Platform Leakage:
Incentive Conflicts in Two-Sided Markets

Yingkang Xie*

Northwestern University†

Huaiyu Zhu

Lalamove

April 18, 2022

Abstract

Leakage happens when buyers and sellers coordinate outside the platform to cut out the middleman, usually to avoid paying fees. Although platforms are concerned about losing revenue, leakage—by its very nature—is hard to measure and manage. Using geolocation data from a large on-demand service platform for cargo delivery, we identify offline transactions that are typically hard to track in online marketplaces. We exploit a quasi-experiment that gradually introduced driver commissions, thereby generating variation in participants' incentives for leakage. The introduction of this commission increased leakage by nearly four percentage points, doubling the percentage of offline transactions we detected. We leverage the variation in commission fees to estimate price sensitivities and transaction costs in a structural model. The likelihood of leakage increases as the quoted price of the delivery increases, as the drivers' potential savings in the commission exceed the costs of offline coordination. Our model estimates suggest that customers typically receive half of the commission savings from drivers to rationalize their agreement to leakage. Counterfactuals show that a stronger bargaining power of customers would exacerbate platform leakage. To conclude, we discuss how targeting coupons, monitoring technology, and matching policy can mitigate leakage by better aligning the incentives of different parties in two-sided markets.

Keywords: disintermediation, platform design, incentives, transaction costs, bargaining.

*Kellogg School of Management, Evanston, IL 60208. Email: yingkang.xie@kellogg.northwestern.edu.

†I want to thank my committee chair Brett Gordon for his exceptional mentorship and constant support. I also thank Eric Anderson, Hemant Bhargava, Kent Grayson, Gaston Illanes, Lakshman Krishnamurthi, Tai Lam, Angela Lee, Ilya Morozov, Daniel Spulber, Artem Timoshenko, Anna Tuchman, Caio Waisman, Jingyi Wang, Joonhyuk Yang, and the seminar participants at Kellogg for their insightful comments.

1 Introduction

Platform businesses (intermediaries) help buyers and sellers find each other and engage in convenient and trustworthy transactions (Einav et al., 2016). However, platforms face the challenge of disintermediation – buyers and sellers can transact directly to circumvent the intermediary's fees. For example, Uber and Lyft drivers may ask clients to cancel the trip on the app and pay them offline to avoid the commission (Bellotti et al., 2017). Transactions under the table are known as "leakage" (Hagiu and Wright, 2022; Ladd, 2022). A Harvard business case documents that approximately 90% of transactions started in a freelance marketplace are conducted offline (Zhu et al., 2018). Given the potential loss of revenue, intermediaries are highly motivated to monitor leakage and design incentives to reduce it. Many practitioners, marketing researchers, and economists[1] have discussed the implications of leakage on firms' strategies. However, by its very nature, leakage is hard to measure, which likely explains little empirical progress on this topic despite its importance.

This paper uses proprietary data from China's largest on-demand cargo delivery platform. The unique data allow us to identify disintermediated transactions, characterize the extent of leakage, and assess how leakage may vary in response to changes in platform fees. The data include a pricing experiment with a staggered rollout design that launched a driver-side commission with a 15% fee in different cities at different times. The average cancellation rates of the 33 treatment cities went up by about 5% (from 23.57% to 28.54%) after the drivers were charged a fee. The potential extent of leakage motivated the development of detection algorithm (Xie et al., 2022), which combines both geolocation and job cancellation data to flag[2] disintermediation at the transaction level. As far as we know, our work is one of the few, if not only, studies that uses a direct measure of disintermediation, rather than indirect measures such as the intentions to disintermediate or the reduction in platform engagement

[1]Theorists study disintermediation with analyses on leakage in online marketplaces (Chaves, 2018; He et al., 2020; Hagiu and Wright, 2022; Peitz and Sobolev, 2022) and showrooming in retail (Wu et al., 2004; Balakrishnan et al., 2014; Jing, 2018; Kuksov and Liao, 2018; Mehra et al., 2018; Wang and Wright, 2020)

[2]A transaction was disintermediated if (1) the assigned driver passed the origin and destination of the canceled trip around the time of service, and (2) the customer did not request and complete similar trips.

(Gu and Zhu, 2021; Zhou et al., 2022). We then estimate a structural model to quantify the underlying factors that motivate or discourage leakage, leveraging the quasi-experimental variation in driver-side fees and customer-side coupons.

We provide insights into proactive retention, product design, and matching algorithms in two-sided markets. They are ex-ante alternatives to punishments that are commonly used to mitigate leakage. For example, many platforms threaten to ban accounts that initiate off-platform transactions (Uber, 2022; Airbnb, 2022; eBay, 2022). However, it is unclear whether platforms have efficient ways to recover all offline transactions and verify who in the buyer-seller pair is at fault for proposing leakage, given that coordination typically happens under the table. Moreover, platforms that rely on punishments may not only antagonise users but also lose future revenue from banned accounts[3]. Lastly, recurring expenses are likely to incur for policy enforcement in the game of fraud detection[4] where people learn to hide their activities from the platform to avoid punishments.

A deeper understanding of the incentive mechanisms affecting leakage can help platforms identify the appropriate economic levers to prevent disintermediation before it happens. Drivers and customers are less likely to collude if their offline transaction costs outweigh the part of commission savings they each retain or receive after bargaining. Our model estimates suggest that the platform may want to allocate marketing efforts toward drivers who are sensitive to commission. In contrast, giving coupons to customers might not be an effective tool to reduce leakage. To encourage customers to stay, the platform may develop services[5] to provide standalone value or reduce the costs of on-platform transaction. Lastly, to prevent leakage, the platform can strategically match drivers and customers such that their joint costs of offline transactions are larger than the fees they pay to the platform.

Since customers were not charged any platform fees, drivers might offer them a discount

[3]In a random sample, 2/3 of drivers had at least one disintermediated transaction over 137 days. Some frequent cheaters were also big contributors to the platform economy. Hence, we may not want to ban them.

[4]With any technology becoming weaponized, there is an inevitable race between the countermeasures to that technology and the development of counters to the countermeasures.(Williamson and Scrofani, n.d.)

[5]For example, the platform can grant customers access to the remote camera in cargo space to monitor their goods in transit. This reduces the need for customers to send someone to accompany the goods.

for offline transactions. To better understand how drivers and customers (two parties) might share the commission fees (surplus) recouped from the platform (intermediary), we adopt a commonly used solution concept from the bargaining literature (Nash, 1950; Sieg, 2000; Zhang et al., 2021; Jiang, 2022) to microfound our model. We find that the two parties split the commission savings in half, on average, to rationalize the joint decision of leakage in our sample. Our counterfactuals show that the likelihood of leakage is higher when customers have stronger bargaining power. Since the bargaining power depends on whether the parities have outside options (Backus et al., 2020) and sufficient time (Rubinstein, 1982) for negotiation, a focus on fast and last-minute matching might mitigate leakage.

Our model estimates also tell us which and when platform services[6] can justify the fees of having the platform as the guardian of trust (Shapiro, 1987). We find that the cost is about ¥6 ($1) for an average driver-customer pair to give up the digital escrow payment service provided by the platform. Moreover, suggestive evidence shows that customers are more likely to give up the convenience of tracking drivers' location in the app if they decide to send someone[7] to accompany the goods in transit. In total, the average offline transaction cost for a typical transaction is between ¥20($3) and ¥25($4), and is higher than the average commission fee of ¥16.5($2.5) the platform receives. These estimates not only help us to understand the value of the platform, but also provide the basis for the platform to evaluate alternative pricing strategies (e.g., a higher commission rate) and potential investment opportunities in new products (e.g., cargo monitoring technology).

Our investigation contributes to the literature of two-sided markets. We not only provide direct evidence that leakage exists, but also show that leakage is subject to heterogeneity in price sensitivity and transaction costs across the two sides of the market. Researchers in industrial organization often assume these problems away when analyzing fees and subsidies (Rochet and Tirole, 2006; Weyl, 2010). Without taking into account leakage, platforms may fail to maximize profits at equilibrium, because optimal fees and subsidies depend not

[6]Platforms provide escrow, transaction monitoring, and dispute settlements (Edelman and Hu, 2016).
[7]Having one(two) passenger(s) is associated with lower offline frictions by roughly ¥1 (¥2).

only on the price elasticity but also on transaction costs (Spulber, 2019; Hagiu and Wright, 2022). We conduct one of the few, if not only, empirical studies that take into account both differential price elasticities and heterogeneous transaction costs to guide platform design. We quantify transaction costs (Coase, 1937; Williamson, 1987) that prevent individuals from coordinating without the platform due to the hassle, inconvenience, and additional efforts.

To the best of our knowledge, this is the first empirical work to study the effect of platform fees on leakage with discussion about its boundary conditions. The most relevant studies to our work are by Gu and Zhu (2021) and Zhou et al. (2022), which investigate how disintermediation increases with trust achieved by reputation systems or repeated interactions. They focus on continuous transactions and do not study how monetary incentives play a role in leakage. Our application in on-demand services demonstrates that disintermediation can happen even in one-off transactions, as long as the commission savings exceed the costs of offline transactions. Our research in the gig economy[8] can motivate new strategies of other businesses (e.g., retail, e-commerce, the sharing economy) that involve buyers and sellers who make decisions on a daily basis about whether or not to engage in direct sales.

2 Empirical Setting

The setting is a mobile app for on-demand cargo delivery services (like Uber for trucks and cargo vans). The app[9] focuses on intra-city delivery in China and serves 363 cities throughout the whole country. Figure 1 illustrates the layout of the app and a transaction from beginning to end. The customer picks the size of the vehicle, chooses the pick-up time, and sets the origin and destination for the delivery. The platform will assign an available driver to the job based on the driver's distance to the pick-up location. The driver is advised by the platform to call or text the customer on the app to communicate delivery details. The customer can track the driver's location in real time as long as the order is not cancelled.

[8]The gig economy connects customers with independent contractors who work on their own schedules.

[9]According to the IPO prospectus of a competitor, the market share of our focal platform is 54.7% in 2020, which is ten times the size of the second largest platform with a share of 5.5% in mainland China.

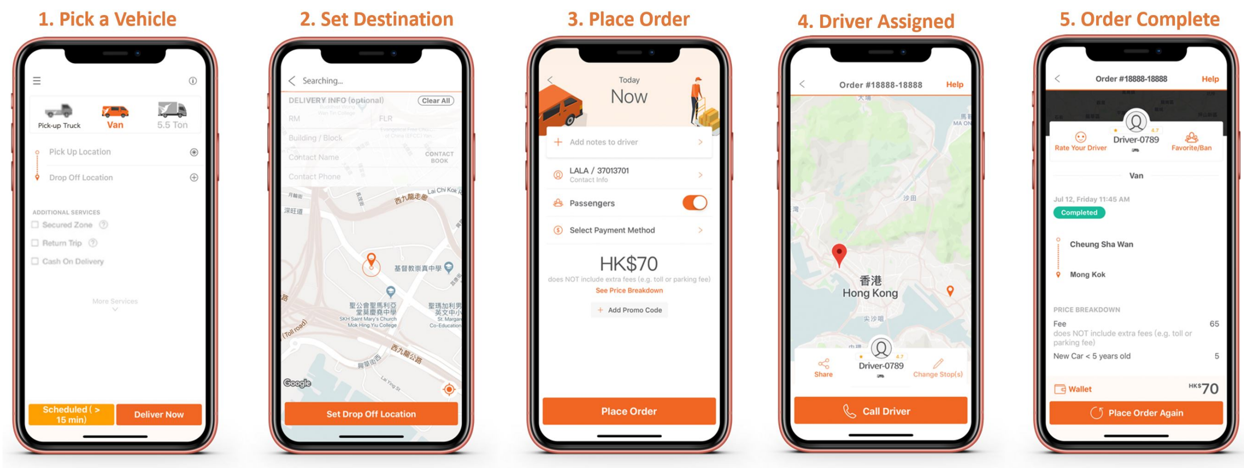


Figure 1: The Order Transaction Flow for Cargo Delivery in the App

This paper uses data in 2019 before the COVID-19 pandemic. The platform had 4 million registered drivers (300,000 monthly active drivers) and 28 million registered users (4 million monthly active customers) in China. In our data, the platform provided 400,000 to 500,000 matches on a daily basis in 2019, but more than 20% of them were canceled.

Pricing Scheme (The Two-Part Tariff)

Our on-demand logistics platform uses a subscription-based model. To improve the margin, the platform has implemented a commission fee. Cities that launched a commission have a pre-intervention period (no commission) and a post-intervention period (15% commission).

The platform is free for customers to use. However, drivers choose from a menu (see Table 1) to use the free service as Non-VIPs or pay for the premium service as Super-VIPs. Non-VIPs were free to use the platform, but they could only take up to two jobs per day. If drivers wanted more than two jobs in a day, they would need to pay a subscription fee to become Super-VIPs. In other words, drivers could pay a monthly membership fee upfront, which varies from ¥399 to ¥799 in different cities, to get unlimited job assignments.

In 2019, the platform gradually rolled out a 15% commission rate to Non-VIP drivers in a random 33 cities. Super-VIP drivers who paid the membership fee could still enjoy unlimited jobs per day without paying any commission. Drivers had equal opportunities to get jobs based on their distance to the pick-up location regardless of their membership tier.

Table 1: The Price Menu for Driver Participants

Tier	Max Daily Jobs	Monthly Membership	Commission Rate	
			Pre-Intervention	Post-Intervention
Non-VIP	2	¥0 (\$0)	0%	15%
Super-VIP	∞	¥399 ∼¥799 (\$60 ∼ \$120)	0%	0%

The average cancellation rates went up from 23.57% to 28.54%, or about 5%, for Non-VIPs who were charged commission after the intervention. Cancellation rates remain the same for Super-VIPs. The contrast provides preliminary evidence of leakage. However, the cancellation is not a direct measure of disintermediation, and the changes in cancellation rates cannot be fully attributed to leakage without strong assumptions on how the commission fee played a role in our application. In short, we do not know if a given transaction was disintermediated or not. To obtain a direct measure, we combine detailed transaction-level data with geolocation information to trace disintermediation at the individual level.

3 Leakage Detection and Description

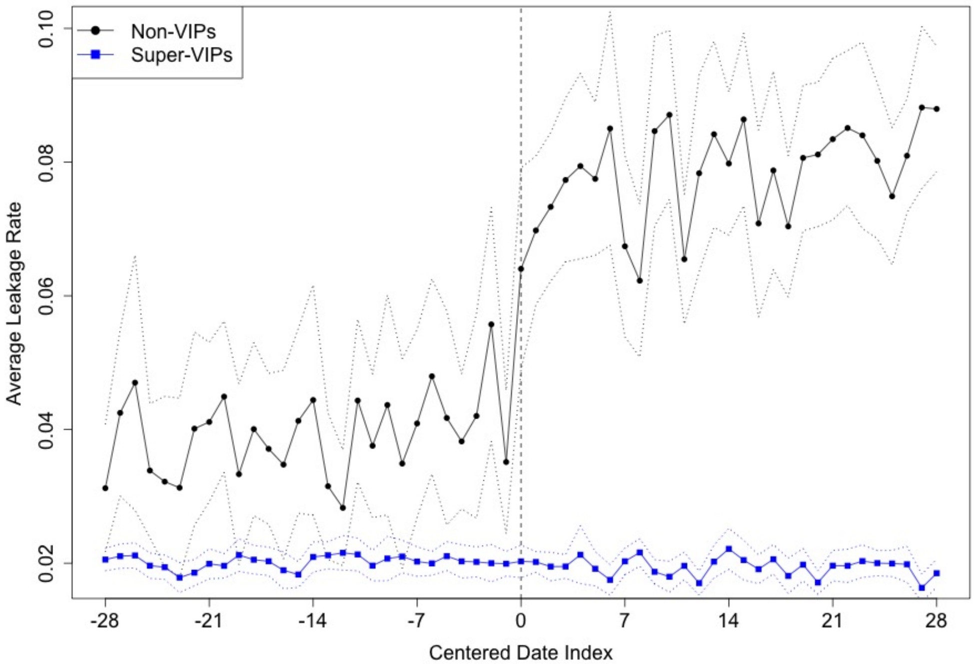
We use geolocation data to check whether a transaction was disintermediated by tracing the GPS footprints of an assigned driver, and use the transaction data to study how leakage responds to platform incentives such as driver-side fees and customer-side subsidies.

Our detection algorithm checks whether the driver has GPS track points within the radius of the origin and destination during the time window of a canceled trip, conditional on the customer not having other completed trips that share similar characteristics. The assumption is that, if a job was taken offline, the driver would still visit the origin and destination within the previously agreed upon time window for the requested trip.

The GPS detection algorithm can achieve both high recall and precision when evaluated against the human labels as ground truth. This study is one of the few, if not only, studies to use such a direct measure of leakage. We will use the GPS detected disintermediation as our dependent variable for the remaining part of the paper.

Figure 2 report the extent of disintermediation. The impact of commission on leakage is

Figure 2: The Detected Leakage Rate for 33 Treated Cities



Note: Leakage rates are averaged across the 33 cities by centering their time series at the launch date of commission. Only Non-VIPs are charged for the 15% commission fee. Super-VIPs have fee waivers. The grey and blue dotted lines are confidence intervals of the leakage rates.

evident without any modeling assumptions or control variables. For Non-VIPs, We see an increase in the disintermediation across the 33 cities from 3.92% to 7.87%. The increase is prominent: the leakage rate increased by 100.7%, or by 3.95 percentage points. In contrast, we do not see visible changes in the average leakage rate for Super-VIP drivers.

The simple pre-post comparison only makes sense when we assume a stable leakage rate without any trending or structural changes other than the commission launch. To provide better suggestive evidence, we implement the synthetic control method (SCM[10]) to evaluate the 33 treated cities. We find that the leakage rate increases by 3.2% and the cancellation rate increases by about 4.4% on average after charging the fee, using the counterfactual leakage or cancellation rates for each treated city that would have occurred had the city not charged a commission fee. The statistical distribution of the SCM estimates are reported in Figure 11 and Table 3 in Appendix A.2.

[10]Appendix A.1 uses Beijing as an example to illustrate how the SCM estimates treatment effects

Descriptive Statistics

We construct the leakage dataset based on a random sample of anonymous drivers with at least one job assignment within the 137 days, tracking all their assigned jobs, cancellations, and VIP status. We randomly draw 1971 drivers[11] from the 144 cities and find that they interacted with 239,057 customers, generating a total of 269,921 matches for our study.

We find that 2/3 of drivers were involved in at least one disintermediated transaction over the 137 days. Jobs were characterized by distance, vehicle size, the time of service, the number of passengers, and the payment method. In our sample, 4.2% of transactions were charged with a 15% commission on drivers. The average commission fee for these transactions was ¥16.5, approximately the minimum hourly wage (\$2∼\$4) in China.

Figure 3 shows that more leakage happened in the transactions with a 15% commission than in transactions without commission. Moreover, the natural variation in quoted price demonstrates that leakage was more likely to occur for high-value jobs than for low-value jobs when there was a commission fee. In other words, increasing the quoted price, even if the commission rate were unchanged, might increase leakage. The commission shock and natural variation in the quote price allow us to identify the effects of commission on leakage.

Figure 5 shows that less leakage happened for subsidized transactions. Coupons might discourage leakage. In our sample, 23.9% of transactions are subsidized on the customer side. Most coupons offered a ¥5 discount. The average subsidy for these transactions was ¥6.5, which was approximately \$1. These customer subsidies were valid disincentive shocks to leakage because coupons were not distributed to customers for leakage reduction. They were part of marketing experiments to vary quoted prices for customer acquisition.

[11]We sample individuals across geographical regions to obtain a sample with good representation of heterogeneity in driver types making decisions in different market conditions.

Figure 3: Leakage Rate by Price

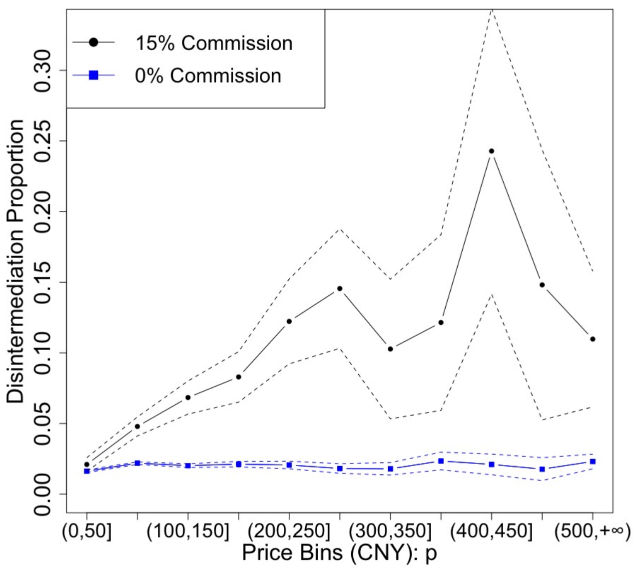


Figure 4: Histogram of Price

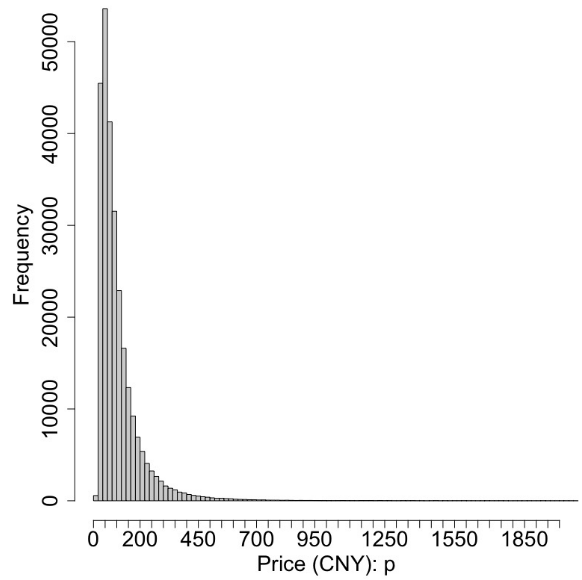


Figure 5: Leakage Rate by Coupon

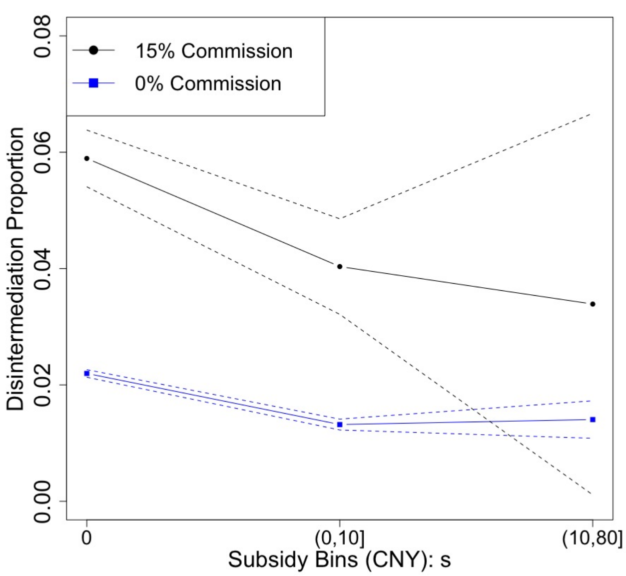
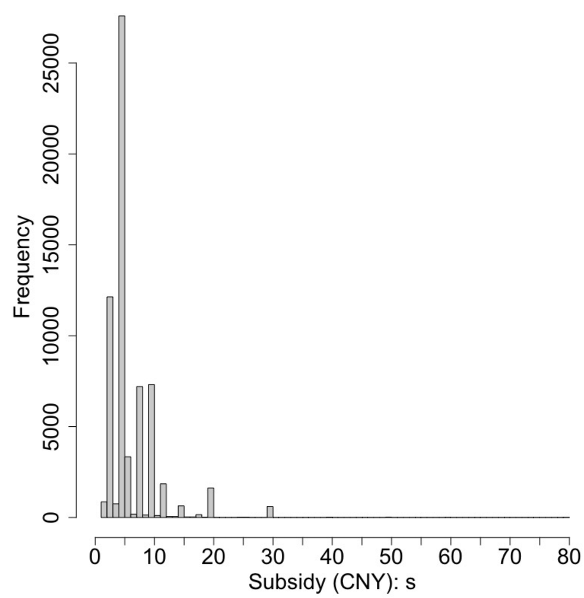


Figure 6: Histogram of Coupon

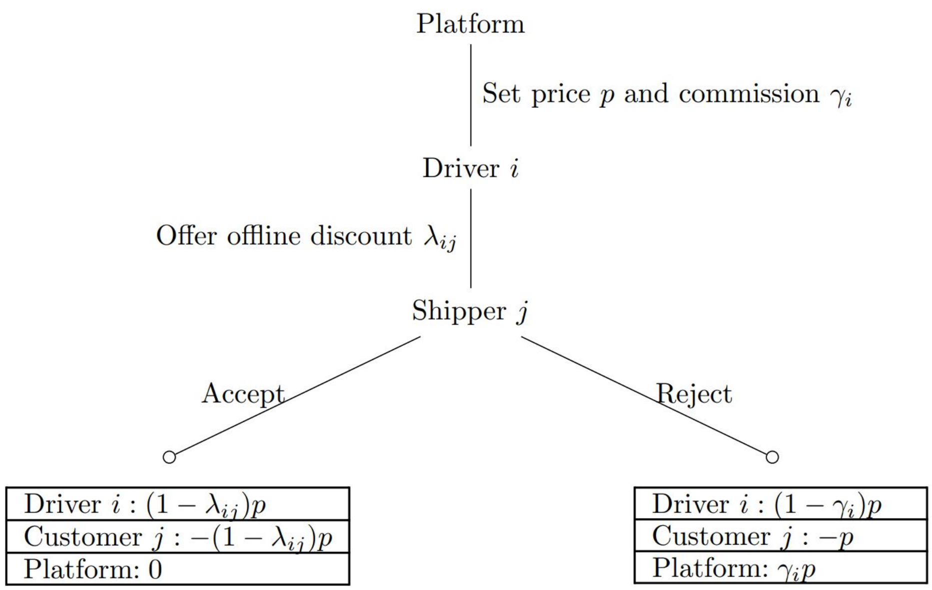


Note: The average leakage rates are computed on the bins of quoted price or subsidy, conditional on whether the transaction was charged a 15% commission. The dotted lines are confidence intervals.

4 Model

The platform matches i_{th} driver with j_{th} customer and determines a quoted price p. Figure 7 demonstrates the monetary transfers between different parties. For on-platform transactions, the driver pays commission fee $\gamma_i p$ to the platform. One can consider $\gamma_i p$ as the bid-ask spread set by the platform. The bid-ask spread depends on the price elasticities and transaction costs (Spulber, 2019). We will formally set up the utility functions that take into account the price elasticities and transaction costs in Section 4.1. For off-platform transactions, the driver can offer the customer a personalized discount λ_{ij} to settle on a price under the table.

Figure 7: The Monetary Transfers



Note: Driver i offers an offline discount to propose leakage. Customer j can accept the driver discount λ_{ij} and go off the platform, or reject the proposal and stay on the platform where the driver is responsible to pay the commission fee $\gamma_i p$ to the platform.

4.1 Utility Functions

To be consistent with the data, we index each match by t that helps us to locate the driver-customer pair using $ij(t)$. The quoted price p_t for the transaction t are non-negative. The platform set a driver-side commission rate $\gamma_{i(t)} \in \{0\%, 15\%\}$ for a subset of drivers.

We denote $\Pi_i^{L_t}(\gamma_{i(t)}, p_t)$ as the driver's utility and $U_j^{L_t}(p_t)$ as the customer's utility, which share the superscript $L_t \in \{1, 0\}$ as the leakage outcome of a joint decision. We make the parametric assumption that the utility functions have the following linear forms:

1. In an on-platform transaction, the payoff functions for $L_t = 0$ are:

$$\begin{aligned} \Pi_{i(t)}^0 &= \beta_i \cdot (1 - \gamma_{i(t)}) p_t \\ U_{j(t)}^0 &= u_{j(t)} - \beta_j \cdot (p_t - s_{j(t)}) \end{aligned} \tag{1}$$

   where β_i and β_j that reflect the heterogeneous marginal utility[12] for money (Dworczak et al., 2021). The platform offers coupon $s_{j(t)}$ to customers to make it cheaper to use the platform. Customer obtain the baseline utility $u_{j(t)}$ if the driver fulfill the job t.

2. In an off-platform transaction, the payoff functions for $L_t = 1$ are:

$$\begin{aligned} \Pi_{i(t)}^1 &= \beta_i \cdot (1 - \lambda_{ij(t)}) p_t - h_{i(t)} \\ U_{j(t)}^1 &= u_{j(t)} - \beta_j \cdot (1 - \lambda_{ij(t)}) p_t - h_{j(t)} \end{aligned} \tag{2}$$

   where $h_{i(t)}$ and $h_{j(t)}$ are the relative hassle for the driver[13] and customer[14] to transact outside the platform, respectively. Without the platform governance, driver i can adjust the quoted price p_t with customer j via a bargained term $\lambda_{ij(t)}$ (Nash, 1950; Rubinstein, 1982). When $\lambda_{ij(t)} \in [0, 1]$, the driver offers a discount[15] to the customer. In other words, there exists an offline trading price at $(1 - \lambda_{ij(t)}) p_t$.

[12]The utility changes given a unit change in the money the driver receives or the customer pays.

[13]Driver's hassle includes the expected efforts on persuasion, communication, and getting payments.

[14]Customer's hassle includes the inconvenience to track the delivery progress, the efforts to get a proof of payment, and the expected cost of dispute settlement if the goods are damaged or stole.

[15]The driver and the customer can bargain and reach an agreement on how to share the commission savings. For example, $\lambda_{ij(t)} = 0.5\gamma_{i(t)}$ indicates a split of the commission in half.

4.2 Discrete Choice Models

Driver i is more likely to prefer leakage if $\Pi^0_{i(t)} < \Pi^1_{i(t)}$. Customer j may want leakage if $U^0_{j(t)} < U^1_{j(t)}$. For driver i on transaction t, the latent utility gain from leakage is

$$\begin{aligned}\Delta\pi_{ij(t)} &= \Pi^1_{i(t)} - \Pi^0_{i(t)} \\ &= \left[\beta_i \cdot (1-\lambda_{ij(t)})p_t - h_{i(t)}\right] - \left[\beta_i \cdot (1-\gamma_{i(t)})p_t\right] \\ &= \beta_i \cdot \underbrace{(\gamma_{i(t)} - \lambda_{ij(t)})p_t}_{\text{Offline Savings}} - h_{i(t)}\end{aligned} \tag{3}$$

Similarly, customers j has the following latent utility gain from leakage:

$$\begin{aligned}\Delta u_{ji(t)} &= U^1_{j(t)} - U^0_{j(t)} \\ &= [u_{j(t)} - \beta_j \cdot (1-\lambda_{ij(t)})p_t - h_{j(t)}] - [u_{j(t)} - \beta_j \cdot (p_t - s_{j(t)})] \\ &= \beta_j \cdot (\underbrace{\lambda_{ij(t)}p_t}_{\text{Transfer}} - s_{j(t)}) - h_{j(t)}\end{aligned} \tag{4}$$

4.3 The Joint Decision of Disintermediation

Leakage happens if and only if both the driver and customer have non-negative indirect utility gain. For unobservable $\epsilon_{ij(t)}$ and $\epsilon_{ji(t)}$, leakage follows a probabilistic distribution:

$$Pr[L_t = 1] = Pr[\Delta\pi_{ij(t)} + \epsilon_{ij(t)} \geq 0, \Delta u_{ji(t)} + \epsilon_{ji(t)} \geq 0] \tag{5}$$

Bargaining (Nash, 1950; Rubinstein, 1982) happens when two agents can create a surplus together but requires a solution to split the surplus. Assume that the driver-customer pair ij can reach an agreement of $\lambda^*_{ij(t)}$[16] as long as the joint utility gains from leakage is non-negative as shown in the Equation (6). At the equilibrium, the $\lambda^*_{ij(t)}$ redistributes the joint surplus $(\Delta\pi_{ij(t)} + \Delta u_{ji(t)})$ between the the driver and the customer to ensure that both individual

[16] I use Nash Bargaining to microfound $\lambda_{ij(t)}$ in Appendix that describes the the agreement process. The solutions of $\lambda^*_{ij(t)}$ can exist for the binding Individual Rationality (IR) constraints.

utility gains ($\Delta\pi^*_{ij(t)}$ and $\Delta u^*_{ji(t)}$) are non-negative[17] before the idiosyncratic shocks kick in. We can thus rewrite Equation (5) into Equation (6) when such $\lambda^*_{ij(t)}$ exists.

$$\begin{aligned} Pr[L_t = 1] &= Pr[\Delta\pi^*_{ij(t)} + \epsilon_{ij(t)} + \Delta u^*_{ji(t)} + \epsilon_{ji(t)} \geq 0] \\ &= Pr[\beta_i \cdot (\gamma_{i(t)} - \lambda^*_{ij(t)})p_t - h_{i(t)} + \epsilon_{ij(t)} \\ &\qquad + \beta_j \cdot (\lambda^*_{ij(t)}p_t - s_{j(t)}) - h_{j(t)} + \epsilon_{ji(t)} \geq 0] \\ &= F_\epsilon \left[\beta_i \cdot \gamma_{i(t)}p_t - (\beta_i - \beta_j)\lambda^*_{ij(t)} \cdot p_t - \beta_j \cdot s_{j(t)} - h_{i(t)} - h_{j(t)}\right] \end{aligned} \tag{6}$$

The $\epsilon_{ij(t)}$ and $\epsilon_{ji(t)}$ are structural error terms that enter into the decisions of driver and customer, respectively. Structural errors (McFadden and Train, 2000) are typically introduced to account for idiosyncratic shocks (e.g., new information) and unobservables (e.g., traffic conditions). In our application, quoted prices are orthogonal to the error terms (e.g., the platform does not use dynamic pricing[18] based on real-time market or traffic conditions).

5 Identification, Estimation, and Results

A proper econometric setting requires that we carefully distinguish what the econometrician can observe from unobserved heterogeneity. The econometrician cannot observe all the determinants of the pre-transfer utilities of driver i and customer j (Galichon and Salanié, 2021). Specifically, we don't know the individual-level β_i, β_j, $h_i(t)$, and $h_j(t)$ in additional to the structural errors. However, we observe types $d \in D$ for drivers (service providers) and $s \in S$ for shippers[19](customers). Richer data converts unobserved heterogeneity into types.

In the following, we denote $\beta_i = \beta_d$ and $h_i(t) = h_d(t)$ if driver i is of type d, and $\beta_j = \beta_s$

[17]Drivers and customers can agree when the $\lambda^*_{ij(t)}$ makes one party indifferent between transacting offline or online. The bargaining power determines whether the driver or customer's IR constraint is binding.

[18]The on-demand cargo delivery service platform determine the pricing rule at the city level, which is a linear function of the trip distance conditional on the requested vehicle size

[19]Customers are called shippers into the industry because they are the individual or legal entity who enters in a contract of carriage with a driver and pays the driver for delivery.

and $h_j(t) = h_s(t)$ if customer j is of type s. Equation 6 becomes Equation 7:

$$Pr[L_t = 1] = F_\epsilon \underbrace{\left[\beta_d \cdot \gamma_{i(t)} p_t - (\beta_d - \beta_s)\lambda^*_{ds} \cdot p_t - \beta_s \cdot s_{j(t)} - (h_{d(t)} + h_{s(t)})\right]}_{x'_{ij(t)}\theta} \tag{7}$$
$$\text{where} \quad d \in D, s \in S$$

Variation in the driver-side commission $\gamma_{i(t)}p_t$ (Figure 3) identifies β_d, the drivers' marginal utility for money. Variation in the customer-side subsidy $s_{j(t)}$ (Figure 5) identifies β_s, customers' marginal utility for money. Since we observe natural variation in the quoted price p_t (Figure 4), we can back out the type-specific offline discount $\lambda_{ds(t)}$ from the price coefficient $-(\beta_d - \beta_s)\lambda^*_{ds(t)}$ given that β_d and β_s are identified.

The remaining part is the offline hassle $h_t = (h_{d(t)} + h_{s(t)})$. They are constant terms that can be decomposed by types of transactions (e.g., on-demand vs. scheduled delivery, escrow vs. cash payment, furniture vs. non-furniture, number of passengers). Another identification strategy of $h_{d(t)}$ is to utilize the within-individual variation of commission fee. By introducing driver-fixed effects, I can back out h_i given the intuition that driver i are more likely to be involved in disintermediated transaction when $\gamma_{i(t)}p_t > h_i$ but stay on the platform when $\gamma_{i(t)}p_t < h_i$ after controlling for other confounding variables. Observing multiple transactions within a driver enables the identification. In my sample, about 2/3 of drivers were involved in at least one offline transactions.

We use data $x_{ij(t)} = [\gamma_{i(t)}, s_{j(t)}, p_t, x_t, x_{ij}]$ to estimate $\theta = [\beta_d, \beta_s, \lambda^*_{ds(t)}, h_{d(t)} + h_{s(t)}]$. Following McFadden and Train (2000), we assume that $\epsilon_t = \epsilon_{ij(t)} + \epsilon_{ji(t)}$ are independently and identically distributed (IID) logit errors[20] . We can use maximum likelihood estimation (MLE) to estimate the standard discrete choice model depicted in the Equation (7) with the data $x_{ij(t)} = [\gamma_{i(t)}, s_{j(t)}, p_t, x_t, x_{ij}]$. We recover θ by maximizing the following log likelihood:

$$\mathcal{L}(\theta) = \sum_{t=1}^{N} L_t \cdot ln\left[\frac{exp(x'_{ij(t)}\theta)}{1 + exp(x'_{ij(t)}\theta)}\right] + (1 - L_t) \cdot ln\left[\frac{1}{1 + exp(x'_{ij(t)}\theta)}\right] \tag{8}$$

[20]Alternatively, we can use partial identification (Manski, 2003; Manski, 2007) to compute bounds that summarize what the data say about the parameters without Type-I error assumption. See Appendix **??**.

To obtain insights, we will start with estimating a model that assumes homogeneous price sensitivity of driver and customer and a universal $\bar{\lambda}$ across different transaction types. We will then introduce driver fixed-effects and customer fixed-effects to understand how the market responds to platform incentives for a subset of active drivers and customers. Lastly, we will introduce the heterogeneity of price sensitivity in different jobs.

5.1 Homogeneous Price Sensitivity

The simplest model to estimate is to assume that there is one type of driver (only one $d \in D$) and one type of customer (only one $s \in S$). In this model, we will only use $x_{ij(t)} = [\gamma_{i(t)}, s_{j(t)}, p_t, x_t]$ to estimate $\theta = [\beta_d, \beta_s, \bar{\lambda}, h_t]$ with a universal offline discount. The MLE of Equation (7) can thus be estimated using the entire sample of transactions with

$$x'_{ij(t)}\theta = -h - x'_t\beta_{h(T)} + \beta_d \cdot \gamma_{i(t)}p_t - \beta_s \cdot s_{j(t)} - (\beta_d - \beta_s)\bar{\lambda} \cdot p_t \tag{9}$$

where $\gamma_{i(t)}p_t$ is the driver-side commission fee, p_t is the quoted price for the job, and $s_{j(t)}$ is the customer-side subsidy, x_t are transaction-specific covariates.

Results for assuming IID transactions are reported in the Column (1) and Column (2) in Table 2. Using the model that controls for VIP status, we identify $\hat{\beta}_d = 0.195$ and $\hat{\beta}_s = 0.025$ from the variation in $\gamma_{i(t)}p_t$ and $s_{j(t)}$, respectively. Since $\hat{\beta}_d > 0$, leakage increases in driver commission in our sample. However, leakage is insensitive to customer coupons[21]. Although leakage decreases in the subsidy, customer coupons might not be an effective lever.

The above estimates help us to back out the surplus division rule in rational expectation. On average, drivers and customers agree on an offline discount that roughly split the commission savings in half. We obtain this insight by leveraging the differential price sensitivities of drivers and customers. Given that we identify $\hat{\beta}_d$ and $\hat{\beta}_s$, the coefficient for price, which is $-(\hat{\beta}_d - \hat{\beta}_s)\bar{\lambda} = -0.013$, tells us that $\hat{\bar{\lambda}} = 0.074$.

[21]The managerial implications of $\beta_d > \beta_s$: it might be more effective to reduce leakage by subsidizing drivers than customers because drivers were in general more price sensitive

Table 2: Homogeneous Price Sensitivity with a Uniform Offline Discount

	Dependent variable: Disintermediation			
	(1)	(2)	(3)	(4)
commission_fee ($\gamma_{i(t)}p_t$)	0.244***	0.195***	0.205***	0.175***
	(0.015)	(0.015)	(0.032)	(0.031)
subsidy_to_customer ($s_{j(t)}$)	-0.024	-0.025	-0.029	-0.027
	(0.049)	(0.049)	(0.051)	(0.051)
transaction_price (p_t)	-0.015***	-0.013***	-0.014***	-0.013***
	(0.002)	(0.002)	(0.003)	(0.002)
is_cash	1.01***	1.02***	1.02***	1.02***
	(0.029)	(0.029)	(0.048)	(0.048)
is_furniture	-3.08***	-3.08***	-3.14***	-3.14***
	(0.173)	(0.173)	(0.175)	(0.175)
is_scheduled	0.458***	0.435***	0.390***	0.389***
	(0.042)	(0.042)	(0.047)	(0.047)
is_intercity	-0.202***	-0.220***	-0.151***	-0.162***
	(0.045)	(0.045)	(0.057)	(0.055)
is_bus_ep	-0.148	-0.170	-0.168	-0.187
	(0.107)	(0.107)	(0.114)	(0.114)
passenger1	0.199***	0.200***	0.165***	0.166***
	(0.051)	(0.051)	(0.054)	(0.054)
passenger2	0.327***	0.334***	0.305***	0.308***
	(0.062)	(0.062)	(0.068)	(0.068)
vehicleTruck_M	0.497***	0.417***	0.372**	0.336*
	(0.047)	(0.047)	(0.186)	(0.184)
vehicleTruck_S	0.261***	0.219***	0.102	0.087
	(0.035)	(0.035)	(0.108)	(0.107)
vehicleVan_M	0.117***	0.113***	0.036	0.030
	(0.042)	(0.042)	(0.070)	(0.070)
vipDriver		-0.473***		-0.546***
		(0.036)		(0.083)
(Intercept)	-4.08***	-3.68***		
	(0.027)	(0.040)		
Fixed-effects: driver_id			✓	✓
Average Discount ($\bar{\lambda}$)	0.068	0.074	0.079	0.085
Observations	269,921	269,911	248,556	248,556
- Unique Drivers	1971	1962	1280	1280
- Unique Users	239,057	239,048	220,920	220,920
Pseudo R^2	0.05059	0.05341	0.10887	0.11021
BIC	53,092.2	52,946.8	64,863.2	64,802.0

Leakage would occur when the total commission savings exceed the offline transaction costs. The estimates, $-\hat{h}/(\hat{\beta}_d - \hat{\beta}_s)$, show that the offline costs in transactions with Non-VIPs[22] are about ¥21.62 ($3) for the on-demand and intra-city delivery of goods requested by non-enterprise customers that use digital escrow payment service. The estimated transaction costs are higher than the average commission fee, ¥16.5, the platform receives in our sample.

By exploiting the within-driver variation for a subset of active drivers, we also recover the heterogeneous transactions costs across drivers in the Column (3) and Column (4) in Table 2. These active drivers seem to be more aggressive in offering an average discount of 8.5%, which is backed out by the model that identifies $\hat{\beta}_d = 0.175$ and $\hat{\beta}_s = 0.027$. Again, we observe that leakage increases in driver commission fees. The average relative hassle for offline transactions is -3.519, which translates to ¥23.77 ($3.5) for Non-VIP drivers. The Super-VIP drivers have an additional ¥3.68 ($0.5) cost to transact offline.

Relative transaction costs can inform us about the value of platform services. For example, transactions are more likely to be disintermediated when customers choose to pay cash instead of using the digital escrow payment service. Only 21.8% of transactions specified cash as the payment method. The relative cost of offline transaction is lower by ¥6 ($1) for these cash-paying jobs. We can interpret this dollar value as how much the driver-customer pairs were implicitly paying for using the platform payment system.

Model estimates suggest that less moral hazard within the driver-customer pair may encourage leakage. Estimates show that the cost of coordination outside the platform is lower when customers have at least one passenger on board with the delivery. The likelihood of disintermediation is higher if customers send two passengers instead of one passenger. According to the transportation law in China, customers can send up to two custodians to accompany their goods and supervise the driver, but they cannot use vans and trucks for rideshare service without loading the cargo.

We find that on-demand delivery (ship now) is less likely to be disintermediated while non-urgent (scheduled) delivery is positively associated with leakage. It is interesting to

[22] The average offline transaction cost of Super-VIPs is ¥2.78 higher than that of Non-VIPs.

see that intercity delivery has lower leakage than intracity delivery. A larger load size ($Truck_M > Truck_S > Van_M > Van_S$) is associated with a higher leakage rate. Lastly, delivery requests from large enterprise customers (about 3%) or furniture moving (about 10% of total transactions) are associated with a lower probability of disintermediation.

5.2 Heterogeneous Price Sensitivity

Previously, we assumed homogeneous price sensitivity as we considered only single driver and customer type. In this section, we relax the assumption to use observed characteristics of drivers and customers to define types $d \in D$ and $s \in S$ with variables in $x_{ij(t)}$. We interact the transaction covariates with $\gamma_{i(t)} p_t$, $s_{j(t)}$, and p_t to allow heterogeneity in β_d and β_s.

- Assume heterogeneous price sensitivity for drivers $\beta_d = \bar{\beta}_d + x'_{ds}\beta_{d_{(ds)}} + x'_t \beta_{d_{(T)}}$
- Assume heterogeneous price sensitivity for customers $\beta_d = \bar{\beta}_d + x'_{ds}\beta_{d_{(ds)}} + x'_t \beta_{d_{(T)}}$
- Assume linearly separable hassle $h_t = h_i + x'_{ds}\beta_{h_{(ds)}} + x'_t \beta_{h_{(T)}}$
- Assume type specific discount $\lambda^*_{ds(t)} = \lambda + x'_{ds}\beta_{\lambda_{(ds)}} + x'_t \beta_{\lambda_{(T)}}$ that drivers and customers agree on in the market to split the surplus from offline transaction

We can estimate $\theta = [\bar{\beta}_d, \bar{\beta}_s, \beta_{d_{(ds)}}, \beta_{d_{(T)}}, \bar{\beta}_s, \beta_{s_{(ds)}}, \beta_{s_{(T)}}, h_i, \beta_{h_{(ds)}}, \beta_{h_{(T)}}, \lambda, \beta_{\lambda_{(ds)}}, \beta_{\lambda_{(T)}}]$ using the data $x_{ij(t)} = [\gamma_{i(t)} p_t, s_{j(t)}, p_t, x_{ij}, x_t]$ with interactions between the variables.

The interaction terms provide fruitful insights into the leakage responses. For example, interactions between commission fees and transaction-specific characteristics can capture the heterogeneity of drivers' price sensitivities in different contexts. Drivers are less likely to disintermediate a furniture delivery or intercity delivery as the amount of commission fees goes up. This negative relationship contradicts regular cases where leakage is more likely to happen when the savings in commission fees are higher. Such contradiction indicates that drivers are willing to pay fees to the platform when they handle specific types of jobs, such as moving furniture or driving to another city, perhaps due to their wants for dispute settlement or protection provided by the platform.

Higher commission fees are not always positively associated with a higher likelihood of leakage, specifically, when drivers are involved in cash-paying jobs (negative coefficients for $commission\_fee \times is\_cash$). The risk of losing the payment (negative coefficient for $transaction\_price \times is\_cash$) may outweigh the potential savings in commission fees (positvie coefficient for $commission\_fee$). Drivers may be concerned about being defaulted on high-value jobs when customers want to use cash instead of the escrow payment service on the platform. Adverse selection could explain why drivers are less price sensitive to platform services when customers specify cash payment.

6 Implications for Platform Design

Platforms want to neutralize disintermediation to capture the value they have created and retain complete transaction data. To forestall leakage, we focus on ex-ante approaches that better align the incentives between the platform and the driver-customer pairs. We do not consider ex-post punishments in this research, because of the difficulty in verifying which party is at fault in two-sided markets as well as dealing with rebuttals. Preventive measures may provide a better long-term outcome for platforms without losing future revenue from banning accounts or triggering antagonism that reduces platform engagement.

6.1 Marketing Interventions

One way to manage leakage is to provide coupons to get the commission fee just below the offline transaction costs between drivers and customers. The platform can target individuals that are sensitive to fee reduction and personalize the amount of compensation. The optimal targeting rule might differ from targeting on predictive churn. Platforms typically use machine learning to target individuals with the highest probability of leaving, but such retention efforts might be futile when they fail to consider individuals' sensitivity to the intervention (Ascarza, 2018). We might not be able to persuade drivers who will disintermediate anyway. Instead, we can compensate drivers who are more sensitive to commission fees and have a moderate offline transaction costs. This alternative retention strategy might result in a more significant reduction in leakage given the same marketing budget.

In future research, we will conduct a counterfactual analysis to test whether targeting drivers with moderate offline transaction costs and high commission sensitivity is more cost-effective than targeting drivers with the highest risk of leakage. The key objective is to use less money to retain more transactions by converting drivers at the boundary of leakage.

6.2 Monitoring Technology

Platforms can use monitoring technology to increase offline transaction costs or decrease online transaction costs. One straightforward way is to actively listen to conversations between buyers and sellers to block contact exchanges[23]. However, monitoring conversations faces many limitations in on-demand services. It not only triggers privacy concerns[24], but it is also evadable when buyers and sellers can have their conversations outside the platform (e.g., meet in person) or shut down the monitoring devices (e.g., turn off the phone).

A potentially better alternative is to invest in platform technology that can reduce the transaction costs for on-platform transactions. For example, the platform can compensate drivers for installing monitoring devices (e.g., video cameras) in the cargo space (e.g., back truck, trailer, or cargo bed). The customer can request access to the live remote video camera to monitor their goods in transit. This service creates a new incentive for customers to stay on the platform. It may also reduce the need for customers to send someone to accompany the goods, which facilitates leakage[25]. Since 2021, the on-demand cargo delivery service app has been actively promoting Internet of Things (IoT) devices on vehicles to help protect customers' safety and assets. The monitoring devices can also provide evidence to resolve disputes regarding damaged goods. However, no efforts have been made to provide customers with video streams or photo snapshots. The platform can take advantage of the

[23] Airbnb detects and blocks contact exchanges by replacing emails and phone numbers with "(Hidden by Airbnb)" to stop people from dealing directly with the guest or host.

[24] Uber's China counterpart, Didi Chuxing, launched a mandatory audio recording as a safety feature. However, passengers are not buying the feature that trades privacy for safety (Shen, 2019).

[25] Table **??** shows that more passengers are associated with the lower hassle and higher leakage.

existing technology to reduce the monitoring cost for customers.

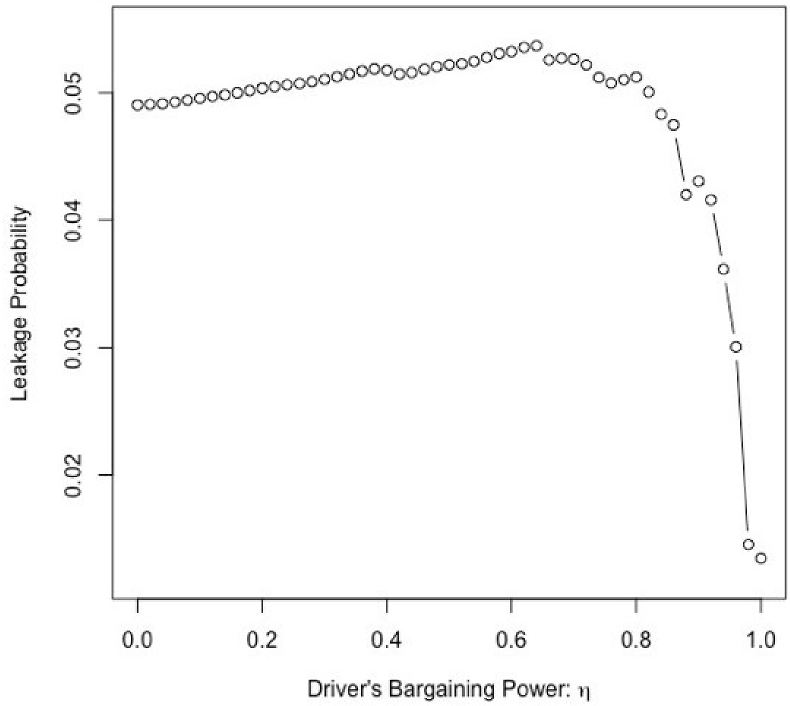
6.3 Matching Policy

The platform can strategically match a driver and a customer who have large enough offline transaction costs together as a pair. The variation in commission fee can recover drivers' hassle, and the transaction characteristics of job requests can index the type-specific customers' hassle (see Section **??** for estimates). Given the historical information[26], platforms can set a restriction to make sure that the pair-specific transaction costs for offline transactions are lower than the commission fee they charge for on-platform transactions.

A focus on fast and last-minute matching might also help the platform mitigate leakage. Our counterfactual analysis in Appendix B shows that the likelihood of leakage is higher when drivers have just slightly stronger bargaining power than customers (see Figure 8). The platform would be better off, in terms of less leakage, having full bargaining power on the driver's side. The platform might suffer from more leakage in a market if customers have a stronger bargaining power. The intuition is that customers with a strong bargaining power might actively ask for an offline discount which drivers cannot decline. For example, customers can pay less offline, rather than the full quoted price, when they have patience (e.g., sufficient time for negotiation) and have outside options (e.g., other competitive drivers). Patience and outside options can empower customers in the bargaining process (Rubinstein, 1982; Backus et al., 2020). Additional descriptive evidence supports the implications of this counterfactual analysis. For example, scheduled jobs have higher leakage, perhaps because the customer has more time to find alternative drivers or negotiate a better deal. However, leakage is lower if the job is scheduled in the early morning for the next day, perhaps due to the lack of supply. Appendix B.3 discusses how market conditions might affect the bargaining power. With the insights from our counterfactual analysis, the platform can experiment with assigning or disclosing drivers to customers at the last minute to mitigate leakage.

[26]For new drivers, the platform can use machine learning to predict hassle based on their characteristics at registration. The screening process of hassle can also inform better driver acquisition or retention.

Figure 8: Counterfactual Leakage Rates from Different Bargaining Power



Note: The likelihood of leakage is highest at $\eta = 0.65$ when drivers have slightly stronger bargaining power than customers (see Appendix B for the micro-founded model that uses Nash Bargaining as the solution concept).

7 Discussion

Cancellation rates went up after Lalamove charged a 15% commission rate. Intrigued by this finding, we leverage geolocation data to identify offline transactions that are typically hard to track in online marketplaces. We find that, on average, the likelihood of leakage increases as the driver commission fees go up, but it is insensitive to the customer coupons in our sample. The result is not uniform and depends on the types of transactions.

To provide insights into preventive measures, we study how leakage responds to platform incentives. We estimate the price sensitivity of leakage and transaction costs in a structural model by exploiting the quasi-experimental variation in incentives (e.g., driver fees and customer subsidies) for leakage. One source of variation comes from the changes in commission fees, which are generated by the staggered rollout of a 15% driver commission across cities and the variation in quoted prices across transactions. Another source of variation comes from the user coupons, which generates external disincentives for leakage.

Our novel data and model estimates provide unique insights into preventive measures. The platform can prioritize on target drivers with high fee sensitivity and compensate them based on transaction costs. Charging them a fee right below the costs for off-platform transaction can be more cost-effective than targeting drivers with the highest churning risk who are not responsive to retention offers (Ascarza, 2018). The platform can also leverage the recent rollout of monitoring technology (IoT devices on vehicles) to mitigate leakage. Granting customers access to the live remote video camera in the cargo space to see their goods can create standalone value for using the platform. This service may also reduce the need for customers to send someone to accompany the goods in transit which facilitate leakage. Lastly, the platform can experiment with assigning drivers to customers at the last minute and continue advertising on fast matching. A focus on instant delivery can prevent drivers and customers from having sufficient time to negotiate, and reduce the chance for them to find competitive outside options as leverage to bargain for a favorable offline price that is different from the price quoted by the platform.

Given the nature of the economic problem, our empirical framework can inspire analyses of other intermediaries in marketing and financial applications, including, but not limited to, retailers in between manufacturers and consumers, housing agents who mediate homeowners and homebuyers, brokers for private equity and investors, and travel agencies searching airlines and hotels for travelers. Similar to our case, these other applications involve buyers and sellers who make decisions on a daily basis about whether or not to engage in direct sales. We hope this study motivates new platform designs and pricing strategies for intermediaries to mitigate leakage when enforcing minimum advertised pricing (MAP) is impossible.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010). "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program". *Journal of the American Statistical Association* 105.490, 493–505.

Airbnb (2022). *Airbnb's Off-Platform Policy.* URL: `https://www.airbnb.com/help/article/2799/airbnbs-offplatform-policy`.

Ascarza, Eva (2018). "Retention futility: Targeting high-risk customers might be ineffective". *Journal of Marketing Research* 55.1, 80–98.

Athey, Susan and Guido W Imbens (2017). "The State of Applied Econometrics: Causality and Policy Evaluation". *Journal of Economic Perspectives* 31.2, 3–32.

Backus, Matthew et al. (2020). "Sequential bargaining in the field: Evidence from millions of online bargaining interactions". *The Quarterly Journal of Economics* 135.3, 1319–1361.

Balakrishnan, Anantaram, Shankar Sundaresan, and Bo Zhang (2014). "Browse-and-switch: Retail-online competition under value uncertainty". *Production and Operations Management* 23.7, 1129–1145.

Bellotti, Victoria et al. (2017). "Why Users Disintermediate Peer-to-Peer Marketplaces". *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* CHI '17. Denver, Colorado, USA: Association for Computing Machinery, 4370–4382.

Binmore, Ken, Ariel Rubinstein, and Asher Wolinsky (1986). "The Nash bargaining solution in economic modelling". *The RAND Journal of Economics*, 176–188.

Chaves, Isaıas N (2018). "Pricing Responses to Platform Leakage: Optimal Platform Design When Matches Are Irrevocable". Unpublished manuscript.

Coase, Ronald Harry (1937). "The nature of the firm". *Economica* 4.16, 386–405.

Dworczak, Piotr, Scott Duke Kominers, and Mohammad Akbarpour (2021). "Redistribution through markets". *Econometrica* 89.4, 1665–1698.

eBay (2022). *Buy or sell outside of eBay policy.* URL: `https://www.ebay.com/help/policies/payment-policies/offering-buy-sell-outside-ebay-policy?id=4272`.

Edelman, Benjamin and Philip Hu (2016). "Disintermediation in two-sided marketplaces". *Harvard Business School Technical Note* 917.4.

Einav, Liran, Chiara Farronato, and Jonathan Levin (2016). "Peer-to-Peer Markets". *Annual Review of Economics* 8.1, 615–635.

Fong, Jessica (2020). "Search, selectivity, and market thickness in two-sided markets: Evidence from online dating". *SSRN 3458373*.

Galichon, Alfred and Bernard Salanié (2021). "Structural estimation of matching markets with transferable utility". *arXiv preprint arXiv:2109.07932*.

Gu, Grace and Feng Zhu (2021). "Trust and disintermediation: Evidence from an online freelance marketplace". *Management Science* 67.2, 794–807.

Hagiu, Andrei and Julian Wright (2022). "Marketplace Leakage". Unpublished manuscript.

He, Eryn Juan et al. (2020). "Off-platform Threats in On-Demand Services". *SSRN 3550646*.

Jiang, Zhenling (2022). "An empirical bargaining model with left-digit bias: A study on auto loan monthly payments". *Management Science* 68.1, 442–465.

Jing, Bing (2018). "Showrooming and Webrooming: Information Externalities between Online and Offline Sellers". *Marketing Science* 37.3, 469–483.

Kuksov, Dmitri and Chenxi Liao (2018). "When showrooming increases retailer profit". *Journal of Marketing Research* 55.4, 459–473.

Ladd, Ted (2022). "The Achilles' heel of the platform business model: Disintermediation". *Business Horizons* 65.3, 277–289.

Manski, Charles F (2003). *Partial identification of probability distributions*. Vol. 5. Springer.

— (2007). "Partial identification of counterfactual choice probabilities". *International Economic Review* 48.4, 1393–1410.

McFadden, Daniel and Kenneth Train (2000). "Mixed MNL models for discrete response". *Journal of applied Econometrics* 15.5, 447–470.

Mehra, Amit, Subodha Kumar, and Jagmohan S Raju (2018). "Competitive strategies for brick-and-mortar stores to counter showrooming". *Management Science* 64.7, 3076–3090.

Nash, John F (1950). "The bargaining problem". *Econometrica: Journal of the econometric society*, 155–162.

Peitz, Martin and Anton Sobolev (2022). "Inflated recommendations". CEPR Discussion Paper No. DP17260.

Rochet, Jean-Charles and Jean Tirole (2006). "Two-sided markets: a progress report". *The RAND journal of economics* 37.3, 645–667.

Rubinstein, Ariel (1982). "Perfect equilibrium in a bargaining model". *Econometrica: Journal of the Econometric Society*, 97–109.

Shapiro, Susan P (1987). "The social control of impersonal trust". *American journal of Sociology* 93.3, 623–658.

Shen, Jill (2019). *Didi faces new privacy pressures with mandatory audio recording*. URL: `https://technode.com/2019/12/24/didi-carpool-audio-recording`.

Sieg, Holger (2000). "Estimating a bargaining model with asymmetric information: Evidence from medical malpractice disputes". *Journal of Political Economy* 108.5, 1006–1021.

Spulber, Daniel F (2019). "The economics of markets and platforms". *Journal of Economics & Management Strategy* 28.1, 159–172.

Uber (2022). *Street Hails and Off-platform Pickups*. URL: `https://www.uber.com/legal/en/document/?name=general-community-guidelines&country=united-states`.

Wang, Chengsi and Julian Wright (2020). "Search platforms: showrooming and price parity clauses". *RAND Journal of Economics* 51.1, 32–58.

Weyl, E. Glen (2010). "A Price Theory of Multi-sided Platforms". *American Economic Review* 100.4, 1642–72.

Williamson, Oliver E (1987). "Transaction cost economics: The comparative contracting perspective". *Journal of economic behavior & organization* 8.4, 617–625.

Williamson, William and James Scrofani (n.d.). "Trends in detection and characterization of propaganda bots". *Proceedings of the 52nd International Conference on System Sciences*.

Wu, Dazhong et al. (2004). "Implications of reduced search cost and free riding in e-commerce". *Marketing Science* 23.2, 255–262.

Xie, Yingkang, Jingyi Wang, and Huaiyu Zhu (2022). "Monitoring Disintermediation: Actions Matter More Than Words". Working Paper, Kellogg School of Management.

Zhang, Xu, Junhong Chu, and Puneet Manchanda (2021). "Online Healthcare Platform Evolution: The Interplay of Bargaining and Network Effects". *SSRN 3965007*.

Zhou, Qiang (Kris) et al. (2022). "Platform Exploitation: When Service Agents Defect with Customers from Online Service Platforms". *Journal of Marketing* 86.2, 105–125.

Zhu, Feng, Weiru Chen, and Shirley Sun (2018). "ZBJ: Building a Global Outsourcing Platform for Knowledge Workers". *Harvard Business School Case* 618-044.

Appendix

A Synthetic Control Methods (SCM)

A.1 The SCM Estimates of Beijing

We estimate the city-specific treatment effects on Non-VIPs using synthetic controls for 33 cities in Section A.2. The synthetic control method (Abadie et al., 2010) is "arguably the most important innovation in the evaluation literature in the last fifteen years" (Athey and Imbens, 2017). The data-driven approach constructs a measure of the counterfactual leakage rate for each treated city that would have occurred had the city not charged a commission fee. In this section, I will use Beijing as an example to illustrate how I implement the synthetic controls estimation.

First, I define the set of candidate controls as the set of all cities without any treatment changes within the 28 days before and 28 days after the launch date of Beijing. Then, using the data in the pre-period, I estimate the parameters of synthetic controls using a constrained linear regression. Let y_{ct} represent the daily leakage rate of the treated city c (e.g., Beijing) on day t, and let T_0 denote the set of days in the pre-period. Let $K \in K$ index the set of candidate control cities. Finally, I estimate a vector of weights $w = \{w_k\}_{k \in K}$ by minimizing the following objective function:

$$\begin{aligned} & \underset{w}{min} \sum_{t \in T_0} \left(y_{ct} - \sum_{k \in K} w_k y_{kt} \right)^2 \\ & s.t. \sum_{k \in K} w_k = 1, w_k \geq 0 \ \forall k \in K \end{aligned} \tag{10}$$

Weights $\hat{w}$ are chosen so that the synthetic control group's pre-period leakage rate closely matches the average of the treated groups. The intuition is that some cities (e.g., neighboring cities) are more similar to the treated city than others in the entire country, and those cities should contribute more to the estimate of the counterfactual leakage rate in the treated city.

Using the weights in $\hat{w}$, I can construct the measure of the fitted leakage rate in the pre-period T_0 and the measure of the counterfactual leakage rage had Beijing not charged a commission fee in post-period T_1. The dash line in Figure 9 demonstrates the synthetic controls counterfactual $\sum_{k \in K} \hat{w}_k y_{kt}$ for each t in both T_0 and T_1. The gaps between the counterfactual and actual leakage rate inform us about the treated effect each day. The city-specific treatment effect for the post-period with $|T_1| = 28$ days is thus:

$$\frac{1}{|T_1|} \sum_{t \in T_1} \left(y_{ct} - \sum_{k \in K} \hat{w}_k y_{kt} \right)^2 \tag{11}$$

The basic idea of the synthetic control method is to utilize pre-period data to construct weighted averages of the non-treated units that fit the treated unit well, and then use those weights to construct the counterfactual for each treated unit in the post-period. Figure 9 shows that we have a good fitting with the average difference between the fitted and actual leakage rate at zero for $|T_0| = 28$ days before the intervention.

The estimated Beijing-specific treatment effect on leakage rate is 3.2% in the 28 days

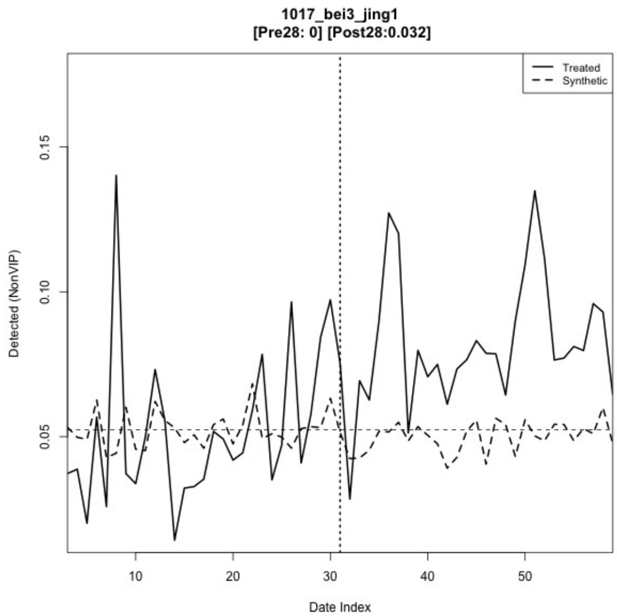


Figure 9: Detected Rate

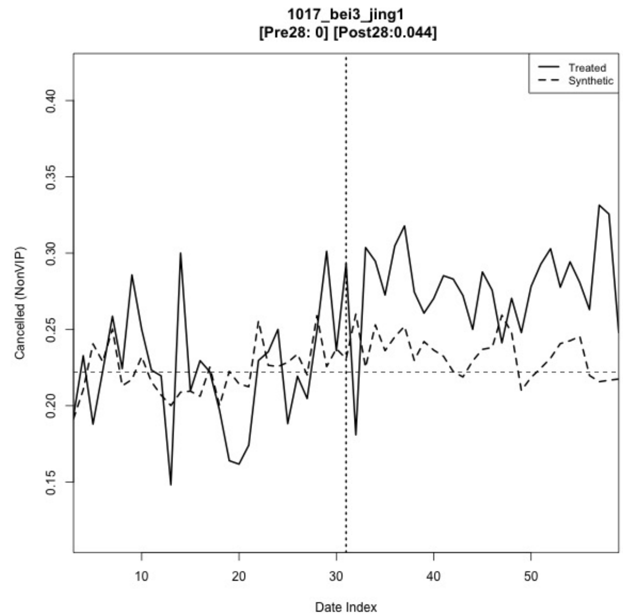


Figure 10: Cancellation Rate

(see Figure 9) after the intervention. We can conduct the same exercise for cancellation rate and find that 4.4% additional transactions are canceled in the 28 days (see Figure 10)after charging the commission fee.

A.2 The Distribution of SCM Estimates of All Treated Cities

We are interested in an overview of the treatment effects across different cities. The information helps inform the sampling strategy for the structural model in Section 4. If we see obvious heterogeneity across cities, we should randomly draw drivers from all over the country to obtain a representative sample to generate insights for the platform. I choose synthetic control methods over difference-in-difference regressions because I can run the estimations in a loop without manually finding adjacent cities for the control group.

We estimate the city-specific treatment effects by creating synthetic controls using Equation (10) for all 33 cities that are treated in the 137 days. Figure 11 demonstrates the city-specific treatment effects on leakage rate and cancellation rate for Non-VIPs. The treatment effects are the gaps between the counterfactual and actual leakage rate. They are calculated based on Equation (11) in the post-period. The top right panel shows that treatment effects on the two metrics are positively correlated and mostly non-negative.

We can conduct similar exercise for Super-VIPs. Table 3 documents the summary statistics of city-specific treatment effects for both the Non-VIPs and Super-VIPs in the 33 cities. The cancellation rate is a nosier metric than the leakage rate according to the standard deviation. This observation can be confirmed by looking into the histograms (Figure 11) of treatment effects on leakage rate are more concentrated than the treatment effects on cancellation rate, or by looking into the wider confidence intervals in Figure 13 than the ones in Figure 12.

Moving on from demonstrating the heterogeneity across cities, Figure 12 averages across

Figure 11: The City-Specific Treatment Effects for Non-VIPs from Synthetic Control Method

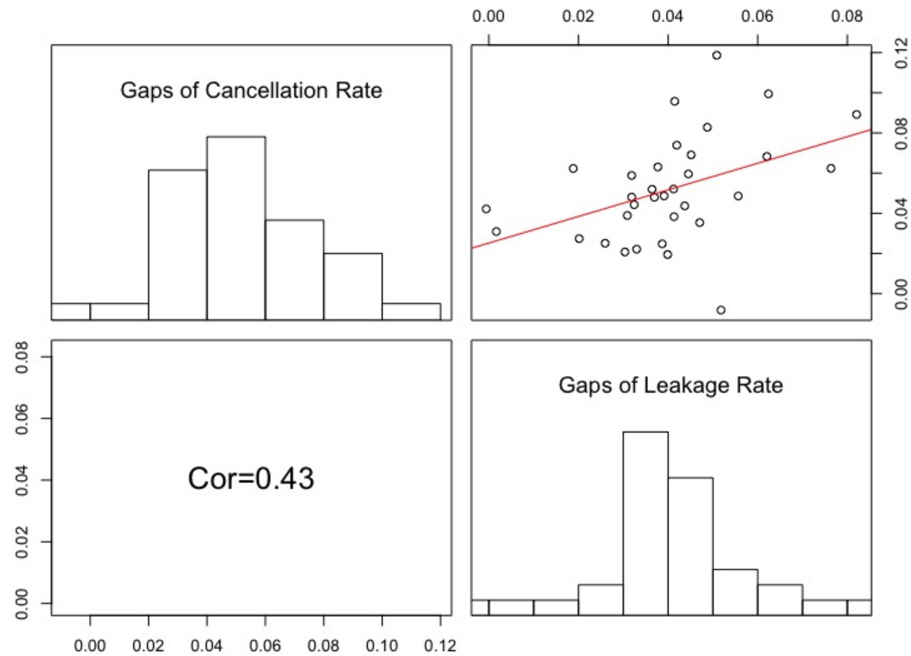


Table 3: Fitted and Predicted Gaps of Synthetic Controls

	Metric	Fitted Gaps Pre28 Mean	Predicted Gaps Post28 Mean (Std Dev)
Non-VIP	% Detected	0%	+4.00% (1.71%)
	% Cancelled	0%	+5.17% (2.65%)
Super-VIP	% Detected	0%	-0.19% (0.34%)
	% Cancelled	0%	-1.19% (1.24%)

the 33 cities and shows the mean detected ratio per day by centering their time series at the launch date. The increase in average leakage rate for Non-VIP drivers after the commission launch is prominent. In contrast, we do not see any visible changes in the average leakage rate for VIP drivers who were not charged the 15% commission rates. Alternatively, we can check how cancellations change after the launch of the commission. Figure 13 shows the average cancellation rate per day across the 33 cities by centering their time series at the launch date. The mean pre- and post-intervention cancellation rates are prominent again for Non-VIP drivers but not for VIP drivers who had fee waiver.

The histograms of estimates in Figure 11 from synthetic control methods demonstrate the variation of commission effects. While the variation of the effects on cancellation rates is larger than the ones on leakage rate, both city-level effects are positively correlated and scattered around the diagonal line. Using additional geolocation data seems to reduce the noise for verifying disintermediated transaction.

Another source of the variation in city-level effects might come from the heterogeneity

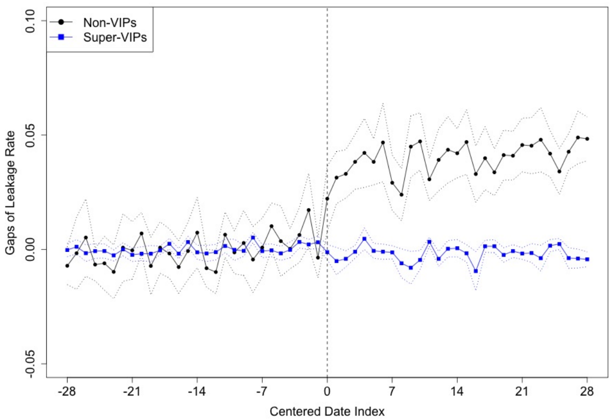


Figure 12: Effects on Detected Rate

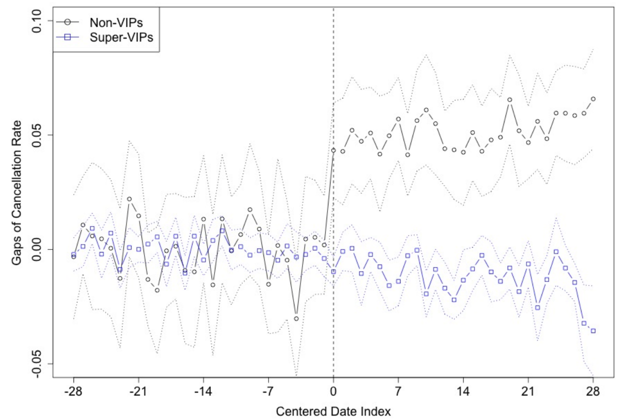


Figure 13: Effects on Cancellation Rate

across cities - different markets feature a different mix of driver composition. With the suggestive evidence in mind, we decide to randomly draw samples of drivers from the entire country to leverage the potential heterogeneity in fee sensitivity across geographical regions because it provides identification for the primitives in the structural model in Section 4.

B Microfounded Bargaining

B.1 Nash Bargaining

Eq (6) assumes that driver i and customer j can reach an equilibrium outcome $\lambda^*_{ij(t)}$ when the joint utility gain from leakage is non-negative. The $\lambda^*_{ij(t)}$ enables agreement as long as both of their individual latent utility gains are non-negative.

This section uses the Nash bargaining model(Nash, 1950) to characterize the static equilibrium. The Nash bargaining solution is a convenient way to characterize the outcome when we do not observe the bargaining process(Jiang, 2022). Given the payment division rule $\lambda_{ij(t)} \in [0, 1]$, the latent utility gains from leakage for driver i and customer j are:

$$\begin{aligned} \Delta\pi_{ij(t)} &= \Pi^1_{i(t)} - \Pi^0_{i(t)} \\ &= \beta_i \cdot (\gamma_{i(t)} - \lambda_{ij(t)}) p_t - h_{i(t)} \end{aligned} \tag{12}$$

and

$$\begin{aligned} \Delta u_{ji(t)} &= U^1_{j(t)} - U^0_{j(t)} \\ &= \beta_j \cdot (\lambda_{ij(t)} p_t - s_{j(t)}) - h_{j(t)} \end{aligned} \tag{13}$$

The Nash bargaining solution (Sieg, 2000; Zhang et al., 2021) is given by maximizing the generalized Nash product specified by the bargaining weight η:

$$\begin{aligned} \max_{\lambda_{ij(t)}} \ & (\Delta\pi_{ij(t)})^{\eta} \cdot (\Delta u_{ji(t)})^{1-\eta} \\ \text{s.t.} \quad & \Delta\pi_{ij(t)} \geq 0 \\ & \Delta u_{ji(t)} \geq 0 \\ & 0 \leq \eta \leq 1 \end{aligned} \tag{14}$$

where $\eta \in [0,1]$ represent the relative bargaining power of driver.

We assume homogeneity of bargaining power in this analysis to start with a parsimonious model, which is micro-founded with fewer primitives. Future analyses can account for the heterogeneity in η across different types of players in the market (Zhang et al., 2021; Jiang, 2022), or estimate a hierarchical Bayesian η that is internally consistent with the data.

The Nash bargaining solution needs to satisfy equality based on the bargaining power:

$$\frac{\eta}{1-\eta} = \frac{\Delta\pi_{ij(t)}}{\Delta u_{ji(t)}} \tag{15}$$

We can analytically solve for the optimal $\lambda^*_{ij(t)}$ after rearranging the function:

$$\lambda^*_{ij(t)} = \frac{(1-\eta)\beta_i\gamma_{i(t)} + \eta\beta_j\frac{s_{j(t)}}{p_t} + \eta\frac{h_{j(t)}}{p_t} - (1-\eta)\frac{h_{i(t)}}{p_t}}{(1-\eta)\beta_i + \eta\beta_j} \tag{16}$$

There exists a unique solution to maximize joint surplus according to the Nash bargaining. The analytical solution indicates that the discount would be larger if the platform commission rate is higher, the subsidy to the consumer is higher, the hassle of the customer is higher, and the hassle of the driver is lower.

B.2 Estimation and Simulations

We can plug the analytical solution of Equation (16) back into Equation (6) to simulate the choice of leakage. We use the model specification in Section 5.2 for estimation.

$$\begin{aligned}
&Pr[L_t = 1|\eta, \gamma_{i(t)}, s_{j(t)}, h_{i(t)}, h_{j(t)}] = Pr[\Delta\pi^*_{ij(t)} + \Delta u^*_{ji(t)} + \epsilon_{ij(t)} + \epsilon_{ji(t)} \geq 0] \\
&= Pr[\beta_i \cdot \gamma_{i(t)}p_t - \beta_j \cdot s_{j(t)} - (\beta_i - \beta_j)\cdot\lambda^*_{ij(t)}\cdot p_t - h_{i(t)} - h_{j(t)} + \epsilon_{ij(t)} + \epsilon_{ji(t)} \geq 0] \\
&= Pr[\beta_i \cdot \gamma_{i(t)}p_t - \beta_j \cdot s_{j(t)} - \frac{\beta_i - \beta_j}{(1-\eta)\beta_i + \eta\beta_j}\cdot[(1-\eta)\beta_i\gamma_{i(t)}p_t + \eta\beta_j s_{j(t)} \\
&+ \eta h_{j(t)} - (1-\eta)h_{i(t)}] - h_{i(t)} - h_{j(t)} + \epsilon_{ij(t)} + \epsilon_{ji(t)} \geq 0]
\end{aligned} \tag{17}$$

If $\eta = 1.0$, drivers have the full bargaining power. Drivers will provide a take-it-or-leave-it-offer that makes customers indifferent between off-platform and on-platform transactions.

$$Pr[L_t = 1|\eta = 1.0] = Pr\left[\beta_i\cdot\left(\gamma_{i(t)}p_t - s_{j(t)}\right) - h_{i(t)} - \frac{\beta_i}{\beta_j}h_{j(t)} + \epsilon_{ij(t)} + \epsilon_{ji(t)} \geq 0\right] \tag{18}$$

Equation (18) shows that the probability of leakage depends on the customers' subsidy and hassle even though customers have no bargaining power (i.e., $\eta = 1.0$). In this extreme case, customers' individual rationality constraint is binding at the equilibrium.

We find that $\eta = 0.975$ yields the maximum log-likelihood in our sample when we use Equation (17) to simulate the choice of leakage. The result suggests that drivers almost own the full bargaining power. It is very likely for drivers to provide a take-it-or-leave-it offer: the discount makes offline transactions weakly preferred by customers.

Given the institutional details, it is not a surprise that drivers have extreme bargaining power. Most drivers are experienced professionals that take jobs regularly. As a result, they have better negotiation skills than customers who are only on the platform for a one-off transaction. Moreover, drivers have complete information about the commission fee the platform charges them, but customers may not know such exact details. However, Table **??** show that experienced users who have repeated transactions on the platform can obtain an offline discount ranged from 16% to 21.6% (see Section **??**). We will discuss what determines bargaining power in Appendix B.3.

The model estimates allow us to simulate the outcomes of counterfactuals by doing a grid of $\eta \in \{0, 0.1, ..., 0.9, 1.0\}$ (see Figure 8) using Equation (17). The worse leakage outcome happens at $\eta = 0.65$ when drivers have slightly stronger power than customers. The simulation shows that the platform would prefer full bargaining power on the driver side ($\eta = 1$) rather than on the customer side ($\eta = 0$).

B.3 More on Bargaining Power

The η is an underlying primitive of a bargaining model. Players with more bargaining power obtain a bigger share of the surplus. This power is captured differently in different contexts. Rubinstein (1982) describes bargaining power as a player's patience (e.g., the discount factor in dynamic models). In other bargaining models (Binmore et al., 1986), bargaining power can represent concepts such as negotiation skills or experience of the player.

The bargaining power is also related to the market thickness for a player (Fong, 2020). Having more available outside options and less competition (Backus et al., 2020) increases the bargaining power. The positive externality of a player might plays a role in determining the bargaining power (Zhang et al., 2021).

With better data, researchers can estimate the bargaining power under different supply-demand relationships, which differs across space, time, and product popularity:

- When the customer has a hard time finding a vehicle, the driver has stronger power.
- When the driver has difficulty getting a job, the customer has high power.
- A scheduled trip on the next day gives a patient customer more time to find a backup driver and thus a potential stronger power in the market.
- A driver with a medium truck is competitive because the vehicle dominates the medium and small van for being able to load more items.
- Experienced drivers or users with repeated transactions on the platform might have better negotiation skills and thus stronger bargaining power.

In our context, the platform can potentially leverage or affect the bargaining power to mitigate leakage by changing market conditions or making strategic matching.