

Identifying Algorithmic Price Discrimination

Nan Wang* Yang Yu†

October 13, 2026

Abstract

Modern pricing algorithms draw on high-dimensional consumer information, which makes price discrimination subtle and hard for consumers and regulators to detect. This paper proposes a disparity metric to test for algorithmic price discrimination. Our method uses double machine learning (DML) and data on both accepted and unaccepted price quotes to address the false positives caused by model misspecification and consumers' selection bias. We apply the method to a major Chinese ride-hailing platform to test whether riders are charged differently based on their ordering history. Our results highlight the need for this method, as naive regression-based methods produce false positives.

*Sauder School of Business, University of British Columbia. Email: nan.wang@sauder.ubc.ca.

†School of Economics and Management, China University of Petroleum. Email: yangyu@cup.edu.cn.

1 Introduction

Advancements in computational capabilities and the increasing availability of granular consumer data have enabled firms to adopt algorithmic pricing strategies across a range of sectors, including but not limited to retail (Brown and MacKay, 2023), gasoline (Assad et al., 2024), housing (Calder-Wang and Kim, 2024), and transportation (Castillo, 2024). The opacity and complexity of the pricing process make it easy for firms to tailor prices to each individual order without consumers noticing. Such behavior was observed even before the widespread use of algorithms, mostly in financial markets, where the decision-making process is opaque and complex. Evidence suggests that groups that are protected today were heavily discriminated against on the basis of race (Charles et al., 2008), gender (Alesina et al., 2013), and ethnicity (Stefan et al., 2018). Although there is no solid evidence that firms use algorithmic pricing to discriminate unlawfully, reports from mainstream media (Martin, 2019; Mahdawi, 2018) and regulators (Federal Trade Commission, 2025) suggest that it exists.

Regulatory authorities across numerous jurisdictions have enacted measures to prohibit discrimination based on protected characteristics, such as gender, race, or age, in the deployment of automated systems.¹ Despite the growing prevalence of algorithmic pricing, there remains a lack of credible empirical methodologies for identifying discrimination against protected groups in these settings.

This paper introduces, to our knowledge, the first disparity metric specifically designed to detect algorithmic price discrimination. Our metric is closely aligned with the legal concept of disparate impact, which is widely adopted in the regulation of automated decision-making systems (Barocas et al., 2023). From an economic perspective, the metric is interpretable as the average treatment effect (ATE) of protected group status on the price received, thus bridging legal and causal inference frameworks.

We highlight two key empirical challenges that complicate the detection of algorithmic

¹For example, in the United States, the Consumer Financial Protection Bureau (CFPB), the Department of Justice’s Civil Rights Division (DOJ), the Equal Employment Opportunity Commission (EEOC), and the Federal Trade Commission (FTC) jointly issued a statement on April 25, 2023, to prevent discrimination in automated systems. In the European Union, both the General Data Protection Regulation (GDPR, effective May 25, 2018) and the Artificial Intelligence (AI) Act (adopted June 13, 2024) explicitly prohibit discrimination on “special categories of personal data,” including ethnicity, religion, and sexual orientation.

price discrimination. First, the complexity and high dimensionality of modern pricing algorithms render standard linear modeling approaches inadequate, leading to potential model misspecification bias in causal effect estimation. Second, reliance on observed transactional data introduces sample selection bias, as such data are conditional on consumers’ unobserved utility and purchase decisions.

To address these challenges, we leverage recent advances in causal machine learning to develop a robust framework for detecting interpersonal price discrimination, net of individual-level heterogeneity in price responsiveness. Flexible machine learning methods mitigate model misspecification by accommodating complex functional forms in both outcome and treatment models. However, fully addressing sample selection bias is inherently difficult without access to data on unrealized transactions, cases in which consumers received an offer but did not complete a purchase. We argue that credible detection of algorithmic price discrimination requires information on the universe of potential transactions, not merely those that culminate in observed purchases.

We implement our proposed framework in the context of a major Chinese ride-hailing platform. Since 2018, concerns about so-called “Big Data Swindling” have proliferated on Chinese social media (Xu, 2024), with widespread reports alleging that loyal or frequent users are systematically charged higher prices than new users for identical services. In response, Chinese regulators issued explicit guidelines prohibiting such practices as an abuse of market power (Gov.cn, 2021). Leveraging a large-scale proprietary dataset from the platform, we empirically assess whether algorithmic pricing results in systematic disparities between frequent and infrequent users. Our baseline analysis, which addresses both model misspecification and sample selection bias, finds no statistically significant price disparity between frequent and infrequent users. In contrast, estimates based solely on realized transactions, ignoring selection bias, suggest that frequent users pay, on average, 0.2% more than non-frequent users. This discrepancy highlights the importance of accounting for selection into observed transactions. This is in line with our theoretical prediction that naive estimates are biased upward if frequent users have a higher willingness to pay or inferior outside options. We further assess the robustness of our findings using a range of machine learning and linear estimators. Across specifications, we find that failure to address sample selection bias consistently produces spurious evidence of discrimination against frequent users. Conversely, linear models such as OLS and LASSO,

which are ill-suited to capture the complex, nonlinear pricing rules of modern algorithms, tend to underestimate any potential disparity. Collectively, our results underscore the necessity of both flexible modeling and careful attention to sample selection in credible empirical assessments of algorithmic price discrimination.

The remainder of the paper is structured as follows. Section 2 reviews the related literature. Section 3 presents our conceptual framework, introduces the disparity metric, and discusses the key empirical challenges in its implementation. Section 4 describes the empirical application, including the institutional background, data, regression model, and estimation strategy. Section 5 reports the results, including descriptive statistics and estimation findings.

2 Related Literature

We build on a large literature that develops empirical tests for discriminatory treatment and outcomes. Evidence of disparate outcomes has been documented in credit and finance (e.g., Charles et al., 2008; Alesina et al., 2013; Stefan et al., 2018). In regulatory and legal settings, enforcement often relies on *disparate impact*—whether a protected group experiences an overall adverse effect—rather than intent (Barocas et al., 2023). When protected-class labels are *unobserved*, recent methods infer or combine auxiliary data to assess fairness and disparate impact (Kallus et al., 2022), complementing our setting where group membership (frequent vs. infrequent users) is observed. Our contribution is to operationalize a *price* disparity metric aligned with disparate impact and to interpret it causally as an ATE of group status on prices, connecting legal standards to modern causal inference.

Firms increasingly deploy algorithmic pricing across sectors, with documented effects on price levels, dispersion, and competition in retail, gasoline, housing, and transportation (Brown and MacKay, 2023; Assad et al., 2024; Calder-Wang and Kim, 2024; Castillo, 2024). In particular, Assad et al. (2024) show that adoption of pricing software in German retail gasoline raised margins and altered competitive dynamics. While most work studies efficiency and competition, explicit tests for *interpersonal* price disparities are scarce. We fill this gap by auditing a ride-hailing platform’s pricing algorithm for systematic differences across user groups using universe-of-quotes data, complementing anecdotal

concerns about “personalized pricing” in the media and policy discourse (Martin, 2019; Mahdawi, 2018; Federal Trade Commission, 2025).

Methodologically, we rely on causal ML to handle high-dimensional, nonlinear pricing rules while delivering valid average effects. We implement double/debiased machine learning (DML) to construct Neyman-orthogonal moments in a partially linear model (Robinson, 1988; Chernozhukov et al., 2018, 2025), mitigating regularization bias from flexible learners (Athey, 2018). We also highlight that transactional data introduce classic sample-selection concerns (Heckman, 1979; Little and Rubin, 2019; Davey, 2009); access to *potential* (offered but not necessarily accepted) prices is crucial for credible detection of algorithmic disparities. Our setting shows how causal ML and appropriate data jointly enable enforceable tests for algorithmic price discrimination.

3 Conceptual Framework

In this section, we lay out our conceptual framework. We first describe the data generating process (DGP), which specifies how firms use algorithms to generate prices. The DGP implies that as long as consumers from distinguishable groups respond differently to price, the firm has an incentive to discriminate. We then describe our causal framework, closely following the potential outcomes literature (Neyman and Iwazskiewicz, 1935; Rubin, 2005; Holland, 1986), and define our disparity metric based on this framework and the legal definition of discrimination.

3.1 Data Generating Process

We consider a monopoly market with a unit mass of myopic consumers. With this limited scope, we rule out strategic behavior such as platform competition and consumers manipulating the algorithm. Products are differentiated and their features are denoted by $W \in \mathbb{R}^k$. Consumers are heterogeneous and can be distinguished by two variables: $D \in \{0, 1\}$, and $Z \in \mathbb{R}^l$. Here D denotes the consumer’s protected feature, which we assume is binary for simplicity, and Z denotes the other features. To simplify notation, write $X = (W, Z)$. We model consumer utility with a very general random utility model and write the probability of purchasing an offered product at price p as $\pi(p, X, D)$. We assume the firm can observe all consumer features that are relevant for the purchase

decision, i.e., D and X .

Non-discriminatory Algorithmic Pricing refers to the case in which the firm does not discriminate on the protected consumer characteristic. There are two algorithms in the pricing process. The first is a prediction machine that computes expected demand from the observable features given the price p :

$$\pi(p, X) = E[\pi(p, X, D) | p, X]$$

In practice, a firm uses experimental data to estimate the expected demand function $\hat{\pi}(p, X)$ in order to avoid endogeneity issues. We therefore focus on the case in which the firm generates prices with the true model. The second algorithm is an optimization solver that finds the price that maximizes expected revenue given the expected demand:

$$p(X) = \arg \max_p p\pi(p, X)$$

Discriminatory Algorithmic Pricing refers to the case in which the firm differentiates between consumers of different types. The first algorithm then estimates a more flexible demand function $\pi(p, X, D)$, which is closer to consumers' actual decision rule. The optimization algorithm then generates the price by solving:

$$p(X, D) = \arg \max_p p\pi(p, X, D)$$

Given the setup above, the revenue generated by a discriminatory algorithm is no less than the revenue generated by a non-discriminatory algorithm. As long as consumers from different protected groups behave differently, i.e., there exist (p, X) such that $\pi(p, X, 0) \neq \pi(p, X, 1)$, the discriminatory algorithm should generate higher revenue.

3.2 Disparity Metric

There are two major theories of discrimination in the legal framework: disparate treatment and disparate impact. Under disparate treatment, it is discriminatory to intentionally treat individuals differently. However, regulating on the basis of intent can be fruitless, especially for algorithms. It requires extensive effort to examine the logic of the algorithm closely, not to mention that firms can use proxies to discriminate against protected groups without directly using the protected features. As a result, lawmakers

favor disparate impact in regulating automated systems (Barocas et al., 2023). Disparate impact defines discrimination as a protected group receiving an overall de facto adverse impact, whether intended or not. We construct our disparity metric closely in line with this definition. It captures the expected price disparity with respect to the protected consumer feature and is given by

$$\theta = E[E[P_i|D_i = 1, X] - E[P_i|D_i = 0, X]]$$

In practice, when regulators request data from platforms to test for potential unlawful price discrimination, X_i typically includes all covariates that the firm claims are used in the pricing algorithm. The disparity metric remains valid regardless of whether the firm is fully transparent in its disclosures. If the estimated disparity metric is statistically significant, this indicates either (1) that the firm has not accurately disclosed all variables used in pricing, or (2) that the algorithm produces disparate impact without directly utilizing the protected characteristic. The latter scenario may arise if the algorithm employs unprotected variables as proxies for the protected characteristic, or if algorithmic bias is present in the pricing process. While our disparity metric is motivated by legal definitions, it is also substantively meaningful from a causal inference perspective. Specifically, it can be interpreted as the average treatment effect (ATE) of the protected characteristic (*treatment*) on price (*outcome*), under certain assumptions.

To formalize this, we adopt the potential outcomes framework (Neyman and Iwazskiewicz, 1935; Rubin, 2005; Holland, 1986). For each order i , let $P_i^{(d)}$ denote the potential price that would be assigned if the protected characteristic were set to $d \in \{0, 1\}$. We can then write the ATE as:

$$ATE = E[P_i^{(1)} - P_i^{(0)}]$$

For our disparity metric to identify the ATE, we need two additional assumptions, which we now state formally.

Assumption 1 (Conditional Ignorability/Unconfoundedness).

$$P_i^{(d)} \perp\!\!\!\perp D_i \mid X$$

This assumption implies that, conditional on the observed covariates X , the assignment of the protected characteristic D is independent of the potential outcomes for price.

In other words, after accounting for X , the distribution of D is as if randomized with respect to the potential prices. It is important to note that, in our context, the (individual) treatment effect is defined at the order level rather than the individual level, as our primary objective is to identify price disparities attributable to group membership rather than to specific individuals.

Assumption 2 (Overlap). *Write the probability that an order with features X comes from a consumer in the protected group, i.e., the propensity score, as $\Pr(D = 1|X) = p(X)$. At each value x in the support of X , we have:*

$$\Pr(0 < p(X) < 1) = 1$$

This assumption requires that, for every value x in the support of X , there is a nonzero probability of observing both $D = 0$ and $D = 1$. In other words, the distribution of D conditional on X is not degenerate at any x . Without this overlap, it is impossible to disentangle the effect of the protected characteristic from the effect of the covariates at those values of X , rendering the conditional and unconditional average treatment effects unidentified in those regions. It is important to emphasize that this assumption pertains to the underlying data-generating process, not to the empirical realization in a finite sample. In high-dimensional settings like ours, it is common for the observed data to be sparse, so that for many values of x in the sample, only one value of D is observed. However, this does not violate the overlap assumption, as the requirement is at the population level. Empirically, we address sparsity by employing machine learning methods to estimate the propensity score and by adopting recent advances in debiased machine learning (Chernozhukov et al., 2018) to ensure that our ATE estimates are robust to local misspecification of nuisance parameters such as the propensity score.

With the two assumptions above, we have the theoretical equivalence

$$\theta = E[E[P_i|D_i = 1, X] - E[P_i|D_i = 0, X]] = ATE$$

A limitation of this disparity metric is that it may fail to detect instances of price discrimination that do not result in a significant average effect. Specifically, it cannot rule out the possibility that members of the protected group experience both favorable and unfavorable price differences across different contexts, leading to an overall average impact that is close to zero. Such a scenario can arise when the direction of the disparity varies

with other covariates; formally, there may exist x, x' such that $\pi(p, x, 1) > \pi(p, x, 0)$, while $\pi(p, x', 1) < \pi(p, x', 0)$. Nevertheless, this limitation is generally not problematic in the context of legal standards, which focus on the existence of a net overall adverse impact rather than context-specific disparities.

3.3 Empirical Challenges

3.3.1 Misspecification

While we have established that, under mild assumptions, our disparity metric is theoretically equivalent to the average treatment effect (ATE), our empirical context differs fundamentally from standard policy evaluation settings. In conventional policy evaluation, researchers typically observe only a limited set of covariates, and the true data-generating process is unknown, making omitted variable bias a primary concern for ATE estimation. In contrast, in our setting, where the data relevant to price generation are proprietary, regulators must collaborate with firms to detect unlawful algorithmic price discrimination. Firms are legally obligated to disclose the variables used in pricing, and dishonest disclosure carries potential legal consequences. As a result, omitted variable bias is less critical in this context. However, the complexity and high dimensionality of the pricing algorithm, which constitutes the true data-generating process, make model misspecification a significant concern for the estimation of the disparity metric.

To address this, we employ machine learning methods that require minimal functional form assumptions and are computationally efficient for large-scale data. While machine learning does not fundamentally alter the identification of the ATE, it provides the necessary flexibility to model complex outcome functions (Athey, 2018). Nevertheless, the use of flexible machine learning models introduces the risk that ATE estimation may be biased due to errors in estimating nuisance parameters. As demonstrated by Chernozhukov et al. (2018), naive application of machine learning can lead to such bias. We mitigate this issue by leveraging recent advances in causal machine learning, specifically the double machine learning (DML) framework. The key idea, analogous to the approach of Robinson (1988), is to use a residual-on-residual regression that constructs Neyman-orthogonal moment equations. This approach ensures that the ATE estimator is locally robust to small errors in the estimation of nuisance parameters, thereby improving the

reliability of our causal inference.

3.3.2 Selection Bias in Transactional Data

The second empirical challenge is the selection bias in typical transactional data, which contain only the transactions that actually took place. It is noteworthy that, in the previous section, we addressed selection bias arising from non-random assignment of the treatment variable by imposing additional identification assumptions. The selection bias of transactional data is a distinct form of selection bias that results from consumers' purchasing decisions. Specifically, this bias, commonly referred to as sample selection bias or missing data bias in the literature (Heckman, 1979; Davey, 2009; Little and Rubin, 2019), arises because observed prices are conditional on the occurrence of a transaction, which itself depends on unobserved consumer utility and decision processes. As a result, the observed transactional data may not be representative of the full population of potential transactions, leading to biased estimates if not properly addressed.

To formalize the selection bias inherent in transactional data, we introduce a more explicit structure for the consumer utility function. While the preceding discussion relied on a general random utility framework, we now specify the utility function to clarify the mechanism underlying consumer purchase decisions and the resulting selection process. In particular, we make two key assumptions: (1) the total utility is additively separable into a valuation component and the disutility of payment, and (2) consumers from different groups differ in their valuations for the same product. A consumer will choose to purchase if and only if their total utility is positive. Formally, we define the utility function as

$$U(p, X, D) = (1 - D)u_0(X) + D u_1(X) - p$$

where $u_0(X)$ denotes the deterministic utility for consumers in group 0, and $u_1(X)$ for those in group 1. The disutility of payment is normalized to have a unit coefficient for interpretability. We can now write the estimand based only on observed transactional data as

$$\theta^{obs} = E[P_i | D_i = 1, X_i, m_i = 1] - E[P_i | D_i = 0, X_i, m_i = 1]$$

where m_i indicates whether the user makes the purchase. Substituting the consumer's purchase rule, we have

$$\begin{aligned}\theta^{obs} &= E\{E[P_i|D_i = 1, X_i, U(P_i, X_i, D_i) > 0] - E[P_i|D_i = 0, X_i, U(P_i, X_i, D_i) > 0]\} \\ &= E\{E[P_i|D_i = 1, u_1(X_i) > P_i] - E[P_i|D_i = 0, u_0(X_i) > P_i]\}\end{aligned}$$

In the case of a non-discriminatory pricing algorithm, where $P_i = p(X_i)$, the estimand becomes

$$\theta^{obs} = E\{E[p(X_i)|u_1(X_i) > p(X_i)] - E[p(X_i)|u_0(X_i) > p(X_i)]\}$$

In the case of a discriminatory pricing algorithm, where $P_i = p(X_i, D_i)$, the estimand becomes

$$\theta^{obs} = E\{E[p(X_i, 1)|u_1(X_i) > p(X_i, 1)] - E[p(X_i, 0)|u_0(X_i) > p(X_i, 0)]\}$$

This expectation has two important implications. First, due to the interplay between consumers' utility functions and the firm's profit-maximizing strategy, only a subset of order features will result in observed transactions—specifically, those for which both the consumer's willingness to pay and the firm's pricing decision align to permit a purchase. Thus, θ^{obs} is computed over a selected sample that is not representative of the full population of potential transactions. Second, θ^{obs} conflates the effect of group membership on price with differences in underlying consumer valuations across groups. As a result, regardless of whether the pricing algorithm is discriminatory or non-discriminatory, this observed metric does not identify the true average treatment effect (ATE) of interest. In fact, under a non-discriminatory pricing algorithm, estimation based solely on transactional data can yield spurious evidence of discrimination—potentially resulting in false positives that expose non-discriminatory firms to unwarranted regulatory or legal scrutiny.

4 Empirical Application

In this section, we present the empirical foundation of our analysis, applying the disparity metric and addressing the empirical challenges outlined previously. We investigate algorithmic price discrimination using data from a major Chinese ride-hailing platform. Due to confidentiality agreements and ongoing data clearance, we refer to the provider as

Firm A throughout the paper. This section is organized as follows. We first provide an overview of the ride-hailing industry, including the operational structure, pricing mechanisms, and the relevance of algorithmic price discrimination. Next, we describe the data sample used in our analysis and present key descriptive patterns. We then introduce our regression model, which incorporates the features outlined in the preceding theoretical section. Finally, we detail our use of double machine learning techniques to estimate the disparity metrics, addressing the empirical challenges inherent in this setting.

4.1 Background

We test empirically for algorithmic price discrimination on ride-hailing platforms. This type of platform is widely available around the world and has radically changed the structure of urban transportation.

Price Structure. In line with industry standards, the price structure on **Firm A** consists of two primary components: the *base fare* and a dynamic *multiplier*. The base fare is determined by a transparent, non-algorithmic formula:

$$\text{base fare} = \text{fixed fee} + p^{\text{time}} \times \text{predicted duration} + p^{\text{mileage}} \times \text{predicted mileage}$$

where p^{time} and p^{mileage} denote the unit prices for time and distance, respectively. These unit prices vary by city and time period (e.g., weekday versus weekend, morning versus afternoon) and are publicly disclosed in the rider app. Notably, the base fare does not adjust in response to real-time fluctuations in demand or supply; however, it does incorporate anticipated traffic conditions through the predicted duration and mileage components. The second key component of the price structure is the *multiplier*, which is determined in real time by the platform’s proprietary algorithm based on prevailing demand and supply conditions. The focal platform typically applies a downward-adjusting multiplier, functioning as a real-time discount or promotion. This contrasts with the upward-adjusting “surge pricing” multipliers commonly used by other ride-hailing platforms, which increase fares during periods of excess demand. Despite these differences, both mechanisms serve to dynamically balance market conditions and optimize platform revenue by adjusting prices in response to demand-supply imbalances.

Upon entering the origin and destination in the app, the rider is immediately presented with the total fare, calculated as $\text{base fare} \times \text{multiplier}$. Additionally, the platform displays

the dynamic promotion, which is computed as $base\ fare \times (1 - multiplier)$. The rider then decides whether to request a ride based on these numbers.

Concerns Regarding Price Discrimination. The main concern about potential unlawful price discrimination stems from the opacity of the algorithmically determined multiplier. Since this multiplier is dynamically adjusted and each consumer observes only their own fare, it is challenging for consumers to detect systematic price disparities across individuals or groups. There is a risk that the platform could use individual-level data to implement personalized pricing, offering different prices for the same service to different users. While similar practices exist in industries such as airlines, housing, and finance, and are generally permitted in jurisdictions like the US and EU for their efficiency benefits, they remain subject to anti-discrimination regulations. In this context, our proposed disparity metric provides a rigorous framework for detecting and quantifying pricing disparities among protected groups, supporting compliance and policy evaluation.

In this study, we focus on a particular form of price discrimination that is explicitly prohibited under Chinese anti-discrimination regulations: the differential pricing of frequent users. Since 2018, the phenomenon known as “Big Data Swindling” (*da shu ju sha shu*) has attracted widespread public attention in China, with numerous reports on social media alleging that established users are systematically charged higher prices than new users for identical services (Xu, 2024). In response to these concerns, the Anti-Monopoly Commission of the State Council issued the “Guidelines of the Anti-Monopoly Commission of the State Council for Anti-monopoly in the Platform Economy Sector” in February 2021 (Gov.cn, 2021). These guidelines explicitly prohibit platforms with dominant market positions from engaging in practices that result in disparate impact, including the use of big data or algorithms to offer different prices based on consumers’ ability to pay, consumption preferences, or usage habits. Our empirical analysis is therefore directly motivated by both the legal framework and ongoing policy debates regarding algorithmic price discrimination against frequent users in the Chinese platform economy.

As discussed in the previous sections, a firm has an economic incentive to charge higher prices to frequent users if these users exhibit a higher probability of purchase than non-frequent users for identical services at the same price. There are two primary theoretical explanations for why such differences in purchase probability may arise. First, frequent users may have inferior outside options compared to non-frequent users. Even if

their intrinsic valuation of the service is similar, the lack of viable alternatives increases their likelihood of making a purchase. For example, consider two users with the same origin and destination: one owns a private vehicle while the other does not. The user with a car may rely less on ride-hailing services and thus is less likely to become a frequent user. When presented with the same price offer, the user without a car—having fewer alternative transportation options—is more likely to request a ride, resulting in a higher purchase probability among frequent users. Second, switching cost theory posits that frequent users face higher costs when considering alternative platforms or transportation modes. Familiarity with the current platform, accumulated user-specific benefits, or a preference for certain platform features can increase the cost of switching. As a result, frequent users are less price-sensitive and more likely to continue using the platform, further increasing their purchase probability relative to non-frequent users.

While our empirical analysis centers on the specific case of differential pricing against frequent users, a practice explicitly prohibited under China’s current legal framework, the disparity metric and estimation methodology developed here are broadly applicable. The proposed framework is sufficiently general to facilitate the detection of algorithmic price discrimination with respect to any observable consumer characteristic, contingent upon the availability of appropriate data. Our focus on frequent users is motivated by the particular legal and policy relevance of this group in the context of our study; nevertheless, the methodological approach can be straightforwardly extended to examine potential algorithmic price discrimination based on other protected characteristics, such as gender, ethnicity, race, or religion, across diverse institutional settings and industries where such characteristics are of legal or regulatory concern.

4.2 Data

Our empirical analysis utilizes data from **Firm A** in Xi’an, China, covering the period from May 1 to May 31, 2025. Xi’an is a major city in central China, with a population of over 12 million. The ride-hailing market is well-established and active, making it a representative setting for this project. The dataset focuses on a primary product category analogous to UberX, which represents the standard service tier on the platform.

Firm A was one of several major competitors operating in the region during the study period. While the presence of platform competition introduces additional com-

plexity to firms’ pricing strategies, this does not undermine our primary objective of detecting algorithmic price discrimination. In fact, multi-platform competition has been a persistent feature of most medium-sized and large Chinese cities in recent years, making it impractical to study a monopoly setting. Moreover, restricting the analysis to earlier periods would not accurately capture the platform’s current pricing practices. Our sample is therefore representative of the contemporary competitive landscape and relevant for evaluating current algorithmic pricing behavior.

Each time a rider inputs an origin and a destination, hereafter referred to as the initiation of a *bubble*, the platform instantaneously generates and displays a price quote. A *bubble* concludes either when the consumer submits a ride request or after 10 minutes of inactivity. For every *bubble*, we observe a comprehensive set of variables, including standard order attributes such as origin, destination, and fare, as well as all features that **Firm A** claims to use in computing the dynamic multiplier. While the precise set of variables employed in the pricing algorithm is proprietary and cannot be disclosed, it is important to note that all observed features pertain to time, location, or their interaction.

The key variable of interest, whether a consumer is classified as a frequent user by the platform, is not directly observable, as the platform’s internal criteria may be proprietary or arbitrary. To address this, we construct a binary indicator for user frequency based on observed order activity. Specifically, we calculate the total number of order requests for each rider over the sample period and assign frequent-user status to orders placed by riders whose cumulative order count exceeds the median (i.e., above the 50th percentile) within the platform. This operationalization aligns with our methodological framework, which is designed for binary treatment analysis. While this approach does not capture potential discrimination along a continuous spectrum of user frequency, we conduct robustness checks by varying the threshold for frequent-user classification to assess the sensitivity of our results.

4.3 Regression Model

Empirically, we use a partially linear model to describe the pricing process as in Robinson (1988) and Chernozhukov et al. (2018). For an order i :

$$\begin{aligned}
P_i &= \theta D_i + g_0(X_i) + u_i, & E[u_i|D_i, X_i] &= 0 \\
D_i &= m_0(X_i) + v_i, & E[v_i|X_i] &= 0
\end{aligned}$$

where D_i is the indicator for rider frequency, P_i is the dynamic multiplier, and X_i represents other order features that **Firm A** claims to use in generating the price. We cannot disclose the details of X_i because they are proprietary, but importantly, X_i includes common order features such as time, location, base fare, demand and supply conditions, and other time- and location-specific features.

The first equation characterizes the platform’s algorithmic pricing process, where $g_0(X_i)$ represents the baseline pricing function for non-frequent users, corresponding to $p(X, 0)$ as defined in Section 3. By adopting this partially linear specification, we impose two key restrictions. First, we assume that the price can be decomposed additively into a component attributable to the protected characteristic and a component determined by other observed features. Second, we impose a homogeneous treatment effect, such that $p(X, 1) - p(X, 0) = \theta$ for all X . An alternative approach would be to estimate a fully nonparametric model, $P_i = g_0(D_i, X_i) + u_i$, which allows for arbitrary heterogeneity in the treatment effect. However, we favor the partially linear model for its interpretability and empirical tractability, with minimal loss of flexibility in high-dimensional settings, as discussed in Chernozhukov et al. (2025).

The second equation does not admit a causal interpretation; rather, it captures the potentially complex dependence between D_i and X_i through an unknown function $m_0(\cdot)$. This reflects the possibility that assignment to the protected group is systematically related to observed covariates. The identifying assumption $E[u_i|D_i, X_i] = 0$ requires that, conditional on D_i and X_i , there are no omitted variables that jointly affect the treatment and the outcome. In our context, this assumption is fundamentally untestable, as the true pricing algorithm is proprietary and known only to the firm.

Nevertheless, it is reasonable to expect that firms disclose the relevant covariates in good faith, given the potential legal consequences of misrepresentation. Moreover, if any omitted variable is correlated with the protected characteristic, its effect will be subsumed in the estimated disparate impact, which firms have a strong incentive to avoid.

4.4 Estimation

We employ the double machine learning (DML) framework proposed by Chernozhukov et al. (2018) as our primary estimation strategy. In the context of our partially linear model, this approach extends the classic method of Robinson (1988) by leveraging flexible machine learning algorithms to estimate the nonparametric components, rather than relying on traditional kernel regression.

For comparison, a naive “single” machine learning approach would estimate the non-linear component of the model using machine learning techniques and then substitute these estimates directly into the original model to construct the moment condition for the parameter of interest. However, this approach is generally inconsistent in high-dimensional settings, as regularization bias and overfitting in the machine learning step can contaminate the subsequent estimation of the treatment effect.

The DML framework addresses these concerns by employing a residual-on-residual regression, which constructs Neyman-orthogonal moment conditions. This orthogonality ensures that the estimator for the parameter of interest is locally robust to small errors in the estimation of nuisance parameters. Furthermore, the DML procedure mitigates overfitting bias through sample splitting and cross-fitting. We now formally describe the estimation procedure.

The estimation has two stages. In the first stage, we construct two residuals. We split the sample into K folds and use machine learning techniques to estimate the cross-fitted predictions:

$$\hat{g}(X_i) = \hat{E}[P_i|X_i] \quad \text{and} \quad \hat{m}(X_i) = \hat{E}[D_i|X_i]$$

We then form the residuals $\tilde{P}_i = P_i - \hat{g}(X_i)$ and $\tilde{D}_i = D_i - \hat{m}(X_i)$. In the second stage, we form the moment condition using the following regression model:

$$\tilde{P}_i = \theta \tilde{D}_i + u_i$$

This regression identifies the ATE because in expectation we have

$$P_i - E[P_i|X_i] = \theta(D_i - E[D_i|X_i]) + u_i$$

5 Results

5.1 Descriptive Statistics

Table 1: Summary Statistics by Rider Frequency

Parameter	Rider frequency					
	< 50%ile		50%ile – 75%ile		>75%ile	
	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)
Duration (Min.)	19.71	(11.60)	18.33	(10.23)	17.80	(9.74)
Mileage (KM)	9.19	(7.71)	8.07	(5.94)	7.53	(5.02)
Base Fare	25.58	(19.15)	23.09	(14.87)	22.06	(12.83)
Total Fare	17.97	(13.19)	16.04	(10.21)	15.20	(8.84)
Multiplier	0.739	(0.10)	0.741	(0.10)	0.746	(0.11)
Number of Obs.	91,102		45,900		46,465	

Table 1 compares ride characteristics by user frequency. High-frequency users take *shorter* trips on average, in both duration and distance (19.71 vs. 17.80 minutes and 9.19 vs. 7.53 km for the lowest and highest frequency groups), yielding lower total fares in levels (¥17.97 vs. ¥15.20). The dynamic multiplier averages are very similar across groups (about 0.74) with only a marginally higher mean for frequent users (0.746 vs. 0.739). Figures 1–2 show that frequent users account for a disproportionate share of requests and are more active at peak times and locations, which can expose them more often to higher real-time multipliers. These patterns motivate conditioning on rich time and location covariates when testing for price disparities.

5.2 Estimation Results

Table 2 reports our disparity metric θ , interpreted as the ATE of frequent-user status on price (the multiplier). Using the *universe of potential requests* (accepted and not accepted), the DML estimate is statistically indistinguishable from zero ($\hat{\theta} = -0.0398$, s.e. 0.0374), implying no systematic price penalty for frequent users after conditioning on observed features. In contrast, restricting to *observed transactions* yields a positive and

Figure 1: Distribution of User Frequency in Orders

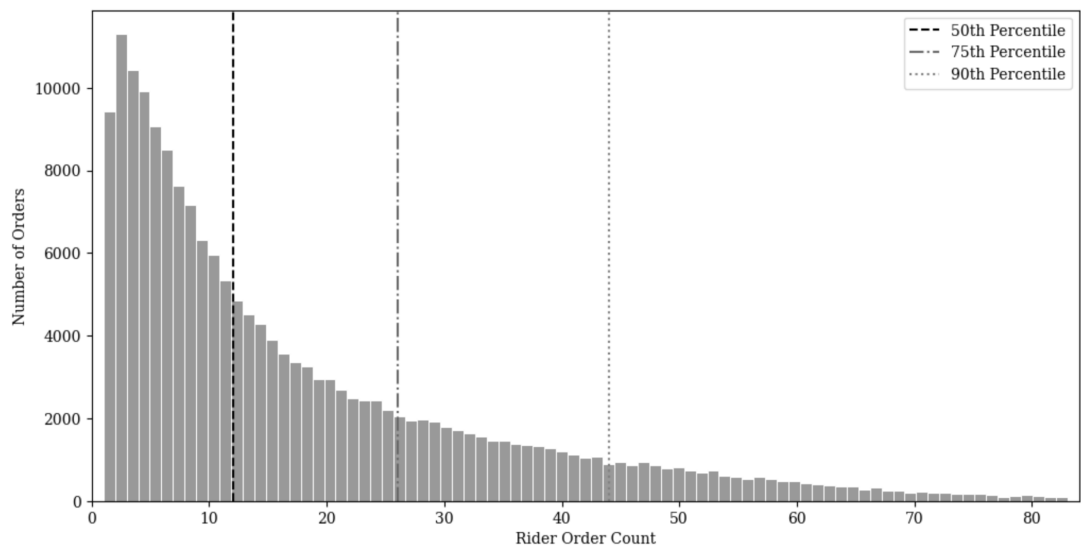
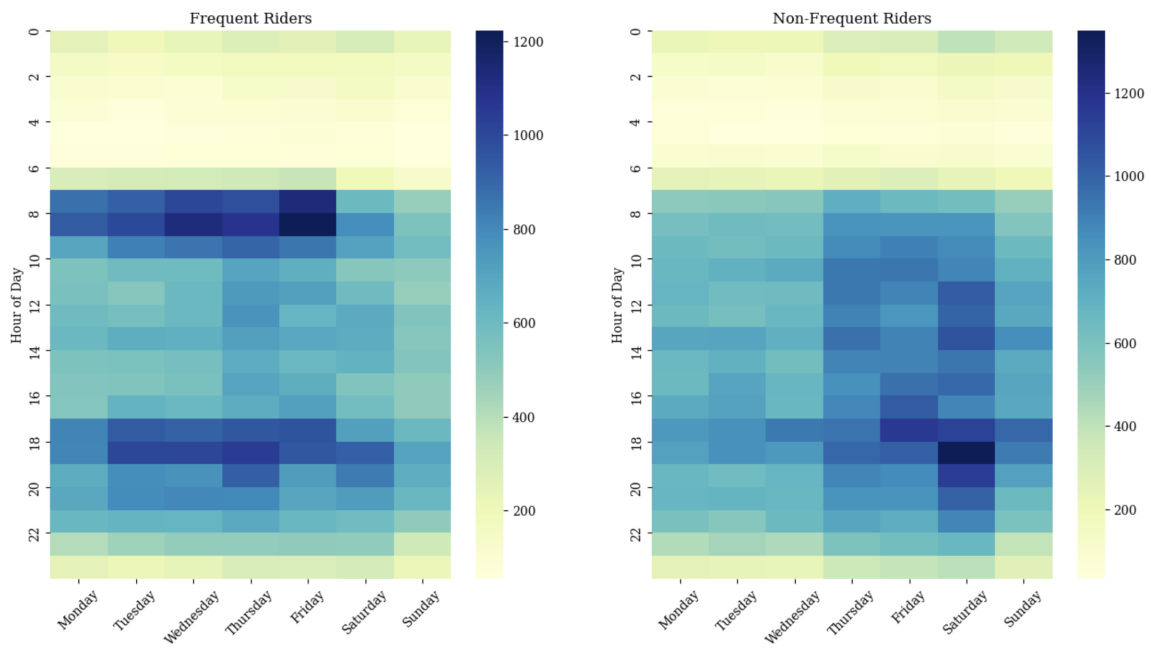


Figure 2: Request Pattern by User Frequency



highly significant estimate ($\hat{\theta}_{\text{obs}} = 0.2279$, s.e. 0.0293). This divergence is classic sample selection: frequent users, with higher willingness to pay or weaker outside options, are more likely to accept higher quoted prices, so transaction-only data overrepresent high-price acceptances and spuriously suggest discrimination. Including unaccepted quotes corrects this bias and eliminates the apparent disparity.

Table 2: Baseline Model

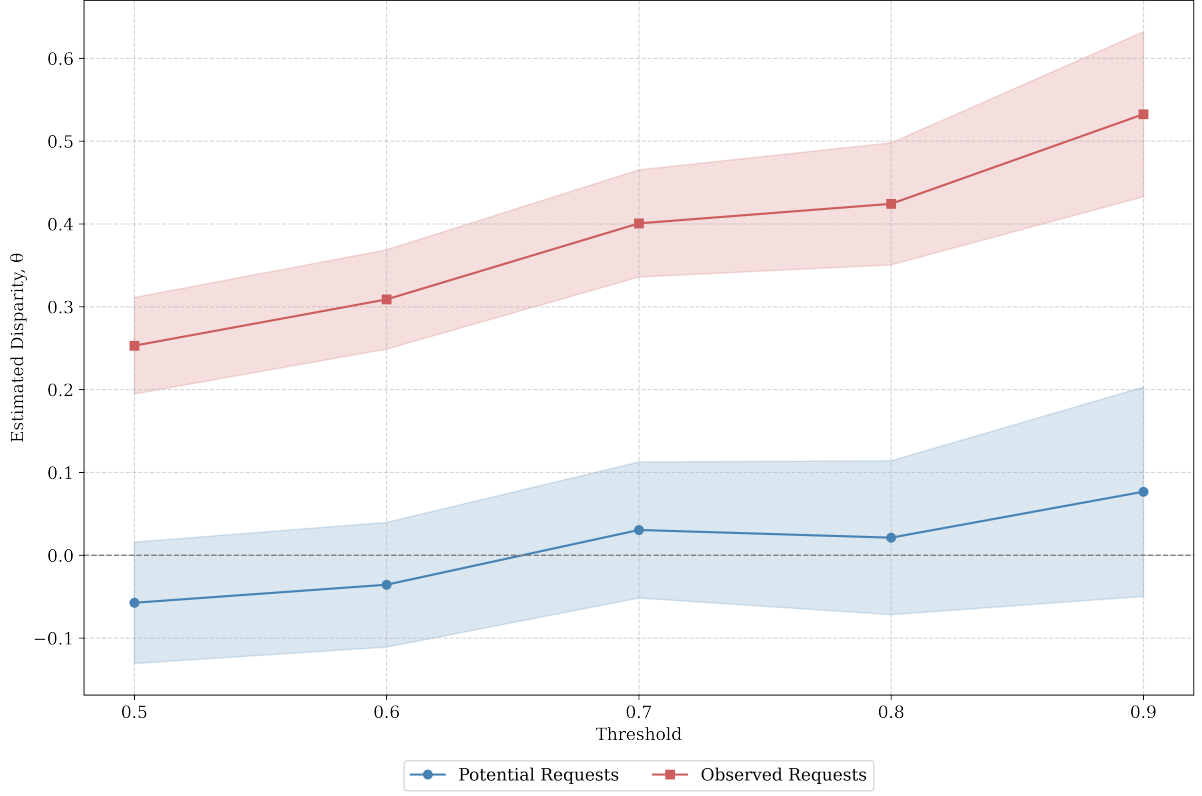
Parameter	All Potential Requests, θ	Only Observed Transactions, θ^{obs}
	(1)	(2)
θ	-0.0398 (0.0374)	0.2279 (0.0293)
Number of Obs.		200,000
ML		XGBoost
MAPE Outcome		0.0598
MAE Treatment		0.465

Note: This table reports baseline estimates of the disparity metric θ , interpreted as the average treatment effect (ATE) of frequent user status on price. Column (1) uses a random sample drawn from all potential ride requests, while column (2) uses a random sample drawn from only observed transactions. The difference in estimation results highlights the sample selection bias in observed transactional data. Both samples contain 200,000 observations. Frequent users are defined as riders above the median order count. We use $K = 5$ cross-fitting folds and median aggregation over $R = 2$ cross-fitting repetitions. Robust standard errors are in parentheses. We report the MAE of the treatment learner because the MAPE is not suitable for a binary outcome.

Figure 3 varies the frequency cutoff used to define the “frequent” group. Point estimates remain near zero and confidence intervals uniformly cover zero across thresholds, indicating robustness to alternative group definitions. Figure 4 compares learners (XGBoost, Random Forest, OLS, LASSO) within the DML/cross-fitting framework on the full potential-request sample. Flexible ML learners (XGBoost, RF) deliver tightly centered estimates around zero; linear specifications exhibit greater variability and occasional attenuation, consistent with misspecification in a complex pricing rule. Across specifications, we find no evidence of disparate pricing once selection is addressed.

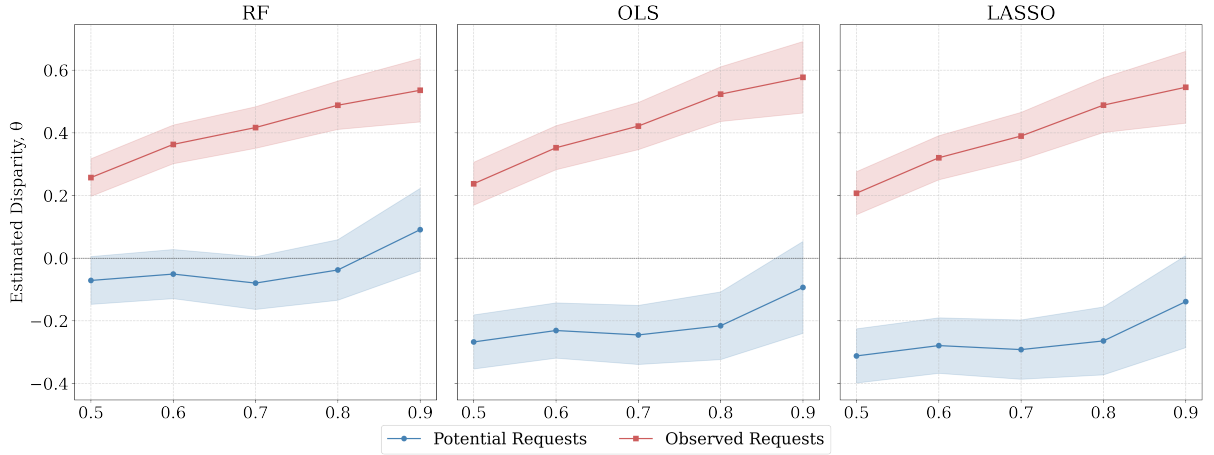
Naive audits based on completed transactions alone can falsely infer discrimination

Figure 3: Estimation Results with Different Frequency Thresholds



Note: This figure presents point estimates and 95% confidence intervals for the treatment effect across varying thresholds of user frequency. Estimation procedures and covariate adjustments are consistent with the baseline specification.

Figure 4: Estimation Results with Alternative Learners



Note: This figure presents point estimates and 95% confidence intervals for the treatment effect across varying thresholds of user frequency for alternative learners. The learners used for the three subplots are Random Forest (RF), OLS, and LASSO.

in algorithmic pricing. Credible enforcement requires access to the *universe of quotes* (or valid proxies) and methods robust to high-dimensional nuisance structure. In our application, the platform’s algorithm does not impose a net price premium on frequent users; observed differences reflect when and where they ride, not who they are.

References

- Alberto F Alesina, Francesca Lotti, and Paolo Emilio Mistrulli. Do women pay more for credit? evidence from italy. *Journal of the European Economic Association*, 11 (suppl_1):45–66, 2013.
- Stephanie Assad, Robert Clark, Daniel Ershov, and Lei Xu. Algorithmic pricing and competition: empirical evidence from the german retail gasoline market. *Journal of Political Economy*, 132(3):723–771, 2024.
- Susan Athey. The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda*, pages 507–547. University of Chicago Press, 2018.
- Solon Barocas, Moritz Hardt, and Arvind Narayanan. *Fairness and machine learning: Limitations and opportunities*. MIT press, 2023.
- Zach Y Brown and Alexander MacKay. Competition in pricing algorithms. *American Economic Journal: Microeconomics*, 15(2):109–156, 2023.
- Sophie Calder-Wang and Gi Heung Kim. Coordinated vs efficient prices: The impact of algorithmic pricing on multifamily rental markets. SSRN Working Paper No. 4403058, 2024. URL <https://ssrn.com/abstract=4403058>.
- Juan Camilo Castillo. Who benefits from surge pricing? *Econometrica*, 2024. Forthcoming.
- Kerwin Kofi Charles, Erik Hurst, and Melvin Stephens Jr. Rates for vehicle loans: race and loan source. *American Economic Review*, 98(2):315–320, 2008.
- Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, pages C1–C68, 2018.
- Victor Chernozhukov, Christian Hansen, Nathan Kallus, Martin Spindler, and Vasilis Syrgkanis. Applied causal inference powered by ML and AI. Online manuscript, 2025.
- Adam Davey. *Statistical power analysis with missing data: A structural equation modeling approach*. Routledge, 2009.

- Federal Trade Commission. Ftc surveillance pricing study indicates wide range of personal data used to set individualized consumer prices. Press release, January, 2025. URL <https://www.ftc.gov/news-events/news/press-releases/2025/01/ftc-surveillance-pricing-study-indicates-wide-range-personal-data-used-set-individ>
- Gov.cn. Guidelines of the anti-monopoly commission of the state council for anti-monopoly in the platform economy sector, 2021. URL https://www.gov.cn/xinwen/2021-02/07/content_5585758.htm.
- James J Heckman. Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, pages 153–161, 1979.
- Paul W Holland. Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960, 1986.
- Nathan Kallus, Xiaojie Mao, and Angela Zhou. Assessing algorithmic fairness with unobserved protected class using data combination. *Management Science*, 68(3):1959–1981, 2022.
- Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*. John Wiley & Sons, 2019.
- Arwa Mahdawi. Is your friend getting a cheaper uber fare than you are? The Guardian, April 13, 2018. URL <https://www.theguardian.com/commentisfree/2018/apr/13/uber-lyft-prices-personalized-data>.
- Nicole Martin. Uber charges more if they think you’re willing to pay more. Forbes, March 30, 2019. URL <https://www.forbes.com/sites/nicolemartin1/2019/03/30/uber-charges-more-if-they-think-youre-willing-to-pay-more/>.
- Jerzy Neyman and Karolina Iwaskiewicz. Statistical problems in agricultural experimentation. *Supplement to the Journal of the Royal Statistical Society*, 2(2):107–180, 1935.
- Peter M Robinson. Root-n-consistent semiparametric regression. *Econometrica: journal of the Econometric Society*, pages 931–954, 1988.

- Donald B Rubin. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American statistical Association*, 100(469):322–331, 2005.
- Matthias Stefan, Felix Holzmeister, Alexander Müllauer, and Michael Kirchler. Ethnical discrimination in europe: Field evidence from the finance industry. *PloS one*, 13(1): e0191959, 2018.
- Jian Xu. Opening the ‘black box’ of algorithms: regulation of algorithms in china. *Communication Research and Practice*, 10(3):288–296, 2024.