Leveling Down: Competition and Discrimination in Service Markets

Kwabena Donkor*

February 2, 2026

Abstract

Does competition reduce discrimination? Yes, but this paper identifies *leveling down*, a mechanism less benign than the standard selection story. Competition narrows discrimination gaps by degrading service for everyone, with the largest degradation for those previously treated best. I model discrimination as in-group favoritism rather than out-group hostility. Providers favor same-group customers, but the favoritism requires surplus that competition erodes. Preferential treatment is a luxury good. In the New York City taxi market, airport queues randomly assign passengers to drivers. Before new competitors entered in 2013, drivers charged higher fares and took longer routes for passengers whose race differed from the driver's. Competition compressed these disparities by two-thirds through leveling down. The cost of degraded service exceeded the benefit of reduced discrimination by a factor of 2.3.

Keywords: Discrimination, competition, leveling down, in-group favoritism, service quality
JEL Codes: J71, L13, R41

*Stanford Graduate School of Business. Correspondence: donkor@stanford.edu. I thank Dean Eckles, Robert Gibbons, Joseph Doyle, Catherine Tucker, and Drazen Prelec, as well as other faculty in the Marketing and Applied Economics groups at MIT Sloan, for valuable guidance and feedback. I also thank Benjamin Enke, Jeff Perloff, Isaac Sorkin, Catherine Wolfram, Kenneth Wilber, and Andrey Simonov for feedback, and Johnny Tang for generously sharing data on taxi driver licenses. I am grateful to Peter Blair for early discussions that helped shape this project, and to Gifty Asare for her support and conversations throughout. Emily Zhang, Rikhil Paresh, Shanshan Liu, and Khanh Vu provided excellent research assistance. I thank colleagues at Stanford GSB and participants at the Stanford GSB work-in-progress seminar for helpful comments and feedback. Above all, I thank God.

1 Introduction

Does competition reduce discrimination? The standard answer, following Becker (1957), is yes: competitive pressure raises the cost of refusing profitable customers and eventually drives discriminators from the market. This paper shows that competition can also reduce discrimination through a different and less benign mechanism—*leveling down*: service quality degrades for everyone, and those treated best lose the most. A competitive shock to the New York City taxi market narrowed discrimination gaps by more than half. Service declined for everyone, especially the previously favored, while outcomes for disfavored passengers did not improve. The costs of this degradation exceeded the benefits of reduced discrimination by a factor of 2.3.

In Becker (1957)'s model, prejudice is a fixed preference that does not respond to competitive pressure. Competition reduces discrimination only through selection: providers who turn away profitable customers are outcompeted by those who do not (Arrow, 1973; Guryan and Charles, 2013). Under selection, discriminators exit the market; discrimination vanishes because the discriminating providers are replaced, not because they change behavior. Drawing on Goldberg (1982), who models discrimination as in-group favoritism (*nepotism*) rather than out-group hostility (*animus*), I derive a different prediction for quality provision. Favoritism can be costly even without any mistreatment of others. A driver might route around congestion and take the most direct path for a same-group passenger—better service, but a lower metered fare. For other passengers, the driver takes the standard or less efficient route and collects the higher fare. The quality gap between favored and disfavored customers depends on the surplus available to fund preferential treatment: when margins are high, drivers can afford the revenue sacrifice; when competition erodes margins, the sacrifice is the first thing cut. Preferential treatment is a luxury good.

I study the NYC taxi market,[1] using the August 2013 entry of "green cabs" (Boro Taxis),

[1]Taxi discrimination is a documented policy concern in New York City. During the paper's sample period, a yellow cab driver was fined $25,000 for bypassing a Black family to pick up white passengers (Flegenheimer, 2015). The TLC subsequently created an Office of Inclusion and launched a citywide anti-

a new class of taxis authorized to pick up street hails in the outer boroughs and northern Manhattan, as a competitive shock.[2] Green cabs operated under identical fare structures and were near-perfect substitutes from the passenger's perspective. Their entry reduced yellow cab driver earnings by 2.4 percent per shift and doubled exit rates.

Three institutional features make this setting particularly informative. First, fares are regulated. Competition therefore operates not through prices but through service: where drivers search for passengers and how efficiently they route trips. Second, a queue system at NYC airports dispatches passengers to drivers in strict first-come-first-served order; the driver-passenger match is effectively random. Third, GPS trip records provide an objective, transaction-level measure of service quality I call *route efficiency*: whether drivers take shorter, faster routes or longer, slower ones. Under metered pricing, inefficient routing directly increases passenger fares and travel time. JFK-to-Manhattan trips carry a flat $52 fare while LaGuardia trips are metered. Route padding therefore increases driver revenue only at LaGuardia. The fare structure variation separates discriminatory intent from other routing confounds.[3]

The results confirm discrimination on two margins: service quality (intensive) and location choice (extensive). Competition affects each differently. Before green cab entry, passengers who traveled to neighborhoods racially different from the driver paid $0.40 more per trip and experienced 40-second longer rides. Drivers were 1.9 percentage points more likely to stay in neighborhoods matching their own racial background after a dropoff, a 4.2 percent increase relative to the baseline stay rate. Competition compressed the service quality gap by half to two-thirds at airports and reduced geographic sorting by one-third. Crucially, the quality gap narrowed through leveling down: after green cab entry, fares and trip times

discrimination campaign after receiving 2,748 service refusal complaints in 2019 (New York City Taxi and Limousine Commission, 2020).

[2]Green cabs were restricted to the "green zone" (outer boroughs and above 96th/110th Street in Manhattan), but enforcement was imperfect and their entry in outer boroughs pushed yellow cabs toward Manhattan, which increased competition citywide.

[3]A within-driver design compares the same driver's routing across fare structures, demographic matches, and competitive environments. This approach relates to Liu et al. (2021), who use route efficiency to study moral hazard in ride-hailing.

increased for all passengers, with the largest increases concentrated among the previously favored. The effect was strongest among drivers most directly exposed to green cab competition and those with the lowest pre-entry earnings, consistent with the prediction that discrimination narrows when the surplus that funds it shrinks. Market-wide, service degradation totaled $42.5 million annually, while discrimination reduction totaled $18.5 million, a leveling-down ratio of 2.3 to 1. Measurement error in the demographic proxy attenuates these estimates toward zero, so the reported magnitudes are conservative.

This paper makes three contributions. First, it documents and quantifies the leveling-down mechanism. Discrimination gaps can narrow because service deteriorates for everyone, with the largest deterioration concentrated among previously favored groups. The pattern challenges the default interpretation that declining discrimination gaps reflect progress. Gap reduction alone is an incomplete measure of anti-discrimination policy.[4] Second, the paper provides the first service-quality test distinguishing the Becker (1957) and Goldberg (1982) models. At any fixed level of competition, the two models produce identical predictions and are indistinguishable in the data. The models diverge in how the gap responds to competitive pressure: constant under Becker, narrowing under Goldberg when the budget constraint binds. The finding that the gap narrows is inconsistent with a pure animus mechanism and consistent with nepotism-driven discrimination that depends on available surplus. Third, the results reveal that competition operates through different channels on each margin. On the intensive margin, quality gaps narrow through leveling down. On the extensive margin, geographic sorting declines through opportunity-cost discipline. The welfare consequences of pro-competitive interventions therefore depend on which margin discrimination primarily operates through.

This paper contributes to the empirical literature testing Becker's prediction that competition reduces discrimination. Evidence spans wage gaps (Charles and Guryan, 2008),

[4]This finding relates to theoretical work on competition and quality provision (Spence, 1975; Shaked and Sutton, 1982) and empirical evidence on service markets (Matsa, 2011; Fréchette et al., 2019; Buchholz, 2022). The leveling-down result suggests the positive competition-quality relationship documented in other settings can reverse when discrimination creates quality heterogeneity that competition compresses.

natural experiments in banking deregulation (Levine et al., 2014), firm survival (Weber and Zulehner, 2014; Pager, 2016), and online markets (Doleac and Stein, 2013). In each case, competition disciplines discrimination through selection: discriminators exit or lose market share. This paper documents an alternative channel—leveling down—in which the same providers discriminate less because they have less surplus to allocate preferentially. Competition compresses the gap from above rather than lifting outcomes from below.

The paper also relates to a growing literature measuring discrimination across multiple margins. Most empirical work targets the extensive margin, whether access is granted at all, through correspondence studies (Bertrand and Mullainathan, 2004; Kline et al., 2024), administrative records on policing (Goncalves and Mello, 2021), and ride cancellations in transportation (Ge et al., 2020). The intensive margin, how quality varies conditional on access, is understudied, in part because service quality is difficult to observe objectively (Nelson, 1970). This research gap matters for policy: interventions that equalize access (algorithmic dispatch, anti-refusal regulations) leave the quality margin unconstrained, and discrimination may migrate accordingly. I use route efficiency to document intensive-margin discrimination and show that competition narrows quality gaps through degradation. This mechanism parallels Hill and Stein (2025), who find that competition to publish first degrades research quality in science.

The paper proceeds as follows. Section 2 develops a model in which nepotistic preferences generate leveling down under competition. Section 3 describes the data and institutional setting. Sections 4–6 document the competitive shock, present the identification strategy, and report the main findings. Sections 7–8 address robustness and quantify economic magnitudes. Section 9 discusses implications and concludes.

2 Theoretical Framework

Differential treatment in the canonical Becker (1957) model reflects a *discrimination coefficient*, a fixed preference that does not respond to market conditions. Discrimination is not an equilibrium object. I adapt the nepotism framework of Goldberg (1982) to quality provision. Employers in Goldberg's model favor same-group workers in hiring; here, providers favor same-group customers through service quality. Nepotistic preferences create a quality gap that depends on the surplus available to fund preferential treatment. Competition erodes surplus. The discrimination gap becomes an *equilibrium object*: it depends on the margin, not just on the provider's preference. The central prediction is *leveling down*: competition narrows the quality gap because favored-group quality falls faster than disfavored-group quality. The gap closes not because the disfavored gain, but because the favored lose their advantage.

2.1 Setup and Preferences

A provider serves customers of two groups, favored (F) and disfavored (D), with population shares μ and $1-\mu$. The provider observes the customer's group and chooses quality $q \geq 0$.[5] All providers share the same technology: quality costs $C(q) = \frac{c}{2}q^2$ ($c > 0$), and each unit of quality earns revenue $m > 0$, so monetary profit per transaction is $mq - \frac{c}{2}q^2$. The margin m falls as competition intensifies.[6] Consumers pay a regulated fare p and value quality at rate $v > 0$ per unit, so welfare per trip is $vq - p$. Since p is fixed, welfare comparisons reduce to quality comparisons.

The provider's utility from serving a customer of group g with quality q is

$$\Pi_g(q; m) = m \cdot q - \frac{c}{2}\, q^2 + \phi \cdot \mathbf{1}[g = F] \cdot q, \tag{1}$$

[5] In the taxi setting, quality maps to route efficiency, the directness of the path taken for the passenger.

[6] In the taxi setting, m captures the profit margin per trip net of vehicle costs and opportunity cost of time. Green cab entry reduces m by lowering trip frequency and fare revenue for incumbent drivers.

where $\phi \geq 0$ is the nepotism parameter: the psychic benefit per unit of quality provided to a favored customer. When $\phi = 0$, the provider is taste-neutral. When $\phi > 0$, the provider experiences the effective marginal return to quality as $m + \phi$ for favored customers but only m for disfavored customers. Disfavored customers receive indifference, not hostility: the provider simply has no psychic reason to deviate from profit maximization when serving them. A nepotistic taxi driver takes a more efficient route for a same-group passenger, a gift of effort rather than a penalty imposed on others. First-order conditions yield optimal quality for each group.

Proposition 1 (Quality choice)**.** *Optimal quality is $q_F = (m + \phi)/c$ for favored customers and $q_D = m/c$ for disfavored customers. The discrimination gap is $\Delta = \phi/c$, independent of margins.*

Proofs appear in Appendix A. The gap ϕ/c does not depend on m as long as the provider can afford the preferred quality level. The next subsection shows what happens when the provider can no longer afford the preferred level.[7]

2.2 Competition and Leveling Down

Favored-customer quality $q_F = (m + \phi)/c$ is costly to provide: the provider earns less than the profit-maximizing level on those transactions. The generosity is sustainable only when margins are high enough to cover the loss. The zero-profit constraint on each transaction requires

$$m \cdot q - \frac{c}{2} q^2 \geq 0 \quad \Longrightarrow \quad q \leq \frac{2m}{c}.$$

This cap tightens linearly as m falls. The cap binds for favored customers when $\phi > m$, that is, when the psychic benefit exceeds the margin and the provider wants to be more generous than the budget allows.

[7] Under Becker (1957), the taste term enters with the opposite sign for the disfavored group: $\Pi_D^B = mq - \frac{c}{2}q^2 - d \cdot q$, yielding $q_D^B = (m - d)/c$ and a gap of d/c. Setting $d = \phi$ makes the two models observationally equivalent at any fixed m. The critical difference is how the gap responds to changes in m, derived below.

Proposition 2 (Leveling down)**.** *When the budget constraint binds ($\phi > m$), constrained quality is $q_F^* = 2m/c$ and $q_D^* = m/c$. The discrimination gap $\Delta^* = q_F^* - q_D^* = m/c$ is strictly increasing in m, with $\Delta^* \to 0$ as $m \to 0$.*

The result is leveling down: the gap narrows because favored customers lose their advantage, not because disfavored customers gain. Favored-group quality falls at rate $2/c$ while disfavored-group quality falls at rate $1/c$. Equality is achieved through degradation, not improvement.

The mechanism is straightforward. Preferential treatment under nepotism is a luxury good: it requires surplus to fund. High margins let the provider sacrifice profit for the psychic reward of serving favored customers. When margins fall, the generosity is curtailed. The gap depends on the margin (m), not just on preferences (ϕ).

The Becker (animus) gap, by contrast, is d/c, a constant that does not depend on margins. Competition reduces quality for both groups at the same rate, and the gap is unchanged. The provider bears no cost from withholding effort, so there is no budget constraint to tighten.[8] The empirical finding that the gap narrows under competition is inconsistent with a pure animus mechanism. Figure 1 illustrates the distinction.

Competition also reduces aggregate consumer welfare. Total welfare under the binding constraint is $W^*(m) = vm(1+\mu)/c - p$, which is strictly increasing in m. Competition therefore narrows the discrimination gap *and* reduces welfare simultaneously, the defining tension of leveling down. Appendix A.4 characterizes the equity-efficiency tradeoff formally.

Nepotistic preferences also generate geographic sorting. A provider who derives psychic benefit from encounters with favored customers prefers locations with higher concentrations of the favored group. Competition reduces geographic sorting: as margins fall, providers can no longer afford the opportunity cost of bypassing denser areas in favor of locations with more favored customers. Geographic selectivity, like preferential treatment on the

[8]Formally: $\partial\Delta^B/\partial m = 0$ under Becker, while $\partial\Delta^G/\partial m = 1/c > 0$ under Goldberg with a binding constraint. Competition (lower m) reduces the gap under Goldberg but not under Becker. The full Becker derivation appears in Appendix A.3.

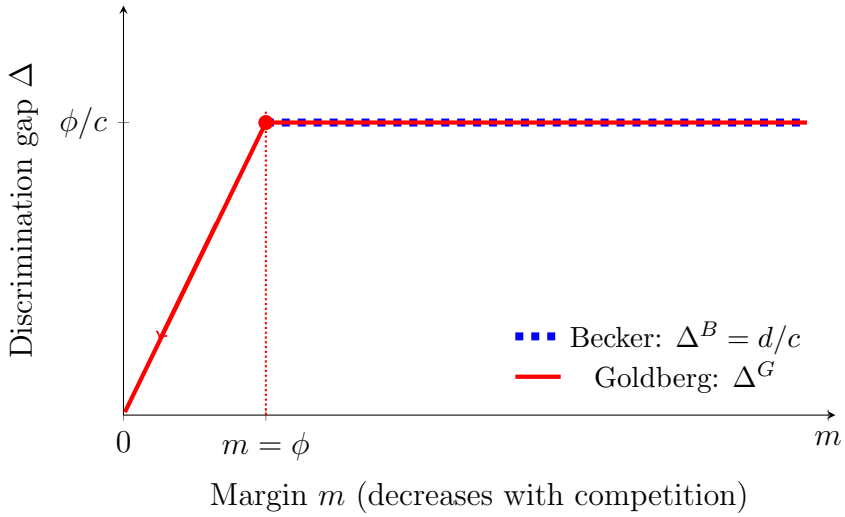


Figure 1: Discrimination gap under competition. Parameters: $c = 2$, $d = \phi = 1$. Under Becker (blue), the gap d/c is constant for all margins. Under Goldberg (red), the gap equals ϕ/c when unconstrained ($m \geq \phi$) but narrows linearly to m/c when the budget constraint binds ($m < \phi$). Discrimination is an equilibrium object under Goldberg: it responds to market structure.

intensive margin, is a luxury good. The Becker (animus) model generates the same sorting prediction: providers avoid areas with more disfavored customers. The two mechanisms are observationally equivalent in location data alone. The models diverge in *how* sorting responds to competition: under Becker, reduced sorting reflects selection (discriminators exit the market); under Goldberg, reduced sorting reflects budget constraints (same providers, tighter budgets). Appendix A.2 formalizes this result.

The model maps directly to the taxi setting: quality corresponds to route efficiency, the nepotism parameter ϕ captures the psychic benefit drivers derive from serving same-group passengers, and the margin m reflects per-trip profitability. Green cab entry provides the competition shock that lowers m. Sections 3–6 use the margin variation to test the model's predictions.

3 Data

I use the universe of 171 million New York City yellow taxi trips recorded by the Taxi and Limousine Commission (TLC) in 2013. I combine trip records with Census demographics and probabilistic driver race classifications to construct measures of racial similarity between drivers and passenger neighborhoods.

3.1 Institutional Setting and Data

New York City's taxi market operates under a medallion system: in 2013, roughly 51,000 licensed drivers operated 13,237 medallions (New York City Taxi and Limousine Commission, 2014). TLC regulations standardize vehicle specifications: model, color, signage, and interior features are uniform across all medallion cabs. Route efficiency is therefore the primary margin of within-trip quality variation.

Service quality on the intensive margin is difficult to measure. Service quality varies across transactions, and much of what constitutes good service is unverifiable (Nelson, 1970). Survey-based measures (Parasuraman et al., 1988) and online ratings (Chevalier and Mayzlin, 2006; Luca, 2016) are subjective and exhibit systematic biases by provider race and gender (Hannák et al., 2017; Teng et al., 2023). Route efficiency sidesteps these limitations. Trip records capture the fare, distance, and duration for every ride, and under metered pricing, an inefficient route directly increases the passenger's fare and travel time. Route efficiency is therefore an objective, consequential, and transaction-level measure of service quality.

The TLC tripsheet records capture every yellow taxi ride through each taxicab's payment meter and GPS system. Each observation includes pickup and dropoff times, GPS coordinates, fare amounts, trip distances, and durations. Standard-rate fares follow a uniform pricing structure: a \$2.50 base fare, a \$0.50 MTA tax, and \$0.50 per fifth of a mile or per minute below 12 mph, with time-of-day surcharges.

Two features of the airport taxi market are central to identification. First, taxis at JFK

and LaGuardia queue in designated holding areas and are dispatched in strict first-come-first-served order. Passengers cannot choose their driver, and drivers cannot refuse passengers. The queue system ensures quasi-random assignment of passengers to drivers and eliminates the selection that would otherwise confound identification. Second, the fare structure differs across airports: LaGuardia trips use metered fares, while JFK-to-Manhattan trips carry a flat $52 fare regardless of route. The fare structure variation creates differential incentives for route padding (Section 5).

3.2 Measuring Demographic Similarity

I construct probabilistic race classifications for drivers from names in the TLC license registration dataset. I apply Bayesian Improved Surname Geocoding (BISG) (Elliott et al., 2009; McCartan et al., forthcoming), which uses 2010 Census surname distributions to assign each driver probabilities of being White, Black, Asian, Hispanic, or Other. I retain the full probability distribution rather than modal assignments, as the probabilistic approach naturally accounts for classification uncertainty in the similarity measures. I match 30,645 drivers who appear in the airport trip sample; the broader green cab entry analysis (Section 4) uses all drivers regardless of race classification.[9]

Passenger race is not directly observed. I proxy for passenger race using Census block demographics at the dropoff location. Airport passengers' destinations provide a reasonable signal of their own demographics. I spatially join each trip's dropoff GPS coordinates to 2010 Census data, which provide population shares by race at the block level (with tract-level imputation where block data are missing; see Appendix B.4). All trips originate at airports, so passenger racial composition is inferred entirely from destination neighborhoods. The geographic proxy introduces measurement error that likely attenuates estimates toward

[9]The trip data capture roughly 40,000 of the approximately 51,000 licensed yellow cab drivers (New York City Taxi and Limousine Commission, 2014) during 2013. The gap reflects inactive license holders and drivers who worked exclusively outside the sample period. Appendix B.1 discusses classification details and robustness to high-confidence subsamples.

zero; significant effects are therefore conservative lower bounds (Appendix B.2).[10]

The primary explanatory variable is *own-group exposure*, which measures whether a driver is in a neighborhood where their own racial group is overrepresented relative to the citywide average:

$$\text{OwnGroupExposure}_{ik}^{\ell} = p_{i,g^*} \times (s_{k,g^*}^{\ell} - \bar{s}_{g^*}) \tag{2}$$

where $g^* = \arg\max_r p_{ir}$ is driver i's most likely racial group, p_{i,g^*} is the probability of belonging to that group, s_{k,g^*}^{ℓ} is the share of group g^* in geographic unit k at level $\ell \in \{\text{block}, \text{CDTA}\}$, and $\bar{s}_{g^*}$ is the citywide share. The citywide subtraction ensures the measure captures active sorting rather than background segregation. The measure is positive when drivers are in neighborhoods where their own group is overrepresented.

I supplement own-group exposure with two alternative similarity measures. *Expected match* computes the probability-weighted alignment between a driver's full racial distribution and the neighborhood demographic profile. The measure ranges from 0 to 1 with a natural probability interpretation. *Cosine similarity* captures the angular similarity between demographic vectors. The measure is scale-invariant and robust to extreme demographic concentrations. Both measures adapt the isolation and exposure indices from the residential segregation literature (Bell, 1954; Massey and Denton, 1988) to individual-level analysis with probabilistic race assignment. Formal definitions appear in Appendix B.9; results are robust across all three measures.

The geographic level of aggregation varies by analysis. For route efficiency, I use Census *block*-level demographics at the dropoff location. For spatial sorting, I use CDTA-level measures that capture broader neighborhood characteristics across New York City's 71 Community District Tabulation Areas (see Appendix B.6).

[10] A natural concern is that Census demographics at commercial destinations poorly reflect actual passenger demographics. I verify in Section 6 that estimates are nearly identical whether trips end in residential areas (green zone-eligible outer boroughs) or the Manhattan commercial core. The proxy therefore performs similarly regardless of destination type.

3.3 Sample and Summary Statistics

The discrimination analysis focuses on trips originating at JFK or LaGuardia during 2013. The raw airport sample contains 5.16 million trips. I exclude shift-ending trips, trips terminating at airports, and observations with missing demographic data; the analysis sample comprises 4.78 million trips. Appendix B.5 documents retention at each stage.

Table 1 summarizes the racial composition of neighborhoods and drivers. The average CDTA is 34% White, 28% Hispanic, 24% Black, and 11% Asian. The driver population differs markedly: under BISG average probabilities, drivers are predominantly Asian (42%), followed by Black (25%), White (17%), and Hispanic (13%). The gap between driver and neighborhood demographics generates the variation in racial similarity that identifies the estimates.

Table 1: Neighborhood Racial Composition and Driver Demographics

	Neighborhood		Driver	
Race/Ethnicity	Mean (%)	SD (%)	Avg. Prob. (%)	Modal (%)
White	33.9	26.7	16.7	30.0
Black	24.1	23.8	25.3	20.5
Hispanic	28.4	20.7	13.0	4.6
Asian	10.9	11.0	42.2	44.8
Other	2.7	2.3	2.8	0.1
N	72 CDTAs		32,357 drivers	

Notes: This table reports the racial and ethnic composition of New York City neighborhoods and taxi drivers. Columns 1–2 show population shares at the Community District Tabulation Area (CDTA) level from the 2010 U.S. Census. Column 3 reports the average probability of each racial category across all drivers, computed from Bayesian Improved Surname Geocoding (BISG) which assigns each driver a probability distribution over racial categories based on surname. Column 4 shows the distribution when each driver is assigned to their modal (highest probability) race. The average probability measure better reflects the underlying uncertainty in racial classification and is consistent with the probability-weighted similarity measures used in subsequent analyses.

Figure 2 displays the racial plurality across neighborhoods. Geographic segregation is substantial: Hispanic neighborhoods concentrate in the Bronx and northern Manhattan, Black neighborhoods in central Brooklyn and southeast Queens, White neighborhoods in southern Brooklyn and Staten Island, and Asian neighborhoods in Flushing and parts of

Queens.

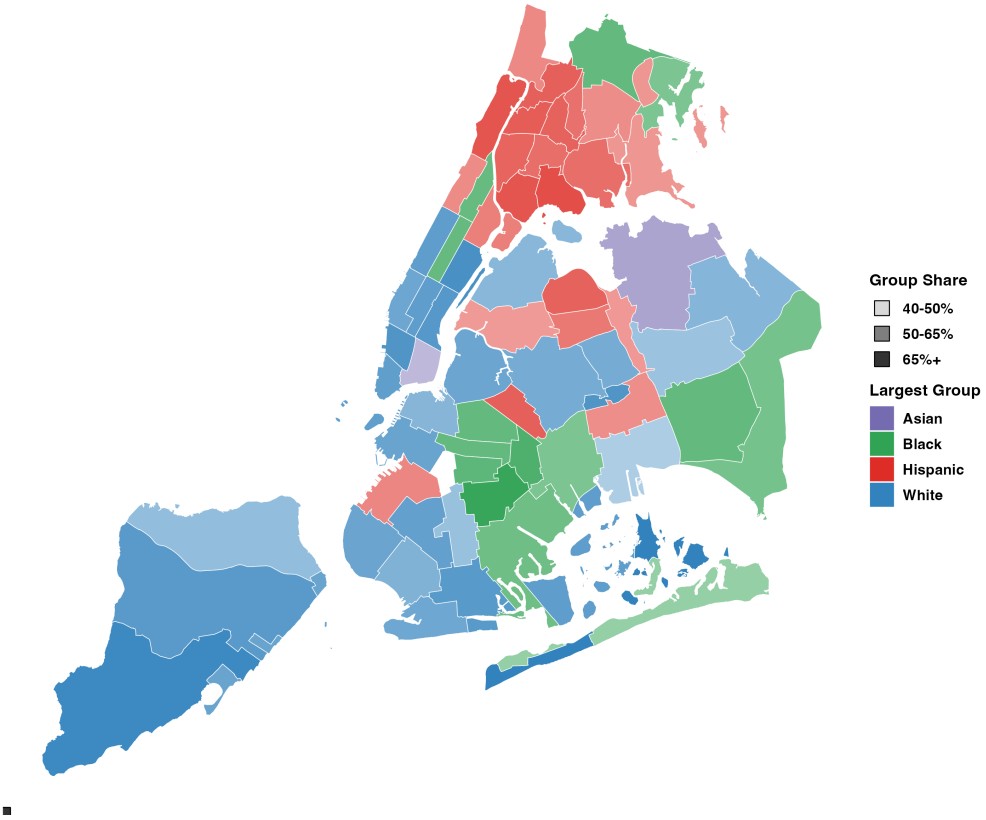


Figure 2: Racial Plurality by Neighborhood in New York City

Map shows the largest racial/ethnic group in each CDTA, with color intensity indicating the group's population share (lighter = 40–50%, darker = 65%+). Blue = White plurality, Green = Black plurality, Red = Hispanic plurality, Purple = Asian plurality. Source: 2010 U.S. Census.

Table 2 presents trip-level summary statistics by airport. LaGuardia trips average 9.7 miles and 27 minutes with metered fares averaging $30; JFK trips average 16.3 miles and 38 minutes with mean fares of $46. Approximately 60 percent of trips originate at LaGuardia. Own-group exposure averages 0.012–0.013 across geographic levels. Values close to zero indicate that drivers on average encounter neighborhood compositions similar to citywide averages. The similarity measures are nearly identical across airports: both airports serve comparable destination distributions.

Table 2: Summary Statistics: Airport Trip Sample

	Total		LaGuardia		JFK	
Variable	Mean	SD	Mean	SD	Mean	SD
Panel A: Trip Characteristics						
Trip distance (miles)	12.30	5.02	9.66	2.88	16.26	4.94
Trip duration (minutes)	31.3	14.3	26.7	10.8	38.3	15.9
Fare amount ($)	36.78	12.53	30.36	8.19	46.41	11.72
Tip amount ($)	4.34	4.53	4.22	3.87	4.54	5.37
Credit card (%)	59.5	—	65.1	—	51.1	—
Panel B: Racial Similarity Measures						
Own-group exposure (CDTA)	0.012	0.120	0.013	0.115	0.010	0.126
Own-group exposure (block)	0.013	0.154	0.014	0.149	0.011	0.162
Expected match (CDTA)	0.195	0.121	0.193	0.116	0.199	0.128
Expected match (block)	0.196	0.151	0.194	0.145	0.199	0.159
Cosine similarity (CDTA)	0.420	0.295	0.413	0.291	0.432	0.301
Cosine similarity (block)	0.400	0.310	0.396	0.306	0.405	0.316
Observations	4,783,489		2,871,062		1,912,427	

Notes: This table reports summary statistics for the airport trip analysis sample. Panel A describes trip characteristics; Panel B reports racial similarity measures between drivers and dropoff neighborhoods. Own-group exposure measures driver exposure to their own racial group relative to the citywide average, weighted by confidence in driver race assignment. Expected match is the probability that a randomly drawn resident from the dropoff neighborhood matches the driver's modal race. Cosine similarity measures the vector alignment between driver race probabilities and neighborhood racial composition. CDTA denotes Community District Tabulation Area. LaGuardia trips have metered fares; JFK trips use a flat $52 fare (plus tolls and tips). Sample includes trips originating at JFK or LaGuardia in 2013, excluding shift-ending trips and trips terminating at airports. Data from NYC TLC administrative records.

4 The Competition Shock

4.1 The Green Cab Program

The analysis uses the entry of green cabs (officially "Boro Taxis") as a shock to competition. The NYC Taxi and Limousine Commission authorized green cab operation in August 2013 (New York City Taxi and Limousine Commission, 2013), restricting pickups to the outer boroughs and northern Manhattan—above 96th Street on the East Side and 110th Street on the West Side. Yellow taxis retained exclusive street-hail rights in the Hail Exclusionary Zone (HEZ), comprising core Manhattan and airports. Figure 3 maps the demarcation.

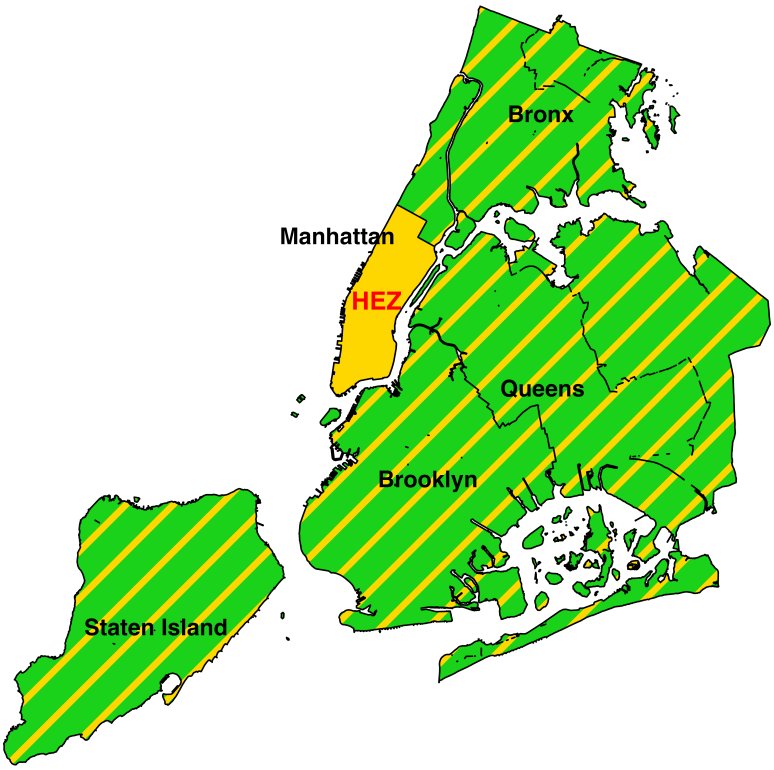


Figure 3: NYC Taxi Zones: Hail Exclusionary Zone (HEZ) Demarcation

Yellow shaded area shows the Hail Exclusionary Zone (HEZ) where only yellow taxis can pick up street hails. Light green background represents areas where green taxis can operate. Source: NYC TLC shapefile data (2013).

Despite these restrictions, green cab entry affected competition citywide. Enforcement

was imperfect; Figure 4 shows substantial green cab activity within the HEZ after entry. Outer-borough entry also induced spatial reallocation of yellow taxis to core Manhattan. Green cabs operated under identical fare structures. From the consumer's perspective, yellow and green cabs were near-perfect substitutes. Approximately 2,106 boro taxis were providing service by mid-December 2013 (NYC Taxi and Limousine Commission, 2013); monthly trip volume grew rapidly (Figure 13 in Appendix C). Green cab trip records do not include driver license information, so BISG race estimation is not possible. The discrimination analysis focuses exclusively on yellow taxi drivers.

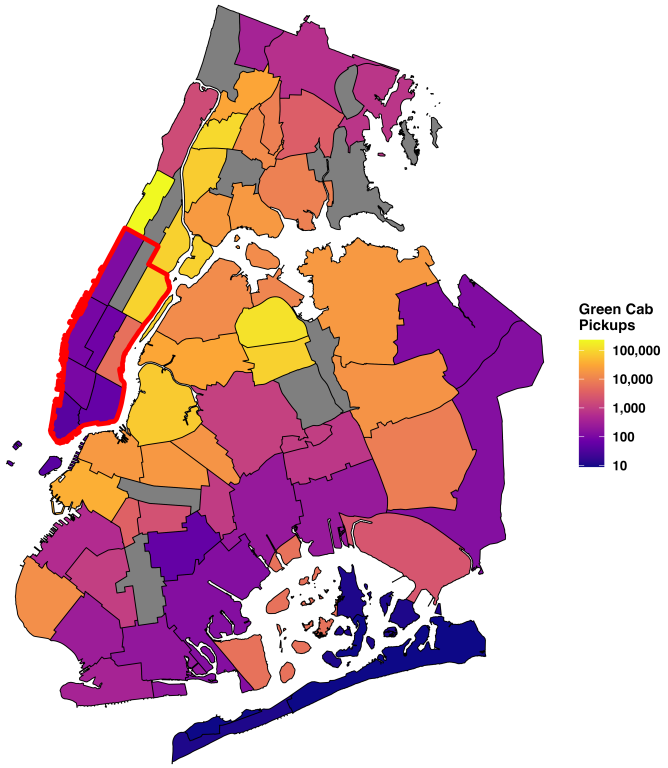


Figure 4: Green Cab Pickup Activity After Entry (August–December 2013)

Map shows green cab pickups by CDTA from August through December 2013. Red outline indicates the HEZ where green cabs were legally prohibited. Staten Island excluded due to minimal activity. Source: NYC TLC trip data.

4.2 Effects on Driver Outcomes

Table 3 reports shift-level regressions of driver outcomes on a post-entry indicator with driver and day-of-week fixed effects.[11] Green cab entry reduced shift income by $6.37 (2.4%). The decline reflects fewer available passengers: trips per shift fell 5.5% while wait time between trips rose 5.1%. Income per trip increased 3.4% despite lower trip volume. Drivers may have grown more selective or padded routes to raise metered fares under earnings pressure. Drivers also worked 0.18 fewer shifts per month and earned $169 less in total monthly income, a 3.1% decline.

The shock also induced substantial market exit. Before entry, the industry added an average of 543 drivers per month; after entry, it lost 532 per month as entries dropped 53% and exits doubled (Figure 15 in Appendix C). Among the 38,115 drivers active before August 2013, 9.2% had exited by November. Pre-entry green zone exposure predicted exit; higher earners were more likely to persist (Appendix C).

4.3 Evidence of Discrimination

I present visual and formal evidence that drivers sort toward demographically similar neighborhoods. I then show descriptive patterns in route efficiency that motivate the formal test in Section 5.

Figure 5 plots the probability of staying in the same neighborhood against own-group exposure. The positive slope indicates drivers are more likely to remain where their own racial group is overrepresented; the slope flattens after green cab entry. The figure residualizes on CDTA fixed effects to hold location fixed. The regression below adds driver fixed effects to isolate within-driver sorting preferences.

[11] The specification is $Y_{it} = \theta \cdot \text{PostEntry}_t + \alpha_i + \mu_d + \varepsilon_{it}$, where Y_{it} is the outcome for driver i on shift t, α_i are driver fixed effects, and μ_d are day-of-week fixed effects. Standard errors are clustered at the driver level.

Table 3: Effect of Green Cab Entry on Yellow Cab Driver Outcomes

	(1)	(2)	(3)
Panel A: Earnings			
	Shift Income	Hourly Wage	Income/Trip
Post entry	-6.37***	-0.41***	0.46***
	(0.13)	(0.01)	(0.01)
Mean dep. var.	269.02	31.26	13.57
Observations	7,860,971	7,860,971	7,860,971
R^2	0.366	0.302	0.450
Panel B: Work Patterns			
	Trips/Shift	Wait Time	Shift Length
Post entry	-1.18***	1.44***	-0.09***
	(0.01)	(0.02)	(0.00)
Mean dep. var.	21.59	28.38	8.71
Observations	7,860,971	7,860,971	7,860,971
R^2	0.420	0.417	0.339
Panel C: Spatial Patterns			
	Trip Distance	GZ Share	
Post entry	0.109***	-0.0015***	
	(0.003)	(0.0004)	
Mean dep. var.	3.30	0.143	
Observations	7,860,971	7,860,971	
R^2	0.453	0.320	
Panel D: Extensive Margin (Driver-Month)			
	Shifts/Month	Monthly Income	
Post entry	-0.18***	-169***	
	(0.02)	(6)	
Mean dep. var.	20.11	5,409	
Observations	390,945	390,945	
R^2	0.640	0.652	

Notes: This table reports the effect of green cab entry on yellow cab driver outcomes. All specifications include driver fixed effects; Panels A–C also include day-of-week fixed effects. Standard errors clustered at the driver level in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

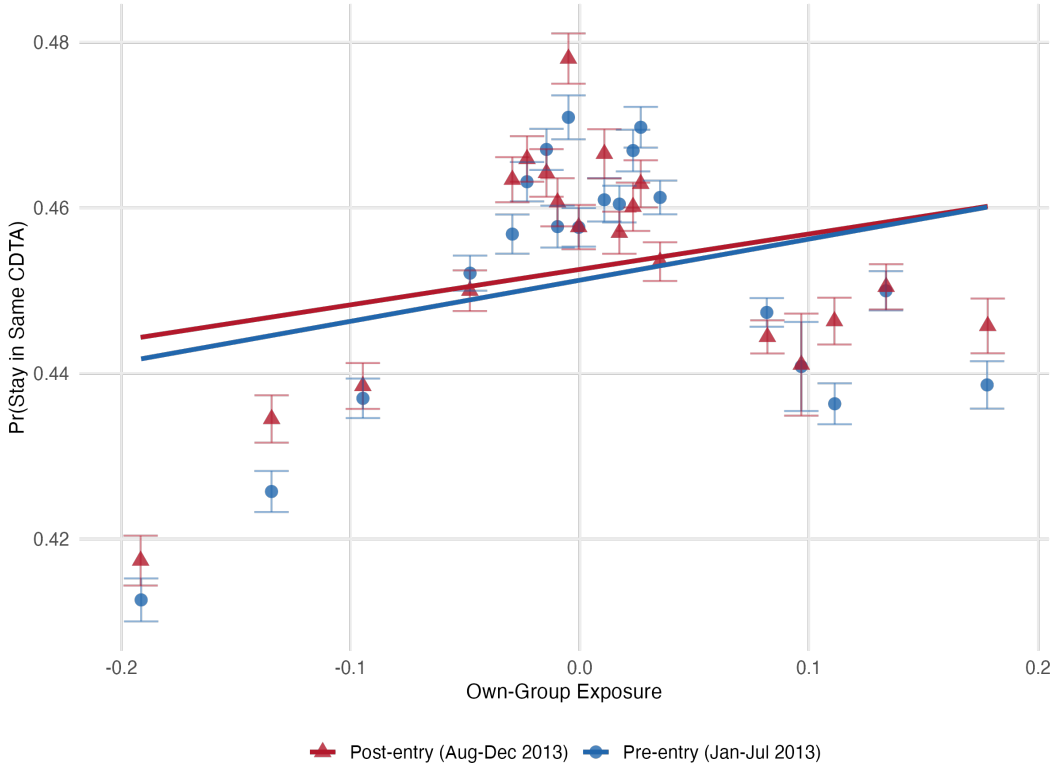


Figure 5: Spatial Sorting: Location Choice and Own-Group Exposure

Binned scatter plot of own-group exposure against the probability of staying in the same CDTA for the next pickup. Each point represents a ventile bin mean; error bars show 95% confidence intervals. Both variables are residualized on CDTA fixed effects. Own-group exposure is trimmed at the 5th and 95th percentiles.

A driver drops off an airport passenger in CDTA k and chooses whether to remain or relocate. I estimate:

$$\begin{aligned} \text{Stay}_{ikt} = \beta_1 \cdot \text{Similarity}_{ik} + \beta_2 \cdot \text{Post}_t \\ + \beta_3 \cdot (\text{Similarity}_{ik} \times \text{Post}_t) + \alpha_i + \delta_k + \lambda_t + \varepsilon_{ikt} \end{aligned} \tag{3}$$

where Similarity_{ik} measures racial similarity between driver i and CDTA k using own-group exposure, expected match, or cosine similarity (Section 3.2). Driver fixed effects (α_i) absorb time-invariant driver characteristics; CDTA fixed effects (δ_k) control for permanent neighborhood features including the distribution of destinations typically reached from that CDTA.[12]

[12]This addresses a concern that drivers might stay in a CDTA not because of its demographics but because passengers from that CDTA tend to travel to demographically favorable destinations (Buchholz, 2022). If the

Standard errors are clustered at the driver level.

Table 4 reports the results. The baseline similarity coefficient is positive and significant across all three measures (0.016–0.019): drivers systematically sort into demographically similar neighborhoods. For expected match, a shift from zero to perfect alignment increases the probability of staying by 1.9 percentage points, a 4.2% increase relative to the baseline rate of 45%. The post-entry interaction is uniformly negative (-0.004 to -0.007): competition reduced discriminatory sorting by approximately one-third.[13]

The sorting results establish that discrimination exists on the extensive margin and responds to competition. Proposition 3 predicts both patterns.

Figure 6 presents analogous descriptive evidence for the intensive margin. The figure plots average fare against own-group exposure for metered airport trips. The negative slope indicates drivers take more efficient routes for passengers from neighborhoods with higher own-group concentrations. Post-entry, the slope flattens and the intercept shifts upward: fares increased for all passengers, but the discrimination gap narrowed. The flattened slope and upward shift are consistent with leveling down: competition degrades service for everyone, and favored passengers lose more of their previous advantage.

conditional distribution of destinations given pickup location is stable over time, this trajectory is absorbed by δ_k. The identifying variation comes from comparing how different drivers respond to the same CDTA: an Asian driver and a Hispanic driver face identical expected destinations but differ in which provide high demographic similarity, analogous to gravity equations in trade.

[13]Columns (2), (4), and (6) use continuous competition measures exploiting cross-CDTA variation in yellow cab concentration changes. The interaction coefficients are strongly negative (-0.007 to -0.032), confirming the binary results.

Table 4: Location Choice: Spatial Sorting by Racial Similarity

	Own-Group Exp.		Expected Match		Cosine Sim.	
	(1)	(2)	(3)	(4)	(5)	(6)
Similarity	0.0156***	0.0132***	0.0189***	0.0169***	0.0106***	0.0097***
	(0.0044)	(0.0043)	(0.0047)	(0.0046)	(0.0025)	(0.0025)
Sim. × Post	-0.0053		-0.0068*		-0.0036**	
	(0.0038)		(0.0038)		(0.0016)	
Post	-0.0019***		-0.0006		-0.0004	
	(0.0005)		(0.0009)		(0.0008)	
Sim. × Log(GC)		-0.0003		-0.0012		-0.0007*
		(0.0010)		(0.0009)		(0.0004)
Sim. × ΔLog(YC)		-0.0317***		-0.0252***		-0.0070***
		(0.0063)		(0.0040)		(0.0014)
Log(GC)		0.0007***		0.0009***		0.0009***
		(0.0002)		(0.0002)		(0.0002)
ΔLog(YC)		0.0025***		0.0078***		0.0063***
		(0.0006)		(0.0010)		(0.0010)
Driver FE	Yes	Yes	Yes	Yes	Yes	Yes
Hour FE	Yes	Yes	Yes	Yes	Yes	Yes
Day-of-week FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	No	Yes	No	Yes	No	Yes
CDTA FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	0.450	0.450	0.450	0.450	0.450	0.450
Observations	4,782,804	4,782,804	4,782,804	4,782,804	4,782,804	4,782,804

Notes: This table reports estimates of the effect of racial similarity on driver location choice. The dependent variable is an indicator for whether the driver's next pickup occurs in the same CDTA as the current dropoff. Columns (1), (3), and (5) use a binary post-entry indicator (August 2013 onward). Columns (2), (4), and (6) use continuous CDTA-level competition measures: Log(GC) is log average daily green cab pickups; ΔLog(YC) is the change in log yellow cab pickups from pre- to post-entry. Own-group exposure is defined as $p_{d,g^*} \times (q_{c,g^*} - \pi_{g^*})$. Standard errors clustered at driver level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

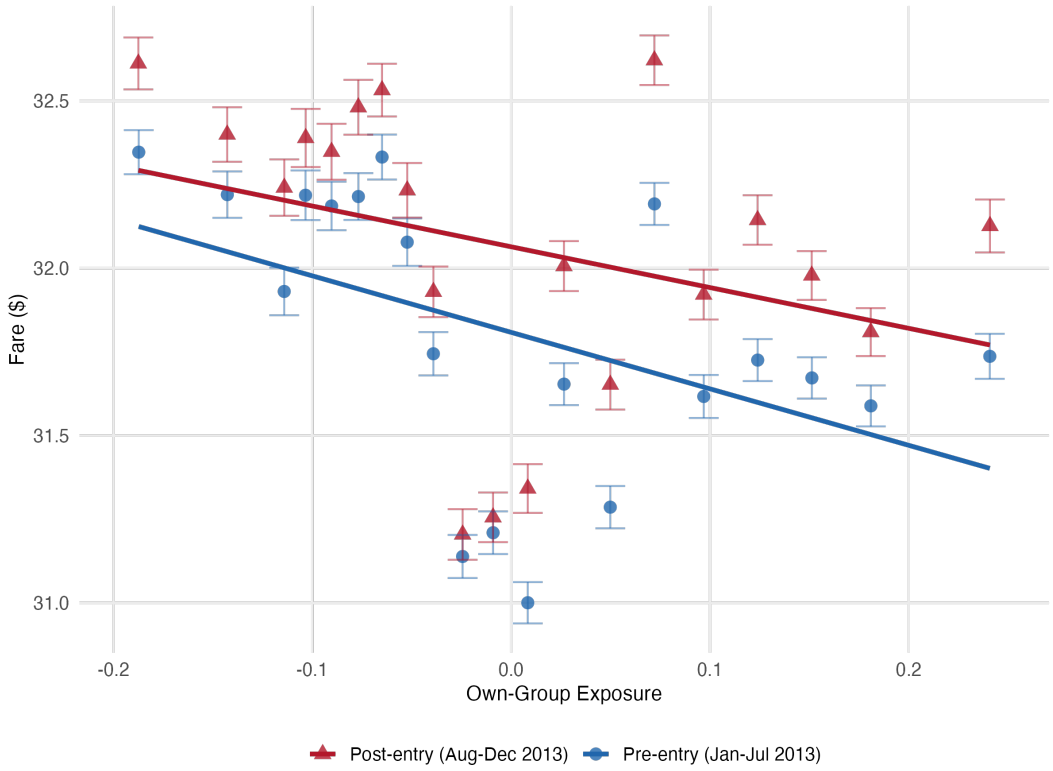


Figure 6: Route Efficiency: Fares and Own-Group Exposure

Binned scatter plot of own-group exposure against average fare for metered airport trips. Lower fares indicate more efficient routing. Own-group exposure is measured at the census block level and trimmed at the 5th and 95th percentiles. Sample excludes JFK flat-fare trips ($52) where route padding does not affect driver revenue.

The route efficiency figure hints at the mechanism, but it does not control for driver heterogeneity, destination geography, or fare structure. The next section uses fare structure variation and within-driver comparisons to isolate the leveling-down prediction.

5 Empirical Strategy

The central question is how drivers adjust when competition compresses profit margins. Does service improve for disfavored passengers (leveling up), or does service degrade for everyone, with favored passengers losing more of their previous advantage (leveling down)? Proposition 2 predicts leveling down.

I use three sources of variation in a within-driver comparison:

$$\begin{aligned} y_{ijt} = {} & \gamma_1(\text{Similarity}_{ij} \times \text{Metered}_{jt}) + \gamma_2(\text{Metered}_{jt} \times \text{Post}_t) \\ & + \gamma_3(\text{Similarity}_{ij} \times \text{Metered}_{jt} \times \text{Post}_t) \\ & + \alpha_i + \delta_j + \pi_{\text{airport}} + \lambda_{\text{hour}} + \mu_{\text{dow}} + \varepsilon_{ijt} \end{aligned} \tag{4}$$

where y_{ijt} measures route characteristics (fare, distance, or time) for a trip by driver i to destination block j at time t. Similarity_{ij} captures racial similarity between driver i and block j's demographic composition. Metered_{jt} equals one for metered-fare trips (all LaGuardia and non-flat-rate JFK trips) and zero for flat-rate JFK-to-Manhattan trips ($52 regardless of route). Post_t indicates trips after July 2013.

The fixed effects structure addresses the concern that route efficiency reflects traffic conditions or geography rather than driver behavior. Destination block fixed effects (δ_j) ensure comparisons are made among trips to the same location and hold constant the optimal route. Driver fixed effects (α_i) absorb all time-invariant characteristics (skill, local knowledge, racial attitudes), so estimates reflect within-driver behavioral responses rather than cross-sectional selection.[14] Hour-of-day and day-of-week fixed effects absorb systematic traffic variation. The flat-rate fare provides a built-in placebo: since the $52 fare is fixed regardless of route, coefficients on flat-rate trips should equal zero. Table 16 in Appendix E validates this restriction.

Green cab entry affected all yellow cab drivers simultaneously. The design does not compare treated and untreated drivers; it compares how the *same driver's* routing behavior varies across fare types and demographic matches before and after the competitive shock. The identifying assumption is that the within-driver relationship between demographic similarity and route efficiency on metered trips would not have shifted in August 2013 absent green cab entry. The institutional features, random passenger assignment and the flat-rate

[14] In 2013, 87.8% of drivers completed trips from both JFK and LaGuardia, so within-driver comparisons across fare regimes are possible. The within-person design addresses concerns about cross-sectional comparisons in discrimination research (Guryan and Charles, 2013).

placebo, substitute for the pre-trend comparisons that a standard difference-in-differences design would require.

The three coefficients map directly onto the model's predictions. γ_1 (Similarity $\times$ Metered) captures baseline discrimination: negative values mean drivers provide better service (shorter, cheaper routes) to similar passengers (Proposition 1). γ_2 (Metered $\times$ Post) captures overall quality changes after green cab entry. The key test is γ_3 (Similarity $\times$ Metered $\times$ Post): a positive coefficient combined with negative γ_1 means the gap narrowed. If $\gamma_2 > 0$ simultaneously, indicating overall quality degradation, then the gap narrowed through leveling down, as Proposition 2 predicts.

6 Results

Table 5 reports the main estimates. Each panel uses a different similarity measure (own-group exposure, expected match, cosine similarity); results are consistent across all three. Drivers provided measurably better service to demographically similar passengers before green cab entry. The baseline coefficient on Similarity $\times$ Metered is negative and significant for fares and trip time across all panels: using expected match (Panel B), similar passengers paid \$0.43 less and experienced trips 40 seconds shorter. Distance coefficients are small and mostly insignificant: discrimination operates through driving speed and route directness rather than total distance. Service quality deteriorated for all metered passengers after green cab entry. The coefficient on Metered $\times$ Post is positive across all panels: fares rose \$0.13–\$0.18 and trip times increased 0.58–0.64 minutes. Drivers padded routes as competitive pressure eroded margins.

The interaction coefficient $\hat{\gamma}_3$, the paper's central test, is positive and significant across all panels: competition compressed the discrimination gap. For expected match, $\hat{\gamma}_3$ is +\$0.29 for fares and +0.35 minutes for time. The baseline fare gap of \$0.43 narrowed to \$0.15 post-entry, a 66% reduction; the time gap of 0.67 minutes narrowed to 0.32 minutes, a 52%

Table 5: Discrimination in Route Efficiency

	Fare ($)	Distance (mi)	Time (min)
Panel A: Own-Group Exposure			
Similarity × Metered	−0.389***	-0.009	−0.438***
	(0.059)	(0.024)	(0.058)
Metered × Post	0.177***	−0.024***	0.644***
	(0.010)	(0.005)	(0.013)
Similarity × Metered × Post	0.294***	0.068***	0.354***
	(0.055)	(0.021)	(0.070)
Panel B: Expected Match			
Similarity × Metered	−0.431***	-0.014	−0.671***
	(0.064)	(0.026)	(0.066)
Metered × Post	0.125***	−0.030***	0.581***
	(0.015)	(0.007)	(0.019)
Similarity × Metered × Post	0.284***	0.035	0.349***
	(0.057)	(0.023)	(0.072)
Panel C: Cosine Similarity			
Similarity × Metered	−0.108***	0.040**	−0.321***
	(0.037)	(0.016)	(0.044)
Metered × Post	0.131***	−0.026***	0.591***
	(0.015)	(0.008)	(0.019)
Similarity × Metered × Post	0.123***	0.009	0.145***
	(0.029)	(0.012)	(0.037)
Observations	4,780,527	4,780,527	4,780,527
Driver FE	Yes	Yes	Yes
Hour/Day-of-week FE	Yes	Yes	Yes
Destination FE	Block	Block	Block

Notes: This table reports estimates from Equation (4). The sample includes all JFK and LaGuardia airport trips in 2013. Metered equals one for trips with metered fares (all LGA trips and JFK trips with fare $\neq$ \$52) and zero for JFK flat-rate trips. Post indicates trips after August 2013 (green cab entry). Panel A uses weighted own-group exposure, defined as $p_{d,g^*} \times (q_{c,g^*} - \pi_{g^*})$. Panel B uses expected match (probability-weighted racial alignment). Panel C uses cosine similarity. All similarity measures are at the block level. Flat-rate trips serve as the control group where route padding does not affect driver revenue. Standard errors clustered at driver level in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

reduction. Crucially, the gap compression occurred alongside overall quality degradation ($\gamma_2 > 0$): service worsened for everyone, but the deterioration was largest for previously favored passengers. Competition did not improve outcomes for disfavored passengers. It eroded the preferential treatment that similar passengers had enjoyed. The pattern confirms leveling down, as Proposition 2 predicts: the discrimination gap narrows because the budget constraint tightens, not because providers become less discriminatory.

6.1 Where Does the Budget Constraint Bind?

Proposition 2 predicts that the route efficiency response concentrates among drivers whose budget constraint binds most tightly after green cab entry. Two predetermined driver characteristics test this prediction: the share of pre-entry dropoffs in green cab-eligible zones, and average pre-entry shift income. Both are computed from January–July 2013 shift data, before green cab entry in August 2013.

The green zone exposure gradient is monotonic (Figure 7a). Drivers in the lowest tercile of pre-entry green zone dropoff share show $\hat{\gamma}_3 = 0.17$; the middle tercile, 0.23; the highest tercile, 0.35. A fully interacted specification confirms the gradient: the quadruple interaction Similarity $\times$ Metered $\times$ Post $\times$ High GZ Exposure yields a coefficient of 0.26 (s.e. = 0.11, $p < 0.05$). The gradient is robust across all three similarity measures (Appendix F).

Pre-entry shift income yields a complementary pattern (Figure 7b). Drivers in the top income tercile exhibit a near-zero $\hat{\gamma}_3 = 0.03$. Drivers in the lower two terciles show substantial effects (0.37 and 0.48). High-income drivers had the largest surplus buffer before entry; they remain unconstrained afterward and leave routing behavior unchanged.

Both patterns point to the same conclusion: the route efficiency response varies across drivers and concentrates where the leveling-down mechanism predicts: among drivers most exposed to the competitive shock and those operating with the thinnest margins. Splitting instead by destination type (commercial core versus residential outer boroughs) yields no meaningful difference (0.21 versus 0.19). The demographic proxy performs similarly whether

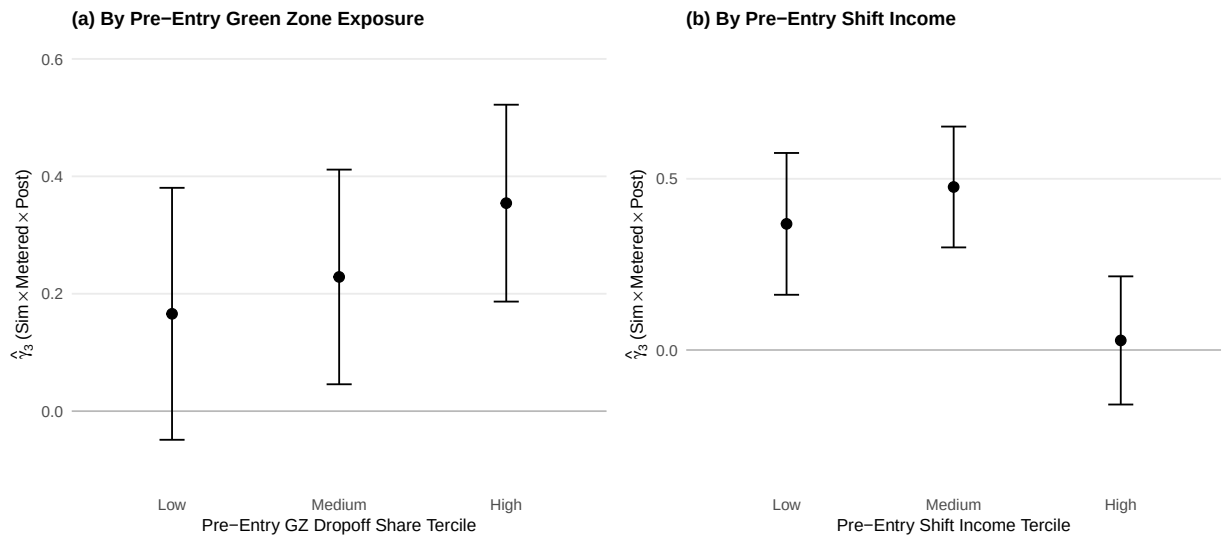


Figure 7: Heterogeneity in Route Efficiency Response

Each point reports the interaction coefficient $\hat{\gamma}_3$ (Similarity × Metered × Post) from Equation (4) estimated separately by tercile. The outcome is fare; the similarity measure is own-group exposure. Error bars show 95% confidence intervals with standard errors clustered at the driver level. Panel (a) splits drivers by pre-entry share of dropoffs in green cab-eligible zones. Panel (b) splits by pre-entry average shift income. Pre-entry characteristics are computed from January–July 2013 shift data and are predetermined relative to green cab entry in August 2013. Full results across all similarity measures and outcomes appear in Appendix F.

trips end in residential areas or the Manhattan commercial core.

7 Robustness Checks

7.1 Statistical Discrimination

Statistical discrimination predicts that drivers use neighborhood demographics as a signal for unobserved passenger characteristics and update beliefs with experience (Arrow, 1973; Phelps, 1972). Taste-based discrimination predicts stable preferences regardless of information. I interact own-group exposure with CDTA-specific experience (cumulative dropoffs in that neighborhood) and general experience (total shifts). The sample includes only new drivers, for whom experience is measured without left-censoring. CDTA experience is plausibly exogenous since passengers choose destinations. The coefficient on Own-Group × CDTA

Experience is essentially zero (0.0007) and statistically insignificant in both sorting (Table 22 in Appendix G) and route efficiency (Table 23) regressions. General experience is also insignificant (-0.0089). Both tests reject statistical discrimination in favor of taste-based preferences that persist regardless of information. The null experience result also rules out the asymmetric-learning channel in Lepage (2024), which predicts that even experience-based discrimination should attenuate as drivers accumulate trips in a neighborhood.

7.2 Differential Tipping

If demographically similar passengers tip more generously, profit-maximizing drivers would sort toward similar neighborhoods and provide better service even absent taste-based preferences. I test the tipping explanation using credit card transactions (54 percent of trips). The similarity-tip interaction is positive and significant, but the magnitude is negligible: a shift from the 25th to 75th percentile of similarity increases tips by approximately \$0.01 per trip. The \$0.01 differential cannot rationalize the \$0.43 baseline fare gap or the 1.9 percentage point sorting effect (Appendix H).

7.3 Non-Airport Trips

Airport trips may not generalize. I replicate the analysis using 19 million randomly selected non-airport trips. The baseline sorting coefficients are positive and significant across all three similarity measures (Table 25 in Appendix I), and the post-entry interactions are negative: competition attenuates sorting outside airports. For route efficiency (Table 26), the baseline similarity coefficients are negative (higher similarity predicts more efficient routes), and the post-entry interactions are positive. The pattern is consistent with leveling down. The pattern documented at airports extends to the full market.

8 Economic Magnitudes

Table 6 translates the regression estimates into aggregate dollar magnitudes. The calibration combines the airport coefficients (Table 5) with the non-airport pre-post estimates (Table 26 in Appendix I). The 2013 yellow taxi market generated 171 million trips: 4.6 million metered airport trips and 165 million non-airport. The airport analysis identifies the mechanism; the non-airport extension enables market-wide calibration. Time costs use the U.S. Department of Transportation value of $12.25 per hour for personal local travel (U.S. Department of Transportation, 2016).

Three findings emerge. First, market-wide discrimination totaled $60.8 million annually before green cab entry: $51.4 million in fare transfers and $9.4 million in time costs (765,000 hours). Per-trip gaps are larger at airports ($0.39–$0.43) than non-airport trips ($0.29–$0.30). Non-airport volume is 36 times larger, so non-airport discrimination accounts for 96 percent of the total.

Second, competition reduced discrimination by $18.5 million, a 30 percent decline. Fare discrimination fell 70 percent at airports but 25 percent market-wide; time discrimination fell 49 percent at airports and 61 percent market-wide.

Third, and central to this paper, the equality gain came at a steep cost. Total service degradation reached $42.5 million, 2.3 times the $18.5 million discrimination reduction. Competition narrowed the gap not by improving outcomes for disfavored passengers, but by eroding preferential treatment as quality declined for everyone. The leveling-down cost must be weighed against a countervailing benefit: green cab entry expanded service availability in previously underserved outer-borough neighborhoods, a welfare improvement on the access margin even as quality within the yellow cab market declined. Quantifying the access benefit requires data on wait times and unserved demand that the trip records do not contain, so the net welfare effect of green cab entry remains an open question. Because the underlying coefficients are attenuated by measurement error in the demographic proxy (Section 3.2), these calibrations represent lower bounds on the true economic costs of discrimination and

Table 6: Calibration of Economic Magnitudes: Airport vs. Non-Airport Trips

	Airport	Non-Airport	Total
Trips (millions)	4.6	165	169.6
Per-Trip Discrimination Gap (Fare)			
Pre-entry	\$0.39–\$0.43	\$0.29–\$0.30	—
Post-entry	\$0.10–\$0.15	\$0.22–\$0.24	—
Aggregate Fare Transfers			
Pre-entry	\$1.9M	\$49.5M	\$51.4M
Post-entry	\$575K	\$38.0M	\$38.6M
Reduction	\$1.3M (70%)	\$11.5M (23%)	\$12.8M (25%)
Per-Trip Discrimination Gap (Time)			
Pre-entry	0.60–0.67 min	0.25–0.26 min	—
Post-entry	0.30–0.35 min	0.10 min	—
Aggregate Time Costs			
Pre-entry (hours)	50,000	715,000	765,000
Pre-entry (\$)	\$610K	\$8.8M	\$9.4M
Post-entry (hours)	25,000	275,000	300,000
Post-entry (\$)	\$310K	\$3.4M	\$3.7M
Reduction	\$300K (49%)	\$5.4M (62%)	\$5.7M (61%)
Summary			
Total discrimination (pre-entry)	**\$2.5M**	**\$58.3M**	**\$60.8M**
Total reduction (post-entry)	**\$1.6M**	**\$16.9M**	**\$18.5M**
Leveling Down (Service Degradation)			
Per-trip fare increase	\$0.13–\$0.18	\$0.17	—
Aggregate fare extraction	\$690K	\$28.1M	\$28.8M
Per-trip time increase	0.58–0.64 min	0.39 min	—
Aggregate time cost (hours)	46,000	1,070,000	1,116,000
Aggregate time cost (\$)	\$565K	\$13.1M	\$13.7M
Total leveling down	**\$1.26M**	**\$41.2M**	**\$42.5M**

Notes: Airport calculations use coefficients from Table 5; non-airport calculations use pre-post coefficients from Table 26 (Appendix I). Time costs are valued at \$12.25 per hour, following U.S. Department of Transportation guidance (U.S. Department of Transportation, 2016), which recommends valuing personal local travel time at 50% of the median wage rate (\$24.50 in 2013). Non-airport estimates should be interpreted with some caution: unlike airports where flat-rate trips provide a control group, the non-airport Post coefficient may capture seasonal effects in addition to competitive pressure, though fixed effects absorb much of this variation.

leveling down.

9 Discussion and Conclusion

Becker's prediction that competition disciplines discrimination has shaped theory and policy for over half a century. This paper confirms that competition reduces discrimination in the NYC taxi market, but the mechanism complicates the default interpretation. The gap narrowed not because disfavored passengers received better service, but because competitive pressure eroded the preferential treatment that similar passengers had enjoyed. For every dollar of discrimination reduced, $2.30 in service quality was sacrificed.

The leveling-down result distinguishes two models of discrimination. Under Becker (1957), prejudice is a fixed preference and the quality gap does not respond to competition. Under Goldberg (1982), providers favor in-group customers using available surplus, and competition erodes that surplus. The finding that the gap narrows under competition is inconsistent with pure animus and consistent with nepotism-driven discrimination that depends on market conditions. The effect concentrates among drivers most exposed to the competitive shock and those with the lowest pre-entry earnings, reinforcing that the mechanism operates through the budget constraint.

Gap reduction alone is an incomplete measure of anti-discrimination policy. Discrimination operates across multiple margins, and constraining one may push differential treatment to another. A ride-hailing platform that implements algorithmic matching eliminates driver selection of passengers but leaves routing effort discretionary. Policies focused on access may miss discrimination that operates through quality. Evaluation requires monitoring both margins.

Four limitations deserve mention. The setting is a single regulated market; generalization requires caution. Airport trips involve drivers who choose to queue, potentially a selected sample, though the non-airport analysis (Section 7.3) yields qualitatively similar results.

The test for statistical discrimination may miss belief updating that occurs early in driver tenure. Route efficiency captures one dimension of service quality, though driver fixed effects and TLC vehicle regulations mitigate concerns about unobserved dimensions.

Leveling down may operate wherever providers retain discretion over service quality and competition compresses the surplus that funds preferential treatment. Understanding discrimination requires examining not only whether gaps close, but how they close and at what cost. Ride-hailing platforms offer a natural test: algorithmic dispatch may eliminate extensive-margin discrimination while leaving intensive-margin quality gaps intact. Competition can reduce discrimination, but the pathway matters.

References

Arrow, Kenneth J., "The Theory of Discrimination," in Orley Ashenfelter and Albert Rees, eds., *Discrimination in Labor Markets*, Princeton, NJ: Princeton University Press, 1973, pp. 3–33.

Athey, Susan, David Blei, Rachel Donnelly, Francisco J. R. Ruiz, Tobias Schmidt, and Jeffrey N. Shapiro, "Estimating Experienced Racial Segregation in US Cities Using Large-Scale GPS Data," *Proceedings of the National Academy of Sciences*, 2021, *118* (46), e2026160118.

Becker, Gary S., *The Economics of Discrimination*, Chicago, IL: University of Chicago Press, 1957.

Bell, Wendell, "A Probability Model for the Measurement of Ecological Segregation," *Social Forces*, 1954, *32* (4), 357–364.

Bertrand, Marianne and Sendhil Mullainathan, "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, 2004, *94* (4), 991–1013.

Buchholz, Nicholas, "Spatial Equilibrium, Search Frictions and Dynamic Efficiency in the Taxi Industry," *The Review of Economic Studies*, 2022, *89* (2), 556–591.

Charles, Kerwin Kofi and Jonathan Guryan, "Prejudice and Wages: An Empirical Assessment of Becker's *The Economics of Discrimination*," *Journal of Political Economy*, 2008, *116* (5), 773–809.

Chevalier, Judith A. and Dina Mayzlin, "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 2006, *43* (3), 345–354.

Doleac, Jennifer L. and Luke C.D. Stein, "The Visible Hand: Race and Online Market Outcomes," *The Economic Journal*, 2013, *123* (572), F469–F492.

Elliott, Marc N., Allen Fremont, Peter A. Morrison, Philip Pantoja, and Nicole Lurie, "A New Method for Estimating Race/Ethnicity and Associated Disparities Where Administrative Records Lack Self-Reported Race/Ethnicity," *Health Services Research*, 2009, *44* (5p1), 1622–1638.

Flegenheimer, Matt, "Cabdriver Violated Human Rights by Ignoring Black Family, Judge Rules," *The New York Times*, August 2015. https://www.nytimes.com/2015/08/07/nyregion/cabdriver-violated-human-rights-by-ignoring-black-family-judge-rules.html.

Fréchette, Guillaume R., Alessandro Lizzeri, and Tobias Salz, "Frictions in a Competitive, Regulated Market: Evidence from Taxis," *American Economic Review*, 2019, *109* (8), 2954–2992.

Ge, Yanbo, Christopher R. Knittel, Don MacKenzie, and Stephen Zoepf, "Racial and Gender Discrimination in Transportation Network Companies," *Journal of Public Economics*, 2020, *190*, 104205.

Goldberg, Matthew S., "Discrimination, Nepotism, and Long-Run Wage Differentials," *The Quarterly Journal of Economics*, 1982, *97* (2), 307–319.

Goncalves, Felipe and Steven Mello, "A Few Bad Apples? Racial Bias in Policing," *American Economic Review*, 2021, *111* (5), 1406–1441.

Guryan, Jonathan and Kerwin Kofi Charles, "Taste-Based or Statistical Discrimination: The Economics of Discrimination Returns to Its Roots," *The Economic Journal*, 2013, *123* (572), F417–F432.

Hannák, Anikó, Claudia Wagner, David Garcia, Alan Mislove, Markus Strohmaier, and Christo Wilson, "Bias in Online Freelance Marketplaces: Evidence from TaskRabbit and Fiverr," in "Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing" 2017, pp. 1914–1933.

Hill, Ryan and Carolyn Stein, "Race to the Bottom: Competition and Quality in Science," *The Quarterly Journal of Economics*, 2025, *140* (2), 1111–1185.

Kline, Patrick, Evan K. Rose, and Christopher R. Walters, "A Discrimination Report Card," *American Economic Review*, 2024, *114* (8), 2472–2525.

Lepage, Louis-Pierre, "Experience-Based Discrimination," *American Economic Journal: Applied Economics*, 2024, *16* (4), 288–321.

Levine, Ross, Alexey Levkov, and Yona Rubinstein, "Bank Deregulation and Racial Inequality in America," *Critical Finance Review*, 2014, *3* (1), 1–48.

Liu, Meng, Erik Brynjolfsson, and Jason Dowlatabadi, "Do Digital Platforms Reduce Moral Hazard? The Case of Uber and Taxis," *Management Science*, 2021, *67* (8), 4665–4685.

Luca, Michael, "Reviews, Reputation, and Revenue: The Case of Yelp.com," *Harvard Business School Working Paper*, 2016, (12-016). Forthcoming, *Management Science*.

Massey, Douglas S. and Nancy A. Denton, "The Dimensions of Residential Segregation," *Social Forces*, 1988, *67* (2), 281–315.

Matsa, David A., "Competition and Product Quality in the Supermarket Industry," *The Quarterly Journal of Economics*, 2011, *126* (3), 1539–1591.

McCartan, Cory, Robin Fisher, Jacob Goldin, Daniel E. Ho, and Kosuke Imai, "Estimating Racial Disparities When Race is Not Observed," *Journal of the American Statistical Association*, forthcoming. Available at `https://imai.fas.harvard.edu/research/birdie.html`.

Nelson, Phillip, "Information and Consumer Behavior," *Journal of Political Economy*, 1970, *78* (2), 311–329.

New York City Taxi and Limousine Commission, "Street Hail Livery ("Boro Taxi") Program," `https://www.nyc.gov/site/tlc/businesses/street-hail-livery-702.page` 2013. Accessed: 2024.

_ , "2014 TLC Factbook," `https://www.nyc.gov/assets/tlc/downloads/pdf/2014_tlc_factbook.pdf` 2014. Accessed: 2024.

_ , "TLC Launches Citywide Campaign to Combat Illegal Service Refusals," `https://www.nyc.gov/assets/tlc/downloads/pdf/press_releases/press_release_02_18_2020.pdf` February 2020. Press release, February 18.

NYC Taxi and Limousine Commission, "HAIL Market Analysis: Boro Taxi Market Study," Technical Report, New York City December 2013.

Pager, Devah, "Are Firms That Discriminate More Likely to Go Out of Business?," *Sociological Science*, 2016, *3*, 849–859.

Parasuraman, A., Valarie A. Zeithaml, and Leonard L. Berry, "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality," *Journal of Retailing*, 1988, *64* (1), 12–40.

Phelps, Edmund S., "The Statistical Theory of Racism and Sexism," *American Economic Review*, 1972, *62* (4), 659–661.

Shaked, Avner and John Sutton, "Relaxing Price Competition Through Product Differentiation," *The Review of Economic Studies*, 1982, *49* (1), 3–13.

Spence, A. Michael, "Monopoly, Quality, and Regulation," *The Bell Journal of Economics*, 1975, *6* (2), 417–429.

Tang, Johnny Jiahao, "Individual Heterogeneity and Cultural Attitudes in Credence Goods Provision," *European Economic Review*, 2020, *126*, 103442.

Teng, Fei, Tristan Botelho, and K. Sudhir, "Can Customer Ratings Be Discrimination Amplifiers? Evidence from a Gig Economy Platform," 2023. Yale School of Management Working Paper, revision requested at *Marketing Science*.

U.S. Department of Transportation, "The Value of Travel Time Savings: Departmental Guidance for Conducting Economic Evaluations, Revision 2," Technical Report, Office of the Secretary of Transportation, Washington, DC 2016. Recommends valuing personal local travel time at 50 percent of the wage rate.

Weber, Andrea and Christine Zulehner, "Competition and Gender Prejudice: Are Discriminatory Employers Doomed to Fail?," *Journal of the European Economic Association*, 2014, *12* (2), 492–521.

Appendix

A Theoretical Derivations

A.1 Proofs of Propositions 1 and 2

Proof of Proposition 1. For favored customers ($g = F$), the provider maximizes

$$\Pi_F(q; m) = (m + \phi)\, q - \frac{c}{2}\, q^2.$$

The objective is strictly concave in q. The first-order condition $(m + \phi) - c\, q = 0$ yields $q_F = (m + \phi)/c$.

For disfavored customers ($g = D$), the provider maximizes

$$\Pi_D(q; m) = m\, q - \frac{c}{2}\, q^2.$$

The first-order condition $m - c\, q = 0$ yields $q_D = m/c$. Both solutions are interior for $m > 0$ and $\phi \geq 0$.

The discrimination gap is $\Delta = q_F - q_D = (m + \phi)/c - m/c = \phi/c$, which is independent of m. □

Proof of Proposition 2. When $\phi > m$, the unconstrained favored quality $q_F = (m + \phi)/c$ violates the zero-profit constraint $q \leq 2m/c$. To verify: $R(q; m) - C(q) = mq - \frac{c}{2}q^2 = q(m - \frac{c}{2}q) \geq 0$ requires $q \leq 2m/c$. Since $(m + \phi)/c > 2m/c$ when $\phi > m$, the constraint binds: $q_F^* = 2m/c$. Disfavored quality $q_D^* = m/c < 2m/c$ remains unconstrained. The constrained gap is $\Delta^* = 2m/c - m/c = m/c$, which is strictly increasing in m. Differentiating, $\partial q_F^*/\partial m = 2/c$ and $\partial q_D^*/\partial m = 1/c$, confirming that favored-group quality falls faster as margins decrease. In the limit, $\Delta^* = m/c \to 0$ as $m \to 0$. □

A.2 Sorting

This section formalizes the sorting prediction discussed informally in Section 2. Consider a city with two types of location: favored-concentrated (F) and disfavored-concentrated (D), with favored customer shares $\alpha > \beta$ and trip arrival rates λ_F and λ_D respectively.

Proposition 3 (Sorting)**.** *A provider with $\phi > 0$ strictly prefers locations with higher favored-customer shares when arrival rates are equal. When the favored location has lower arrival rates ($\lambda_F < \lambda_D$), competition reduces sorting by tightening the budget constraint that makes geographic selectivity affordable.*

Proof. A provider with nepotism parameter $\phi > 0$ choosing location $j \in \{F, D\}$ earns expected utility per period

$$\bar{U}_j = \lambda_j \left[\gamma_j(\pi + \phi) + (1 - \gamma_j)\,\pi\right] = \lambda_j \left[\pi + \gamma_j\,\phi\right],$$

where γ_j is the favored customer share in location j and π denotes expected monetary profit per trip.

Sorting preference. Holding $\lambda_F = \lambda_D = \lambda$, we have $\bar{U}_F - \bar{U}_D = \lambda\,\phi\,(\alpha - \beta) > 0$ since $\alpha > \beta$ and $\phi > 0$. The provider strictly prefers the location with a higher favored share.

Competition and sorting. When locations differ in density ($\lambda_F < \lambda_D$), sorting toward F carries an opportunity cost. The provider requires minimum earnings $\bar{E}$ per period to remain active. If $\lambda_F \cdot \pi < \bar{E} \leq \lambda_D \cdot \pi$, the favored location alone cannot sustain the provider. As π falls with competition, this constraint tightens: the provider must spend more time in the high-density (but less favored) location, reducing sorting intensity. The budget constraint—not a change in preferences—drives the reduction. □

A.3 Becker (Animus) Benchmark

Under the Becker mechanism, the provider experiences disutility from effort on behalf of disfavored customers:

$$\Pi_g^B(q; m) = mq - \frac{c}{2}q^2 - d \cdot \mathbf{1}[g = D] \cdot q,$$

where $d > 0$ is the animus parameter. Maximizing over q: $q_F^B = m/c$ and $q_D^B = (m - d)/c$ (interior for $m > d$). The discrimination gap is

$$\Delta^B = q_F^B - q_D^B = \frac{d}{c},$$

which is independent of margins. Differentiating with respect to m:

$$\frac{\partial q_F^B}{\partial m} = \frac{\partial q_D^B}{\partial m} = \frac{1}{c}.$$

Both qualities fall at the same rate, so $\partial\Delta^B/\partial m = 0$. Competition does not narrow the discrimination gap under Becker.

Observational equivalence at baseline. Setting $d = \phi$, both models yield the same discrimination gap: $\Delta^B = d/c = \phi/c = \Delta^G$. The quality *levels* differ ($q_F^B = m/c \neq (m+\phi)/c = q_F^G$), but a researcher observing only the gap cannot distinguish the two mechanisms at any fixed m. The models diverge in how the gap responds to changes in m: $\partial\Delta^B/\partial m = 0$ while $\partial\Delta^G/\partial m = 1/c > 0$ when the budget constraint binds ($\phi > m$). A competition shock that lowers m is therefore informative about the underlying mechanism.

A.4 Welfare Comparison: Goldberg vs. Becker

Under Becker, consumer welfare is

$$W^B(m) = v\left[\mu \cdot \frac{m}{c} + (1 - \mu) \cdot \frac{m - d}{c}\right] - p = \frac{v}{c}\left[m - (1 - \mu)d\right] - p.$$

Under Goldberg with a binding budget constraint ($\phi > m$):

$$W^G(m) = v\left[\mu \cdot \frac{2m}{c} + (1-\mu) \cdot \frac{m}{c}\right] - p = \frac{vm(1+\mu)}{c} - p.$$

The welfare derivatives are $\partial W^B/\partial m = v/c$ and $\partial W^G/\partial m = v(1+\mu)/c$. Since $\mu > 0$:

$$\frac{\partial W^G}{\partial m} = \frac{v(1+\mu)}{c} > \frac{v}{c} = \frac{\partial W^B}{\partial m}.$$

For a given margin reduction $\Delta m < 0$, the Goldberg welfare loss strictly exceeds the Becker welfare loss. Under Becker, both groups lose quality at rate $1/c$. Under Goldberg, favored customers lose at $2/c$ and disfavored at $1/c$; the additional loss for favored customers makes total degradation larger.

The *leveling-down ratio*—total quality degradation per unit of gap reduction—has a closed form. For a marginal reduction in m:

$$\rho = \frac{\mu \cdot (2/c) + (1-\mu) \cdot (1/c)}{2/c - 1/c} = \frac{(1+\mu)/c}{1/c} = 1 + \mu.$$

Since $\mu \in (0,1)$, we have $\rho \in (1,2)$: for every unit of discrimination gap reduction, between 1 and 2 units of total quality are sacrificed. Leveling down is not merely a different path to reduced discrimination—it is a costlier path than uniform quality decline under Becker, where the gap does not change at all.

Equity-efficiency frontier. Define efficiency as average quality $\bar{q} = \mu q_F + (1-\mu)q_D$ and equity as the negative of the gap $-\Delta = q_D - q_F$ (higher values indicate more equal treatment). As competition reduces m, the two models trace different paths through this space. Under Becker, efficiency falls but equity is unchanged—the path is vertical. Under Goldberg with a binding constraint, efficiency falls *and* equity increases—the path is downward-sloping. Figure 8 illustrates.

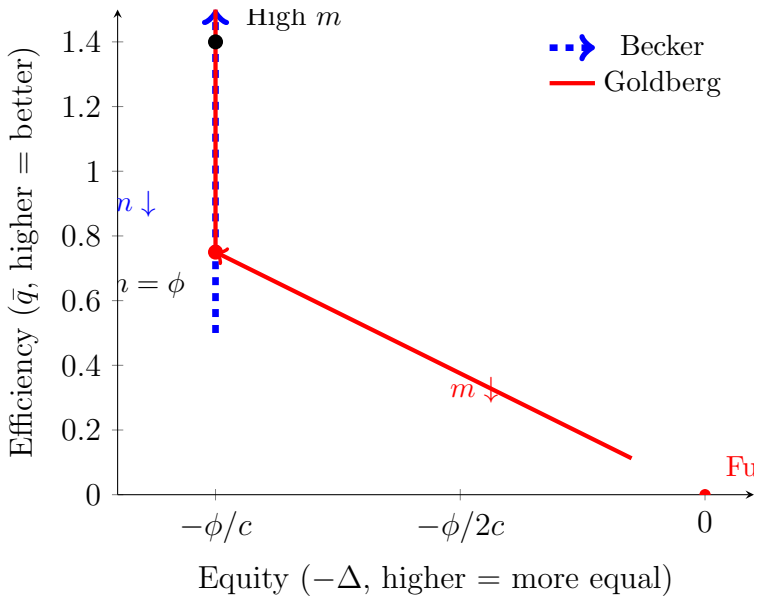


Figure 8: Equity-efficiency frontier under competition. As margins m fall, both models reduce efficiency (average quality). Under Becker (blue, dashed), equity is unchanged—the path is vertical. Under Goldberg (red, solid), the paths coincide while unconstrained; once the budget constraint binds at $m = \phi$, the Goldberg path slopes toward the origin as equity increases. Full equality ($\Delta = 0$) is reached only as $m \to 0$, at which point efficiency is also zero. Parameters: $c = 2$, $d = \phi = 1$, $\mu = 0.5$.

B Data and Measurement

This appendix discusses data sources, measurement approaches, and limitations of the analysis.

B.1 Driver Race Classification

Driver race classification relies on probabilistic surname matching using Bayesian Improved Surname Geocoding (BISG) rather than self-reported race. While this introduces measurement error, the approach offers several advantages. First, it provides objective, replicable classifications free from survey response bias or social desirability effects. Second, BISG has been validated in prior research and shown to produce classifications that align well with self-reported race at the aggregate level (Elliott et al., 2009; McCartan et al., forthcoming). Third, the probabilistic nature of BISG naturally propagates measurement uncertainty into

the similarity measures, which incorporate the full probability distributions rather than hard classifications. More broadly, the ability to link behavior to persistent driver identifiers in NYC taxi administrative data has enabled recent work documenting stable heterogeneity in detouring behavior (Tang, 2020).

Nevertheless, individual-level misclassification remains a concern. I address misclassification through robustness checks using only drivers with high-confidence classifications (maximum race probability above 0.8). Results are qualitatively similar in the restricted sample: measurement error in driver race does not drive the main findings.

B.2 Passenger Race Inference

Passenger race is not directly observed but inferred from neighborhood demographics at pickup and dropoff locations. Neighborhood-based inference introduces several sources of measurement error. First, passengers at any given location may not be representative of the neighborhood's residential population. Business districts, tourist areas, and transportation hubs draw passengers from across the city. Second, even when passengers are local residents, their individual race may differ from neighborhood averages.

The airport setting partially mitigates the first concern. Airport passengers represent a more random sample of trips across all NYC neighborhoods than street hails, where business travelers and tourists dominate. The queue-based dispatch system further ensures that variation in trip destinations is exogenous to driver behavior.

The second source of error, within-neighborhood heterogeneity, introduces classical measurement error that likely attenuates estimates toward zero. The use of similarity measures based on continuous demographic shares (rather than binary same-race indicators) helps address this issue by incorporating information about the likelihood of racial matches. Nonetheless, estimates should be interpreted as lower bounds on the true effects of racial homophily.

B.3 Sample Selection and External Validity

The analysis focuses on airport trips, which comprise only 8 percent of all Yellow Taxi trips in 2013. This sample restriction raises questions about external validity. Airport trips differ from other taxi trips in several ways: longer average distances, higher fares, more geographically dispersed destinations, and systematic differences in passenger composition (more travelers versus local residents).

I focus on airports precisely because of the quasi-random assignment feature, which is essential for identification. However, the estimated magnitudes may not generalize to street hails, where both driver selection and passenger characteristics differ. The estimates should be interpreted as measuring discrimination in settings where drivers cannot select passengers but can influence route choices and subsequent location decisions.

B.4 Missing Data and Imputation

I address missing data through progressive geographic aggregation. When block-level demographics are unavailable, I impute using tract-level, then zone-level, then CDTA-level demographics. The hierarchical approach preserves as much geographic granularity as possible while minimizing data loss.

Table 7 documents the imputation procedure. New York City contains 38,683 census blocks; 30,131 (77.9%) have complete demographic data at the block level. The remaining blocks, typically industrial areas, parks, or transportation infrastructure with zero or very small populations, require imputation from higher geographic levels. Tract-level imputation resolves 7,991 blocks and brings coverage to 98.6%. Zone-level and CDTA-level imputation address an additional 419 blocks for final coverage of 99.6%. Only 142 blocks (0.4%) remain without demographic data after all imputation steps; trips to these blocks are excluded from analyses requiring block-level similarity measures.

The main specifications use block-level similarity measures where available and CDTA-level measures for robustness. Results are consistent across geographic aggregation levels:

Table 7: Census Block Demographic Imputation Procedure

Imputation Stage	Missing	Imputed	Covered	Coverage (%)
Original block-level data	8,552	—	30,131	77.9
After tract-level imputation	561	7,991	38,122	98.6
After zone-level imputation	222	339	38,461	99.4
After CDTA-level imputation	142	80	38,541	99.6
Final sample	142	8,410	38,541	99.6

Notes: This table documents the hierarchical imputation procedure for census block demographics. "Missing" indicates blocks without demographic data at each stage. "Imputed" shows blocks resolved at each imputation level. "Covered" is the cumulative count of blocks with demographic data. Total census blocks in NYC = 38,683. Missing blocks typically have zero population (industrial areas, parks, transportation infrastructure). Imputation proceeds hierarchically: blocks missing data are assigned demographics from their containing census tract; if tract data are unavailable, zone-level data are used; finally, CDTA-level data serve as the last resort. Source: 2010 U.S. Census.

measurement error from imputation does not substantively affect the conclusions.

B.5 Sample Construction

Table 8 documents the sample construction process and data retention at each stage.

Table 8: Sample Construction

Step	Observations	Dropped	Retained (%)
1. TLC trip data (2013, spatially merged)	171,035,939	—	100.0
2. Merge driver demographics (BISG)	147,381,417	23,654,522	86.2
3. Drop missing similarity measures	144,508,459	2,872,958	84.5
4. Airport pickups (excl. airport dropoffs)	5,159,821	139,348,638	3.0

Notes: This table documents the sample construction process for the analysis. Step 1 begins with the full 2013 NYC Taxi and Limousine Commission trip records after merging with census geographic boundaries. Step 2 merges Bayesian Improved Surname Geocoding (BISG) driver race probabilities based on driver license records. Step 3 drops trips with missing racial similarity measures at pickup or dropoff locations. Step 4 restricts to trips originating at JFK or LaGuardia airports, excluding trips that also terminate at airports.

B.6 Geographic Units

Figure 9 shows the demarcations of all 71 Community District Tabulation Areas (CDTAs) across New York City's five boroughs, including Staten Island.



Figure 9: Community District Tabulation Areas (CDTAs) in New York City

Map shows all 71 Community District Tabulation Areas used as geographic units for analyzing driver location choices. CDTAs aggregate 2020 census tracts to approximate community districts. Source: NYC Department of City Planning.

B.7 Green and Yellow Taxi Comparison

Table 9 compares trip characteristics between green and yellow taxis in 2013.

B.8 Airport Trip Coverage

Figure 10 shows the geographic distribution of the airport trip sample across New York City neighborhoods.

Table 9: Comparison of Yellow and Green Taxi Trips

Variable	Yellow Taxi	Green Taxi
Panel A: Trip Characteristics		
Trip distance (miles)	7.4	3.0
Trip duration (minutes)	12.9	13.3
Fare amount ($)	11.60	12.40
Total amount ($)	13.90	14.30
Panel B: Payment		
Tip amount ($)	1.30	0.90
Tolls amount ($)	0.20	0.10
Surcharge ($)	0.30	0.40
Credit card (%)	54.0	32.6
Panel C: Sample Size		
Total trips	144,508,459	1,205,720

Notes: This table compares trip characteristics between yellow and green taxis in 2013. Green taxis (Boro Taxis) began operations in August 2013 with street-hail rights restricted to northern Manhattan (above 96th/110th Streets) and the outer boroughs. Yellow taxis operate citywide but face competition from green taxis in the designated green zone. Trip duration converted from seconds to minutes. All monetary values in 2013 dollars. Data from NYC Taxi and Limousine Commission administrative records.

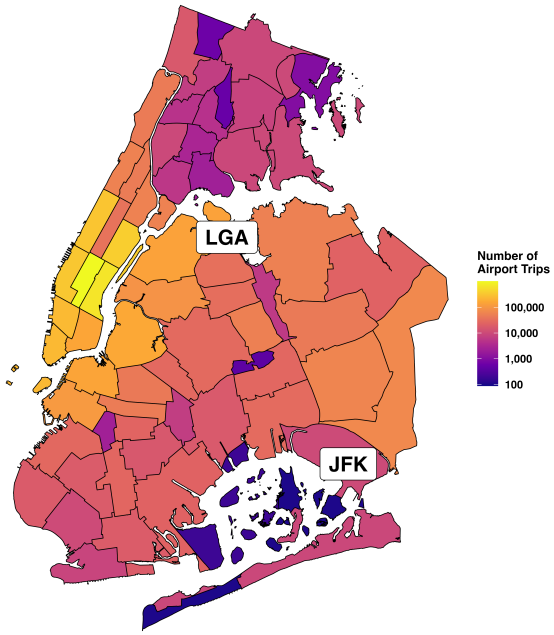


Figure 10: Geographic Coverage of Airport Trip Sample

Map shows airport trips (pickups at JFK or LGA) by destination CDTA in 2013. Manhattan receives the highest volume; coverage declines in the outer boroughs. Staten Island excluded due to minimal taxi activity.

B.9 Similarity Measure Definitions

This appendix provides formal definitions for the expected match and cosine similarity measures used alongside own-group exposure (defined in Section 3.2).

Expected Match. This measure computes the probability-weighted alignment between driver and neighborhood demographics:

$$\text{ExpectedMatch}_{ik}^{\ell} = \sum_{r \in \{W,B,A,H,O\}} p_{ir} \times s_{kr}^{\ell} \tag{5}$$

where p_{ir} is driver i's probability of being race r and s_{kr}^{ℓ} is the population share of race r in geographic unit k at level ℓ. The measure ranges from 0 to 1 and has a natural probability interpretation: it equals the probability that a randomly drawn resident of unit k matches driver i's race. Expected match is analogous to the classic isolation index from the residential segregation literature (Bell, 1954; Massey and Denton, 1988). Athey et al. (2021) recently extended this approach to measure "experienced segregation" using GPS data on where individuals actually spend time.

Cosine Similarity. This measure captures the angular similarity between the driver's racial probability vector and the location's demographic composition vector:

$$\text{CosineSimilarity}_{ik}^{\ell} = \frac{\sum_r p_{ir} \times s_{kr}^{\ell}}{\sqrt{\sum_r p_{ir}^2} \times \sqrt{\sum_r (s_{kr}^{\ell})^2}} \tag{6}$$

This metric equals 1 when the two vectors point in the same direction (perfect demographic alignment) and approaches 0 for orthogonal distributions. Cosine similarity is scale-invariant, making it robust to locations with extreme demographic concentrations where expected match might be mechanically high or low.

Distributional Properties. The raw similarity measures exhibit right-skewness (see Figures 11–12 below). To address potential concerns about outliers, I also construct z-score standardized versions computed within each geographic unit, measuring how much a driver's

similarity deviates from the average of all drivers visiting that location. Results using both raw measures and z-scores appear in the robustness tables.

B.10 Similarity Measure Distributions

Figures 11 and 12 show the empirical distributions of the three similarity measures defined in Appendix B.9 and their z-score transformations.

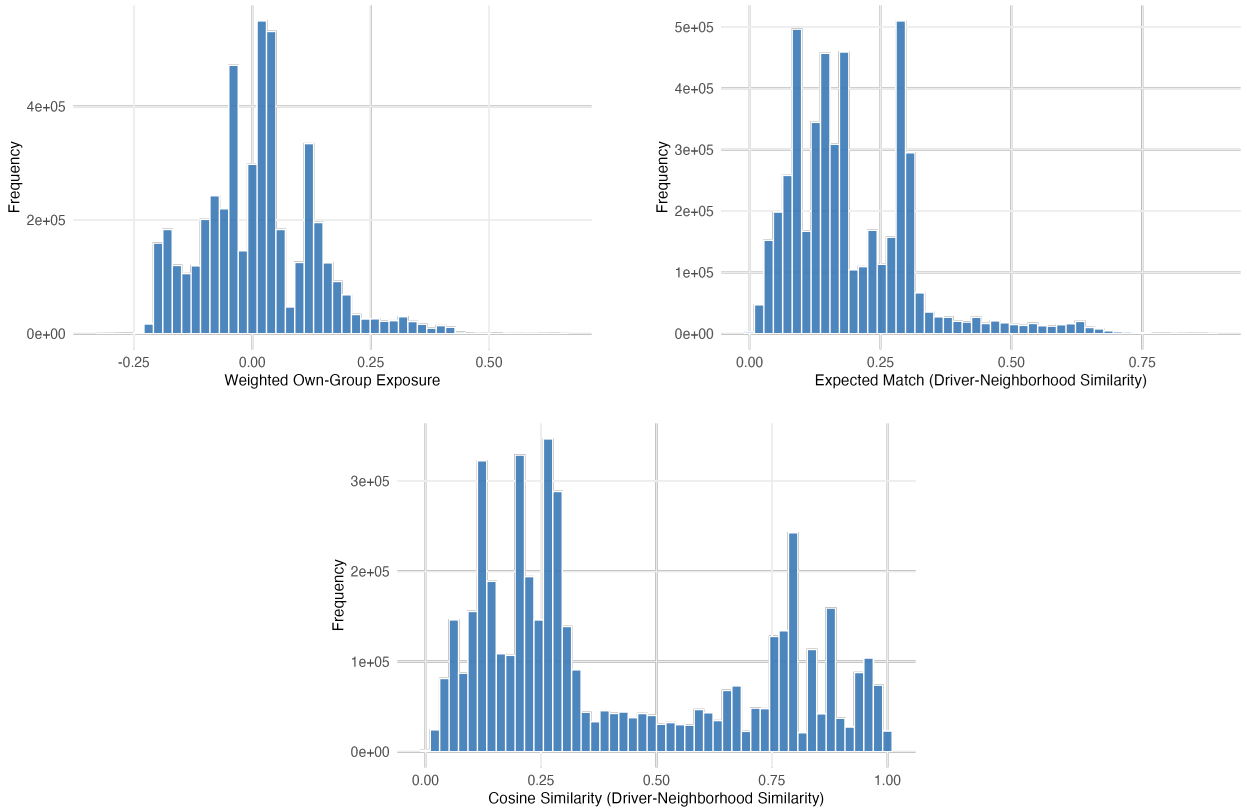


Figure 11: Distributions of Raw Similarity Measures

Histograms of the three similarity measures across all trips in 2013. Top left: weighted own-group exposure. Top right: expected match (probability that a random resident matches driver race; ranges 0–1). Bottom: cosine similarity (angular alignment between driver and neighborhood demographic vectors).

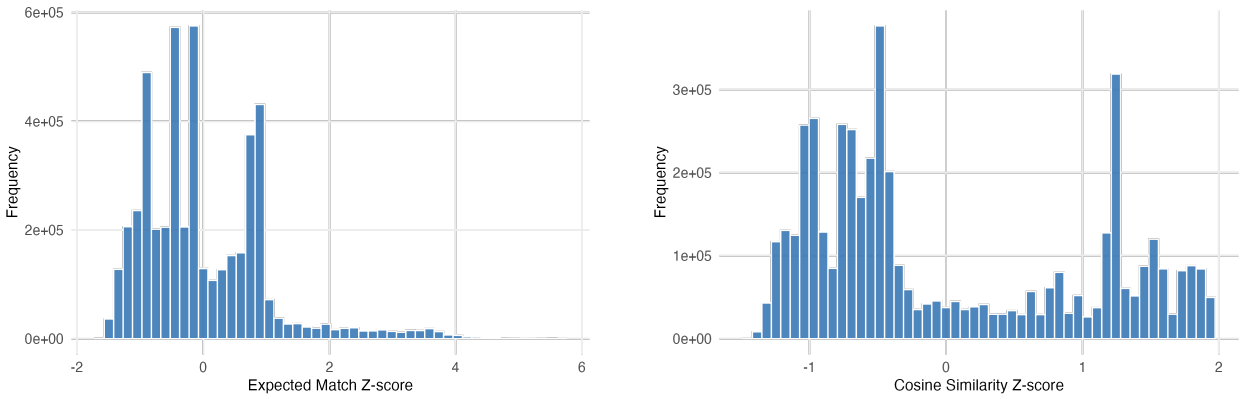


Figure 12: Distributions of Z-Score Similarity Measures

Histograms of z-score standardized similarity measures across all trips in 2013. Z-scores are computed within each CDTA, measuring how much a driver's similarity with a destination deviates from the CDTA average. Left: expected match z-score. Right: cosine similarity z-score.

C Green Cab Entry: Supplementary Evidence

This appendix supplements the green cab entry analysis in Section 4.

C.1 Market Changes

Figure 13 shows the rapid growth of green cab trip volume during the entry period. Table 10 compares yellow cab shift characteristics before and after entry. Figure 14 maps spatial changes in taxi activity.

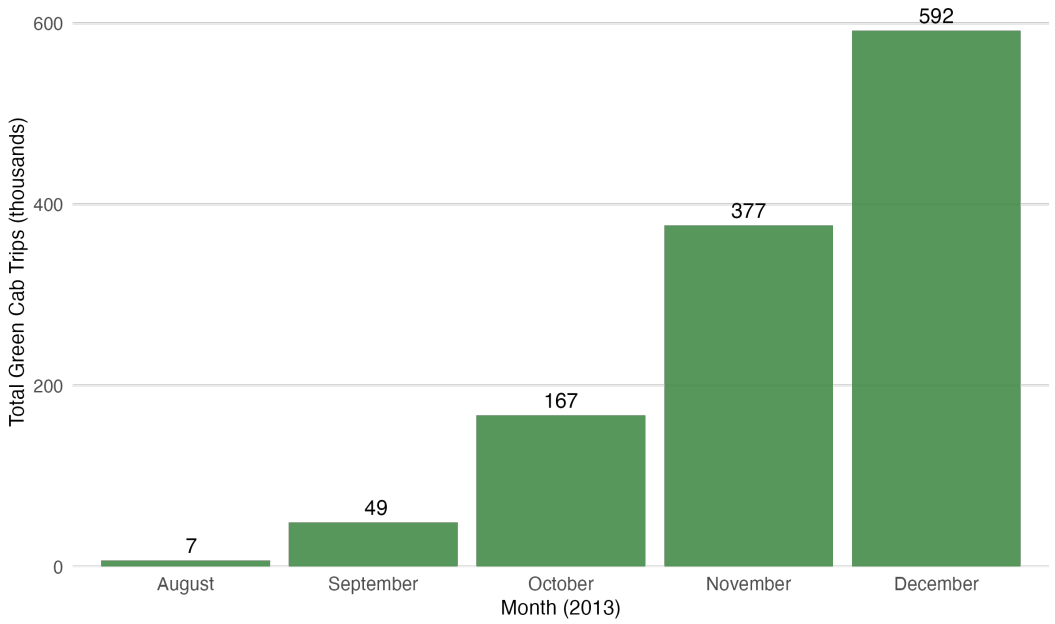


Figure 13: Green Cab Trip Volume During Entry Period (August–December 2013)

Monthly total green cab trips from August through December 2013. The rapid growth reflects both fleet expansion and increasing utilization as the green cab service became established. Source: NYC TLC green cab trip data.

Table 10: Summary Statistics: Yellow Cab Shifts Before and After Green Cab Entry

	Mean			
	Pre-Entry (Jan–Jul 2013)	Post-Entry (Aug–Dec 2013)	Difference	% Change
Panel A: Earnings				
Shift income ($)	271.26	265.88	-5.38	-2.0
Hourly wage ($/hour)	31.47	30.97	-0.50	-1.6
Income per trip ($)	13.39	13.83	0.44	3.3
Panel B: Work Patterns				
Shift length (hours)	8.73	8.69	-0.04	-0.4
Trips per shift	22.05	20.94	-1.11	-5.0
Wait time between trips (min)	27.79	29.21	1.42	5.1
Panel C: Spatial Patterns				
Average trip distance (miles)	3.26	3.35	0.10	2.9
Green zone pickup share	0.144	0.141	-0.0028	-1.9
Panel D: Sample Size				
Number of shifts	4,590,780	3,270,191		
Number of unique drivers	38,115	37,072		

Notes: This table reports mean shift-level characteristics for New York City yellow taxi drivers before and after green cab entry. Pre-entry period covers January 1 through July 31, 2013; post-entry period covers August 1 through December 31, 2013. Green cabs were authorized to begin operations on August 8, 2013.

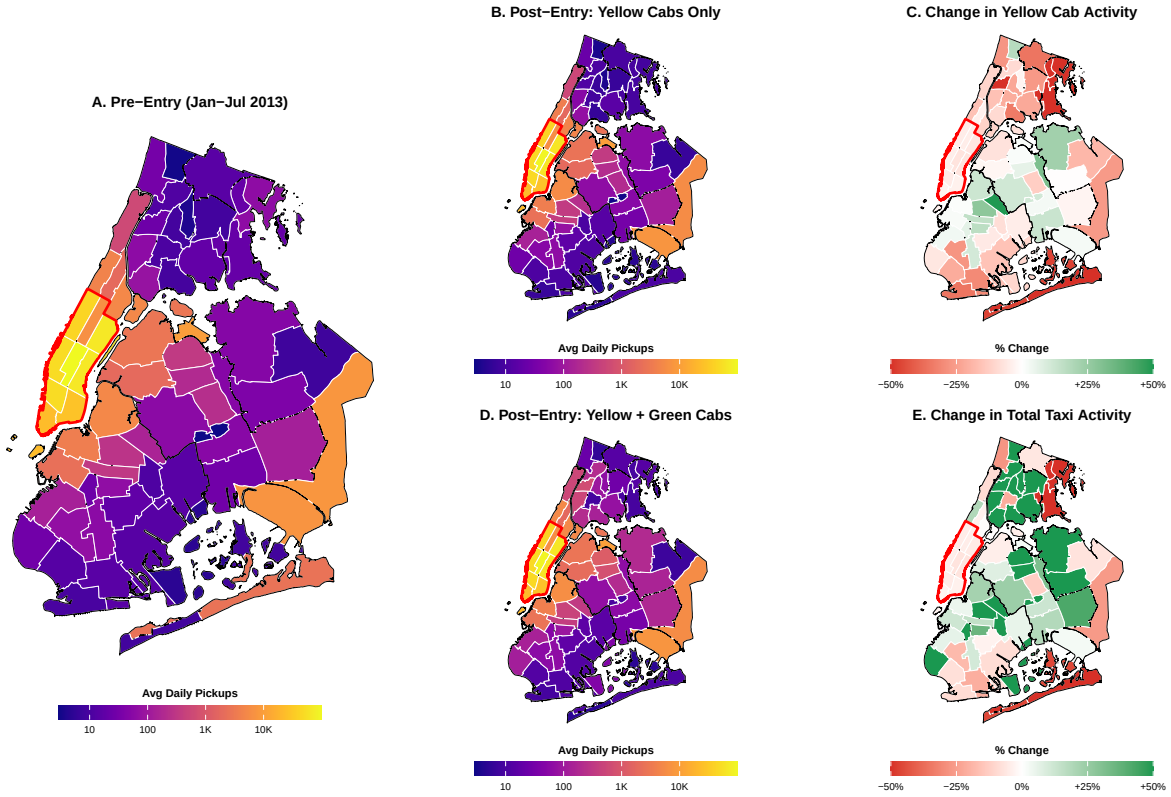


Figure 14: Taxi Activity Before and After Green Cab Entry

Panel A shows average daily taxi pickups by CDTA before green cab entry (January–July 2013). Top row: Yellow cab activity only. Panel B shows post-entry yellow cab pickups (August–December 2013); Panel C shows percentage change in yellow cab activity. Bottom row: Total taxi activity (yellow + green cabs combined). Panel D shows post-entry total pickups; Panel E shows percentage change in total taxi activity. Red indicates decline; green indicates increase. The Hail Exclusionary Zone (HEZ) boundary is outlined in red. Staten Island is excluded due to minimal taxi activity. Yellow cab activity declined (median −5.6%), but total taxi activity increased (median +6.0%) as green cabs expanded service in outer boroughs.

C.2 Incumbent Driver Turnover and Exit

Figure 15 shows monthly driver entry and exit patterns.

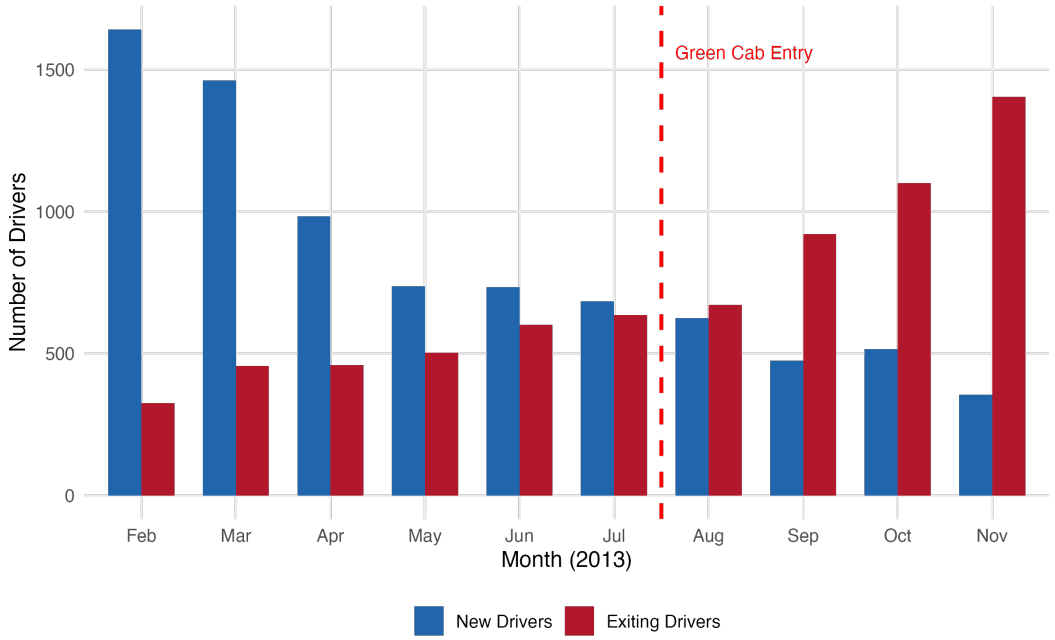


Figure 15: Driver Entry and Exit Before and After Green Cab Entry

Monthly counts of new drivers entering (first observed shift in that month) and drivers exiting (last observed shift in that month, before December) the yellow cab industry, February–November 2013. The dashed vertical line marks green cab entry in August 2013. Before entry, the industry added an average of 543 drivers per month; after entry, it lost an average of 532 drivers per month.

Table 11 presents multinomial logit results for exit timing as a function of pre-entry driver characteristics.

Table 12 presents a complementary Cox proportional hazard model of driver exit following green cab entry.

The Cox model confirms the key findings from the multinomial logit. All regressors are standardized to have mean zero and unit variance, so hazard ratios represent the effect of a one standard deviation increase. Pre-entry green zone exposure strongly predicts exit: a one standard deviation increase in green zone pickup share increases the exit hazard by 60% (HR $= 1.61$). The interaction between log income and green zone share is negative (HR $= 0.62$): the green zone exposure effect is attenuated for higher-income drivers.

Table 11: Exit Timing: Multinomial Logit Results

	Exit Month (Reference: Stayed through December)			
	August	September	October	November
Log pre-entry income	-0.92***	-0.30	0.09	0.24
	(0.21)	(0.19)	(0.17)	(0.16)
Pre-entry GZ share	-0.86	3.74**	7.23***	8.11***
	(2.06)	(1.58)	(1.39)	(1.24)
Log income × GZ share	0.06	-0.68**	-1.46***	-1.57***
	(0.40)	(0.30)	(0.27)	(0.24)
Observations	38,115			

Notes: This table reports multinomial logit estimates of exit timing. The sample includes drivers active before green cab entry (January–July 2013). The dependent variable is the month of last observed shift, with December (stayed through year-end) as the reference category. Standard errors in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 12: Cox Proportional Hazard Model: Incumbent Driver Exit

	Hazard Ratio	95% CI	p-value
Log pre-entry income	0.98	[0.93, 1.02]	0.319
Pre-entry GZ share	1.6	[1.4, 1.9]	0.0000
Log income × GZ share	0.62	[0.53, 0.72]	0.0000
Observations	38,115		
Events (exits)	3,513		

Notes: This table reports Cox proportional hazard model estimates for incumbent drivers (those active before August 2013). The outcome is time to exit (last observed shift before December 2013). All regressors are standardized (mean 0, SD 1), so hazard ratios represent the effect of a one standard deviation increase. Hazard ratios greater than 1 indicate higher exit risk. Pre-entry characteristics are calculated as averages over shifts from January–July 2013.

C.3 New Entrant Survival

I compare drivers who entered before green cab entry (February–July, $N = 6{,}239$) with those who entered after (August–November, $N = 1{,}969$). Figure 16 shows Kaplan-Meier survival curves: post-entry cohort drivers experienced faster attrition, with only 68% remaining active by month 4 versus 78% of pre-entry drivers (log-rank $\chi^2 = 84.4$).

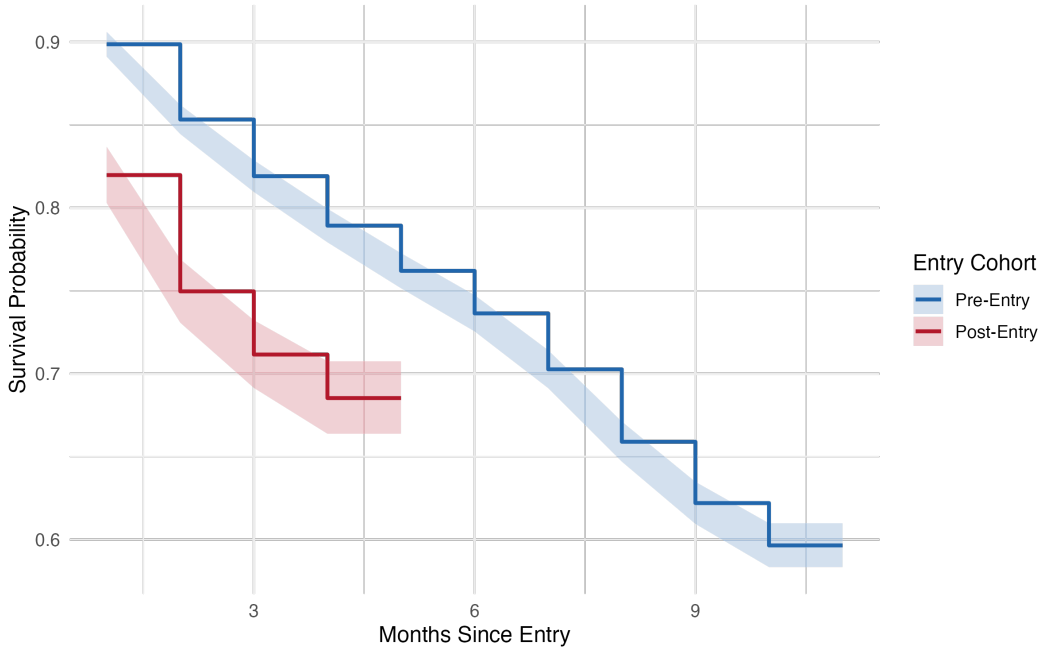


Figure 16: Kaplan-Meier Survival Curves: New Entrants by Entry Cohort

Survival curves for drivers who entered the yellow cab industry during 2013, by entry cohort. Pre-entry cohort includes drivers whose first shift occurred February–July 2013; post-entry cohort includes drivers whose first shift occurred August–November 2013. Shaded regions indicate 95% confidence intervals. Survival is defined as remaining active through December 2013.

Table 13 presents Cox proportional hazard models. Post-entry cohort drivers faced a 58% higher exit hazard (HR = 1.58, Model 1). Green zone exposure independently predicts exit (HR = 1.99, Model 2), while higher earnings strongly predict survival (HR = 0.22, Model 4). The cohort effect attenuates after controlling for income but remains significant (HR = 1.33). The result complements the incumbent exit analysis in Section 4.2.

Table 13: Cox Proportional Hazard Models: New Entrant Exit

	(1)	(2)	(3)	(4)
Post-Entry Cohort	1.580***	1.597***	1.535***	1.329***
	(0.049)	(0.049)	(0.067)	(0.050)
Avg. GZ Share		1.993***	1.872***	0.100***
		(0.138)	(0.157)	(0.126)
Post-Entry × GZ Share			1.334	
			(0.328)	
Log(Avg. Income)				0.216***
				(0.032)
Observations	8,208	8,208	8,208	8,208
Events	2,888	2,888	2,888	2,888

Notes: This table reports hazard ratios from Cox proportional hazard models for exit among drivers who entered the yellow cab industry during 2013 (February–November). Post-Entry Cohort equals one for drivers whose first shift occurred August–November (after green cab entry). GZ Share is the driver's average fraction of pickups in the green zone across all observed shifts. Standard errors in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

D Robustness: Spatial Sorting

D.1 Continuous Competition Measures

Table 14 presents location choice estimates using continuous CDTA-level competition measures instead of the binary entry indicator used in the main text. The specification includes month fixed effects for additional temporal controls while using variation in competition intensity across neighborhoods.

The key variables are: (1) Log(GC Pickups), the log of average daily green cab pickups

Table 14: Location Choice: Continuous Competition Measures

	Levels		Z-scores		Own-Group
	(1)	(2)	(3)	(4)	(5)
Similarity	0.0169***	0.0097***	0.0020***	0.0029***	0.0132***
	(0.0046)	(0.0025)	(0.0006)	(0.0007)	(0.0043)
Sim. × Log(GC)	-0.0012	-0.0007*	-0.0001	-0.0002*	-0.0003
	(0.0009)	(0.0004)	(0.0001)	(0.0001)	(0.0010)
Sim. × ΔLog(YC)	-0.0252***	-0.0070***	-0.0030***	-0.0021***	-0.0317***
	(0.0040)	(0.0014)	(0.0005)	(0.0004)	(0.0063)
Log(GC)	0.0009***	0.0009***	0.0007***	0.0007***	0.0007***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
ΔLog(YC)	0.0078***	0.0063***	0.0028***	0.0034***	0.0025***
	(0.0010)	(0.0010)	(0.0006)	(0.0007)	(0.0006)
Similarity Measure	Exp. Match	Cos. Sim	Exp. Match	Cos. Sim	Own-Grp
Driver FE	Yes	Yes	Yes	Yes	Yes
Hour FE	Yes	Yes	Yes	Yes	Yes
Day-of-week FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Airport indicator	Yes	Yes	Yes	Yes	Yes
CDTA FE	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	0.450	0.450	0.450	0.450	0.450
Observations	4,782,804	4,782,804	4,782,804	4,782,804	4,782,804

Notes: This table reports estimates using continuous CDTA-level competition measures, which allows for month fixed effects. Log(GC) is the log of average daily green cab pickups in the CDTA during the post-entry period (August–December 2013). ΔLog(YC) is the change in log average daily yellow cab pickups from pre-entry (January–July) to post-entry (August–December). CDTAs with no green cab activity have Log(GC) = log(1) = 0. The interaction terms capture how competition intensity moderates the relationship between racial similarity and location choice. All specifications include driver, hour, day-of-week, month, CDTA, and airport fixed effects. Standard errors clustered at the driver level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

in each CDTA during the post-entry period (August–December 2013), and (2) ΔLog(YC), the change in log average daily yellow cab pickups from the pre-entry to post-entry period. CDTAs with no green cab activity have Log(GC Pickups) = log(1) = 0.

The interaction terms (Similarity $\times$ Log(GC Pickups) and Similarity $\times$ ΔLog(YC)) capture how competition intensity moderates the relationship between racial similarity and location choice. The baseline relationship between similarity and staying probability is attenuated in neighborhoods with greater competitive pressure, consistent with the main results, whether measured by higher green cab penetration or by greater yellow cab concentration due to spatial reallocation (higher ΔLog(YC)).

Month fixed effects control for citywide seasonal patterns; competition effects are identified through cross-CDTA variation. Results are qualitatively similar to the binary entry specification: competition reduces discriminatory spatial sorting regardless of how competition is measured.

D.2 Z-Score Similarity Measures

Table 15 presents location choice estimates using Z-score standardized similarity measures. Z-scores are computed within each CDTA, measuring how much a driver's similarity with a destination deviates from the CDTA average. This specification tests whether drivers respond to relative similarity (how they compare to other drivers visiting the same CDTA) rather than absolute similarity levels.

E Robustness: Route Efficiency

E.1 Constrained vs. Unconstrained Specifications

Table 16 compares unconstrained and constrained specifications. The constrained specification excludes coefficients that should theoretically equal zero for the fare outcome (since flat-rate JFK trips have a fixed $52 fare regardless of route).

Table 15: Location Choice: Z-Score Similarity Measures

	Expected Match Z-score		Cosine Similarity Z-score	
	(1)	(2)	(3)	(4)
Similarity	0.0023***	0.0020***	0.0031***	0.0029***
	(0.0006)	(0.0006)	(0.0007)	(0.0007)
Similarity × Post	-0.0008*		-0.0011**	
	(0.0005)		(0.0005)	
Post	-0.0019***		-0.0019***	
	(0.0005)		(0.0005)	
Similarity × Log(GC)		-0.0001		-0.0002*
		(0.0001)		(0.0001)
Similarity × ΔLog(YC)		-0.0030***		-0.0021***
		(0.0005)		(0.0004)
Log(GC)		0.0007***		0.0007***
		(0.0002)		(0.0002)
ΔLog(YC)		0.0028***		0.0034***
		(0.0006)		(0.0007)
Driver FE	Yes	Yes	Yes	Yes
Hour FE	Yes	Yes	Yes	Yes
Day-of-week FE	Yes	Yes	Yes	Yes
Month FE	No	Yes	No	Yes
CDTA FE	Yes	Yes	Yes	Yes
Mean dep. var.	0.450	0.450	0.450	0.450
Observations	4,782,804	4,782,804	4,782,804	4,782,804

Notes: This table reports estimates using Z-score standardized similarity measures. Z-scores are computed within each CDTA, measuring how much a driver's similarity with a destination deviates from the CDTA average. Columns (1) and (3) use binary post-entry indicators. Columns (2) and (4) use continuous competition measures with month fixed effects. Standard errors clustered at driver level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 16: Route Efficiency: Constrained vs Unconstrained Specifications

	Unconstrained			Constrained		
	Fare	Dist.	Time	Fare	Dist.	Time
Panel A: Flat-Rate Baseline (should be ≈ 0 for fare)						
Expected Match	-0.603^{***}	-0.434^{***}	-0.579^{***}	—	—	—
	(0.072)	(0.045)	(0.159)			
Post	-0.048^{***}	0.121^{***}	0.996^{***}	—	—	—
	(0.012)	(0.014)	(0.045)			
ExpMatch × Post	-0.135^{***}	-0.157^{***}	0.421^{**}	—	—	—
	(0.045)	(0.043)	(0.170)			
Panel B: Metered-Specific Effects						
ExpMatch × Metered	0.091	0.360^{***}	-0.138	-0.431^{***}	-0.014	-0.671^{***}
	(0.102)	(0.054)	(0.168)	(0.064)	(0.026)	(0.066)
Metered × Post	0.172^{***}	-0.149^{***}	-0.400^{***}	0.125^{***}	-0.030^{***}	0.581^{***}
	(0.019)	(0.014)	(0.047)	(0.015)	(0.007)	(0.019)
ExpMatch × Met. × Post	0.417^{***}	0.191^{***}	-0.067	0.284^{***}	0.035	0.349^{***}
	(0.074)	(0.045)	(0.183)	(0.057)	(0.023)	(0.072)
Observations	4,780,527	4,780,527	4,780,527	4,780,527	4,780,527	4,780,527
Destination FE	Block	Block	Block	Block	Block	Block

Notes: This table compares unconstrained and constrained specifications. Panel A shows coefficients on terms that should theoretically equal zero for the fare outcome (since flat-rate JFK trips have fixed $52 fare regardless of route). The constrained specification excludes these terms, imposing the theoretical restriction. Panel B shows metered-specific effects that are estimated in both specifications. The similarity of coefficients across specifications validates the theoretical constraints. All regressions include driver, hour, day-of-week, and destination block fixed effects, plus an indicator for JFK (versus LGA) airport. Standard errors clustered at driver level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Panel A shows the unconstrained coefficients on terms that apply only to flat-rate trips. While some are statistically significant due to large sample sizes, the magnitudes are economically small relative to the metered-trip effects (Panel B). This validates the use of flat-rate trips as a control group. The similarity of Panel B coefficients across constrained and unconstrained specifications confirms that imposing the theoretical restrictions does not materially affect the main estimates.

E.2 Z-Score Similarity Measures

Table 17 presents estimates using Z-score standardized similarity measures for route efficiency outcomes. Z-scores are computed within each destination block, measuring how much a driver's similarity with a destination deviates from the block average. The results are qualitatively similar to the main specifications using raw similarity measures, confirming that discrimination operates through relative demographic alignment.

F Route Efficiency Heterogeneity

This appendix reports the full heterogeneity results summarized in Section 6.1. For each heterogeneity dimension, I report (i) $\hat{\gamma}_3$ from sample-split regressions across all three similarity measures and outcomes, and (ii) the fully interacted specification that adds the quadruple interaction to the full sample.

F.1 Pre-Entry Green Zone Exposure

Table 18 reports $\hat{\gamma}_3$ from separate regressions by tercile of pre-entry green zone dropoff share. Drivers are split by the average share of their pre-entry (January–July 2013) shift dropoffs in green cab-eligible zones. The monotonically increasing pattern documented for fares in Figure 7a extends to trip distance for own-group exposure and expected match.

Table 17: Route Efficiency: Z-Score Similarity Measures

	Fare ($)	Distance (mi)	Time (min)
Panel A: Expected Match (Z-score)			
Similarity × Metered	-0.065^{***}	-0.002	-0.101^{***}
	(0.010)	(0.004)	(0.010)
Metered × Post	0.181^{***}	-0.023^{***}	0.649^{***}
	(0.010)	(0.005)	(0.013)
Similarity × Metered × Post	0.043^{***}	0.005	0.053^{***}
	(0.009)	(0.004)	(0.011)
Panel B: Cosine Similarity (Z-score)			
Similarity × Metered	-0.033^{***}	0.012^{**}	-0.100^{***}
	(0.011)	(0.005)	(0.014)
Metered × Post	0.181^{***}	-0.023^{***}	0.649^{***}
	(0.010)	(0.005)	(0.013)
Similarity × Metered × Post	0.038^{***}	0.003	0.045^{***}
	(0.009)	(0.004)	(0.012)
Observations	4,780,527	4,780,527	4,780,527
Driver FE	Yes	Yes	Yes
Hour/Day-of-week FE	Yes	Yes	Yes
Destination FE	Block	Block	Block

Notes: This table reports estimates from Equation (4) using Z-score standardized similarity measures. Z-scores are computed within each block, measuring how much a driver's similarity with a destination deviates from the block average. Panel A uses expected match Z-scores; Panel B uses cosine similarity Z-scores. All other specifications are identical to Table 5. Standard errors clustered at driver level in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 18: Route Efficiency by Pre-Entry Green Zone Exposure Tercile

Similarity	GZ Tercile	Fare (USD)	Distance (mi)	Time (min)	N
Own-Group Exp.	Low	0.166 (0.111)	0.003 (0.037)	0.241 (0.133)	1,579,130
	Medium	0.229* (0.122)	0.061 (0.044)	0.262 (0.140)	1,584,771
	High	0.354*** (0.082)	0.094** (0.038)	0.374** (0.148)	1,582,108
Expected Match	Low	0.153 (0.118)	0.020 (0.039)	0.194 (0.133)	1,579,130
	Medium	0.259** (0.120)	0.058 (0.042)	0.226 (0.139)	1,584,771
	High	0.306*** (0.084)	0.048 (0.037)	0.481*** (0.150)	1,582,108
Cosine Sim.	Low	0.068 (0.054)	-0.003 (0.020)	0.137* (0.074)	1,579,130
	Medium	0.126** (0.063)	0.020 (0.022)	0.067 (0.073)	1,584,771
	High	0.134*** (0.046)	0.015 (0.019)	0.225*** (0.082)	1,582,108

Notes: Each cell reports the coefficient on Similarity × Metered × Post from Equation (4) estimated on the indicated subsample. Standard errors clustered at the driver level in parentheses. All specifications include driver, hour, day-of-week, and destination block fixed effects. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 19 reports the interacted specification. The proxy variable is an indicator for above-median pre-entry green zone dropoff share. The "Base" column reports $\hat{\gamma}_3$ for drivers with below-median GZ exposure; the "Diff" column reports the additional effect for above-median drivers.

Table 19: Interacted Model: High GZ Exposure

Similarity	Outcome	Base ($\hat{\gamma}_3$)	Diff ($\hat{\gamma}_3\times$ High GZ)	N
Own-Group Exp.	Fare	0.105 (0.086)	0.263** (0.112)	4,749,065
	Distance	0.013 (0.033)	0.081* (0.043)	4,749,065
	Time	0.511*** (0.118)	−0.272* (0.147)	4,749,065
Expected Match	Fare	−0.250*** (0.090)	0.797*** (0.117)	4,749,065
	Distance	0.081** (0.039)	−0.073 (0.051)	4,749,065
	Time	2.363*** (0.130)	−3.120*** (0.168)	4,749,065
Cosine Sim.	Fare	−0.158*** (0.044)	0.423*** (0.058)	4,749,065
	Distance	0.013 (0.019)	−0.006 (0.026)	4,749,065
	Time	1.014*** (0.065)	−1.353*** (0.083)	4,749,065

Notes: Each row reports coefficients from a single regression that includes the base interaction terms from Equation (4) plus their interactions with an indicator for above-median pre-entry GZ dropoff share. "Base" is the coefficient on Sim × Metered × Post; "Diff" is the coefficient on Sim × Metered × Post × High GZ. Standard errors clustered at the driver level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

F.2 Pre-Entry Shift Income

Table 20 reports sample-split results by income tercile.

Table 20: Route Efficiency by Pre-Entry Shift Income Tercile

Similarity	Income Tercile	Fare (USD)	Distance (mi)	Time (min)	N
Own-Group Exp.	Low	0.368*** (0.125)	0.086* (0.045)	0.468*** (0.153)	1,578,206
	Medium	0.476*** (0.097)	0.143*** (0.041)	0.499*** (0.151)	1,592,259
	High	0.028 (0.103)	−0.015 (0.038)	0.024 (0.135)	1,575,714
Expected Match	Low	0.347*** (0.120)	0.062 (0.042)	0.423*** (0.142)	1,578,206
	Medium	0.468*** (0.105)	0.098** (0.042)	0.567*** (0.155)	1,592,259
	High	0.041 (0.104)	0.013 (0.037)	−0.042 (0.134)	1,575,714
Cosine Sim.	Low	0.102* (0.060)	0.016 (0.022)	0.192** (0.079)	1,578,206
	Medium	0.204*** (0.055)	0.037* (0.021)	0.261*** (0.083)	1,592,259
	High	0.060 (0.052)	0.004 (0.019)	−0.002 (0.068)	1,575,714

Notes: See notes to Table 18. Pre-entry shift income is the driver's average shift income from January–July 2013. Tercile cutoffs are $247 and $292.

Table 21 reports the interacted specification. The proxy variable is an indicator for below-median pre-entry shift income.

Table 21: Interacted Model: Low Income

Similarity	Outcome	Base ($\hat{\gamma}_3$)	Diff ($\hat{\gamma}_3\times$ Low Inc.)	N
Own-Group Exp.	Fare	0.173** (0.076)	0.235** (0.112)	4,749,065
	Distance	0.036 (0.030)	0.065 (0.043)	4,749,065
	Time	0.117 (0.096)	0.481*** (0.142)	4,749,065
Expected Match	Fare	0.410*** (0.080)	−0.278** (0.120)	4,749,065
	Distance	0.074** (0.033)	−0.084 (0.054)	4,749,065
	Time	−0.311*** (0.106)	1.430*** (0.171)	4,749,065
Cosine Sim.	Fare	0.206*** (0.040)	−0.180*** (0.061)	4,749,065
	Distance	0.033* (0.017)	−0.051* (0.028)	4,749,065
	Time	−0.144*** (0.053)	0.626*** (0.085)	4,749,065

Notes: Each row reports coefficients from a single regression that includes the base interaction terms from Equation (4) plus their interactions with an indicator for below-median pre-entry shift income. "Base" is the coefficient on Sim × Metered × Post (above-median income drivers); "Diff" is the additional effect for below-median income drivers. Standard errors clustered at the driver level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

The income interacted specification shows sign sensitivity across similarity measures for the fare outcome: the quadruple interaction is positive for own-group exposure but negative for expected match and cosine similarity. The sample-split results (Table 20) are more consistent: all three measures show large positive coefficients for low and medium terciles and near-zero coefficients for the top tercile. The interacted specification's sign sensitivity likely reflects the interaction between the binary income proxy and the lower-order terms rather than genuine disagreement about the direction of heterogeneity. The main text therefore emphasizes the sample-split pattern.

G Statistical Discrimination Tests

This appendix presents the statistical discrimination tests referenced in Section 7.1. Statistical discrimination predicts that drivers use neighborhood demographics as a signal for unobserved passenger characteristics and update beliefs with experience. Taste-based discrimination predicts stable preferences regardless of information.

G.1 Spatial Sorting

Table 22 tests whether neighborhood-specific or general experience attenuates the relationship between own-group exposure and location choice.

Table 22: Statistical Discrimination Test: Experience and Spatial Sorting

	Own-Group Exposure
Own-Group Exposure	0.0314**
	(0.0142)
Own-Group × Post	-0.0177
	(0.0177)
Own-Group × CDTA Exp	0.0007
	(0.0022)
Own-Group × General Exp	-0.0089
	(0.0148)
CDTA Experience	0.0049***
	(0.0004)
General Experience	-0.0227***
	(0.0024)
Driver FE	Yes
Hour FE	Yes
Day-of-week FE	Yes
CDTA FE	Yes
Observations	417,867

Notes: This table tests for statistical discrimination by interacting own-group exposure with CDTA-specific and general experience. Sample restricted to new drivers (first observed February 2013+) for whom experience is measured without left-censoring. Under statistical discrimination, Own-Group × CDTA Exp should be negative as drivers update beliefs with neighborhood experience. Standard errors clustered at driver level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The coefficient on Own-Group × CDTA Experience is essentially zero (0.0007) and statistically insignificant. Neighborhood-specific experience does not attenuate discrimination. The interaction with general experience is also insignificant (−0.0089). General driver maturation does not explain the pattern.

G.2 Route Efficiency

Table 23 extends the statistical discrimination test to route efficiency outcomes.

Table 23: Statistical Discrimination Test: Route Efficiency

	Fare ($)	Distance (mi)	Time (min)
Similarity × Metered	-0.052	0.085	-0.064
	(0.226)	(0.094)	(0.274)
Sim × Metered × Post	0.663**	0.102	0.663**
	(0.292)	(0.113)	(0.334)
Sim × Metered × CDTA Exp	0.042	0.016	0.068*
	(0.033)	(0.013)	(0.038)
Sim × Metered × Gen Exp	-0.452	-0.065	-0.633**
	(0.301)	(0.130)	(0.305)
Driver FE	Yes	Yes	Yes
Hour/Day-of-week FE	Yes	Yes	Yes
Destination FE	Block	Block	Block
Observations	417,867	417,867	417,867

Notes: This table tests for statistical discrimination in route efficiency by interacting similarity × metered with CDTA-specific and general experience. Sample restricted to new drivers (first observed February 2013+) for whom experience is measured without left-censoring. Sample includes JFK and LaGuardia airport trips; metered equals one for trips with metered fares, zero for JFK flat-rate ($52) trips. Under statistical discrimination, Sim × Metered × CDTA Exp should be negative. Standard errors clustered at driver level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The coefficient on Similarity × Metered × CDTA Experience is small and insignificant across all outcomes (0.042 for fare, 0.016 for distance, 0.068 for time). Experience does not reduce route padding discrimination. Both tests reject statistical discrimination in favor of taste-based preferences.

H Tipping and Demographic Similarity

This appendix tests whether differential tipping can explain the discrimination patterns documented in Section 7.2. If demographically similar passengers tip more generously, profit-maximizing drivers would have incentive to provide better service even absent taste-based preferences.

H.1 Specification

The sample includes 10.4 million credit card yellow cab trips from 2013 (cash transactions do not record tips; mean tip rate is 18 percent of the $13.99 average fare). I estimate:

$$\text{Tip}_{ijt} = \phi_1 \cdot \text{Fare}_{ijt} + \phi_2 \cdot \text{Similarity}_{ij} + \phi_3 \cdot (\text{Fare}_{ijt} \times \text{Similarity}_{ij}) + \alpha_i + \delta_k + \lambda_h + \mu_d + \nu_m + \varepsilon_{ijt} \quad (7)$$

where Tip_{ijt} is the tip amount for driver i serving a passenger traveling to location j at time t, Fare_{ijt} is the total fare, and Similarity_{ij} measures demographic alignment between driver i and location j. I include driver fixed effects (α_i), pickup CDTA fixed effects (δ_k), hour fixed effects (λ_h), day-of-week fixed effects (μ_d), and month fixed effects (ν_m).

The coefficient of interest is ϕ_3, the interaction between fare and similarity. A positive ϕ_3 indicates that demographically similar passengers tip at higher rates, meaning the marginal tip per dollar of fare is higher when driver and passenger demographics are aligned.

H.2 Results

Table 24 presents tipping regression results using multiple similarity measures. I examine similarity at both the CDTA level (relevant for sorting decisions) and the block level (relevant for within-trip quality decisions).

Panel A reports results using CDTA-level similarity measures. The fare coefficient is approximately 0.17 across all specifications: passengers tip about 17 cents per dollar of fare on average. The interaction coefficients are positive and statistically significant: 0.0051 for

Table 24: Tipping and Demographic Similarity

	Panel A: CDTA-Level Similarity			Panel B: Block-Level Similarity		
	Own-Group	Expected Match	Cosine Sim	Own-Group	Expected Match	Cosine Sim
	(1)	(2)	(3)	(4)	(5)	(6)
Total Fare	0.1685***	0.1668***	0.1671***	0.1621***	0.1615***	0.1609***
	(<0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Similarity	−0.0838***	−0.1339***	−0.0506***	−0.0558***	−0.0828***	−0.0537***
	(0.0076)	(0.0082)	(0.0044)	(0.0068)	(0.0071)	(0.0038)
Fare × Similarity	0.0051***	0.0090***	0.0033***	0.0012***	0.0032***	0.0029***
	(0.0004)	(0.0004)	(0.0002)	(0.0004)	(0.0004)	(0.0002)
Driver FE	Yes	Yes	Yes	Yes	Yes	Yes
Pickup CDTA FE	Yes	Yes	Yes	Yes	Yes	Yes
Hour/Day/Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10,423,008	10,423,008	10,423,008	10,161,626	10,161,626	10,161,626
R^2	0.555	0.555	0.555	0.463	0.463	0.463

Notes: Dependent variable is tip amount ($). Panel A uses CDTA-level similarity measures; Panel B uses block-level measures averaged across pickup and dropoff locations. The coefficient of interest is Fare × Similarity, which captures whether demographically similar passengers tip at higher rates. A positive coefficient indicates that the marginal tip per dollar of fare is higher when driver and passenger demographics are aligned. All specifications include driver, pickup CDTA, hour, day-of-week, and month fixed effects. Sample restricted to credit card transactions from a random subsample of 20 million trips. Standard errors in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

own-group exposure, 0.0090 for expected match, and 0.0033 for cosine similarity. These coefficients imply that moving from the 25th to 75th percentile of demographic similarity increases the tip rate by 0.05–0.09 percentage points.

Panel B reports results using block-level similarity measures (averaged across pickup and dropoff locations). The interaction coefficients are somewhat smaller: 0.0012 for own-group exposure, 0.0032 for expected match, and 0.0029 for cosine similarity. The pattern is consistent with Panel A: demographically similar passengers tip at modestly higher rates.

H.3 Interpretation

The expected match specification (Panel B) yields an interaction coefficient of 0.0032. A shift from the 25th to 75th percentile of similarity changes the tip rate by 0.064 percentage points, approximately $0.01 in additional tips per trip at the average fare. The route efficiency discrimination documented in Table 5 shows a fare gap of $0.43. The tipping differential can rationalize at most 2.3 percent of this gap. Two caveats apply: tips are observed only for credit card transactions, and the analysis conditions on completed trips rather than expected

tips. Despite these limitations, the 40:1 ratio between route efficiency discrimination and tipping differentials makes it unlikely that measurement issues explain the findings.

I Non-Airport Trip Analysis

I analyze approximately 19 million non-airport trips from a random 20 million trip sample (excluding all JFK and LaGuardia pickups and dropoffs) to test whether the airport findings generalize. For non-airport trips, I assume destination demographics are exogenous to driver behavior, a weaker assumption than the airport setting, where queue-based dispatch ensures quasi-random assignment.

I.1 Spatial Sorting: Non-Airport Trips

Table 25 replicates the location choice analysis for non-airport trips. The dependent variable is an indicator for whether the driver's next pickup occurs in the same CDTA as the current dropoff. I use a pre-post specification comparing the pre-entry (January–July 2013) and post-entry (August–December 2013) periods.

Table 25: Location Choice: Non-Airport Trips

	Own-Group	Expected Match	Cosine Sim
	(1)	(2)	(3)
Similarity	0.0149***	0.0190***	0.0111***
	(0.0026)	(0.0029)	(0.0018)
Similarity × Post	-0.0017	-0.0047***	-0.0021***
	(0.0016)	(0.0017)	(0.0007)
Driver FE	Yes	Yes	Yes
Hour/Day FE	Yes	Yes	Yes
CDTA FE	Yes	Yes	Yes
Observations	18,882,332	18,882,332	18,882,332

Notes: This table reports location choice estimates for non-airport trips (N ≈ 19 million). The dependent variable is an indicator for whether the driver's next pickup is in the same CDTA. Post is an indicator for August–December 2013 (after green cab entry). Standard errors clustered by driver in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

The baseline similarity coefficients are positive and highly significant across all three measures: 0.015 for own-group exposure, 0.019 for expected match, and 0.011 for cosine similarity. These magnitudes are comparable to the airport-trip estimates in Table 4: drivers sort toward demographically similar neighborhoods regardless of trip origin. The post-entry interaction coefficients are uniformly negative, expected match (-0.005) and cosine similarity (-0.002) reach statistical significance. Competition reduces discriminatory sorting in non-airport markets, consistent with the airport analysis.

I.2 Route Efficiency: Non-Airport Trips

Table 26 presents route efficiency results for non-airport trips. Without the metered versus flat-rate fare variation available at airports, the flat-rate placebo is not available. Instead, I compare similarity effects before and after green cab entry. A key difference from the airport analysis is that both pickup and dropoff locations vary for non-airport trips. Pickup-dropoff location pair fixed effects control for route-specific factors.

I implement these fixed effects at the census tract level rather than the block level for computational feasibility. New York City contains approximately 30,000 census blocks, and because origin-destination pairs are directional (a trip from block A to block B follows a different route than a trip from B to A), block-level pairs would generate up to $30,000^2 \approx 900$ million fixed effects, far exceeding what is computationally tractable. Census tracts, of which NYC has approximately 2,200, yield a maximum of roughly 4.8 million pairs; in the data, 231,888 unique pickup-dropoff tract pairs are observed. Census tracts provide an appropriate level of aggregation: they are small enough to capture route-specific variation in expected fare, distance, and time, while remaining computationally feasible. Similarity is still measured at the dropoff block level as the passenger demographic proxy, preserving the granularity of the treatment variable.

The baseline similarity coefficients are negative and highly significant across all outcomes and similarity measures. For own-group exposure (Panel A), the coefficients are -0.30 for

Table 26: Route Efficiency: Non-Airport Trips (Diff-in-Diff)

	Fare	Distance	Time
Panel A: Own-Group Exposure			
Similarity	-0.2974***	-0.1056***	-0.2537***
	(0.0123)	(0.0044)	(0.0184)
Post	0.1811***	-0.0059***	0.4186***
	(0.0022)	(0.0008)	(0.0035)
Similarity × Post	0.0610***	0.0088*	0.0969***
	(0.0133)	(0.0046)	(0.0206)
Observations	18,808,320	18,808,320	18,808,320
Panel B: Expected Match			
Similarity	-0.2864***	-0.0992***	-0.2635***
	(0.0135)	(0.0049)	(0.0203)
Post	0.1662***	-0.0071***	0.3881***
	(0.0033)	(0.0012)	(0.0052)
Similarity × Post	0.0786***	0.0067	0.1592***
	(0.0132)	(0.0046)	(0.0209)
Observations	18,808,320	18,808,320	18,808,320
Panel C: Cosine Similarity			
Similarity	-0.0457***	-0.0198***	-0.0355***
	(0.0088)	(0.0032)	(0.0131)
Post	0.1640***	-0.0077***	0.3820***
	(0.0033)	(0.0012)	(0.0052)
Similarity × Post	0.0437***	0.0047*	0.0927***
	(0.0066)	(0.0024)	(0.0104)
Observations	18,808,320	18,808,320	18,808,320
Driver FE	Yes	Yes	Yes
Hour/Day FE	Yes	Yes	Yes
Pickup-Dropoff Tract Pair FE	Yes	Yes	Yes

Notes: This table reports diff-in-diff estimates for route efficiency outcomes in non-airport trips. Similarity is measured at the dropoff block level. Pickup-dropoff tract pair fixed effects control for route-specific factors; this is necessary for non-airport trips because both pickup and dropoff locations vary (unlike airport trips where pickup is always JFK or LGA). The coefficient on Similarity captures baseline discrimination; Post captures overall changes in service quality after green cab entry; Similarity × Post captures how discrimination changed after entry. Standard errors clustered by driver in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

fare, −0.11 for distance, and −0.25 for time. Higher demographic similarity predicts more efficient routes: lower fares, shorter distances, and less trip time. Drivers provide better service, more direct routes, to passengers traveling to demographically similar areas.

The post-entry interaction coefficients are positive and statistically significant in nearly all specifications. For own-group exposure, the coefficients are 0.061 for fare, 0.009 for distance, and 0.097 for time. Competition reduces the efficiency advantage previously enjoyed by similar passengers. The pattern is consistent across all three similarity measures and all three outcome variables.

These results support the "leveling down" interpretation from the main analysis: competition narrows the discrimination gap by reducing preferential treatment for favored passengers rather than improving service for disfavored passengers. The non-airport findings demonstrate that this mechanism operates throughout the taxi market, not only for airport trips with quasi-random destination assignment.

I.3 Equilibrium Supply Effects

Section 8 quantifies the direct welfare costs of discrimination through fare extraction and time costs. Here I note potential second-order equilibrium effects on taxi availability that cannot be directly estimated but that may represent additional welfare losses.

Route Padding and Effective Supply

Route padding reduces effective taxi supply. If each affected trip takes approximately 40 seconds longer, then at any moment more cabs are occupied than necessary. Given average trip durations of 12–15 minutes, this represents roughly a 4–5% increase in trip duration for affected rides. Across the fleet, this is equivalent to 1–2% fewer cabs available at any given time, or roughly 130–260 fewer effective medallions out of the 13,000-cab fleet. This reduced availability could marginally increase passenger wait times citywide.

Sorting and Spatial Mismatch

Sorting behavior may compound these supply effects. If drivers systematically avoid certain neighborhoods in favor of demographically similar areas, passengers in avoided neighborhoods face reduced taxi availability beyond what aggregate fleet size would predict. This spatial mismatch between supply and demand could generate longer wait times and lower match rates for passengers in neighborhoods that drivers tend to avoid.

Limitations

Quantifying these equilibrium effects precisely requires market clearing assumptions and structural modeling beyond the scope of this paper. I note them as potential additional welfare costs while emphasizing that the direct time and fare costs documented in Section 8 represent the primary measurable losses.