Table EC.1 Descriptive statistics of platform usage across users with increasing minimum ride thresholds. The table reports user composition and ride distribution by Uber-only, Lyft-only, and multihoming users as low-activity users are progressively excluded.

Appendix B: Pricing Model

Our dataset includes information on the prices of rides that the customers choose, but it lacks the corresponding prices offered by competing platforms for the same trips. This missing information is crucial for our model, so we have developed a price prediction model to estimate the prices of the competing platform.

Given the trip characteristics of a ride, such as length, duration, rain, and rush hour information, we develop a model to predict Uber's price for the rides taken with Uber. Now, let us consider a Lyft ride for which we have the trip characteristics and the price the customer paid for the Lyft ride. However, we do

not have information on the price Uber would have offered if the customer had checked the fare for the same origin and destination at the same time. In this scenario, we utilize Uber's price prediction model to estimate this price by using the corresponding trip characteristics. We repeat the same process to estimate Lyft's price for the rides taken with Uber.

We now discuss the details of the price prediction model[9] for each platform. We will drop the subscripts i, j, and t for the remainder of this section. The pricing strategy for both Uber and Lyft is based on three components: *base fare*, *surge multiplier*, and *toll* as follows:

$$\textit{trip fare} = (\textit{base fare}) \cdot (\textit{surge multiplier}) + \textit{toll}, \tag{EC.1}$$

where *base fare* is the fixed component that depends only on *trip length* and *trip duration*. Then, a *surge multiplier* is multiplied by the base fare, and finally, the cost of tolls is added to this fare if the ride involves a toll road.

The *base fare* is given by

$$\textit{base fare} = \delta_0 + \delta_1 \cdot \textit{trip length} + \delta_2 \cdot \textit{trip duration}, \tag{EC.2}$$

where δ_0 is the flat fee for every ride that accounts for the pickup cost. The platform then charges a per-mile rate (δ_1) for the distance traveled and a per-minute rate (δ_2) for the trip duration.

Surge pricing is implemented during periods of high demand and limited driver availability, such as peak hours, popular events, heavy traffic, or inclement weather conditions. Since this *surge multiplier* is not directly observable, we model the *surge multiplier*[10] using the following equation:

$$\textit{surge multiplier} = \delta_3 + \delta_4 \cdot \textit{imbalance} + \delta_5 \cdot \textit{rush hour} + \delta_6 \cdot \textit{weekend} + \delta_7 \cdot \textit{rain} + \epsilon^s, \tag{EC.3}$$

where *rush hour* is an indicator variable that takes a value of 1 when the trip starts from 7 a.m. to 9 a.m. and 4 p.m. to 6 p.m. on a weekday and 0 otherwise. *weekend* is an indicator variable that takes a value of 1 when the trip is on the weekend and 0 otherwise. *rain* is an indicator variable that takes a value of 1 if it is raining and 0 otherwise. Since we do not observe special events that lead to high demand, we use the imbalance in observed pickup and drop-off data as a proxy for the demand-supply imbalance. Specifically, we calculate the ratio of pickups to drop-offs within a one-hour time window in specific taxi zones[11] using the TLC dataset. It is important to note that TLC data tracks the pickup and drop-off locations for each ride, allowing us to calculate this ratio.

It is important to note that the trip fare without surge and tolls constitutes a fixed component without uncertainty. The main source of uncertainty in our estimation arises from the surge component, denoted as ϵ^s. Finally, the complete equation for *trip fare* can be written as follows:

$$\begin{aligned}\textit{trip fare} = &(\delta_0 + \delta_1 \cdot \textit{trip length} + \delta_2 \cdot \textit{trip duration}) \\ &\cdot (\delta_3 + \delta_4 \cdot \textit{imbalance} + \delta_5 \cdot \textit{rush hour} + \delta_6 \cdot \textit{weekend} + \delta_7 \cdot \textit{rain} + \epsilon^s) + \textit{toll}.\end{aligned}$$

[9] https://www.ridester.com/uber-rates-cost/

[10] https://www.ridester.com/surge-pricing/

[11] New York City is divided into 262 taxi zones.

Since this model is not linear in parameters, we cannot employ an OLS regression framework. We could potentially estimate this model using Bayesian estimation methods while preserving the model structure, but it would be computationally intensive. Therefore, we estimate a linear model with interaction terms and origin-destination fixed effects. Although this model does not preserve the exact structure of the pricing mechanism, it is more general and theoretically recovers the true parameters given enough data. The final model is given by

$$\begin{aligned} \textit{trip fare} = \omega_0 + \omega_1 \cdot \textit{trip length} + \omega_2 \cdot \textit{trip duration} + \omega_3 \cdot \textit{imbalance} + \omega_4 \cdot \textit{rush hour} \\ + \omega_5 \cdot \textit{weekend} + \omega_6 \cdot \textit{rain} + \textit{interaction terms} + \textit{o-d fixed effects} + \epsilon^p, \end{aligned} \tag{EC.4}$$

where the interaction terms include (*trip length* · *imbalance*), (*trip length* · *rush hour*), (*trip length* · *weekend*), (*trip length* · *rain*), (*trip duration* · *imbalance*), (*trip duration* · *rush hour*), (*trip duration* · *weekend*), and (*trip duration* · *rain*).

Furthermore, we assume *toll* is zero for rides within the same borough, while rides crossing boroughs incur a fixed toll determined by the origin and destination boroughs. We capture this information in the model using fixed effects for each origin-destination pair.

We train a price prediction model separately for Uber and Lyft using data from the first eight months and evaluate its accuracy on the last four months' data. The Mean Absolute Percentage Error (MAPE) for the test data set was 19% for Uber and 18% for Lyft, demonstrating the model's effectiveness in estimating ride prices. Finally, we train these models on the entire dataset to obtain comprehensive pricing models for Uber and Lyft, which we utilize to predict counterfactual prices.

Appendix C: Waiting Time Model

As discussed in Section 5, we do not directly observe the waiting time for both platforms in each ride. However, since the waiting time is an important factor in our modeling, we estimate it using the publicly available TLC data. We estimate the waiting time for each taxi zone in a given hour by treating it as an independent M/M/1 queuing system. This estimation takes into account the arrival rate λ and the state-dependent service rate μ_n, for each taxi zone in a given hour, given by

$$\mu_n = \mu \cdot (n + k), \tag{EC.5}$$

n represents the system's state, indicating the number of customers waiting in the system, and k represents the platform's ability to adjust the service rate through various strategies. When there are no customers waiting in the system ($n = 0$), the service rate of the platform in a specific taxi zone during a given hour is ($\mu \cdot k$). It's worth noting that as more customers wait in a particular area, platforms often attempt to increase the service rate by incentivizing drivers to relocate to high-demand zones. For example, when $k = 1$, the service rate doubles from μ to 2μ as the number of customers waiting increases from 0 to 1.

Figure EC.1 represents the flow diagram of this system where arrows indicate possible transitions from one state to another. Let each node represent the system's state, which indicates the number of riders waiting in the system. The transition from state n to $n+1$ (i.e., arrival) occurs at constant rate λ while transitioning

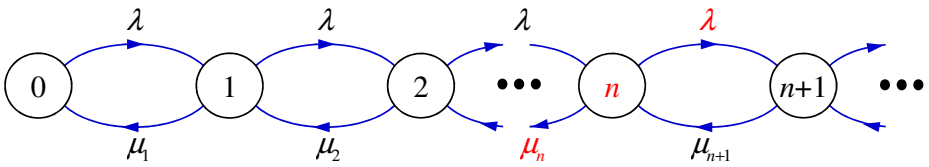

Figure EC.1 Flow diagram for the M/M/1 queuing model with state dependent service rate.

from state $n+1$ to n (i.e., departure) occurs at a state-dependent rate μ_{n+1}. This can be modeled as a birth and death CTMC with birth parameters $\lambda_n = \lambda$ for $n \geq 0$ and death parameters μ_n for $n \geq 1$.

Our goal is to analyze the steady-state behavior of this system and estimate the waiting time at a given state of the system. Assuming the system reaches a steady state, let p_n be the fraction of time the system is in state n, and the flow in and out of the system must be balanced. At the steady state, the flow from state n to $n+1$ is given by the product of p_n and λ, and the flow from $n+1$ to n is given by the product of p_{n+1} and μ_{n+1}. Equating the flow into state n and the flow out of state n, we obtain the following equilibrium equations:

$$\lambda p_0 = \mu_1 p_1, \tag{EC.6}$$

$$(\lambda + \mu_n) p_n = \lambda p_{n-1} + \mu_{n+1} p_{n+1}, \qquad n = 1, 2, \ldots \tag{EC.7}$$

Finally, the probabilities p_n are normalized to one

$$\sum_{j=0}^{\infty} p_j = 1. \tag{EC.8}$$

We solve the system of equations above by recursively expressing all probabilities p_n in terms of p_0 as follows

$$p_n = \frac{\lambda}{\mu_n} p_{n-1} = p_0 \prod_{i=1}^{n} \frac{\lambda}{\mu_i}, \qquad n = 1, 2, \ldots \tag{EC.9}$$

Finally, we solve for p_0 by substituting p_n in Equation (EC.8)

$$\begin{aligned}
1 &= p_0 + \sum_{j=1}^{\infty} p_j \\
1 &= p_0 + \sum_{j=1}^{\infty} p_0 \prod_{i=1}^{j} \frac{\lambda}{\mu_i} \\
1 &= p_0 \left[1 + \sum_{j=1}^{\infty} \prod_{i=1}^{j} \frac{\lambda}{\mu_i} \right] \qquad \text{(From Equation (EC.9))} \\
p_0 &= \left[1 + \sum_{j=1}^{\infty} \prod_{i=1}^{j} \frac{\lambda}{\mu_i} \right]^{-1}
\end{aligned} \tag{EC.10}$$

Substituting $\mu_n = \mu \cdot (n + k)$ in Equations (EC.9) and (EC.10), we get

$$\begin{aligned}
p_n &= p_0 \prod_{i=1}^{n} \frac{\lambda}{\mu(i+k)} \\
&= p_0 \cdot \left(\frac{\lambda}{\mu} \right)^n \prod_{i=1}^{n} \frac{1}{(i+k)}
\end{aligned}$$

$$= p_0 \cdot \left(\frac{\lambda}{\mu}\right)^n \cdot \frac{k!}{(n+k)!} \tag{EC.11}$$

$$\begin{aligned}
p_0 &= \left[1 + \sum_{j=1}^{\infty} \prod_{i=1}^{j} \frac{\lambda}{\mu(i+k)}\right]^{-1} \\
&= \left[1 + \sum_{j=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^j \prod_{i=1}^{j} \frac{1}{i+k}\right]^{-1} \\
&= \left[1 + \sum_{j=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^j \cdot \frac{k!}{(j+k)!}\right]^{-1} \\
&= \left[1 + k! \cdot \left(\frac{\lambda}{\mu}\right)^{-k} \cdot \sum_{j=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^{j+k} \cdot \frac{1}{(j+k)!}\right]^{-1} \\
&= \left[1 + k! \cdot \left(\frac{\lambda}{\mu}\right)^{-k} \cdot \sum_{j=k+1}^{\infty} \left(\frac{\lambda}{\mu}\right)^{j} \cdot \frac{1}{(j)!}\right]^{-1} \\
&\qquad \text{(Transformation of the summation variable)} \\
&= \left[1 + k! \cdot \left(\frac{\lambda}{\mu}\right)^{-k} \cdot \left(\exp\left(\frac{\lambda}{\mu}\right) - \sum_{j=0}^{k} \left(\frac{\lambda}{\mu}\right)^{j} \cdot \frac{1}{j!}\right)\right]^{-1}. \qquad \text{(EC.12)} \\
&\qquad \text{(Using Taylor series expansion of } \exp(\lambda/\mu)\text{)}
\end{aligned}$$

Next we compute the expected queue length, L, for $k \geq 1$ as follows

$$\begin{aligned}
L &= \sum_{j=1}^{\infty} j \cdot p_j \\
&= \sum_{j=1}^{\infty} j \cdot p_0 \cdot \left(\frac{\lambda}{\mu}\right)^j \cdot \frac{k!}{(j+k)!} \\
&= p_0 \cdot k! \cdot \sum_{j=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^j \cdot \frac{j}{(j+k)!} \\
&= p_0 \cdot k! \cdot \left(\frac{\lambda}{\mu}\right)^{-k} \cdot \sum_{j=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^{j+k} \cdot \frac{j}{(j+k)!} \\
&= p_0 \cdot k! \cdot \left(\frac{\lambda}{\mu}\right)^{-k} \cdot \sum_{j=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^{j+k} \cdot \left(\frac{j+k}{(j+k)!} - \frac{k}{(j+k)!}\right) \\
&= p_0 \cdot k! \cdot \left(\frac{\lambda}{\mu}\right)^{-k} \cdot \sum_{j=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^{j+k} \cdot \left(\frac{1}{(j+k-1)!} - \frac{k}{(j+k)!}\right) \\
&= p_0 \cdot k! \cdot \left(\frac{\lambda}{\mu}\right)^{-k} \cdot \left(\sum_{j=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^{j+k} \cdot \frac{1}{(j+k-1)!} - \sum_{j=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^{j+k} \frac{k}{(j+k)!}\right) \\
&= p_0 \cdot k! \cdot \left(\frac{\lambda}{\mu}\right)^{-k} \cdot \left(\sum_{j=k}^{\infty} \left(\frac{\lambda}{\mu}\right)^{j+1} \cdot \frac{1}{j!} - \sum_{j=k+1}^{\infty} \left(\frac{\lambda}{\mu}\right)^{j} \frac{k}{j!}\right) \\
&\qquad \text{(Transformation of the summation variables)} \\
&= p_0 \cdot k! \cdot \left(\frac{\lambda}{\mu}\right)^{-k} \cdot \left(\frac{\lambda}{\mu} \cdot \sum_{j=k}^{\infty} \left(\frac{\lambda}{\mu}\right)^{j} \cdot \frac{1}{j!} - k \cdot \sum_{j=k+1}^{\infty} \left(\frac{\lambda}{\mu}\right)^{j} \frac{1}{j!}\right) \\
&= p_0 \cdot k! \cdot \left(\frac{\lambda}{\mu}\right)^{-k} \cdot \left(\frac{\lambda}{\mu} \cdot \left(\exp\left(\frac{\lambda}{\mu}\right) - \sum_{j=0}^{k-1} \left(\frac{\lambda}{\mu}\right)^{j} \cdot \frac{1}{j!}\right) - k \cdot \left(\exp\left(\frac{\lambda}{\mu}\right) - \sum_{j=0}^{k} \left(\frac{\lambda}{\mu}\right)^{j} \frac{1}{j!}\right)\right) \qquad \text{(EC.13)}
\end{aligned}$$

$$= p_0 \cdot k! \cdot \left(\frac{\lambda}{\mu}\right)^{-k} \cdot \left(\frac{\lambda}{\mu} \cdot \left(\exp\left(\frac{\lambda}{\mu}\right) - \sum_{j=0}^{k-1} \left(\frac{\lambda}{\mu}\right)^j \cdot \frac{1}{j!}\right) - k \cdot \left(\exp\left(\frac{\lambda}{\mu}\right) - \sum_{j=0}^{k-1} \left(\frac{\lambda}{\mu}\right)^j \frac{1}{j!}\right) + \left(\frac{\lambda}{\mu}\right)^k \cdot \frac{k}{k!}\right)$$

$$= p_0 \cdot k! \cdot \left(\frac{\lambda}{\mu}\right)^{-k} \cdot \left(\left(\frac{\lambda}{\mu} - k\right) \cdot \left(\exp\left(\frac{\lambda}{\mu}\right) - \sum_{j=0}^{k-1} \left(\frac{\lambda}{\mu}\right)^j \cdot \frac{1}{j!}\right) + \left(\frac{\lambda}{\mu}\right)^k \cdot \frac{1}{(k-1)!}\right).$$

(Using Taylor series expansion of $\exp(\lambda/\mu)$ when $k \neq 0$)

Note that this expression is valid for $k > 0$. Finally, using Little's law we obtain the waiting time, W, as

$$W = \frac{L}{\lambda} = \frac{p_0 k!}{\lambda} \cdot \left(\frac{\lambda}{\mu}\right)^{-k} \cdot \left(\left(\frac{\lambda}{\mu} - k\right) \cdot \left(\exp\left(\frac{\lambda}{\mu}\right) - \sum_{j=0}^{k-1} \left(\frac{\lambda}{\mu}\right)^j \cdot \frac{1}{j!}\right) + \left(\frac{\lambda}{\mu}\right)^k \cdot \frac{1}{(k-1)!}\right). \tag{EC.14}$$

Recall that we independently estimate the waiting time in each taxi zone in a given hour. From Equation (EC.14), we see that the waiting time is influenced by three factors: λ, μ, and k. Here, λ and μ are specific to each taxi zone in a given hour, while k represents the platforms' ability to adjust their service rate, which remains consistent across taxi zones and time.

We calibrate our waiting time model by setting the arrival rate λ as the number of pick-ups we observe in each taxi zone in a given hour in the TLC data, and the service rate when the queue length is zero, $\mu \cdot k$, to the number of dropoffs we observe in each taxi zone in a given hour. We choose the value of k such that the expected waiting time prediction based on our model is equal to the average waiting time obtained from open sources. Finally, we estimate the waiting time for every ride using the estimated λ, μ, and k.

Appendix D: Proof of Proposition 1

We first rewrite the proposition when the customer considers Uber first in Phase I and decides whether to consider Lyft in Phase II as follows

PROPOSITION EC.1. *The condition* $\mathbb{E}_{P,W}[\max\{U_{\mathcal{L}it} - u_{\mathcal{U}it}, 0\}] \geq c_i$ *is equivalent to* $\hat{u}_{\mathcal{L}it} - u_{\mathcal{U}it} + \sigma_{\nu,\mathcal{L}} \cdot C_i \geq 0$, *where* $\sigma^2_{\nu,\mathcal{L}} = \alpha_1^2 \cdot \sigma^2_{P,\mathcal{L}i} + \alpha_2^2 \cdot \sigma^2_{W,\mathcal{L}i}$; α_1 *and* α_2 *represent the coefficients for price and waiting time in the utility specification, respectively;* C_i *represents the transformed search cost given by* $\phi(C_i) - C_i \cdot (1 - \Phi(C_i)) = c_i/\sigma_{\nu,\mathcal{L}}$, *where* $\phi(.)$ *and* $\Phi(.)$ *are the probability density function (pdf) and cumulative distribution function (cdf) of the standard normal variable, respectively.*

Proof: We can rewrite the condition as

$$\begin{aligned} \mathbb{E}_{P,W}[\max\{U_{\mathcal{L}it} - u_{\mathcal{U}it}, 0\}] &\geq c_i \\ \mathbb{E}_{P,W}[U_{\mathcal{L}it} - u_{\mathcal{U}it} | U_{\mathcal{L}it} > u_{\mathcal{U}it}] \cdot \Pr(U_{\mathcal{L}it} > u_{\mathcal{U}it}) + 0 \cdot \Pr(U_{\mathcal{L}it} \leq u_{\mathcal{U}it}) &\geq c_i \\ \mathbb{E}_{P,W}[U_{\mathcal{L}it} - u_{\mathcal{U}it} | U_{\mathcal{L}it} > u_{\mathcal{U}it}] \cdot \Pr(U_{\mathcal{L}it} > u_{\mathcal{U}it}) &\geq c_i \end{aligned}$$

Note that $U_{\mathcal{L}it}$ is a random variable from the customer's perspective at this stage, where the randomness comes from the uncertainty in the true price and waiting time, and the mean and variance of $U_{\mathcal{L}it}$ are given by $\hat{u}_{\mathcal{L}it}$ and $\sigma^2_{\nu,\mathcal{L}} = \alpha_1^2 \cdot \sigma^2_{P,\mathcal{L}i} + \alpha_2^2 \cdot \sigma^2_{W,\mathcal{L}i}$, respectively. We define $\nu_{\mathcal{L}it}$ such that $U_{\mathcal{L}it} = \hat{u}_{\mathcal{L}it} + \nu_{\mathcal{L}it}$, representing the difference between the realized utility Lyft and the expected utility of Lyft. It follows that $\nu_{\mathcal{L},it}$ is a

normal random variable with mean 0 and variance $\sigma^2_{\nu,\mathcal{L}}$. Finally, we define expected change in utility as $D^{\mathbb{E}}_{\mathcal{L},it} = \mathbb{E}[U_{\mathcal{L}it} - u_{\mathcal{U},it}] = \hat{u}_{\mathcal{L}it} - u_{\mathcal{U},it}$, and rewrite the expected gain as

$$\begin{aligned}
\mathbb{E}_\nu\left[\hat{u}_{\mathcal{L}it} - u_{\mathcal{U},it} + \nu_{\mathcal{L},it} | \hat{u}_{\mathcal{L}it} + \nu_{\mathcal{L},it} > u_{\mathcal{U},it}\right] \cdot \Pr(\hat{u}_{\mathcal{L}it} + \nu_{\mathcal{L},it} > u_{\mathcal{U},it}) &\geq c_i \\
\mathbb{E}_\nu\left[D^{\mathbb{E}}_{\mathcal{L},it} + \nu_{\mathcal{L},it} | \nu_{\mathcal{L},it} > -D^{\mathbb{E}}_{\mathcal{L},it}\right] \cdot \mathrm{P}(\nu_{\mathcal{L},it} > -D^{\mathbb{E}}_{\mathcal{L},it}) &\geq c_i \\
\left[D^{\mathbb{E}}_{\mathcal{L},it} + \mathbb{E}_\nu\left[\nu_{\mathcal{L},it} | \nu_{\mathcal{L},it} > -D^{\mathbb{E}}_{\mathcal{L},it}\right]\right] \cdot \mathrm{P}(\nu_{\mathcal{L},it} > -D^{\mathbb{E}}_{\mathcal{L},it}) &\geq c_i \\
\left[\frac{D^{\mathbb{E}}_{\mathcal{L},it}}{\sigma_{\nu,\mathcal{L}}} + \mathbb{E}\left[\frac{\nu_{\mathcal{L},it}}{\sigma_{\nu,\mathcal{L}}} | \frac{\nu_{\mathcal{L},it}}{\sigma_{\nu,\mathcal{L}}} > -\frac{D^{\mathbb{E}}_{\mathcal{L},it}}{\sigma_{\nu,\mathcal{L}}}\right]\right] \cdot \mathrm{P}\left(\frac{\nu_{\mathcal{L},it}}{\sigma_{\nu,\mathcal{L}}} > -\frac{D^{\mathbb{E}}_{\mathcal{L},it}}{\sigma_{\nu,\mathcal{L}}}\right) &\geq \frac{c_i}{\sigma_{\nu,\mathcal{L}}}
\end{aligned}$$

We denote $(\nu_{\mathcal{L},it}/\sigma_{\nu,\mathcal{L}})$ with a standard normal random variable z, since $\nu_{\mathcal{L},it}$ is a normal random variable with mean zero and variance $\sigma^2_{\nu,\mathcal{L}}$. We rewrite the above equation concisely by denoting $\delta_{\mathcal{L},it} = -(D^{\mathbb{E}}_{\mathcal{L},it}/\sigma_{\nu,\mathcal{L}})$ and $c'_i = (c_i/\sigma_{\nu,\mathcal{L}})$ as

$$\left[-\delta_{\mathcal{L},it} + \mathbb{E}\left[z | z > \delta_{\mathcal{L},it}\right]\right] \cdot \mathrm{P}\left(z > \delta_{\mathcal{L},it}\right) \geq c'_i.$$

We now use the expectation of truncated normal distribution to simplify as follows

$$\begin{aligned}
\left[-\delta_{\mathcal{L},it} + \frac{\phi(\delta_{\mathcal{L},it})}{1 - \Phi(\delta_{\mathcal{L},it})}\right] \cdot (1 - \Phi(\delta_{\mathcal{L},it})) &\geq c'_i \\
-\delta_{\mathcal{L},it}\left(1 - \Phi(\delta_{\mathcal{L},it})\right) + \phi(\delta_{\mathcal{L},it}) &\geq c'_i.
\end{aligned}$$

Denoting the L.H.S. of the above equation by $B(\delta_{\mathcal{L},it})$, we know that if $(1 - \Phi(\delta_{\mathcal{L},it})) > 0$, then there exists a unique pair $\delta_{\mathcal{L},it}, c'_i$ that solves the equation $B(\delta_{\mathcal{L},it}) = c'_i$ and the inverse $\delta_{\mathcal{L},it} = B^{-1}(c'_i)$ exists. Hence, given a search cost, c_i, we can find a C_i such that $B(C_i) = c'_i$, and the condition can be written as $\delta_{\mathcal{L},it} \leq C_i$. Finally, substituting the expression for $\delta_{\mathcal{L},it}$, we get the condition $\hat{u}_{\mathcal{L}it} - u_{\mathcal{U},it} + C_i \cdot \sigma_{\nu,\mathcal{L}} \geq 0$

Appendix E: Mathematical Identification Assumptions

In this section, we discuss the problem of identifying the parameters in our Bayesian learning model. Since the model is not identifiable in its current form, additional restrictions must be imposed to make it identifiable. To this end, we first set the coefficients of trip-specific covariates relative to a reference choice, which is assumed to be the Lyft platform. This is common practice when using choice models that do not have an outside option. We let $\boldsymbol{\beta}_{\mathcal{L}}$ and $\boldsymbol{\alpha}_{0,\mathcal{L}}$ be 0 and replace $\boldsymbol{\beta}_{\mathcal{U}}$ and $\boldsymbol{\alpha}_{0,\mathcal{U}}$ by $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_0$, respectively.

Next, it can be seen from Equation (15) that the separate identification of the variance for aleatoric uncertainty ($\sigma^2_{W,ji}$) and the variance for epistemic uncertainty at the first time period ($\lambda^2_{ji,1}$) is not possible, so that only their ratio $\frac{\sigma^2_{W,ji}}{\lambda^2_{ji1}}$ is identifiable. To address this issue, the value of $\sigma^2_{W,ji}$ is set to 1 for all customers and platforms, which is a standard assumption in the literature (Shin et al., 2012). Similarly, $\sigma^2_{P,ji}$ is also set to 1 for all i and j. We then assume that all customers have the same prior beliefs, that is, $\Lambda_{ji,1} = \Lambda_{j,1}$, $\hat{\boldsymbol{\gamma}}_{ji,1} = \hat{\boldsymbol{\gamma}}_{j,1}$, $\lambda^2_{ji,1} = \lambda^2_{j,1}$, and $\hat{w}_{ji,1} = \hat{w}_{j,1}$. This is because it is not possible to separate the mean and variance of prior beliefs for each customer.

Finally, $\boldsymbol{\alpha}_0$ and $\boldsymbol{C}$ represent the customer-specific parameters for the unobserved brand affinity and search cost, respectively. The estimation of these parameters can also be challenging because of the limited number of rides per customer, which leads to estimation errors and a high variability. One way to overcome this issue is to assume that these parameters are the same across all customers, but this would limit the model flexibility.

A better approach would be to use a hierarchical Bayesian model that relies both on customer-specific and population-level parameters as follows:

$$\alpha_{0,i} \sim \mathcal{N}(\mu_{\alpha_0}, \sigma^2_{\alpha_0}), \qquad C_i \sim \mathcal{N}(\mu_C, \sigma^2_C),$$

where μ_{α_0} and $\sigma^2_{\alpha_0}$ denote the mean and variance at the population level for the unobserved brand affinity; and μ_C and σ^2_C are the mean and variance at the population level for the search cost $\boldsymbol{C}$.

Appendix F: Model-free Evidence: Factors Driving Switching Between Platforms

To support and justify key aspects of our modeling framework, we provide model-free evidence on the factors influencing platform switching behavior. Specifically, we estimate a logistic regression where the dependent variable is a binary indicator representing whether a customer switched platforms from their previous ride. The independent variables include trip characteristics such as trip length, rush hour status, weekend status, airport ride status, and the ride's position in the customer's sequence of rides (i.e., purchase occasion).

As shown in Table EC.2, customers are significantly more likely to switch platforms during rush hour, on weekends, for airport trips, and for longer rides – contexts where price and wait time fluctuations are likely to be more salient. Additionally, the negative coefficient on purchase occasion suggests that customers are more likely to switch platforms earlier in their observed lifecycle, which is consistent with our structural assumption that customers become more certain in their beliefs over time. These findings provide strong justification for including the aforementioned covariates as alternative-specific variables in our model, given their effect on riders' switching behavior.

	Switching
Intercept	-1.409***
	(0.004)
Purchase occasion	-0.015***
	(0.000)
Trip length	0.032***
	(0.000)
Rush hour	0.037***
	(0.005)
Weekend	0.029***
	(0.005)
Airport	0.415***
	(0.009)
N	1,400,304
Log-Likelihood	-660,571

Table EC.2 Impact of trip characteristics on switching behavior.

Appendix G: Empirical Evidence to Support Identification of Consideration Sets

A key feature of our model is its ability to capture how customers form consideration sets, which in turn drives their multihoming behavior. Since consideration sets are latent, that is, unobservable in the data, a natural question arises: is there sufficient variation in the data to infer whether a rider considered only the chosen platform or both platforms before making a ride request?

We address this by examining how customers respond to fare changes on their default versus alternative platforms. If riders always considered both platforms before each request, then a fare increase of $x on the default platform should have the same effect as a fare decrease of $x on the alternative platform. In this symmetric case, only the price difference matters, implying that both platforms were actively considered. In contrast, if the effects of price changes are asymmetric, it suggests riders did not always consider both platforms when choosing. See Abaluck and Adams-Prassl (2021) for more details where the authors used a similar logic to show the identifiability of their model from the data.

To test this, we present model-free empirical evidence showing asymmetric responses to price changes, validating the need to explicitly model consideration sets. Specifically, we define the default platform as the one used in a rider's most recent trip. We then examine two correlations: (1) between the default platform's price and the probability of switching to the alternative platform, and (2) between the rival platform's price and the switching probability. To isolate the effect of fare changes from trip-specific factors, we restrict the analysis to customers with prior rides of similar length (i.e., trip length variation under one mile).

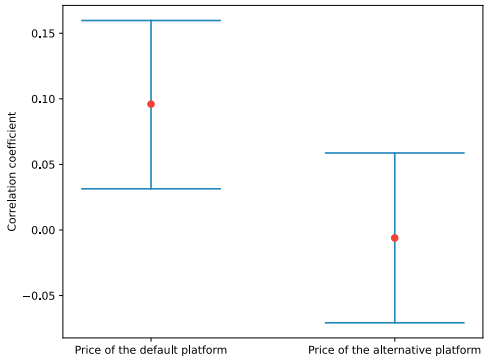

Figure EC.2 Correlation coefficient between the decision to switch and the price of the default platform and the price of the alternative platforms.

In the left part of Figure EC.2, we illustrate the correlation coefficient between the decision to switch and the price of the default platform which is equal to 0.095 (p-value = 0.003). It indicates that the price of the default platform significantly impacts the switching decision. On the other hand, we find a non-significant correlation of -0.006 (p-value = 0.854) between the decision to switch and the price of the alternative platform, as illustrated in the right part of Figure EC.2. These results clearly demonstrate that customers exhibit a higher sensitivity to price changes on their default platform than on the alternative platform, suggesting that they do not consider both platforms for every ride.

Appendix H: Model-Free Evidence for Learning and Multi-Homing Behavior over Time

Our structural model assumes that customers learn about both price and waiting time, thereby reducing the variance of their beliefs about platform attributes over time. This learning is fundamental to understanding why customers become less responsive to price discounts over time and why their ride allocation across platforms evolves as they gain more information.

In our model, as customers learn about platforms' prices and wait times, they become more likely to select and remain with a particular platform over time. Thus, in this section, we present empirical evidence supporting customer learning by examining whether customers consolidate their ride choices onto a single platform as they gain experience. Such consolidation would occur if their uncertainty about platform attributes decreases over time and they sequentially search for the platform to ride.

We focus on customers who have ride activity in both the first and last six months of our observation window. For each of these customers, we calculate the proportion of rides taken with Uber (the "Uber share") for each half-year period.

Based on this, we define the Multi-Homing Degree for each period as

$$\text{Multi-Homing Degree} = 1 - 2 \times |\text{Uber share} - 0.5|$$

This metric ranges from 0 (exclusive use of one platform, i.e., Uber share is 0 or 1) to 1 (equal use of both platforms, i.e., Uber share is 0.5). Comparing the average Multi-Homing Degree between the first and last six months allows us to assess whether increased experience is associated with a decrease in ride-splitting behavior (i.e., with consolidation on a single platform), as predicted by our model.

The observed decline in mean Multi-Homing Degree from 0.36 in the first six months to 0.23 in the last six months demonstrates that as customers gain experience, they are more likely to consolidate onto a single platform. This pattern directly supports our model's core assertion: learning reduces uncertainty and increases single-homing as customers sequentially search for platforms.

Appendix I: Robustness Test: Measurement Error in Competitor Price Predictions

As discussed in Appendix B, our dataset includes the prices of the rides that customers actually choose, but the corresponding prices offered by the competing platform for the same trips are unobserved and must be predicted for our analysis. This introduces measurement error in the competitor price, which can attenuate the estimated price coefficient and potentially affect other structural parameters.

In this section, we conduct a robustness test using a multiple-imputation for measurement error correction (Cole et al., 2006). We draw competitor prices from the predictive distribution obtained by our pricing model and re-estimate our choice model. We focus on the MNL specification with stickiness (Model 2 in Section 6) because running multiple imputations for the full structural model is computationally intensive. The results show that allowing for measurement error in predicted competitor prices slightly reduces the magnitude of the price coefficient while leaving all other coefficients essentially unchanged, indicating that our main conclusions are robust.

We estimate the predictive distribution of competitor prices by first fitting platform-specific price models for Uber and Lyft as described in Appendix B. These models predict the prices as $p_{j,it} = \hat{p}_{j,it} + \hat{\nu}_{j,it}$ where $\hat{p}_{j,it}$ is the predicted price and $\hat{\nu}_{j,it}$ is the residuals for each platform and observation.

To capture heteroskedasticity in prediction error, we fit a parametric variance function for each platform. Specifically, we model the log absolute value of the OLS residuals as a linear function of the log of the predicted price:

$$\log|\hat{\nu}_{j,it}| \approx a_0 + a_1 \log \hat{p}_{j,it} \tag{EC.15}$$

This technique exploits the fact that, under normality, the expected absolute residual is proportional to the standard deviation for a given mean, so the fitted regression defines a price-specific noise scale. Exponentiating both sides yields the fitted standard deviation for any predicted price:

$$\hat{\sigma}_j(\hat{p}_{j,it}) = \exp\left(a_0 + a_1 \log \hat{p}_{j,it}\right) \tag{EC.16}$$

Thus, the error variance increases as a power function of the predicted price, allowing us to reflect observed heteroskedasticity in the pricing data.

In the imputation step, we draw $M = 20$ alternative prices for the unobserved competitor platform of each observation from a normal predictive distribution centered at the fitted price with the estimated heteroskedastic variance,

$$p_{c,it}^{*(m)} \sim N(\hat{p}_{c,it}, \hat{\sigma}_j^2(\hat{p}_{c,it})), \quad m = 1, \ldots, M,$$

where $p_{c,it}^{*(m)}$ denotes the m-th draw of the unobserved competitor price. For each draw m, we re-estimate Model 2 (the multinomial logit with stickiness parameter in Section 6) using imputed competitor prices in place of the point-predicted competitor prices, while keeping the observed prices for the chosen platform unchanged. This yields M sets of parameter estimates $\{\hat{\theta}^{(m)}, \widehat{\mathrm{Var}}(\hat{\theta}^{(m)})\}_{m=1}^{M}$, which we combine using Rubin's rules for multiple imputation to obtain pooled coefficients and standard errors (Rubin, 2018).

	Baseline MNL	Multiple imputation
Price ($)	-0.016***	-0.015***
	(0.002)	(0.001)
Waiting Time (min)	-0.002**	-0.002**
	(0.003)	(0.003)
Stickiness	2.450***	2.450***
	(0.019)	(0.019)
UBA	0.328***	0.327***
	(0.051)	(0.051)
Controls	(Yes)	(Yes)
N	58,918	58,918

Table EC.3 Baseline vs. multiple-imputation estimates with heteroskedastic price predictions. Mean and standard error (in parentheses) of the parameter estimates are obtained through MLE. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

We report the parameter estimates obtained via standard MLE in Table EC.3, comparing the pooled multiple-imputation estimates with baseline estimates based on the competitor's point price predictions. The price coefficient changes from -0.016 in the baseline model to -0.015 under multiple imputation, a reduction in magnitude of less than 10%. All other coefficients, including the stickiness parameter, are virtually unchanged at three decimal places. This pattern is consistent with the modest attenuation expected when explicitly integrating over measurement error in a continuous regressor. Importantly, the main qualitative conclusions are unaffected: customers remain substantially less sensitive to price differences than to loyalty and other non-price factors, and the relative ranking and significance of all key coefficients are preserved.

Overall, the multiple-imputation analysis shows that even when we account for the full predictive uncertainty in competitor prices using a heteroskedastic measurement-error correction, our main results are robust. The estimated price sensitivity becomes slightly more conservative in magnitude, while the behavioral drivers of platform choice and the implied managerial implications remain unchanged.

Appendix J: Hierarchical Bayesian Framework: MCMC Estimation

To estimate the values of the parameters in our structural model, we exploit a hierarchical Bayesian approach, which is described in this subsection. More specifically, we estimate the set of population-level parameters $\{\mu_{\alpha_0}, \sigma_{\alpha_0}, \mu_C, \sigma_C, \boldsymbol{\alpha}, \boldsymbol{\beta}, \Lambda_{j,1}, \hat{\boldsymbol{\gamma}}_{j,1}, \lambda^2_{j,1}, \hat{w}_{j,1}\}$ and individual-level parameters $\{\alpha_{0,i}, C_i\}$ for each rider i. Then, our hierarchical Bayesian specification can be represented as follows:

$$y_{jit}|\boldsymbol{\theta} \sim \text{Bernoulli}(\Pr(\cdot\cdot j)_{it}), \qquad \text{for all } j, i, \text{and } t,$$

where $\boldsymbol{\theta}$ denotes the vector with all the parameters specified above and $\Pr(\cdot\cdot j)$ is the probability that rider i chooses platform j at time t. Note that $\Pr(\cdot\cdot j)_{it}$ depends on $\boldsymbol{\theta}$ as given by Equation (22). We then describe the hierarchical structure of the individual-level parameters as follows:

$$\begin{aligned}
&\alpha_{0,i} \sim \mathcal{N}(\mu_{\alpha_0}, \sigma^2_{\alpha_0}), && C_i \sim \mathcal{N}(\mu_C, \sigma^2_C), && \text{for } i = 1, \ldots, I,\\
&\mu_{\alpha_0} \sim \mathcal{N}(0,1), && \mu_C \sim \mathcal{N}(0,1), &&\\
&\sigma_{\alpha_0} \sim \ln\mathcal{N}(0,1), && \sigma_C \sim \ln\mathcal{N}(0,1), &&
\end{aligned}$$

where $\alpha_{0,i}$ and C_i for each customer i are drawn from normal distributions with population-level mean and variance parameters. The mean parameters have hyperpriors set as normal distributions with mean 0 and variance 1, while the variance parameters have hyperpriors set as log-normal distributions with parameters 0 and 1. The rest of the population-level mean and variance parameters are specified as follows:

$$\begin{aligned}
&\alpha_k \sim N(0,1), && \text{for } k = 1, 2, 3,\\
&\beta_k \sim N(0,1), && \text{for } k = 1, \ldots, |\boldsymbol{\beta}|,\\
&[\hat{\boldsymbol{\gamma}}_{j,1}]_k \sim N(0,1), && \text{for } k = 1, 2,\\
&\hat{w}_{j,1} \sim N(0,1), &&\\
&[\Lambda_{j,1}]_{kk} \sim \ln\mathcal{N}(0,1), && \text{for } k = 1, 2,\\
&\lambda^2_{j,1} \sim \ln\mathcal{N}(0,1). &&
\end{aligned}$$

The parameters of the model are estimated using a hierarchical Bayesian specification, implemented using the Stan package and the Hamiltonian Monte Carlo (HMC) algorithm. The HMC algorithm is a more efficient simulation method compared to random walk algorithms like Metropolis-Hastings, as it leverages the gradient information to better explore the parameter space. The HMC algorithm generates a posterior sample for each iteration, forming a joint posterior distribution of the parameters. For our analysis, we run five HMC driven Markov chains with 2,000 warmup iterations and 200 sampling iterations. We get 200 support points per chain or 1,000 support points that define the posterior distribution of the parameters.

Appendix K: Riders' Heterogeneity in Unobserved Brand Affinity

In this section, we examine the heterogeneity in customer preferences as captured by the intercept term in Model 4. Specifically, we analyze the distribution of estimated customer-specific intercepts, which reflect unobserved brand affinity for one platform over the other. As illustrated in Figure EC.3, the distribution is centered slightly above zero, indicating that, on average, customers exhibit a preference for Uber. That being said, the distribution shows considerable dispersion, revealing substantial latent preferences for both Uber and Lyft across different segments of the population.

To further explore the drivers of this heterogeneity, we compare summary statistics for the top 5% of users with the strongest platform preferences – those at the tails of the intercept distribution – with the full user sample. As shown in Table EC.4, these highly loyal users take substantially more rides: an average of 57.11 rides for Uber-preferring users and 39.2 for Lyft-preferring users, compared to 19.64 rides in the overall sample. Additionally, both groups exhibit strong platform loyalty, completing over 91% of their rides on their preferred platform.

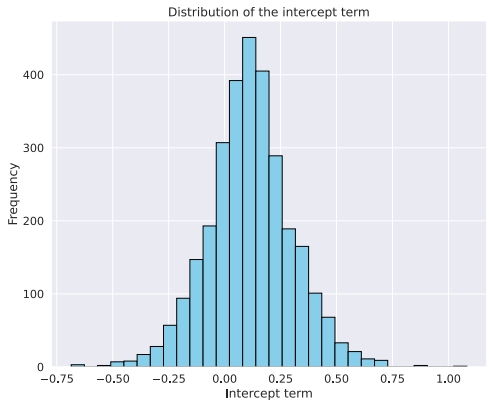

Figure EC.3 Distribution of estimated customer-specific intercepts from Model 4.

	Uber - preferring users	Lyft - preferring users	All users
Average number of rides	57.1	39.2	19.6
% rides on Uber	92.5	8.6	51.9
% rides on Lyft	7.5	91.4	48.1
% Weekend rides	32.2	35.3	38.0
% Rush hour rides	30.0	30.0	30.3
Average trip length (miles)	6.1	5.7	6.3

Table EC.4 Comparison of ride characteristics for the top 5% users with strongest Uber and Lyft preference, relative to the overall sample.

Appendix L: Model Performance over Time

In this section, we evaluate the performance of our model over the customer's lifetime to evaluate if our model performs a better/worse job for earlier/later rides. To assess the performance of our models in explaining the choice behavior of customers over time, we plot the average log-likelihood of the observed model over time in Figure EC.4, which is computed as follows:

$$\frac{1}{|I_t|}\sum_{i\in I_t}\sum_{j\in\{\mathcal{U},\mathcal{L}\}} y_{jit}\log(\Pr(j,m)_{it})$$

where I_t is the set of customers who took at least t rides. Clearly, the model's performance shows a discernible upward trend over time. One key distinction of our model compared to the benchmark is its ability to capture

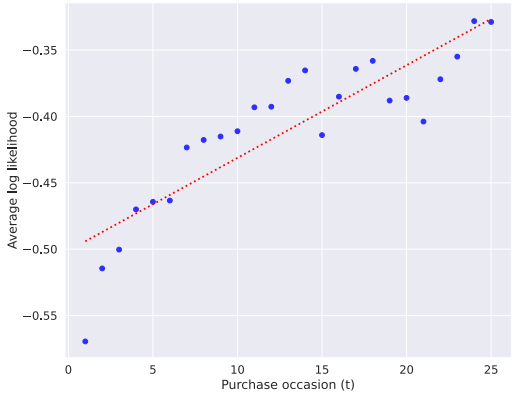

Figure EC.4 Model performance over time.

the dynamic interaction between customers and platforms. For instance, it can incorporate elements such as the updating of the customers' beliefs about price and waiting time, allowing for a more nuanced representation of the customer-platform relationship. Thus, our model gains advantages from sharper customer belief distributions over time, resulting in enhanced accuracy in its predictions. We also acknowledge that there may be an alternative explanation for the improvement in our model's prediction accuracy over time, as the variability in customer choices could decrease over their lifetime.

Appendix M: Robustness Test: Explaining Customer-level Switching Behavior

	Mixed MNL model	Our model
MAE	0.161	0.143
RMSE	0.261	0.226

Table EC.5 Model performance on switching behavior

In Section 6, we present evidence of our model's superiority over the mixed MNL model, as indicated by a substantial enhancement in log likelihood. In this section, we conduct robustness checks to evaluate the

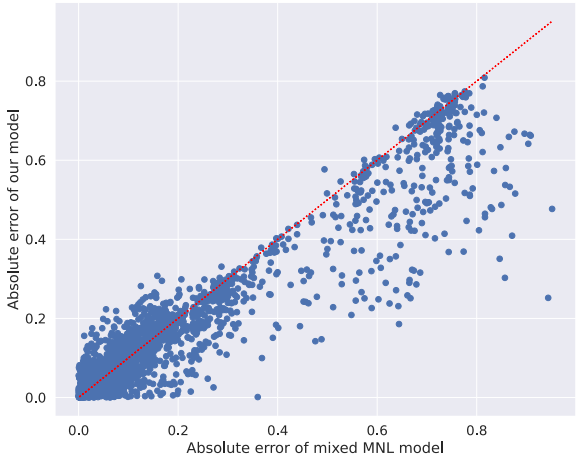

Figure EC.5 **Scatter plot depicting the absolute error of a mixed MNL model relative to our model for every customer.**

performance of our model using different metrics, with a particular focus on its ability to predict customer-level (as opposed to ride-level) switching behavior. Specifically, we evaluate our model on how well it predicts the propensity of a customer to switch to a competing platform ($\mathcal{P}_i$), which is computed as the fraction of times a customer switches between the platforms. We compare the model performance by utilizing two metrics: mean absolute error (MAE) and root mean square error (RMSE), which are defined in the following way.

$$MAE = \sum_{i=1}^{I} \frac{1}{I} |\hat{\mathcal{P}}_i^m - \mathcal{P}_i|, \qquad RMSE = \sqrt{\frac{1}{I} \sum_{i=1}^{I} \left(\hat{\mathcal{P}}_i^m - \mathcal{P}_i\right)^2},$$

where I is the number of customers in our analysis and $\mathcal{P}_i$ is the probability of customer i to switch based on the customers' ride history. This probability is computed by dividing the total number of switches by the total number of rides taken by the customer, i.e.,

$$\mathcal{P}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} \mathcal{S}_{it},$$

where T_i is the number of rides taken by customer i and $\mathcal{S}_{it}$ is an indicator variable that takes the value of 1 if customer i chooses a platform at time period t which is different from her/his choice at time $t-1$. We compute the estimated probability of customer i to switch based on a specific choice model in a similar fashion, i.e.,

$$\hat{\mathcal{P}}_i^m = \frac{1}{T_i} \sum_{t=1}^{T_i} \hat{\mathcal{P}}_{it}^m,$$

where $\hat{\mathcal{P}}_{it}^m$ is the estimated probability of customer i at time t to switch to a different platform from what was chosen at time $t-1$ according to the choice model m (i.e., we compare our model with the Mixed MNL model). We report these metrics in Table EC.5 and it can be seen from the table that our model outperforms the benchmark across both metrics (i.e., the lower the metrics the better).

Figure EC.5 provides additional evidence of the superior performance of our model versus the benchmark. Each dot in that figure corresponds to a particular customer and the x-axis (resp. y-axis) of every dot corresponds to the $|\hat{\mathcal{P}}_i^m - \mathcal{P}_i|$ (i.e., absolute error) of the Mixed MNL model (resp. our model). The fact that 57% of the dots in Figure EC.5 are under the 45-degree line indicates that our model explains the customers' probabilities to switch a platform better than the Mixed MNL model.

Appendix N: Additional Tables

	TLC	Our data
Uber market share (%)	68.59	65.74
Trip duration (minutes)	20.43	20.63
Trip duration - Uber (minutes)	20.24	20.04
Trip duration - Lyft (minutes)	21.14	21.75
% of weekend rides	32.13	34.52
% of rush hour rides	32.23	31.64
% of airport rides	9.02	5.36

Table EC.6 Sample representativeness: Comparison between our data sample collected by a third-party data provider and publicly available TLC data on all rides in New York City.

	Uber	Lyft	Overall
# Rides	33,663	25,255	58,918
# Rides with discounts	7,589	3,054	10,643
% Rides with discounts	22.5%	12.1%	18.1%
# Customers with discounts	1,047	1,190	1,749
% Customers with discounts	34.9%	39.7%	58.3%
Avg. discount value	$5.52	$4.52	$5.23

Table EC.7 Descriptive statistics of the current discounting strategy for 3000 customers used in model estimation.