Leveling Down: Competition and Discrimination in Service Markets

Kwabena Donkor*

July 20, 2026

Abstract

Does competition reduce discrimination? Yes. The standard mechanism is selection, which drives discriminators from the market. But I identify a less benign channel, *leveling down*: competition narrows the gap 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, so discrimination shows up in how a passenger is served, not in whether the driver takes the fare. 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. Market-wide, the cost of degraded service exceeded the benefit of reduced discrimination by more than two to one.

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, Andrey Simonov, Dmitry Taubinsky, Takeaki Sunada, Bryan Bollinger, and Max Muller 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 and the Marketing Science conference 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 turning away profitable customers and eventually drives discriminators from the market. This paper documents a different and less benign path to the same statistic, which I call *leveling down*. A competitive shock to the New York City taxi market narrowed racial disparities in service quality by half to two-thirds. Service did not improve for anyone. It deteriorated for everyone, and it deteriorated most for the passengers previously treated best. The narrowing is an equity-efficiency tradeoff, and a steep one: market-wide, the cost of the degraded service exceeded the benefit of the reduced disparity by more than two to one.

In Becker's model, prejudice is a fixed preference that does not respond to market conditions, and competition reduces discrimination only through selection: providers who sacrifice profit to indulge animus are outcompeted by those who do not (Arrow, 1973; Guryan and Charles, 2013). Drawing on Goldberg (1982), who models discrimination as in-group favoritism (nepotism) rather than out-group hostility (animus), I derive a different prediction for service quality. Favoritism is costly to the provider even when no one is mistreated: disfavored passengers receive indifference, not hostility, because the driver has no reason to depart from profit-maximizing service for them. A driver can route a same-group passenger efficiently around congestion, delivering better service at the cost of a lower metered fare, and take the longer, more profitable route for everyone else. The quality gap then depends on the surplus available to fund it: when margins are high, drivers can afford the revenue sacrifice; when competition erodes margins, the gift is the first thing cut. Preferential treatment is a luxury good, and the discrimination gap becomes an equilibrium object. The gap is constant in market conditions under animus and shrinks with margins under nepotism. At any fixed level of competition the two models are observationally equivalent; a shock to margins is the only instrument that can tell them apart.

I study the New York City yellow taxi market[1] and use the August 2013 entry of green cabs (Boro Taxis), a new class of taxis authorized to pick up street hails in the outer boroughs and northern Manhattan, as the margin shock.[2] Green cabs charged identical fares, so from the passenger's perspective they were near-perfect substitutes. Their entry compressed driver

[1]Taxi discrimination is a documented policy concern in New York City. During the sample period, a yellow cab driver was fined $25,000 for bypassing a Black family to pick up white passengers (Flegenheimer, 2015). The TLC later created an Office of Inclusion and launched a citywide anti-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 outer boroughs and upper Manhattan, but enforcement was imperfect, and their entry pushed yellow cabs toward the core, raising competition citywide (Section 2.2).

margins, cutting monthly income by about $107, roughly a third of a single shift's earnings.

Four institutional features turn this market into a laboratory for discrimination in service quality. First, fares are regulated, so competition operates not through prices but through service. Second, airport queues at JFK and LaGuardia dispatch passengers to drivers in strict first-come-first-served order, so the driver–passenger match is effectively random. Third, GPS trip records provide an objective, transaction-level measure of quality I call *route efficiency*: whether the driver takes a fast route or a slow, padded one.[3] Fourth, and central to the design, the same driver works under two fare regimes: JFK-to-Manhattan trips carry a flat $52 fare, while LaGuardia trips run on the meter. Padding a route raises revenue only on the meter. A driver who favors similar passengers through routing will therefore show a similarity gradient on metered trips and none on flat-rate trips. A confound common to the two trip types cannot produce that asymmetry: traffic, weather, destination geography and skill hit metered and flat trips alike. The favoritism estimate is within driver, holds the destination block, the hour and day of week fixed.

Before green cab entry, drivers favored passengers of their own race. A same-race passenger paid about $0.44 less than a different-race passenger on the same metered trip, and reached the destination 26 seconds sooner. On flat-rate trips, where the route cannot move the fare, the gap vanishes. Favoritism concentrates where the racial makeup of the destination describes who is actually riding: it is strong on trips into residential blocks, where the gap roughly doubles to about a dollar and nearly a minute, and fades to noise in commercial blocks, whose residents are not the passengers. This gradient is not local knowledge, not tip-chasing, and not statistical learning. It survives excluding each driver's home territory and in fact vanishes there; tips do not respond to similarity although fares do; and experience does not attenuate it.

After entry, the gap collapsed from above. Metered fares rose and trips slowed for every passenger, but the previously favored lost the most, and the fare gap compressed by 54 to 69 percent. What marks this as competition rather than seasonality is where the response concentrates. Drivers most exposed to green cab territory compressed the gap most, and low- and middle-income drivers compressed it sharply while high-income drivers did not compress it at all. The same drivers changed their behavior; the narrowing did not come from a change in who was driving. A seasonal or citywide shock cannot sort drivers this way, by exposure and by margin. Aggregated to the market, service degradation reached $44.3 million a year against a $20.0 million reduction in discrimination, and measurement error in the demographic proxy attenuates every estimate toward zero, so these magnitudes are conservative.

[3] Liu et al. (2021) use route efficiency to study moral hazard in ride-hailing.

The paper makes three contributions. The first is to document and quantify leveling down. Discrimination gaps can narrow not because the disfavored gain but because service deteriorates for everyone, with the largest deterioration among the previously favored. A shrinking gap is therefore an incomplete measure of progress: the equity gain comes bundled with a larger efficiency loss, and anti-discrimination policy that scores interventions by gap reduction alone counts that degradation as success. The finding runs against the usual reading of competition and quality. Competition is generally thought to raise quality (Spence, 1975; Shaked and Sutton, 1982; Matsa, 2011; Fréchette et al., 2019; Buchholz, 2022); here it lowers quality for everyone, because discrimination had created a quality premium that competition erodes. The mechanism parallels Hill and Stein (2025), where competition to publish first degrades the quality of scientific work.

The second contribution is the first service-quality test that separates Becker (1957) from Goldberg (1982). The two models predict identical cross-sections, so existing evidence cannot tell them apart; they diverge only in how the gap responds to margins, and the collapse of the gap when margins compressed is the nepotism prediction, not the animus one. This sets the paper apart from the empirical literature testing Becker's selection channel across wage gaps, banking deregulation, firm survival and online markets (Charles and Guryan, 2008; Levine et al., 2014; Weber and Zulehner, 2014; Pager, 2016; Doleac and Stein, 2013). In that literature, competition disciplines discrimination by removing discriminators from the market. Here the same drivers discriminate less because they can no longer afford the preference: the gap compresses from above rather than lifting from below.

The third contribution is to measure discrimination on the intensive margin, quality conditional on service, where evidence is scarce because quality is hard to observe (Nelson, 1970). Correspondence studies, policing records and ride-cancellation data measure the extensive margin of access (Bertrand and Mullainathan, 2004; Kline et al., 2024; Goncalves and Mello, 2021; Ge et al., 2020); GPS route efficiency makes the quality margin observable at the transaction level. The distinction matters for policy: interventions that equalize access, from algorithmic dispatch to anti-refusal rules, leave the quality margin unconstrained, so discrimination can persist through service even where access is equal.

The paper proceeds as follows. Section 2 describes the market, the green cab entry, and the measurement of driver–passenger similarity. Section 3 establishes pre-entry favoritism with the metered-versus-flat design and shows it concentrates in residential destinations, where the demographic proxy has content. Section 4 develops the nepotism model and derives the leveling-down prediction that separates it from animus. Section 5 documents the margin compression and tests the prediction. Section 6 rules out local knowledge, tip recovery and learning. Section 7 converts the coefficients into market-wide dollars. Section 8

concludes.

2 Setting and Data

I use the universe of over 173 million New York City yellow taxi trips recorded by the Taxi and Limousine Commission (TLC) in 2013. I combine the trip records with two demographic inputs: each driver's race, inferred from the name on the driver's license, and the racial composition of the Census block where each trip ends. Together these inputs yield a trip-level measure of racial similarity between the driver and the destination. This section describes the market, the entry of green cabs in August 2013, and the measurement of similarity.

2.1 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). Regulation fixes the other terms of a ride. Fares follow a metered schedule set by the TLC, and vehicle model, color, signage and interior features are standardized across all medallion cabs. A passenger who hails a yellow cab therefore gets the same car at the same price from any driver. What varies is the route the driver chooses.

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.

Service quality is difficult to measure. Quality varies from one transaction to the next, and much of what constitutes good service is unverifiable (Nelson, 1970). The standard alternatives, surveys (Parasuraman et al., 1988) and online ratings (Chevalier and Mayzlin, 2006; Luca, 2016), are subjective; ratings also carry disparities of their own, scoring minority providers lower for comparable service (Hannák et al., 2017; Teng et al., 2023). The trip records overcome these problems. I measure service quality by *route efficiency*: whether the driver takes a fast route to the destination or a slow, padded one. An inefficient route costs the passenger directly: under metered pricing it raises the fare, and it always lengthens the ride. Route efficiency is therefore an objective, consequential and transaction-level measure of service quality.

Measuring discrimination in service quality faces two obstacles. In a street-hail market, drivers choose where to search and whom to stop for, so the driver–passenger match is itself

an outcome of any discriminatory preference, and comparisons of service across passenger types confound treatment with matching. And a slow or expensive trip is hard to interpret on its own: the question is how the same driver would have served a different passenger on the same route, a counterfactual the data rarely supply. Two features of the airport taxi market overcome these obstacles. 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. Padding a route pays only on the meter, and the same driver delivers passengers to the same destination blocks under both regimes. The fare structure therefore varies the incentive to pad while holding the driver and the destination fixed, the contrast on which Section 3.1 builds the research design.

2.2 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 the airports. In particular, green cabs could not pick up passengers at JFK or LaGuardia: the airport queues remained exclusively yellow, so competitive pressure reached airport trips through drivers' overall earnings rather than through rival cabs at the curb (Section 5.1). Figure 1 maps the demarcation: solid yellow marks the yellow-exclusive territory, the HEZ and the two airports, and stripes mark the green cab pickup territory.

Despite these restrictions, green cab entry affected competition citywide. Enforcement was imperfect: green cabs picked up passengers inside the HEZ as well. Green cabs operated under identical fare structures, so from the passenger's perspective yellow and green cabs were near-perfect substitutes. Roughly 2,100 green cabs were in service by mid-December 2013 (NYC Taxi and Limousine Commission, 2013), and their trip volume grew rapidly, from about 7,000 trips in August to over 600,000 by December (Appendix B). Green cab trip records do not identify individual drivers, so the discrimination analysis is confined to yellow taxi drivers.

Figure 1: NYC Taxi Zones: Hail Exclusionary Zone (HEZ) Demarcation



2.3 Measuring Demographic Similarity

I infer each driver's race from the name on the driver's license, recorded in the TLC 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 assigning each driver to a single race, which carries classification uncertainty into the similarity measures instead of discarding it. I classify 32,478 drivers, 30,616 of whom appear in the analysis sample defined in Section 2.4.[4]

Passenger race is not directly observed. I proxy for passenger race using the racial composition of the dropoff Census block: I spatially join each trip's dropoff GPS coordinates to 2010 Census data, which provide population shares by race at the block level, and impute from the surrounding tract where block data are missing (Appendix A.3). All trips originate

[4]The trip data capture roughly 43,000 of the approximately 51,000 licensed yellow cab drivers (New York City Taxi and Limousine Commission, 2014) during 2013; classification succeeds for 32,478. The gaps reflect inactive license holders, drivers who worked exclusively outside the sample period, and names without a usable surname match. Appendix A.1 details the classification.

at airports, so the destination is the only geographic signal of who is in the cab. The signal is strongest where people live: a trip ending in a residential block plausibly carries a resident, while a destination in a commercial core describes no one in particular. Section 3.3 exploits this variation directly and shows that the estimates concentrate exactly where the proxy has content. More generally, measurement error in the proxy attenuates estimates toward zero, so significant effects are conservative lower bounds.

The primary measure of driver–passenger racial similarity is *expected match*: the probability that the driver and a randomly chosen resident of the dropoff block share a race,

$$\text{ExpMatch}_{ij} = \sum_{r} p_{ir}\, s_{jr}, \tag{1}$$

where p_{ir} is the probability that driver i is race r and s_{jr} is the population share of race r in dropoff block j. The measure runs from 0 to 1. A coefficient on expected match therefore reads as the effect of moving from a destination with no chance of a racial match to one where a match is certain.

I complement expected match with *own-group exposure*, which measures whether the dropoff block overrepresents the driver's own racial group:

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

where $g^* = \arg\max_r p_{ir}$ is driver i's most likely racial group, p_{i,g^*} the probability of belonging to it, and $\bar{s}_{g^*}$ the citywide share of that group. The citywide subtraction is the point of this second measure. Some groups are numerous across the whole city, so a driver from a large group would register high exposure almost everywhere. Subtracting the citywide share removes that mechanical component: the index is positive only where the driver's own group is overrepresented relative to the city as a whole. Expected match measures raw alignment; own-group exposure nets out citywide composition. Agreement between the two therefore guards against the concern that estimates reflect where drivers happen to operate rather than similarity itself. Section 3 reports estimates under both measures, with the own-group tables in Appendix C. A third measure, cosine similarity, and formal definitions of all three appear in Appendix A.6, which also relates them to the isolation and exposure indices of the residential segregation literature (Bell, 1954; Massey and Denton, 1988).

2.4 Sample and Summary Statistics

The discrimination analysis focuses on trips originating at JFK or LaGuardia during 2013. The raw airport sample contains 5.50 million trips. I exclude trips terminating at an airport,

trips whose dropoff cannot be placed in a Census block, and trips with an invalid distance, duration or fare or with missing destination demographics. The analysis sample comprises 5.21 million trips. Appendix A.4 documents retention at each stage.

Table 1 summarizes the racial composition of the destinations drivers face and of the drivers themselves, at the unit the analysis uses: the dropoff Census block. Across trips in the analysis sample, the average destination block is 56% White, 15% Hispanic, 14% Asian, and 11% Black. The driver population differs markedly: under BISG average probabilities, drivers are predominantly Asian (42%), followed by Black (25%), White (17%), and Hispanic (13%). The estimates draw their identifying variation not from this difference in averages but from the dispersion around them. Destination composition varies widely across blocks, with standard deviations of 15 to 28 percentage points for the major groups, so the same driver serves destinations that range from demographically unlike to demographically like. Driver fixed effects absorb every level difference across drivers and racial groups.

Table 1: Destination and Driver Racial Composition

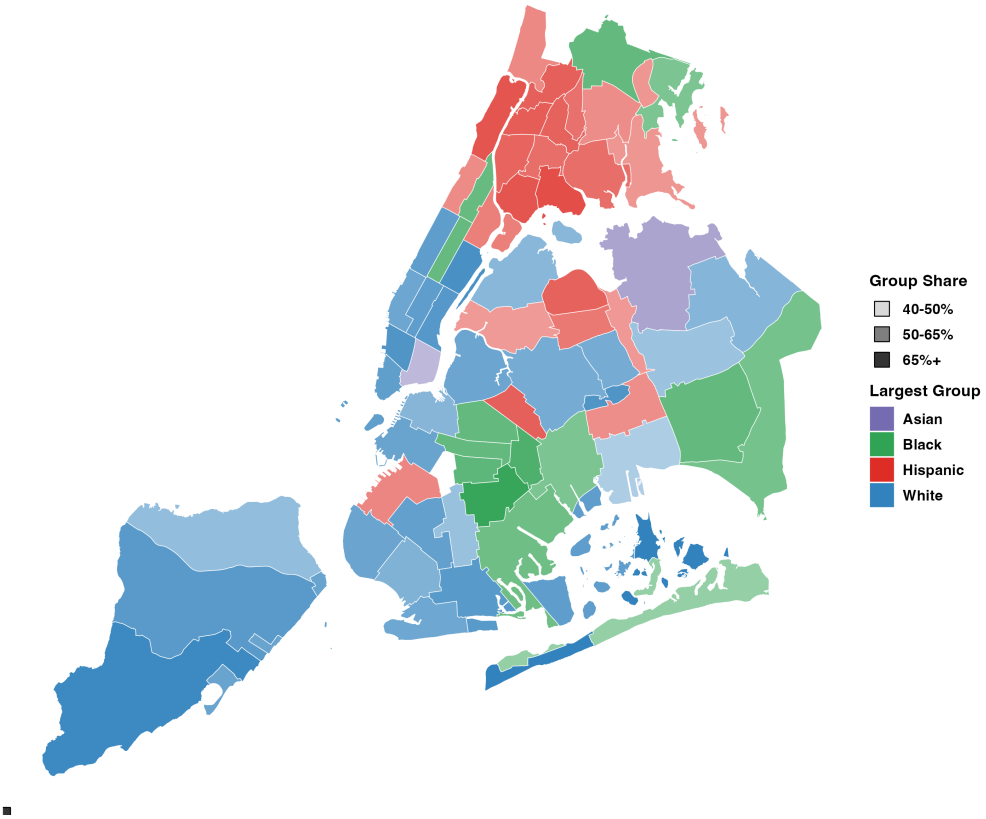
	Destination block		Driver	
Race/Ethnicity	Mean (%)	SD (%)	Avg. Prob. (%)	Modal (%)
White	56.0	27.6	16.7	30.2
Black	11.3	20.6	25.2	20.3
Hispanic	14.9	16.7	13.0	4.8
Asian	14.4	14.8	42.2	44.7
Other	3.3	7.3	2.8	0.1
N	34,537 blocks		32,478 drivers	

Notes: Columns 1–2 report the mean and standard deviation, across trips in the analysis sample (Table A.1), of the 2010 Census racial composition of the dropoff Census block, the same destination shares from which the racial similarity measures are constructed, so the statistics describe exactly the variation the analysis uses. For blocks with no resident population, shares are imputed from the surrounding tract, taxi zone, or community district, in that order. N counts distinct destination blocks. Column 3 reports the average BISG probability of each race across all classified drivers; column 4 assigns each driver to their modal (highest-probability) race. Driver classifications from Bayesian Improved Surname Geocoding (BISG) on the FOIL 2013 license registry.

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.

Table 2 presents trip-level summary statistics for the two airports, with non-airport trips alongside for comparison. LaGuardia trips average 9.7 miles and 27 minutes with metered

Figure 2: Racial Plurality by Neighborhood in New York City



fares averaging $31; JFK trips average 16.2 miles and 38 minutes with mean fares of $46. Roughly 58% of airport trips originate at LaGuardia. A typical non-airport trip is about a third the size of an airport trip, at 2.3 miles, 12 minutes and $11; Section 7 returns to this contrast when extrapolating the airport estimates to the wider market. Destinations are moderately residential on average: the *residential share* of a block, the share of its built floor area in residential use measured from New York City PLUTO tax-lot records, averages 0.55 across destination blocks. At the block level, expected match averages 0.20 and own-group exposure 0.01; the near-zero average own-group exposure reflects the citywide subtraction built into that measure, so the identifying variation comes from deviations around this mean rather than its level. The similarity measures are nearly identical across the two airports.

3 Favoritism in Route Efficiency

Do taxi drivers provide better service to passengers of their own race? Before green cab entry, they did. The same driver charged demographically similar passengers less than different-race passengers going to the same destination, and got them there faster. I call this

Table 2: Summary Statistics: Airport and Non-Airport Trips, 2013

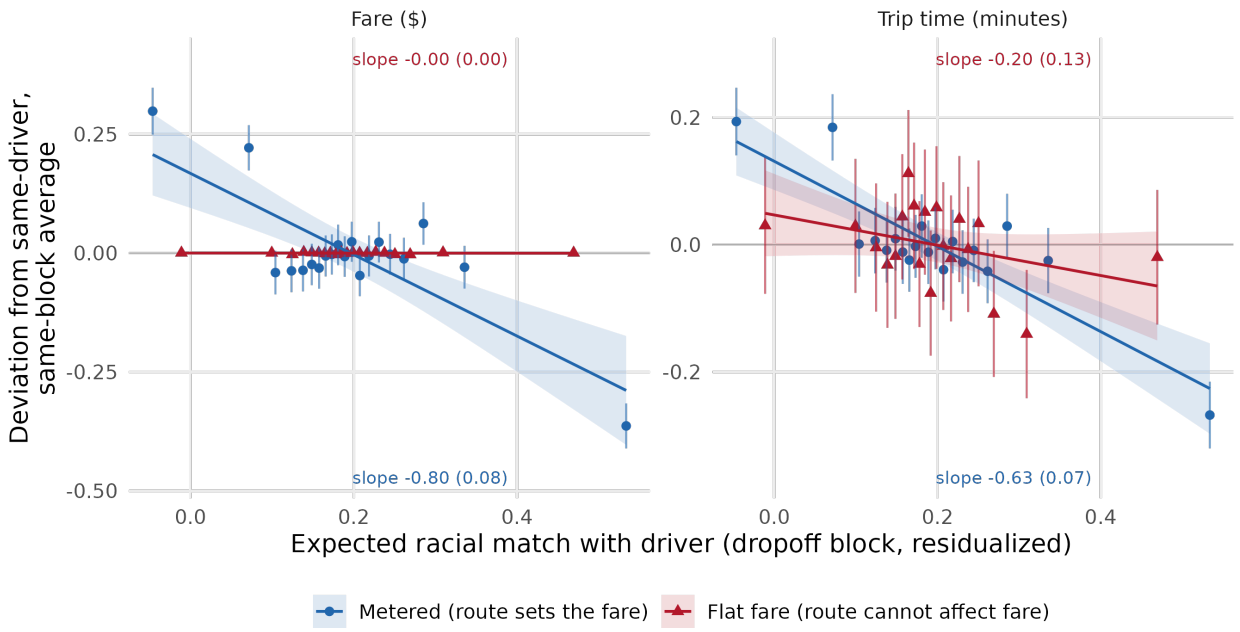
	LaGuardia		JFK		Non-airport	
Variable	Mean	SD	Mean	SD	Mean	SD
Trip characteristics						
Trip distance (miles)	9.68	3.00	16.21	5.06	2.34	2.16
Trip duration (minutes)	26.5	11.2	37.6	15.9	11.6	7.6
Fare amount ($)	30.53	9.03	46.49	12.48	10.77	6.80
Tip amount ($)	4.21	3.90	4.52	5.39	1.16	1.65
Credit card (%)	64.6	—	50.9	—	53.7	—
Residential share of dropoff block	0.52	0.36	0.59	0.35	0.52	0.36
Racial similarity measures (dropoff Census block)						
Expected match	0.194	0.145	0.200	0.160	0.198	0.152
Own-group exposure	0.014	0.149	0.010	0.163	0.011	0.155
Observations	3,026,324		2,179,749		136,788,485	
Unique drivers	30,022		28,389		32,393	

Notes: The LaGuardia and JFK columns describe the analysis sample: 2013 airport-pickup trips, excluding trips that terminate at an airport, trips whose dropoff cannot be placed in a Census block, trips with an invalid distance ($\leq$ 0.1 or $>$ 100 miles), duration ($<$ 1 or $>$ 180 minutes), or fare ($\leq$ 0 or $>$ \$300), and trips with missing destination demographics; Table A.1 documents each exclusion. The Non-airport column applies the same validity requirements to trips picked up elsewhere in New York City. *Expected match* is the probability that the driver and a randomly chosen resident of the dropoff Census block share a race. *Own-group exposure* is the share of the driver's own racial group in the dropoff block minus that group's citywide share, weighted by the driver's classification probability. Community District Tabulation Area versions and a third measure, cosine similarity, appear in Appendix Table A.2. *Residential share* is the share of built floor area in residential use in the dropoff block, from New York City PLUTO tax-lot records, defined for the 73.5% of airport trips and 73.3% of non-airport trips with matched land-use records. LaGuardia trips are metered; JFK-to-Manhattan trips carry a flat \$52 fare. Data from NYC TLC administrative records.

within-driver quality gap *favoritism*, and use the term throughout for the advantage that similar passengers receive; Section 4 formalizes it as a model of in-group nepotism.

Figure 3 shows the result and the research design in a single view. Each panel plots a service outcome, the fare on the left and the trip time on the right, against driver–passenger similarity, separately for two kinds of trip. Metered trips let the route determine the fare; JFK–Manhattan flat-rate trips fix it at \$52 regardless of route. On metered trips, fares and times fall as the passenger grows more similar to the driver. On flat-rate trips, where routing cannot move the fare, both gradients are flat. That contrast is the paper's identification, and Section 3.1 builds the estimator around it.

Figure 3: Favoritism in Route Efficiency: Metered Trips versus the Flat-Fare Placebo



Binned means of the fare (left panel) and trip time (right panel) against expected match (driver–passenger racial similarity), estimated separately for metered trips, where the route sets the fare, and JFK–Manhattan flat-rate trips, where the fare is $52 regardless of route. Pre-entry trips (January–July 2013); 20 equal-count bins per series; bars are 95% confidence intervals. Each series is residualized on its own driver, destination-block, hour-of-day and day-of-week fixed effects, so each point compares trips by the same driver to the same block at the same hour, and both series receive identical treatment. The own-group-exposure version appears in Appendix Figure C.1.

3.1 Empirical Strategy

The design rests on a contrast between two fare regimes within the same driver. Metered trips reward inefficient routing: a longer or slower route raises the fare. Flat-rate trips remove that margin: the JFK–Manhattan fare is $52 regardless of route. A driver who favors similar passengers through routing will show a similarity gradient on metered trips and none on flat trips. A confound common to both trip types cannot produce that asymmetry. On trips before green cab entry (January–July 2013), I estimate the within-driver specification

$$y_{ijt} = \gamma_1(\text{Sim}_{ij} \times \text{Metered}_{jt}) + \beta\,\text{Metered}_{jt} + \eta\,\text{JFK}_{jt} + \alpha_i + \delta_j + \lambda_{\text{hr}} + \mu_{\text{dow}} + \varepsilon_{ijt}, \qquad (3)$$

where y_{ijt} is the fare or time of a trip by driver i to destination block j at time t; Sim_{ij} is the demographic similarity between driver i and block j (Section 2.3); and Metered_{jt} equals one for metered trips, meaning all LaGuardia trips and JFK trips off the flat rate, and zero for JFK–Manhattan flat-rate trips. Fare and time are the two welfare-relevant

outcomes: the fare is a transfer from the passenger to the driver, and the added travel time is deadweight loss. Driver fixed effects α_i absorb every time-invariant driver characteristic, including skill and racial attitudes, so γ_1 reflects how the same driver treats different passengers. Destination-block fixed effects δ_j hold the endpoint, and hence the optimal route, fixed. Hour-of-day and day-of-week fixed effects absorb systematic traffic variation. Metered and JFK-pickup indicators enter as main effects.

The comparison is within driver, and nearly every driver contributes to it: 88.9% of drivers in the analysis sample complete both metered and flat-rate trips, and 90.8% pick up at both airports, so γ_1 rests on the whole driver population rather than a selected subset. Holding the driver fixed also addresses the selection concerns that Guryan and Charles (2013) raise about cross-sectional comparisons in discrimination research. What it does not settle is whether the gap reflects taste or a driver's sharper familiarity with own-group neighborhoods; Section 6.1 takes up that alternative and rejects it.

The flat rate provides a built-in placebo, and I impose it as a restriction. Similarity cannot operate through routing on flat trips, so every coefficient specific to flat trips should equal zero. Equation (3) therefore omits the similarity main effect, which acts only on flat trips. The data validate the restriction: on flat-rate trips the fare slope on similarity is statistically zero, as Figure 3 shows and Appendix Table C.1 estimates.

The coefficient has a precise reading. The favoritism gradient γ_1 measures how much less a passenger pays, in dollars or minutes, per unit of similarity to the driver on metered trips; Section 4 formalizes a negative γ_1 within the nepotism model. The identifying assumption is correspondingly narrow: a confound must move fares or times differentially by trip type for the same driver, at the same destination block, at the same hour. Traffic, weather, destination geography and driver skill all fail that test because both trip types face them identically. The estimator uses no calendar variation and only pre-entry trips, so no part of γ_1 rests on timing.

One might prefer airport×block fixed effects, which would hold the exact route from each airport to each block. They cannot enter this specification: a trip is on the flat rate exactly when it is a JFK pickup to Manhattan, so the airport×block cell determines Metered almost perfectly ($R^2 = 0.95$), and identifying γ_1 off the few cells that contain both fare regimes yields unstable estimates. The JFK main effect η absorbs the first-order JFK-versus-LaGuardia difference in route length instead. The concern that motivates those fixed effects, that the route to a block differs by origin airport, is met directly in Appendix C.1: restricting to metered trips makes Metered constant, airport×block fixed effects become admissible, and the favoritism gradient survives under this strictest destination control (Table C.5).

3.2 Main Results

Table 3 reports the estimates; columns (1)–(2) use every pre-entry trip in the analysis sample. The favoritism gradient is negative, precisely estimated and economically meaningful. Under expected match, a passenger one unit more similar to the driver paid \$0.44 less for the same metered trip and arrived 26 seconds sooner. One unit spans the measure's full range, from no chance to certainty that passenger and driver share a race. Own-group exposure gives the same picture (Appendix Table C.6).

The advantage runs through speed, not distance. On the same pre-entry trips, the distance gradient is a precise null (Appendix Table C.2): favored passengers cover the same miles in less time and at lower cost, consistent with faster streets and less time-padding rather than shorter routes. That the channel is time rather than distance fits both the meter and the setting. The meter charges for distance above 12 mph and for time below it, so a driver can inflate a fare either by adding miles or by taking a slower path over the same miles; to a known destination, extra miles are conspicuous and disputable while slow minutes are neither, and the discretion a driver retains to a fixed block lies mostly in speed. A driver who merely knew faster routes to own-group neighborhoods could produce the same speed advantage, but Section 6.1 rules this out: the gap vanishes on the trips a driver knows best, the reverse of what route knowledge would predict.

3.3 Where Favoritism Concentrates: Residential Destinations

Favoritism concentrates in trips that end where people live. I measure this by the residential share of the destination block (Section 2.4), which runs from zero in a purely commercial block to one in an entirely residential one. Columns (3)–(4) of Table 3 re-estimate Equation (3) with the favoritism term interacted with residential share; δ_1 is the coefficient on that interaction. The lower-order interaction of metered and residential share enters as a control, and the block fixed effects absorb the residential-share main effect. The favoritism gradient at residential share r is $\gamma_1 + \delta_1 r$: it equals γ_1 in a fully commercial block and $\gamma_1 + \delta_1$ in a fully residential one. The pooled gradient of columns (1)–(2) is a weighted average of these endpoints, which is why it sits between them.

Favoritism steepens sharply with residential share on the 73.5% of analysis-sample trips with matched land-use records. The commercial endpoint is weak: on fares, $\gamma_1 = -0.25$ under expected match, and on time there is no negative gradient at all. The interaction δ_1 is large and highly significant for both outcomes, and the implied fully-residential gradient $\gamma_1 + \delta_1$ reaches -0.95 on fares and -0.97 minutes on time, roughly twice the pooled estimate; own-group exposure gives the same endpoints (Appendix Table C.6). A binary split at half-

Table 3: Favoritism in Route Efficiency Before Green Cab Entry

	All destinations		By residential share	
	Fare ($)	Time (min)	Fare ($)	Time (min)
	(1)	(2)	(3)	(4)
Similarity × Metered (γ_1)	-0.445***	-0.429***	-0.252**	0.160
	(0.068)	(0.067)	(0.105)	(0.121)
Similarity × Metered × Residential share (δ_1)			-0.704***	-1.125***
			(0.144)	(0.159)
Metered and JFK controls	Yes	Yes	Yes	Yes
Metered × Residential share control	No	No	Yes	Yes
Driver, block, hour, day-of-week FE	Yes	Yes	Yes	Yes
Observations	2,958,327	2,958,327	2,174,063	2,174,063

Notes: Within-driver estimates of favoritism on pre-entry trips only (January–July 2013), using expected match, the paper's primary similarity measure. The estimating equation (Eq. 3) contains no post-entry terms; the response to green cab entry is estimated in Table 5. The flat-rate placebo is imposed as a restriction: similarity cannot operate through routing on JFK–Manhattan flat-fare trips, so the similarity main effect is omitted and flat trips serve as the within-driver control group. Columns (1)–(2) use all pre-entry route-efficiency trips. Columns (3)–(4) interact the favoritism term with the *residential share* of the dropoff block (residential built floor area over total built floor area, from PLUTO tax-lot records) on the land-use-matched sample; the favoritism gradient at residential share r is $\gamma_1 + \delta_1 r$, so γ_1 in columns (3)–(4) is the gradient at a fully commercial block and $\gamma_1 + \delta_1$ at a fully residential one. The residential-share main effect is absorbed by the block fixed effects. Own-group exposure gives the same picture (Appendix Table C.6). Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

residential gives the same picture (Appendix Table C.9).

I read the concentration primarily as validation of the similarity measure. Similarity is imputed from the resident racial composition of the dropoff block. Almost no one lives at a commercial destination, so resident composition is a poor proxy for who steps out of the cab there, and classical measurement error attenuates the gradient toward zero. The estimates behave exactly as this logic predicts: favoritism is strong where residents plausibly are the passengers and fades to noise where they are not. That pattern is what genuine passenger demographics produce; a spurious block-level correlate has no reason to respect the boundary.

The concentration also admits a behavioral reading: drivers may have more scope or more motive to favor similar passengers on residential trips. The two readings cannot be fully separated, because residential share also proxies for how well similarity is measured. A horse race that adds block population to the interaction rejects the sample-size form of measurement error, since the residential-share gradient survives conditioning on population while population itself enters small and wrong-signed on fares (Appendix C.1). The

proxy-relevance form, under which residents need not represent passengers in commercial blocks, remains observationally equivalent to the behavioral channel. Whether the behavioral channel is taste or an alternative such as familiarity is one of the mechanism questions Section 6 takes up; here I treat the fully-residential gradient of -0.95 as an upper bound on de-attenuated favoritism.

4 A Model of Discrimination as Nepotism

Section 3 established that the same driver, delivering passengers to the same block at the same hour, charges demographically similar passengers less and gets them there faster. What preference produces this gap? Two canonical models fit the pre-entry fact equally well. Under Becker (1957), the driver holds animus toward dissimilar passengers and gives them a slower, costlier trip than profit alone would call for; the gap is a fixed *discrimination coefficient*. Under the nepotism framework of Goldberg (1982), adapted here from hiring to quality provision, the driver instead favors similar passengers, giving them a faster, cheaper trip than profit alone would call for, at a cost he bears himself in forgone fares. Disfavored passengers meet indifference, not hostility. At a fixed level of competition the two models are observationally equivalent: they produce the same quality gap.

The models diverge in one place: how the gap responds to the provider's margin. Both tastes cost the provider profit, but only one is bounded by what the provider can afford. Favoring lifts service above the profit-maximizing level, which the provider can fund only while the margin leaves surplus to spare; as competition compresses the margin, the gift becomes unaffordable and the nepotism gap shrinks. Penalizing lowers service below the profit-maximizing level, which remains feasible at any margin, so the Becker gap holds fixed. This is *leveling down*: competition narrows the gap not because the disfavored gain but because the favored lose their advantage, and discrimination becomes an *equilibrium object*. Green cab entry compressed margins, and Section 5 uses that shock to tell the two models apart.

4.1 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: the provider's monetary profit from a transaction

[5] In the taxi setting, quality maps to route efficiency: the fare and travel time on a trip to a given destination, which the driver sets through the route he takes.

of quality q is $mq - \frac{c}{2}q^2$, where $m > 0$ parameterizes the margin and $c > 0$ governs the cost of quality. Profit is maximized at $q^* = m/c$; providing quality beyond this level sacrifices profit. 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

$$\Pi_g(q; m) = m \cdot q - \frac{c}{2}\, q^2 + \phi \cdot \mathbf{1}[g = F] \cdot q, \tag{4}$$

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 faster, cheaper route for a same-group passenger, giving up fare revenue rather than imposing a penalty 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 D. 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]

4.2 The Leveling-Down Prediction

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

[6] The profit function is a reduced form that captures the net return to quality from all sources. In the taxi setting, more efficient routing saves time and fuel but reduces metered fare revenue; the profit-maximizing quality m/c reflects the point where these forces balance. Green cab entry reduces m by lowering trip frequency and overall driver earnings.

[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.

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 he can afford.

Proposition 2 (Leveling down)**.** *When the zero-profit 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.

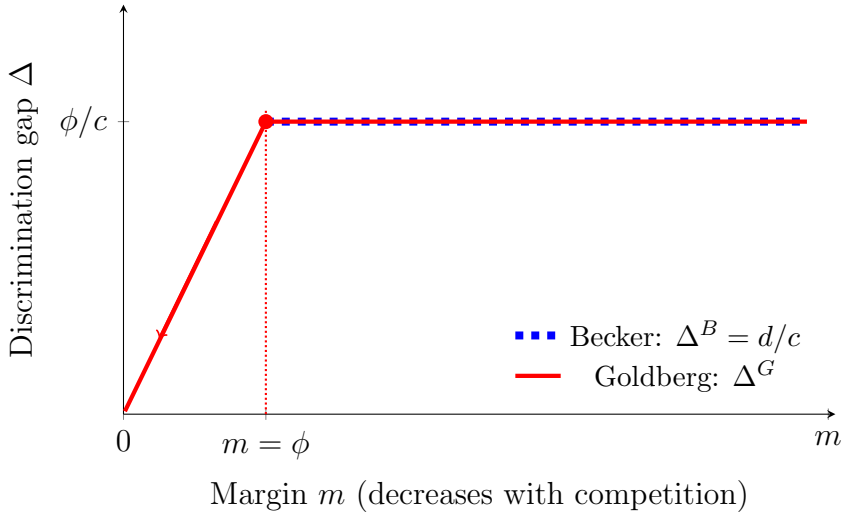
The mechanism is straightforward. Preferential treatment under nepotism is a luxury good: it requires surplus to fund. High margins let the provider sacrifice profit for the psychic reward of serving favored customers. When margins fall, the generosity is curtailed. The gap depends on the margin (m),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

The Becker (animus) gap, by contrast, is d/c, a constant that does not depend on margins. Animus lowers disfavored quality below the profit-maximizing level, a deviation the provider can sustain at any margin, so no such limit ever binds, and competition lowers quality for both groups at the same rate. The leveling-down asymmetry belongs to favoring, which alone pushes quality above the profit-maximizing level and so alone runs into the surplus that competition erodes. Formally, $\partial\Delta^B/\partial m = 0$ while $\partial\Delta^G/\partial m = 1/c > 0$ when the constraint binds ($\phi > m$): a margin shock compresses the nepotism gap but leaves the animus gap untouched (Appendix D.3). The empirical finding that the gap narrows under competition is therefore inconsistent with a pure animus mechanism. Figure 4 illustrates the distinction.

The closed forms come from the quadratic cost, but the leveling-down prediction does not. Under any strictly convex cost the zero-profit cap tightens as margins fall and the nepotism gap narrows, whereas the Becker constant-gap result is the fragile one, holding only for exactly quadratic costs and reversing under more convex costs (Appendix D.5).

The same force that narrows the gap also lowers welfare. As competition compresses the margin, the favored lose their quality advantage and the gap shrinks; but average quality falls at the same time, because the surplus that funded generous routing shrinks for everyone. Aggregate welfare $W^*(m) = vm(1+\mu)/c - p$ rises with the margin, so it falls as competition intensifies. Equity and efficiency move in opposite directions along the margin: the market

Figure 4: Discrimination Gap Under Competition: Becker vs. Goldberg



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 zero-profit constraint binds ($m < \phi$). Discrimination is an equilibrium object under Goldberg: it responds to market structure.

buys a smaller gap only by degrading service for everyone. This equity-efficiency tension is the defining feature of leveling down. Section 7 quantifies it, and Appendix D.4 traces the frontier.

The model also predicts geographic sorting on the extensive margin, that providers favor locations dense with same-group customers and that competition erodes this selectivity; because this paper studies the intensive margin, I develop the sorting prediction and its evidence in Appendices D.2 and E.

The mapping to the setting is direct: quality is route efficiency, the nepotism parameter ϕ is the psychic benefit a driver derives from serving same-group passengers, and the margin m is per-trip profitability, so the favoritism gradient of Section 3 is the model's unconstrained gap $\Delta = \phi/c$ (Proposition 1). Green cab entry lowers m. Section 5 shows that the entry compressed driver margins and tests the leveling-down prediction against its Becker alternative.

5 Competition and Leveling Down

5.1 The Shock to Driver Margins

The model's mechanism runs through margins, so I first document its input: green cab entry compressed what yellow drivers earned. Table 4 reports within-driver comparisons of shift

and monthly outcomes before and after entry.[8] Monthly income fell $107, about a third of a single shift's earnings, and shift income fell $4.47. The squeeze ran through the number of passengers: drivers completed about one fewer trip per shift, a 5.2% drop, and waited 4.9% longer between fares. They partly offset it by taking fewer but longer fares, so income per trip rose 4.2% and the average trip grew 3.9% longer in distance and 5.5% in time, consistent with more selective acceptance and more padding under earnings pressure. Yellow taxis did not retreat from the contestable territory: the green-zone share of their pickups edged up rather than down. These comparisons are descriptive; they size the margin shock but do not carry the paper's identification, which comes below.

The shock also pushed drivers out of the industry. Before entry the yellow-cab workforce grew by about 549 drivers a month on net; after entry it shrank by about 546 a month, as arrivals slowed and departures rose (Appendix B). Among the roughly 39,400 drivers active before August 2013, 9.1% had left by November, and the departures were not random: exit concentrated among the drivers most exposed to the green zone and earning the least, the margin the mechanism operates on. That drivers left, and left selectively, matters for what follows. If the drivers who most favored own-group passengers were also the ones who left, the post-entry narrowing could reflect a changed mix of drivers rather than changed behavior. Section 5.2 confronts this directly, re-estimating on the balanced panel of drivers present both before and after entry. Even after the exits, those who remained completed fewer trips and earned less per shift than before, so the entrants more than offset the thinning of the yellow fleet.

5.2 Leveling Down

Section 3 showed favoritism before entry. The question now is how it responded when entry compressed margins. Figure 5 shows the answer in raw form. It splits metered trips into those carrying more- and less-similar passengers and plots each group's fare and trip time before and after entry, holding the driver and the destination fixed. Before entry, more-similar passengers pay less and arrive sooner, the favoritism of Section 3. After entry, service degrades for both groups, but it degrades more for the previously favored: the fare gap collapses and the time advantage disappears. The gap narrows by leveling down.

Two objections attend a before-and-after comparison like this. The change could be seasonal, common to the second half of any year. Or it could reflect a changed mix of

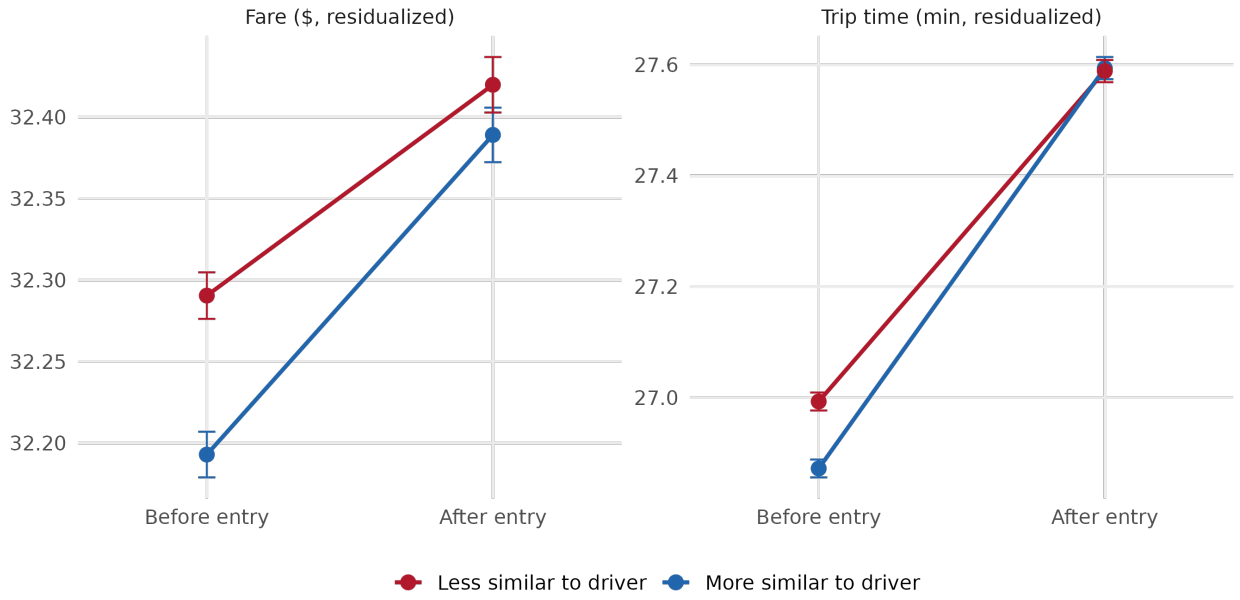
[8] The specification is $Y_{it} = \theta\,\mathrm{PostEntry}_t + \alpha_i + \mu_d + \varepsilon_{it}$, where Y_{it} is the outcome for driver i on shift t, α_i are driver fixed effects, and μ_d day-of-week fixed effects. The sample includes every driver in the 2013 trip records, with or without a race classification; the discrimination analysis of Sections 3 and 5.2 requires classified drivers (Section 2.3).

Table 4: Effect of Green Cab Entry on Yellow Cab Driver Outcomes

	(1)	(2)	(3)
Panel A: Earnings			
	Shift Income	Hourly Wage	Income/Trip
Post entry	-4.47***	-0.11***	0.68***
	(0.23)	(0.03)	(0.01)
Mean dep. var.	317.47	37.63	16.20
Observations	6,840,262	6,840,173	6,840,262
R^2	0.065	0.024	0.147
Panel B: Work Patterns			
	Trips/Shift	Wait Time	Shift Length
Post entry	-1.13***	1.38***	-0.08***
	(0.01)	(0.02)	(0.01)
Mean dep. var.	21.57	28.17	8.70
Observations	6,840,262	6,788,631	6,840,173
R^2	0.399	0.404	0.284
Panel C: Trip Characteristics			
	Trip Distance	Trip Time	GZ Share
Post entry	0.130***	0.750***	0.0072***
	(0.003)	(0.006)	(0.0002)
Mean dep. var.	3.357	13.593	0.1300
Observations	6,840,262	6,840,262	6,760,487
R^2	0.423	0.368	0.345
Panel D: Extensive Margin (Driver-Month)			
	Shifts/Month	Monthly Income	
Post entry	-0.08***	-107***	
	(0.02)	(9)	
Mean dep. var.	20.85	6,619	
Observations	328,071	328,071	
R^2	0.634	0.558	

Notes: This table reports the effect of green cab entry (August 2013) on yellow cab driver outcomes. Each column is a separate regression of the outcome on a post-entry indicator. All specifications include driver fixed effects; Panels A–C also include day-of-week fixed effects. Panels A–C are at the driver-shift level; Panel D is at the driver-month level. In Panel C, Trip Distance and Trip Time are shift-level means of per-trip miles and minutes, and GZ Share is the share of the shift's pickups originating in the green (contestable) zone. Standard errors clustered at the driver level in parentheses. Sample: all 2013 yellow-taxi shifts by classified drivers. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure 5: Leveling Down: Metered Fares and Trip Times by Passenger–Driver Similarity, Before vs. After Entry



Mean metered fare (left panel) and trip time (right panel) for trips carrying more- versus less-similar passengers, before and after green cab entry. Trips are split at the median of expected match; outcomes and similarity are residualized on driver, destination-block, hour-of-day and day-of-week fixed effects, so the comparison holds the driver and the destination fixed. Bars are 95% confidence intervals. Before entry, more-similar passengers pay about \$0.10 less and arrive 0.12 minutes sooner. After entry, service degrades for everyone, but degrades more for the previously favored group: the fare gap collapses to about \$0.03 and the trip-time gap disappears. Within-group mean similarity is identical before and after entry, so the narrowing is behavioral, not a composition shift. The median split is descriptive; Table 5 estimates the gradient and its change (γ_1, γ_3). The own-group-exposure version appears in Appendix Figure C.2.

drivers, since entry pushed the most-exposed drivers out (Section 5.1). The rest of this section answers both, with a within-driver estimator that controls for more than the figure can show and a cross-driver test that seasonality cannot mimic.

I extend Equation (3) to the full year of 2013 and let every metered coefficient shift after entry:

$$\begin{aligned} y_{ijt} = {} & \gamma_1(\text{Sim}_{ij} \times \text{Metered}_{jt}) + \gamma_2(\text{Metered}_{jt} \times \text{Post}_t) + \gamma_3(\text{Sim}_{ij} \times \text{Metered}_{jt} \times \text{Post}_t) \\ & + \beta\,\text{Metered}_{jt} + \eta\,\text{JFK}_{jt} + \alpha_i + \delta_j + \lambda_{\text{hr}} + \mu_{\text{dow}} + \varepsilon_{ijt}, \end{aligned} \tag{5}$$

where Post_t indicates trips after July 2013 and the remaining terms are as in Equation (3). The flat-rate placebo is again imposed as a restriction: the terms that act only on flat trips (the similarity main effect, Sim × Post, Post, and JFK × Post) are omitted. The three coefficients read cleanly. γ_1 is the pre-entry favoritism gradient, close to the estimate from

pre-entry trips alone (Table 3): the fare gradients match to the cent, and the time gradients differ modestly because the fixed effects are estimated on different samples. γ_2 measures the post-entry change in metered service common to all passengers, and γ_3 the change in the favoritism gap. Table 5 reports the estimates. The Full columns use the whole 5.2 million-trip analysis sample; the Balanced columns re-estimate on continuing drivers, those with at least ten metered trips in each period, which holds the set of drivers fixed and removes any role for the compositional change entry induced.

Table 5: Favoritism and Leveling Down in Route Efficiency: Within-Driver Estimates

	Fare ($)		Time (min)	
	Full	Balanced	Full	Balanced
Similarity × Metered (γ_1)	-0.439***	-0.441***	-0.507***	-0.503***
	(0.066)	(0.068)	(0.065)	(0.067)
Metered × Post (γ_2)	0.103***	0.102***	0.546***	0.573***
	(0.015)	(0.015)	(0.018)	(0.018)
Similarity × Metered × Post (γ_3)	0.236***	0.214***	0.334***	0.310***
	(0.060)	(0.061)	(0.070)	(0.071)
Observations	5,206,073	4,951,219	5,206,073	4,951,219
Drivers	30,616	22,076	30,616	22,076
Within R^2	0.534	0.533	0.242	0.241
Metered, JFK main effects	Yes	Yes	Yes	Yes
Driver FE	Yes	Yes	Yes	Yes
Destination-block FE	Yes	Yes	Yes	Yes
Hour-of-day, day-of-week FE	Yes	Yes	Yes	Yes

Notes: Within-driver estimates of Equation 5 for trips to destination block j, using expected match; own-group exposure and cosine similarity appear in Appendix Table C.4. Metered equals one for metered trips (all LaGuardia trips and JFK trips off the flat rate) and zero for JFK–Manhattan flat-rate trips ($52 regardless of route); Post indicates trips after July 2013. Similarity is measured at the dropoff Census block. The Full columns use the full route-efficiency sample; the Balanced columns restrict to continuing drivers with at least ten metered trips in each period, so the stability of γ_3 shows that the post-entry narrowing is within-driver rather than compositional. A negative γ_1 means cheaper and faster service for similar passengers; $\gamma_2 > 0$ means service degrades for all after entry; $\gamma_3 > 0$ means the gap narrows. Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Green cab entry eroded the surplus that funded favoritism, and drivers responded as the model predicts. After entry, metered service degraded for every passenger: $\gamma_2 > 0$, with fares rising $0.10 and trips lengthening about half a minute under expected match ($0.15 and 0.61 minutes under own-group exposure). At the same time, the favoritism gap narrowed: $\gamma_3 > 0$ and highly significant for both similarity measures and both outcomes, with fares +$0.24 and time +0.33 minutes under expected match. The pre-entry fare gap

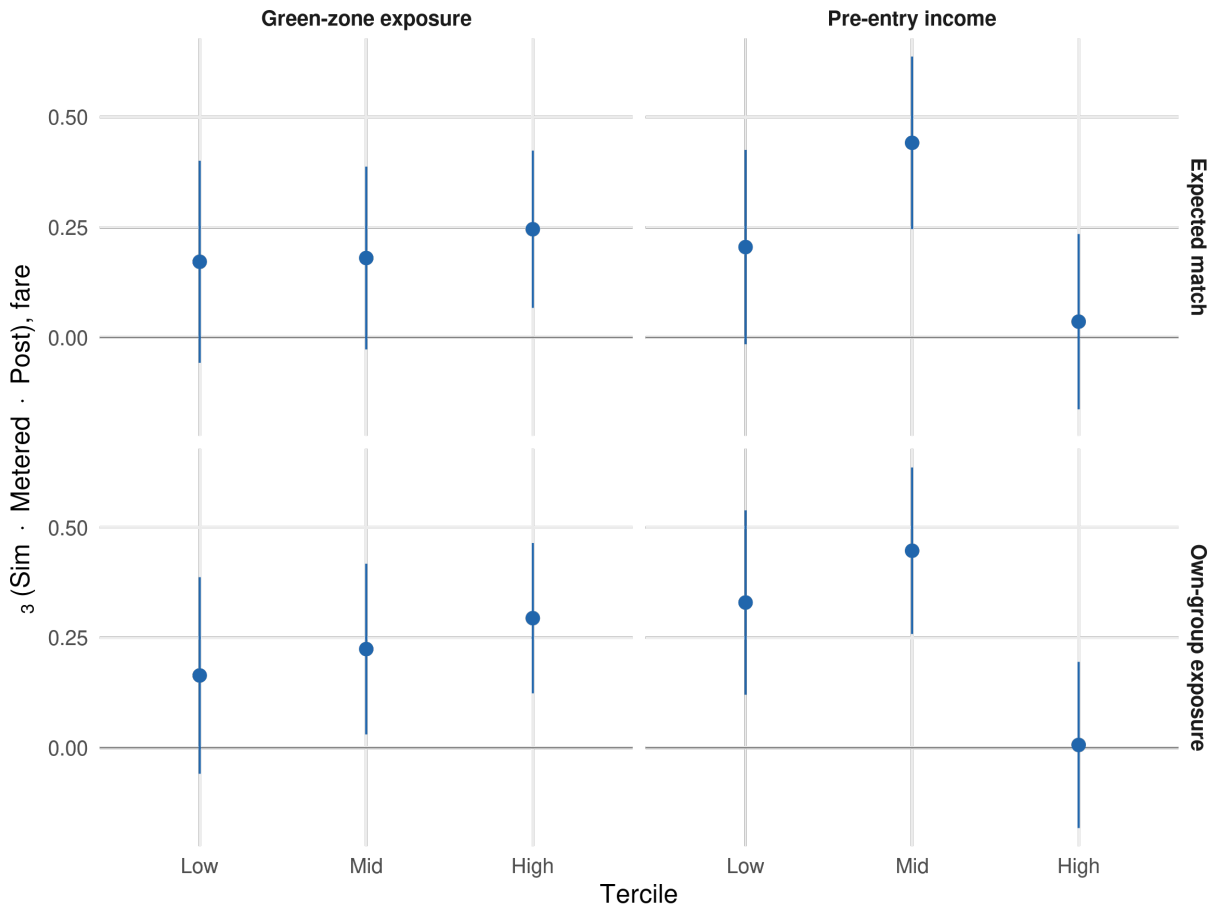
of \$0.44 shrinks to \$0.20, a 54% compression under expected match and 69% under own-group exposure, while the level of service falls for everyone. This is the signature of leveling down (Proposition 2): competition narrows the gap not by improving service for disfavored passengers but by withdrawing the treatment that similar passengers had enjoyed. It also separates the two models. A taste-based account predicts a gap invariant to the driver's margin; nepotism predicts that the gap collapses when the surplus that funds it disappears. The data show the second pattern.

Is the narrowing competition or seasonality? The two predict different things, and the data choose. A common calendar force would move the gap uniformly across drivers. Nepotism predicts the opposite: preferential treatment is a luxury funded by the surplus competition erodes, so the compression must concentrate among the drivers whose margins are thinnest. That concentration identifies the competition effect, and it is what I find. Figure 6 re-estimates γ_3 within terciles of two predetermined driver characteristics, each measured before entry: the share of dropoffs in the contestable green zone, and average shift income (Appendix Table C.11 reports the full estimates). The narrowing rises with green-zone exposure, from an insignificant \$0.16 among the least-exposed drivers to \$0.29 among the most-exposed, and it depends still more sharply on income: low- and middle-income drivers compress the fare gap by \$0.33 to \$0.45, while high-income drivers do not compress it at all, for every similarity measure and outcome. When the surplus that funds favoritism shrinks, the drivers with the least slack cut it first, and those who can still afford it do not. Seasonal forces fall on high- and low-exposure drivers alike, so this cannot be a seasonal artifact; it is the fingerprint of the competitive mechanism, and it doubles as a test of which drivers can still afford to favor their own group. The compression also follows the geography of favoritism itself, concentrating in residential destinations where the pre-entry gradient is steepest (Appendix Table C.7).

A cross-sectional test suits this setting better than a time series. A time-series design would want a sharp entry date and a clean pre-period, and neither exists here. Green cab entry was a ramp, growing roughly twentyfold from August to December rather than switching on. The September 2012 fare increase, which raised the JFK flat fare from \$45 to \$52, contaminates any pre-period reaching into 2012. And the driver records that permit race classification exist only for 2013, so a multi-year discrimination series is unavailable in any case. The heterogeneity design sidesteps all three, because it does not use the timing of entry at all.

Two caveats temper the interpretation. Green-zone exposure and pre-entry income are correlated, so the two splits are not independent confirmations, and the income gradient could partly reflect other differences between high- and low-earning drivers rather than affordability

Figure 6: Competition Response by Driver Characteristic: γ_3 by Green-Zone Exposure and Pre-Entry Income Tercile



alone. But both predetermined splits point the same way, and it is the concentration of the response, not its average level, that distinguishes competition from seasonality.

The narrowing is a behavioral response, not the selective exit of discriminating drivers. Section 5.1 showed that exit fell hardest on the most-exposed, lowest-earning drivers, plausibly those who favored own-group passengers most, so a changed composition could in principle produce the post-entry gap. It does not. The Balanced columns of Table 5 re-estimate on drivers present in both periods, 72% of drivers and 95% of metered trips, and the fare γ_3 barely moves: from +0.24 to +0.21 under expected match, and from +0.27 to +0.25 under own-group exposure. The same drivers narrowed the gap.

6 Mechanisms and Robustness

I have interpreted the favoritism gradient as a taste for serving one's own group. Three other explanations could produce the same within-driver pattern without any such preference: a driver may know own-group neighborhoods better, may be recovering an expected tip, or may be learning a destination's demographics with experience. This section takes up each; none accounts for the gap.

6.1 Local Knowledge

The first alternative is local knowledge. Drivers may route more efficiently to neighborhoods they know well, and through residential homophily those neighborhoods are disproportionately their own group's. The gradient would then reflect route familiarity, not preference. The data do not record what a driver knows, so I test the idea indirectly. I take each driver's home neighborhood to be the modal community district of their first pickup of a shift, and I flag the 3.8 percent of trips that end there. Excluding these home trips leaves the gradient essentially unchanged, and if anything larger: in Panel A of Table 6 it moves from -0.44 to -0.47 in fares and from -0.51 to -0.52 in minutes. Knowledge of the destination does not generate the gap.

Favoritism is weakest exactly where a driver's local knowledge is greatest. If the gradient were route knowledge, it would be largest on trips into the driver's own neighborhood. It is instead smallest there. Panel B interacts the gradient with an indicator for home-neighborhood trips: away from home the gradient is the full -0.46 in fares and -0.53 in minutes, but the Sim$\times$Metered$\times$Home interaction is large and positive, $+0.91$ and $+0.81$, so the gap essentially closes on the driver's home turf. This is the reverse of what route knowledge predicts. The home-neighborhood proxy is coarse, and a driver knows far more streets than the area around a shift's first fare, so it does not measure knowledge directly. But the pattern runs against familiarity rather than with it: where knowledge is deepest, the gap is smallest.

6.2 Rational Tip Extraction

Drivers do not recover the fare premium through tips. A driver who expects dissimilar passengers to tip less could pad the metered fare to make up the difference, so that the higher fare serves profit rather than preference. This story makes a sharp prediction: if the driver were recovering an expected tip, the tip would fall by as much as the fare rose, and the tip gradient would be equal and opposite to the fare gradient, near $+0.40$. Table 7 estimates

Table 6: Ruling Out Local Knowledge: Favoritism Off the Driver's Home Neighborhood

	(1) Fare ($)	(2) Time (min)
Panel A: Favoritism gradient γ_1 (Sim×Metered)		
Full sample	-0.439***	-0.507***
	(0.066)	(0.065)
Excluding home dropoffs	-0.471***	-0.522***
	(0.067)	(0.066)
Panel B: Home-interaction specification (full sample)		
Sim×Metered (non-home trips)	-0.464***	-0.528***
	(0.066)	(0.066)
Sim×Metered×Home	0.914***	0.807***
	(0.188)	(0.226)
Driver, block, hour, day-of-week FE	Yes	Yes
Home-dropoff share of trips	3.8%	
Observations, full sample	5,206,073	
Observations, excluding home dropoffs	5,009,566	

Notes: Tests whether route-efficiency favoritism is local route knowledge of the driver's own neighborhood rather than taste, using expected match, the paper's primary similarity measure (own-group exposure agrees; Appendix Table C.12). *Home* is the driver's modal first-of-shift pickup community district (CDTA); a trip is a home dropoff if it ends there. Panel A reports the favoritism gradient γ_1 from the triple-difference (Eq. 5): the full-sample estimate reproduces Table 5, and excluding home-dropoff trips leaves it essentially unchanged, so favoritism is not driven by home-route familiarity. Panel B instead adds a Sim×Metered×Home interaction to the full-sample specification: the first row is the gradient on non-home trips and the second the additional effect on home trips. The implied home-trip gradient is their sum, +0.45 dollars and +0.28 minutes: the gap closes on the routes the driver knows best, the reverse of the familiarity prediction. Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

both on airport metered credit-card trips, where tips are recorded. The fare gradient is -0.40 (column 1), but the tip gradient is -0.03 and statistically indistinguishable from zero (column 2). Dissimilar passengers pay the full fare premium and recover none of it in tips: across the range of similarity the fare moves about twelve times as much as the tip. The padding is not tip recovery. Nor is it the race-blind extraction a monopsony or fraud story implies, because that would not vary with passenger similarity.

6.3 Learning

Experience does not erode the favoritism gradient. A driver who reads a destination's demographics as a signal of route difficulty or tips should update that belief as trips to a neighborhood accumulate, so the gradient should shrink with experience. This is statistical discrimination through learning. The airport setting is well suited to test it, because the

Table 7: Ruling Out Rational Tip Extraction: Fares Respond to Similarity, Tips Do Not

	(1) Fare ($)	(2) Tip ($)
Similarity (expected match)	-0.402***	-0.033
	(0.111)	(0.043)
Driver, block, hour, day-of-week FE	Yes	Yes
Mean dep. var.	33.24	6.64
Observations	573,224	573,224

Notes: Estimated on airport metered credit-card trips, the same sample in which route extraction is documented (Section 3), using expected match, the paper's primary similarity measure; own-group exposure and cosine similarity give the same picture (Appendix Table C.13). Each column regresses its outcome on similarity within driver and destination block, with hour and day-of-week fixed effects, so the two columns compare the same trips. If the fare gradient were profit-maximizing recovery of an expected tip shortfall, total payments would be equalized across passengers and the tip coefficient would be roughly equal and opposite to the fare coefficient (+0.40); instead it is −0.03 and statistically indistinguishable from zero. Moving from the 10th to the 90th percentile of similarity lowers the fare by $0.12 per trip but the tip by only $0.01, a ratio of 12 to 1: dissimilar passengers pay the full fare premium and recover none of it in tips. The extraction cannot be rationalized as expected-tip recovery, consistent with taste-based favoritism rather than profit-motivated extraction. Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

queue assigns passengers: a driver cannot choose how much contact to have with any group, so the endogenous-contact channel through which beliefs can fail to converge (Lepage, 2024) does not operate here. Table 8 adds the test directly. It restricts the sample to drivers first observed in March 2013 or later, whose neighborhood experience is measured from their first trip, and interacts the gradient with prior visits to the dropoff neighborhood and with cumulative shifts (columns 2 and 4). Both interactions are near zero, and if anything the gradient grows with tenure. This cohort is small, 4,076 drivers and 5.5 percent of trips, so the estimates are less precise than the headline, and they speak to belief-updating within 2013 rather than over a career. Within those limits, experience does not attenuate favoritism, which is what a taste predicts and learning does not.

7 Economic Magnitudes

Table 9 translates the estimates into aggregate dollar magnitudes for 2013. The airport coefficients (Table 5) carry the identification; the market-wide figures extend them and should be read as a calibration, not as causal estimates of their own. The yellow-taxi market generated 171 million trips in 2013: 4.5 million metered airport trips and 167 million non-airport trips. The extension makes one assumption for the discrimination gap and none for the degradation. Off the airports the gap cannot be cleanly identified, because drivers

Table 8: Ruling Out Learning: Experience Does Not Attenuate Favoritism

	Fare ($)		Time (min)	
	(1)	(2)	(3)	(4)
Panel A: Expected match				
Sim×Metered	-0.208	-0.042	-0.385	-0.200
	(0.270)	(0.289)	(0.325)	(0.357)
Sim×Metered×Post	0.537*	0.800**	0.811**	1.182***
	(0.290)	(0.365)	(0.361)	(0.443)
Sim×Metered×CDTA experience		0.024		0.022
		(0.038)		(0.050)
Sim×Metered×General experience		-0.500		-0.565
		(0.340)		(0.417)
Panel B: Own-group exposure				
Sim×Metered	-0.350	-0.153	-0.439	-0.086
	(0.267)	(0.283)	(0.309)	(0.346)
Sim×Metered×Post	0.603**	0.871**	0.857**	1.323***
	(0.277)	(0.357)	(0.347)	(0.421)
Sim×Metered×CDTA experience		0.014		-0.010
		(0.038)		(0.048)
Sim×Metered×General experience		-0.508		-0.769*
		(0.361)		(0.411)
Experience interactions	No	Yes	No	Yes
Driver, block, hour, day-of-week FE	Yes	Yes	Yes	Yes
New drivers		4,076		
Observations		288,067		

Notes: Tests whether neighborhood experience attenuates the favoritism gradient, as statistical discrimination predicts (Arrow, 1973; Phelps, 1972). The sample is new drivers, first observed in March 2013 or later, whose within-2013 experience is measured without left-censoring; it is 5.5% of route-efficiency trips. Columns (1) and (3) estimate Equation 5 on this cohort; columns (2) and (4) interact Sim×Metered with CDTA experience (prior pickups and dropoffs in the dropoff neighborhood) and general experience (cumulative shifts), each per 100 units, so the Sim×Metered row there is the gradient at zero experience. The interactions are near zero, and the test is well powered: erasing the cohort's baseline fare gradient over the median driver's 140 prior visits would take an interaction of +0.15 per 100 visits, which column (2) rejects by more than 3 standard errors. This also rules out the asymmetric learning of Lepage (2024). Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

choose whom to pick up, so I scale the airport per-trip gap to non-airport trip sizes, which are about a third as large. The estimated off-airport gradient brackets this scaled value: a specification-matched dropoff-block control yields a gap as large as the airport one, and a route-pinning tract-pair control yields a smaller one (Appendix Table C.14). The degradation is not extrapolated. It is the non-airport entry effect, a race-blind pre/post shift that is stable across both controls. Time is valued at $12.25 per hour, the Department of Transportation rate for personal local travel (U.S. Department of Transportation, 2016).

Three findings emerge. First, discrimination cost mismatched passengers $31.3 million in the year before green cab entry: $25.0 million in excess fares and $6.4 million in time costs. The per-trip gap is about three times larger at airports ($0.39–$0.44 in fares) than on the shorter non-airport trips ($0.13–$0.15), but non-airport volume is 37 times larger, so those trips account for $29.0 million of the total.

Second, competition reduced discrimination by $20.0 million, a 64 percent decline. At airports, where the flat-rate placebo pins the interpretation, the per-trip fare gap fell from $0.39–$0.44 to $0.12–$0.20.

Third, and central to the paper, the equality gain was purchased with a larger quality loss. Service degradation totaled $44.3 million: $28.9 million in higher fares and $15.4 million in added travel time, about 1.3 million passenger-hours. Degradation therefore exceeded the discrimination reduction by more than two to one. Degradation is the firmer of the two figures: it rests on the entry coefficient, which is stable across every specification in Appendix Table C.14. The reduction is the extrapolated one, and a more conservative route-pinned reading of the off-airport gap would make it smaller and the ratio larger. The main caveat runs the other way. Off airports no flat-rate control group exists, so the entry coefficient may absorb seasonal variation alongside competitive pressure; Section 5 shows that the airport response follows the cross-sectional exposure and income gradients that competition predicts and seasonality does not.

Two considerations frame the welfare reading. Green cab entry expanded service in previously underserved outer-borough neighborhoods, a gain on the access margin that trip records for yellow cabs cannot quantify, so the net welfare effect of entry remains open. And measurement error in the demographic proxy attenuates every underlying coefficient, so the discrimination and leveling-down totals are lower bounds.

8 Discussion and Conclusion

Becker's prediction that competition disciplines discrimination has shaped theory and policy for more than half a century. Competition did reduce discrimination in the New York City

Table 9: Calibration of Economic Magnitudes: Airport vs. Non-Airport Trips

	Airport	Non-Airport	Total
Trips (millions)	4.5	167	171.4
Per-Trip Discrimination Gap (Fare)			
Pre-entry	\$0.39–\$0.44	\$0.13–\$0.15†	—
Post-entry	\$0.12–\$0.20	\$0.04–\$0.07†	—
Aggregate Fare Transfers			
Pre-entry	\$1.9M	\$23.1M	\$25.0M
Post-entry	\$741K	\$9.1M	\$9.8M
Reduction	\$1.1M	\$14.0M	\$15.1M
Aggregate Time Costs			
Pre-entry (\$)	\$383K	\$6.0M	\$6.4M
Post-entry (\$)	\$89K	\$1.4M	\$1.5M
Summary			
Total discrimination (pre-entry)	**\$2.3M**	**\$29.0M**	**\$31.3M**
Total reduction (post-entry)	**\$1.4M**	**\$18.6M**	**\$20.0M**
Leveling Down (Service Degradation)			
Aggregate fare extraction	\$567K	\$28.3M	\$28.9M
Aggregate time cost (\$)	\$534K	\$14.9M	\$15.4M
Total leveling down	**\$1.1M**	**\$43.2M**	**\$44.3M**

Notes: Airport magnitudes use the estimates of Table 5: the per-trip discrimination gap is $|\gamma_1|$ (pre-entry) and $|\gamma_1 + \gamma_3|$ (post-entry), and leveling down is γ_2. †The non-airport discrimination gap scales the airport gap to non-airport trip sizes by the ratio of mean fares and durations, an extrapolation rather than an estimate; the estimated off-airport gradient of Appendix Table C.14 brackets it. Non-airport leveling down is the entry (post) main effect, which is not extrapolated and is stable across specifications. Aggregates apply the midpoint of the own-group and expected-match estimates to all 171 million yellow-taxi trips in 2013 (4.5 million airport-metered, 167 million non-airport). Time is valued at \$12.25 per hour, the U.S. Department of Transportation value for personal local travel (U.S. Department of Transportation, 2016). Measurement error in the demographic proxy attenuates every coefficient, so the magnitudes are conservative.

taxi market, but the mechanism complicates that reading. The gap narrowed not because disfavored passengers received better service, but because competition eroded the preferential treatment that similar passengers had enjoyed. The market bought each dollar of reduced discrimination with more than two dollars of degraded service.

The leveling-down result separates 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 out of available surplus, and competition erodes that surplus. A gap that narrows under competition is inconsistent with pure animus and consistent with nepotism that depends on market conditions. The narrowing concentrates among the drivers most exposed to green cab territory and those with the

lowest pre-entry earnings: preferential treatment is the first thing a driver cuts when the margin that funds it disappears. A seasonal or citywide shock cannot sort by exposure and by margin in this way, which is what identifies the response as competition rather than calendar.

Gap reduction alone is an incomplete measure of anti-discrimination policy. Discrimination operates on more than one margin, and constraining one can push differential treatment onto another. Interventions that equalize access, from algorithmic dispatch to anti-refusal rules, remove a driver's choice of whom to serve but leave routing discretion untouched, so discrimination can persist through service quality even as the access gap closes. A policy that scores itself by the access gap alone will miss this, and may count leveling down as success. Evaluation has to track both margins.

Several limitations bound these conclusions. The setting is a single regulated market, and generalization calls for care. Identification rests on airport trips, where the queue assigns passengers; drivers who choose to work the airport queue may differ from others, and extending the estimates to the wider street-hail market is a calibration rather than an identified result (Section 7). The learning test draws on the small cohort of drivers who entered in 2013, so it speaks to belief-updating within a driver's first months rather than over a career. Route efficiency captures one dimension of service quality, though driver fixed effects and uniform vehicle regulation limit what unobserved dimensions could confound it. A fuller welfare accounting would also weigh the access gains from green cab entry in previously underserved neighborhoods, which the yellow-cab records cannot value, so the net welfare effect of entry remains open.

Leveling down can arise wherever providers keep discretion over service quality and competition compresses the surplus that funds preferential treatment. Understanding discrimination then requires asking not only whether a gap closes, but how it closes and at what cost. In sum, competition can reduce discrimination. The pathway determines whether that is progress.

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 Data and Measurement

This appendix documents the data construction: driver race classification, passenger race inference, the imputation of missing Census data, sample construction, the external validity of the airport sample, and formal definitions of the similarity measures.

A.1 Driver Race Classification

Driver race classification relies on probabilistic surname matching with Bayesian Improved Surname Geocoding (BISG) rather than on self-reports. The approach has three advantages. First, it is objective and replicable, free of survey response and social-desirability bias. Second, BISG is validated: its classifications align well with self-reported race at the aggregate level (Elliott et al., 2009; McCartan et al., forthcoming). Third, BISG returns a probability distribution over races, and the similarity measures use the full distribution, so classification uncertainty propagates into the measures instead of being discarded. The ability to link behavior to persistent driver identifiers in New York City taxi administrative data has also enabled recent work documenting stable heterogeneity in detouring behavior (Tang, 2020).

Of the 43,191 drivers in the 2013 trip records, 32,478 receive a classification; the remainder lack a usable surname match. Individual-level misclassification nonetheless remains. Because the classification enters the similarity measures as probabilities, misclassification acts as classical measurement error in the regressors and attenuates the estimates toward zero; the main-text estimates are conservative in this respect.

A.2 Passenger Race Inference

Passenger race is inferred from the residential composition of the dropoff Census block (Section 2.3). Two sources of error follow. First, the passengers at a destination need not resemble its residents: business districts, tourist attractions and transportation hubs draw people from across the city. Second, even a resident passenger's race can differ from the block average.

The research design limits both errors. The airport setting mitigates the first: airport passengers travel to destinations across the entire city rather than the commercial cores where street hails concentrate, the queue-based dispatch keeps the destination mix exogenous to driver behavior, and Section 3.3 shows the estimates concentrate in residential destinations,

exactly where residents plausibly are the passengers. The second error is classical: within-block heterogeneity attenuates the estimates toward zero, so the main-text effects are lower bounds on the true response to passenger race.

A.3 Missing Census Data and Imputation

Some Census blocks, typically parks, industrial land and transportation infrastructure, have no resident population, so block-level race shares are undefined there. I impute demographics for these blocks by progressive geographic aggregation: from the surrounding Census tract where possible, then the taxi zone, then the community district. The cascade preserves as much geographic granularity as the data allow. After imputation, destination demographics are defined for more than 99.9 percent of otherwise-valid airport trips; the remainder are excluded from the analysis sample (Table A.1).

A.4 Sample Construction

Table A.1 documents retention from the raw 2013 trip records to the analysis sample.

Table A.1: Sample Construction and Retention

Stage	Trips	Drivers
All 2013 yellow taxi trips (TLC records)	173,177,113	43,191
Trips by drivers with a race classification	149,146,820	32,478
Airport pickups (JFK or LaGuardia)	5,496,759	—
Excluding trips terminating at an airport	5,281,545	—
Excluding unlocatable dropoffs	5,242,920	—
Excluding invalid distance, duration, or fare	5,208,417	—
Excluding missing destination demographics	5,206,073	30,616
Analysis sample	5,206,073	30,616
Location-choice sample (appendix sorting analysis)	4,884,367	—

Notes: Each airport-sample row applies its exclusion to the row above it, so counts fall cumulatively from airport pickups to the analysis sample. Race classification uses Bayesian Improved Surname Geocoding on driver names from the license registry (Appendix A.1); unclassified drivers lack a usable surname match. A dropoff is unlocatable when its GPS coordinates fall outside any New York City Census block. Invalid values are a distance ≤ 0.1 or > 100 miles, a duration < 1 or > 180 minutes, or a fare ≤ 0 or $> \$300$. Destination demographics are missing for blocks without Census data after the imputation cascade of Appendix A.3. The location-choice sample conditions on observing the driver's next pickup in the same shift, which excludes shift-ending trips, and on CDTA-level similarity measures; it is used only in the appendix sorting analysis.

A.5 External Validity of the Airport Sample

Airport pickups account for under four percent of 2013 yellow taxi trips, and they differ from street-hail trips: they are longer, costlier and more geographically dispersed, and their passengers include more travelers (Table 2). I focus on airports because the queue makes the driver–passenger match quasi-random, which is essential for identification. The estimated magnitudes need not generalize to street hails, where drivers choose whom to serve. The estimates measure discrimination where drivers cannot select passengers but retain discretion over the route. Table C.14 examines non-airport trips directly, and Section 7 flags the market-wide extrapolation as a calibration rather than a causal estimate.

A.6 Similarity Measure Definitions

Section 2.3 defines the two main measures, expected match and own-group exposure, at the dropoff Census block. The appendix sorting analysis (Appendix E) applies the same formulas at the Community District Tabulation Area level, with district population shares in place of block shares. Expected match is the taxi-market analogue of the isolation index from the residential segregation literature (Bell, 1954; Massey and Denton, 1988); Athey et al. (2021) extend the same idea to measure experienced segregation from GPS data.

The third measure is *cosine similarity*, the angular alignment between the driver's race-probability vector and the destination's composition vector:

$$\text{CosineSimilarity}_{ij} = \frac{\sum_r p_{ir}\, s_{jr}}{\sqrt{\sum_r p_{ir}^2}\,\sqrt{\sum_r s_{jr}^2}}, \tag{6}$$

with p_{ir} and s_{jr} as in Equation (1). The measure equals 1 when the two vectors are proportional and approaches 0 when they are orthogonal. Cosine similarity is scale-invariant, which makes it robust in destinations with extreme demographic concentrations where expected match is mechanically high or low. The robustness tables of Appendix C report all three measures.

A.7 Full Summary Statistics: Similarity Measures

Table A.2 reports all three similarity measures, at both the block and district levels, for the airport analysis sample.

Table A.2: Racial Similarity Measures: Full Set (CDTA and Census Block)

	All airport		LaGuardia		JFK	
Measure	Mean	SD	Mean	SD	Mean	SD
2013 airport trips, analysis sample (similarity measures, CDTA and Census-block level)						
Own-group exposure (CDTA)	0.011	0.120	0.012	0.115	0.009	0.127
Own-group exposure (block)	0.012	0.155	0.014	0.149	0.010	0.163
Expected match (CDTA)	0.196	0.121	0.193	0.116	0.200	0.128
Expected match (block)	0.196	0.152	0.194	0.145	0.200	0.160
Cosine similarity (CDTA)	0.421	0.296	0.413	0.291	0.433	0.302
Cosine similarity (block)	0.400	0.311	0.396	0.307	0.406	0.317
Observations	5,206,073		3,026,324		2,179,749	

Notes: Full set of racial similarity measures summarized in Table 2, reported at both the dropoff Census block and Community District Tabulation Area (CDTA) levels. Own-group exposure is the share of the driver's own racial group in the destination minus that group's citywide share, weighted by the driver's classification probability; expected match is the probability that the driver and a randomly chosen resident of the destination share a race; cosine similarity is the vector alignment between the driver's race probabilities and the destination composition. Sample and exclusions follow Table 2 (airport analysis sample).

B Green Cab Entry: Supplementary Evidence

This appendix supplements the competition-shock analysis in Section 5.1.

B.1 Market Changes

Figure B.1 shows the rapid growth of green cab trip volume during the entry period, from about 7,000 trips in August 2013 to over 600,000 by December. Table B.1 compares yellow cab shift characteristics before and after entry: shift income and trips per shift fell while wait time between trips rose, consistent with a thinner pool of available passengers per yellow driver.

B.2 Incumbent Driver Turnover and Exit

Figure B.2 plots monthly driver entry and exit. Before green cab entry the industry added an average of 549 drivers per month; afterward it lost 546 per month, as new entry slowed and exits accelerated. Among the roughly 39,400 drivers active before entry, 9.1% had exited by November. Table B.2 models exit timing among these incumbents as a multinomial logit on predetermined pre-entry characteristics, and Table B.3 a complementary Cox proportional-hazard model with standardized regressors. Both tell the same story: higher pre-entry earnings predict persistence (a one-standard-deviation increase in log income lowers the exit hazard by 12%), while greater pre-entry green-zone exposure predicts faster exit (hazard

Figure B.1: Green Cab Trip Volume During the Entry Period (August–December 2013)

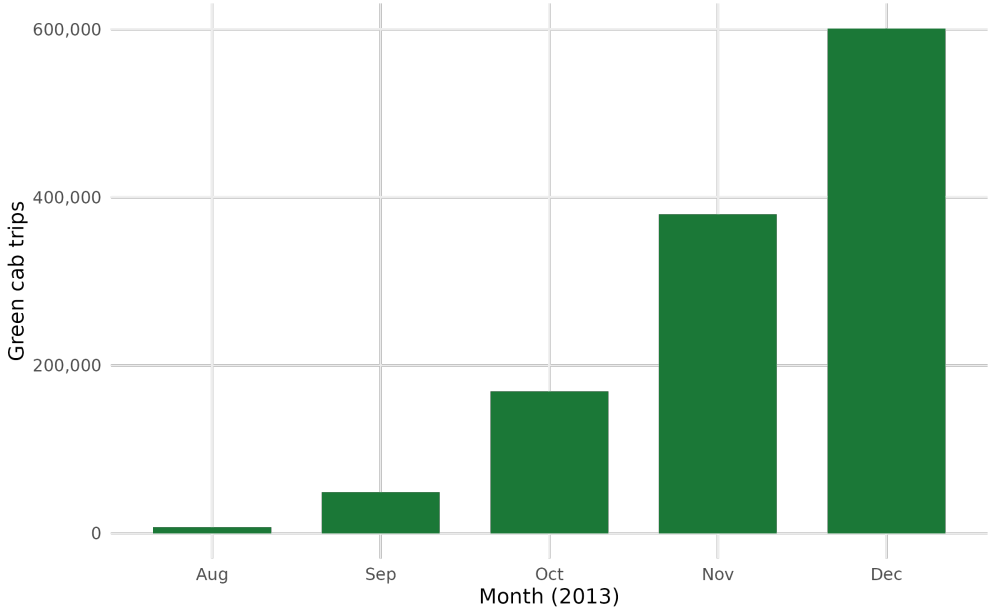


Table B.1: Summary Statistics: Yellow Cab Shifts Before and After Green Cab Entry

	Pre-Entry (Jan–Jul)	Post-Entry (Aug–Dec)	Difference	% Change
Panel A: Earnings				
Shift income ($)	318.34	314.57	-3.77	-1.2
Hourly wage ($/hour)	38.46	37.98	-0.48	-1.2
Income per trip ($)	15.98	16.64	0.66	4.1
Panel B: Work Patterns				
Shift length (hours)	8.71	8.68	-0.03	-0.3
Trips per shift	21.95	20.88	-1.07	-4.9
Wait time between trips (min)	27.76	29.16	1.40	5.0
Panel C: Spatial Patterns				
Average trip distance (miles)	3.32	3.44	0.12	3.6
Green zone pickup share	0.130	0.135	0.006	4.6
Panel D: Sample Size				
Number of shifts	4,654,224	3,320,320		
Number of unique drivers	39,470	38,080		

Notes: Mean shift-level characteristics for all New York City yellow taxi drivers before and after green cab entry (August 2013). Pre-entry covers January–July 2013; post-entry August–December 2013. Sample is all yellow-cab shifts, not restricted to race-classified drivers.

ratio 1.29), with the exposure effect attenuated for higher earners. The drivers the shock pushes out are disproportionately the most exposed and least able to absorb the earnings hit, the same margin on which the mechanism operates.

Figure B.2: Driver Entry and Exit Before and After Green Cab Entry

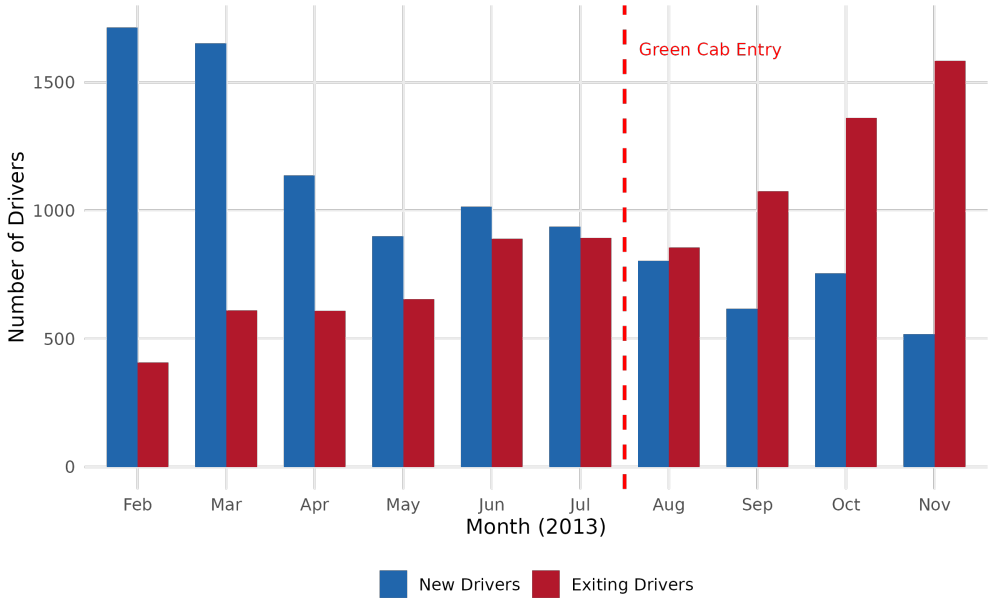


Table B.2: Exit Timing: Multinomial Logit (Incumbent Drivers)

	Aug	Sep	Oct	Nov
Exit month (baseline: active through December)				
Log pre-entry income	-0.782***	-0.630***	-0.586***	-0.258**
Pre-entry GZ share	-0.332	0.757	1.523*	3.050***
Log income × GZ share	-0.088	-0.150	-0.443**	-0.625***
Observations		39,388		

Notes: Multinomial logit of exit-month choice among incumbent drivers active before green cab entry (first shift before August). The baseline outcome is remaining active through December. Each column gives the log relative-risk coefficients for exiting in that month versus staying, as a function of predetermined pre-entry driver characteristics. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

B.3 New Entrant Survival

I compare drivers who first entered the industry before green cab entry (February–July) with those who entered after (August–November). Figure B.3 shows Kaplan–Meier survival curves: the post-entry cohort attrites markedly faster (log-rank $\chi^2 = 249$). Table B.4

Table B.3: Incumbent Driver Exit: Cox Proportional Hazard

	Hazard ratio	(SE)
Log pre-entry income (std.)	0.880***	(0.029)
Pre-entry GZ share (std.)	1.293***	(0.053)
Log income × GZ share (std.)	0.781***	(0.050)
Observations	39,388	
Events (exits)	3,574	

Notes: Cox proportional-hazard model of incumbent-driver exit timing after green cab entry. Regressors are standardized (mean 0, SD 1), so each hazard ratio is the multiplicative effect on the exit hazard of a one-standard-deviation increase. A ratio above one indicates faster exit. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

confirms this in Cox models. Post-entry entrants face a 74% higher exit hazard (Model 1), attenuated but still significant after controlling for income (Model 4); green-zone exposure independently predicts exit, and higher earnings strongly predict survival. Entrants arriving into the more competitive post-entry market were both selected differently and less able to establish themselves.

Figure B.3: Kaplan–Meier Survival Curves: New Entrants by Entry Cohort

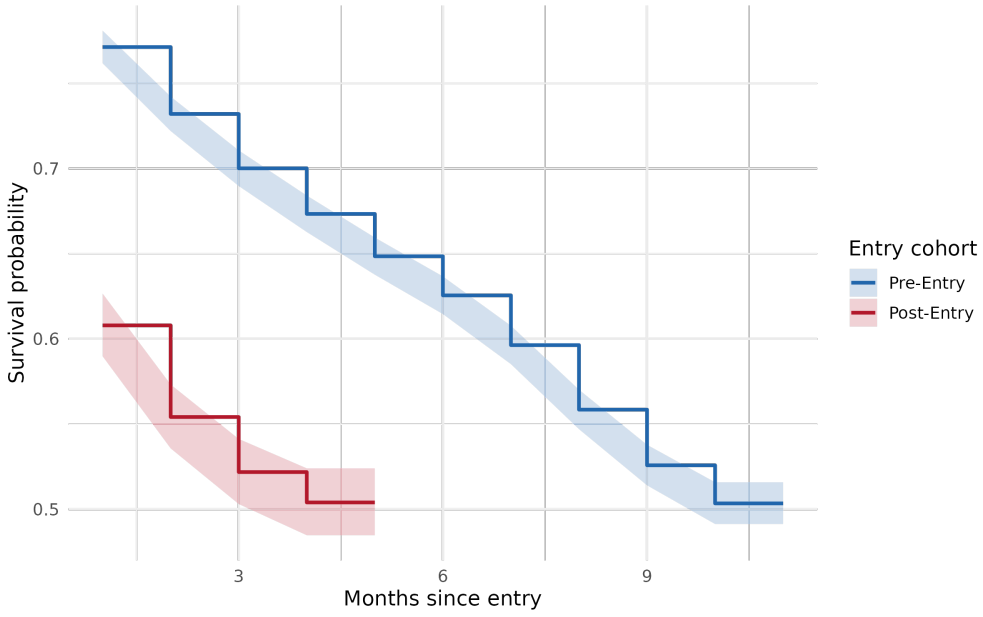


Table B.4: New Entrant Exit: Cox Proportional Hazard Models

	(1)	(2)	(3)	(4)
Post-entry cohort	1.741***	1.700***	1.755***	1.348***
	(0.034)	(0.035)	(0.043)	(0.035)
Avg. GZ share		4.492***	4.765***	0.923
		(0.058)	(0.074)	(0.052)
Post-entry × GZ share			0.860	
			(0.118)	
Log avg. income				0.437***
				(0.011)
Observations	10,056	9,962	9,962	9,962
Events	4,735	4,642	4,642	4,642

Notes: Hazard ratios from Cox models of 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. GZ share is the driver's average green-zone pickup share. A ratio above one indicates faster exit. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

C Additional Robustness

This appendix collects estimates that support the main text: the route-efficiency placebo and the fixed-effects choice, the distance outcome, alternative similarity measures, residential heterogeneity, competition-response heterogeneity, the mechanism tests under alternative measures, and the non-airport analysis.

C.1 Route-Efficiency Specification: The Placebo and Fixed-Effects Choice

The flat-rate placebo holds. Estimating the fare specification of Equation (5) without the constraint, adding back the similarity main effect, leaves the similarity slope on flat-rate trips statistically indistinguishable from zero for every similarity measure (Table C.1), as it must be mechanically: a JFK–Manhattan trip earns $52 regardless of route. This validates the constrained specification and the use of flat trips as a control group.

Why airport×block fixed effects are inadmissible in the triple-difference. Because a trip is on the flat rate if and only if it is a JFK pickup to a Manhattan block, Metered is nearly collinear with the airport×block cell: a regression of Metered on airport×block fixed effects yields $R^2 = 0.95$, against $R^2 = 0.16$ for additive block fixed effects. Only 8.95% of airport×block cells contain both flat and metered trips, and these hold just 7.4% of all metered trips. Imposing airport×block fixed effects therefore identifies the metered

Table C.1: The Flat-Rate Placebo: Similarity Does Not Move the Fare Where Routing Cannot

	Own-Group	Expected Match	Cosine
Metered gradient (γ_1)	-0.391***	-0.439***	-0.100***
	(0.061)	(0.066)	(0.038)
Flat-rate slope	-0.002	-0.005	-0.003
	(0.004)	(0.003)	(0.002)
Metered trips		5,206,073	
Flat-rate trips		1,296,066	
Driver, block, hour, day-of-week FE	Yes	Yes	Yes

Notes: The fare gradient on driver–passenger similarity. *Metered gradient* is γ_1, the Similarity×Metered coefficient of the headline specification (Table 5): on metered trips a more similar passenger pays less. *Flat-rate slope* is from a separate regression of the fare on similarity, within driver, destination block, hour, and day of week, on JFK–Manhattan flat-rate trips, where the $52 fare is fixed regardless of route; it is statistically indistinguishable from zero for every measure, so similarity cannot move the fare where routing cannot. The contrast between the two is the paper's identification. Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

main effect, γ_1, and γ_2 off that small, selected set, producing estimates that are unstable across similarity measures and in which γ_2 (Metered × Post) flips negative, an artifact of near-collinearity rather than evidence of "leveling up." The additive-block specification of Equation (5), with the JFK pickup main effect absorbing the JFK-versus-LaGuardia route-length difference, is the correct design.

Distance is the wrong margin. The headline table reports fare and time. Distance is a mechanism diagnostic, and its pre-entry gradient (Section 3.2) is a precise null across all three similarity measures (Table C.2). Favored passengers reach the same destination block in less time and at a lower fare over essentially the same distance, so favoritism operates through travel time, the slower routes that accrue the meter's time-based charge, not through longer-mileage detours. The null is not an artifact of over-controlling: the fixed effects absorb a similar share of the distance and fare variance (72% each), and distance retains ample residual variation, a within-cell standard deviation of 2.7 miles. The channel is speed in traffic, the dimension over which a driver retains discretion to a fixed destination.

The full-year distance estimates, which add the response to green cab entry, appear in Table C.3. The pre-entry gradient γ_1 there is likewise a precise null, and the distance response to entry is second-order: the leveling-down of Section 5 runs through time and fares, not mileage.

Table C.2: Route-Efficiency Favoritism, Distance Outcome Before Green Cab Entry (Mechanism)

	Own-group exposure (1)	Expected match (2)	Cosine similarity (3)
Similarity × Metered (γ_1)	-0.010	-0.039	0.023
	(0.024)	(0.027)	(0.017)
Driver, block, hour, day-of-week FE	Yes	Yes	Yes
Observations	2,958,327	2,958,327	2,958,327

Notes: Within-driver estimates of the favoritism gradient on trip *distance* (miles), pre-entry trips only (January–July 2013), under the baseline specification of columns (1)–(2) of Table 3 with distance as the outcome. The gradient is a precise null: favored passengers cover the same distance, so favoritism operates through travel time (slower routes), not shorter mileage. Full-year distance estimates, including the response to green cab entry, appear in Table C.3. Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Alternative similarity measures. The headline table reports expected match, the paper's primary measure; Table C.4 repeats the triple-difference, on the full sample and on the balanced panel of continuing drivers, for all three similarity measures. Own-group exposure nets out background segregation and corroborates the headline; cosine similarity weights the full racial distribution and is less directly interpretable. The estimates are qualitatively consistent throughout (favoritism on fares and time, leveling down, post-entry narrowing).

Within-metered robustness admits the stricter control. Restricting the sample to metered trips makes Metered constant, so airport×destination-block fixed effects become admissible. Table C.5 estimates $y_{ijt} = \beta_1 \text{Sim}_{ij} + \beta_2(\text{Sim}_{ij} \times \text{Post}_t) + \alpha_i + \delta_{aj} + \lambda_{\text{hr}} + \mu_{\text{dow}}$ on metered trips, where δ_{aj} are airport×block fixed effects. The baseline favoritism slope β_1 remains negative (-0.17 own-group, -0.31 expected match, -0.17 cosine for fare) and the post-entry interaction β_2 positive, mirroring the headline triple-difference under the strictest available destination control.

Pre-entry favoritism: own-group exposure. Table C.6 repeats the pre-entry favoritism table of Section 3.2, baseline and residential-share columns, with own-group exposure in place of expected match.

Residential heterogeneity: full-year estimates and robustness. Table C.7 extends the pre-entry residential interaction of Section 3.3 to the full year within Equation (5): the favoritism gradient (δ_1) and its post-entry change (δ_3) both steepen with residential share, so

Table C.3: Route Efficiency, Distance Outcome (Mechanism)

	Distance (mi)
Panel A: Own-Group Exposure	
Similarity × Metered (γ_1)	-0.001
	(0.024)
Metered × Post (γ_2)	-0.022***
	(0.005)
Similarity × Metered × Post (γ_3)	0.055**
	(0.022)
Observations	5,206,073
Within R^2	0.571
Panel B: Expected Match	
Similarity × Metered (γ_1)	-0.006
	(0.026)
Metered × Post (γ_2)	-0.025***
	(0.007)
Similarity × Metered × Post (γ_3)	0.016
	(0.024)
Observations	5,206,073
Within R^2	0.571
Panel C: Cosine Similarity	
Similarity × Metered (γ_1)	0.049***
	(0.016)
Metered × Post (γ_2)	-0.021***
	(0.008)
Similarity × Metered × Post (γ_3)	-0.002
	(0.013)
Observations	5,206,073
Within R^2	0.571
Metered, JFK main effects	Yes
Driver, block, hour, day FE	Yes

Notes: Same specification as Table 5 with trip distance as the outcome. The baseline favoritism coefficient γ_1 is a precise null for own-group exposure and expected match, and economically negligible though statistically significant for cosine. The distance null reflects a genuine feature, not weak identification: discrimination operates through travel time, the slower routes that accrue the meter's time component, not through longer-mileage detours (Appendix C.1). Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.4: Route Efficiency Across Similarity Measures

	Fare ($)		Time (min)	
	Full	Balanced	Full	Balanced
Panel A: Expected Match (headline)				
Similarity × Metered (γ_1)	-0.439***	-0.441***	-0.507***	-0.503***
	(0.066)	(0.068)	(0.065)	(0.067)
Metered × Post (γ_2)	0.103***	0.102***	0.546***	0.573***
	(0.015)	(0.015)	(0.018)	(0.018)
Similarity × Metered × Post (γ_3)	0.236***	0.214***	0.334***	0.310***
	(0.060)	(0.061)	(0.070)	(0.071)
Observations	5,206,073	4,951,219	5,206,073	4,951,219
Drivers	30,616	22,076	30,616	22,076
Within R^2	0.534	0.533	0.242	0.241
Panel B: Own-Group Exposure				
Similarity × Metered (γ_1)	-0.391***	-0.391***	-0.318***	-0.308***
	(0.061)	(0.063)	(0.056)	(0.058)
Metered × Post (γ_2)	0.146***	0.141***	0.607***	0.630***
	(0.010)	(0.010)	(0.013)	(0.013)
Similarity × Metered × Post (γ_3)	0.268***	0.245***	0.338***	0.306***
	(0.057)	(0.058)	(0.067)	(0.068)
Observations	5,206,073	4,951,219	5,206,073	4,951,219
Drivers	30,616	22,076	30,616	22,076
Within R^2	0.534	0.533	0.242	0.241
Panel C: Cosine Similarity				
Similarity × Metered (γ_1)	-0.100***	-0.100**	-0.167***	-0.163***
	(0.038)	(0.040)	(0.043)	(0.044)
Metered × Post (γ_2)	0.112***	0.112***	0.558***	0.584***
	(0.015)	(0.016)	(0.019)	(0.019)
Similarity × Metered × Post (γ_3)	0.095***	0.079**	0.134***	0.126***
	(0.030)	(0.031)	(0.036)	(0.037)
Observations	5,206,073	4,951,219	5,206,073	4,951,219
Drivers	30,616	22,076	30,616	22,076
Within R^2	0.534	0.533	0.242	0.241
Metered, JFK main effects	Yes	Yes	Yes	Yes
Driver, block, hour, day FE	Yes	Yes	Yes	Yes

Notes: The headline triple-difference (Table 5) reproduced for all three similarity measures, on the full route-efficiency sample and on the balanced panel of continuing drivers (at least ten metered trips in both periods). Panel A repeats the headline expected-match estimates for comparison. Own-group exposure nets out background segregation and corroborates the headline; cosine similarity weights the full racial distribution and is less directly interpretable but qualitatively consistent. Fixed effects and standard errors as in Table 5. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.5: Route Efficiency, Robustness: Within Metered Trips, Airport×Block Fixed Effects

	Fare ($)	Time (min)
Panel A: Own-Group Exposure		
Similarity (β_1, baseline)	-0.168***	-0.336***
	(0.032)	(0.046)
Similarity × Post (β_2)	0.288***	0.629***
	(0.041)	(0.065)
Observations	3,910,007	3,910,007
Within R^2	0.000	0.000
Panel B: Expected Match		
Similarity (β_1, baseline)	-0.311***	-1.102***
	(0.032)	(0.047)
Similarity × Post (β_2)	0.590***	2.354***
	(0.030)	(0.044)
Observations	3,910,007	3,910,007
Within R^2	0.000	0.001
Panel C: Cosine Similarity		
Similarity (β_1, baseline)	-0.173***	-0.557***
	(0.020)	(0.029)
Similarity × Post (β_2)	0.279***	1.148***
	(0.015)	(0.022)
Observations	3,910,007	3,910,007
Within R^2	0.000	0.001
Driver FE	Yes	Yes
Airport × block FE	Yes	Yes
Hour-of-day, day-of-week FE	Yes	Yes

Notes: Estimated on metered trips only, so *Metered* is constant and airport×destination-block fixed effects (which control for the route from each airport to each block) are admissible without the collinearity that rules them out in the headline triple-difference (Table 5). β_1 is the within-route baseline favoritism slope; β_2 (Similarity × Post) is the change after green-cab entry. Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.6: Favoritism Before Green Cab Entry: Own-Group Exposure

	All destinations		By residential share	
	Fare ($)	Time (min)	Fare ($)	Time (min)
	(1)	(2)	(3)	(4)
Similarity × Metered (γ_1)	-0.364***	-0.246***	-0.027	0.280**
	(0.062)	(0.058)	(0.099)	(0.113)
Similarity × Metered × Residential share (δ_1)			-0.900***	-1.027***
			(0.141)	(0.153)
Metered and JFK controls	Yes	Yes	Yes	Yes
Metered × Residential share control	No	No	Yes	Yes
Driver, block, hour, day-of-week FE	Yes	Yes	Yes	Yes
Observations	2,958,327	2,958,327	2,174,063	2,174,063

Notes: Within-driver estimates of favoritism on pre-entry trips only (January–July 2013), using own-group exposure. The estimating equation (Eq. 3) contains no post-entry terms; the response to green cab entry is estimated in Table 5. The flat-rate placebo is imposed as a restriction: similarity cannot operate through routing on JFK–Manhattan flat-fare trips, so the similarity main effect is omitted and flat trips serve as the within-driver control group. Columns (1)–(2) use all pre-entry route-efficiency trips. Columns (3)–(4) interact the favoritism term with the *residential share* of the dropoff block (residential built floor area over total built floor area, from PLUTO tax-lot records) on the land-use-matched sample; the favoritism gradient at residential share r is $\gamma_1 + \delta_1 r$, so γ_1 in columns (3)–(4) is the gradient at a fully commercial block and $\gamma_1 + \delta_1$ at a fully residential one. The residential-share main effect is absorbed by the block fixed effects. This is the own-group-exposure twin of Table 3. Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

the post-entry compression of Section 5 concentrates exactly where the pre-entry favoritism is steepest. Table C.8 repeats the full-year specification for own-group exposure and cosine similarity, and Table C.9 replaces the continuous residential share with a binary indicator for blocks at least half residential by built floor area. All reproduce the main-text pattern, favoritism near zero in commercial blocks and strong in residential ones, confirming it is not specific to a similarity measure or to the linear functional form.

Mechanism vs. measurement: residential share against block population. Table C.10 interacts the favoritism term with both standardized residential share and standardized log block population. The sample-size form of measurement error predicts the gradient should load on population (more residents anchor a better-measured composition). It does not: residential share remains large and highly significant conditional on population, while population enters small and, for fares, with the opposite sign to the measurement prediction. The sample-size channel is therefore rejected; the proxy-relevance channel (residents need not represent passengers in commercial blocks) remains observationally equivalent to a behavioral residential mechanism and cannot be separated with these data.

Table C.7: Route-Efficiency Favoritism by Residential Share of Destination

	Fare ($)	Time (min)
Expected Match		
Sim × Metered (γ_1)	-0.250**	0.197*
	(0.101)	(0.116)
Sim × Metered × Residential share (δ_1)	-0.703***	-1.207***
	(0.141)	(0.156)
Sim × Metered × Post (γ_3)	-0.107	-0.289*
	(0.125)	(0.156)
Sim × Metered × Post × Residential share (δ_3)	0.444**	0.716***
	(0.176)	(0.214)
Observations	3,828,940	3,828,940
Metered, JFK, Post controls	Yes	Yes
Driver, block, hour, day FE	Yes	Yes

Notes: The favoritism term (γ_1) and its post-entry change (γ_3) each interacted with the residential share of the dropoff block (δ_1, δ_3), using expected match; own-group exposure and cosine similarity agree (Appendix Table C.8). Residential share runs from 0 (commercial) to 1 (fully residential), so the favoritism gradient $\gamma_1 + \delta_1 r$ equals γ_1 in a commercial block and $\gamma_1 + \delta_1$ in a residential one; the fully residential fare gradient is roughly twice the pooled estimate of Table 5. Lower-order Metered, JFK, and Metered×residential-share terms are included; the residential-share main effect is absorbed by the block fixed effects. I read the concentration primarily as validation: similarity is built from each block's residents, so it is well measured where blocks are residential and noisy where they are commercial. A behavioral residential channel is consistent but not separately identified; Appendix Table C.10 rules out a sample-size measurement artifact and Table C.9 confirms the pattern with a binary split. Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Competition-response heterogeneity: full estimates. Table C.11 reports the estimates plotted in Figure 6 in Section 5, adding the trip-time outcome and exact standard errors for both similarity measures.

Local knowledge and tipping: alternative measures. Table C.12 repeats the drop-home local-knowledge test of Section 6.1 with own-group exposure; Table C.13 repeats the tip-outcome test of Section 6.2 for all three similarity measures. Both reproduce the main-text conclusions.

Non-airport route efficiency. The airport queue is what identifies the favoritism gradient, so the main-text estimates speak to airport trips. Table C.14 asks whether the same pattern appears on the non-airport street hails that make up the rest of the market, and supplies the bracket the calibration of Section 7 relies on. Off the airports the gradient

Table C.8: Route-Efficiency Favoritism by Residential Share: Alternative Similarity Measures

	Fare ($)	Time (min)
Panel A: Own-Group Exposure		
Sim × Metered (γ_1)	-0.059	0.302***
	(0.095)	(0.109)
Sim × Metered × Residential share (δ_1)	-0.891***	-1.081***
	(0.139)	(0.150)
Sim × Metered × Post (γ_3)	0.024	-0.196
	(0.119)	(0.150)
Sim × Metered × Post × Residential share (δ_3)	0.262	0.579***
	(0.171)	(0.209)
Observations	3,828,940	3,828,940
Panel B: Cosine Similarity		
Sim × Metered (γ_1)	-0.157***	0.091
	(0.057)	(0.064)
Sim × Metered × Residential share (δ_1)	-0.234***	-0.461***
	(0.071)	(0.079)
Sim × Metered × Post (γ_3)	-0.051	-0.056
	(0.060)	(0.076)
Sim × Metered × Post × Residential share (δ_3)	0.196**	0.169
	(0.083)	(0.104)
Observations	3,828,940	3,828,940
Metered, JFK, Post controls	Yes	Yes
Driver, block, hour, day FE	Yes	Yes

Notes: As Table C.7 (continuous residential share) with own-group exposure and cosine similarity in place of expected match. Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

cannot be cleanly identified, because drivers choose whom to pick up and where to cruise: a dropoff in an own-group block usually follows a nearby pickup, so a dropoff-block control (columns 1 and 3) mixes favoritism with pickup geography and returns a gradient as large as the airport one, while pickup-by-dropoff tract-pair fixed effects (columns 2 and 4) hold the route fixed and return a smaller lower bound. The airport-identified gap, scaled to non-airport trip sizes, falls between these two readings. The leveling-down (Post) coefficient, by contrast, is a race-blind pre/post shift and is stable across both designs, which is why Section 7 extrapolates the gap but not the degradation. Table C.15 repeats the estimates for all three similarity measures and adds the distance outcome; the pattern is measure-invariant.

Table C.9: Route-Efficiency Favoritism, Binary Residential Indicator (Robustness)

	Fare ($)	Time (min)
Panel A: Own-Group Exposure		
Sim × Metered (γ_1, commercial)	-0.175**	0.134
	(0.083)	(0.094)
× Residential ($\geq$ 50%)	-0.605***	-0.730***
	(0.099)	(0.109)
Sim × Metered × Post (γ_3, commercial)	0.070	-0.062
	(0.100)	(0.127)
× Residential	0.178	0.349**
	(0.124)	(0.153)
Observations	3,828,940	3,828,940
Panel B: Expected Match		
Sim × Metered (γ_1, commercial)	-0.321***	0.009
	(0.088)	(0.101)
× Residential ($\geq$ 50%)	-0.489***	-0.810***
	(0.100)	(0.113)
Sim × Metered × Post (γ_3, commercial)	-0.022	-0.105
	(0.106)	(0.133)
× Residential	0.286**	0.395**
	(0.127)	(0.156)
Observations	3,828,940	3,828,940
Panel C: Cosine Similarity		
Sim × Metered (γ_1, commercial)	-0.148***	0.030
	(0.050)	(0.057)
× Residential ($\geq$ 50%)	-0.177***	-0.312***
	(0.050)	(0.056)
Sim × Metered × Post (γ_3, commercial)	-0.009	0.021
	(0.051)	(0.065)
× Residential	0.119**	0.049
	(0.060)	(0.075)
Observations	3,828,940	3,828,940
Metered, JFK, Post controls	Yes	Yes
Driver, block, hour, day FE	Yes	Yes

Notes: As Table C.7 but with a binary indicator for blocks at least 50% residential by built floor area replacing the continuous share, confirming the residential concentration is not an artifact of the linear functional form. γ_1, γ_3 are the commercial-block coefficients; the indented rows are the residential increments. Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.10: Residential Share vs. Block Population: Mechanism or Measurement?

	Fare ($)	Time (min)
Panel A: Own-Group Exposure		
Sim × Metered × Residential share (std.)	-0.396***	-0.328***
	(0.048)	(0.051)
Sim × Metered × log Block population (std.)	0.255***	0.071
	(0.043)	(0.050)
Sim × Metered (γ_1, at means)	-0.511***	-0.349***
	(0.065)	(0.063)
Observations	3,828,940	3,828,940
Panel B: Expected Match		
Sim × Metered × Residential share (std.)	-0.333***	-0.364***
	(0.049)	(0.053)
Sim × Metered × log Block population (std.)	0.292***	0.089*
	(0.044)	(0.051)
Sim × Metered (γ_1, at means)	-0.584***	-0.509***
	(0.070)	(0.072)
Observations	3,828,940	3,828,940
Metered, JFK, Post controls	Yes	Yes
Driver, block, hour, day FE	Yes	Yes

Notes: The favoritism interaction Sim × Metered is interacted with both the dropoff block's residential share and its log resident population, each standardized. If the residential gradient were a sample-size measurement artifact, it would load on population, since more residents give a better-measured racial composition. It does not: the residential-share interaction stays large and significant conditional on population, while the population interaction is small and, for fares, significant with the opposite sign to that prediction. The sample-size form of measurement error therefore does not explain the residential gradient. What remains is proxy relevance, that residents need not represent the passengers dropped in a commercial block, which cannot be separated from a behavioral residential channel. The two regressors are correlated, so this is a partial test. Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

C.2 Motivating Figures: Own-Group Exposure

Figures C.1 and C.2 repeat the two motivating figures (Figures 3 and 5) with own-group exposure in place of expected match. Both give the same picture. The metered gradients are -0.75 on fares and -0.58 on trip time, against flat-trip slopes indistinguishable from zero; the median-split fare gap narrows from \$0.11 to \$0.06 after entry, and the trip-time gap closes from 0.12 to 0.02 minutes.

Table C.11: Heterogeneity in the Competition Response (γ_3 by Tercile)

	Green-zone exposure			Pre-entry income		
	Low	Mid	High	Low	Mid	High
Exp. match, Fare	0.172	0.180*	0.246***	0.205*	0.442***	0.036
	(0.117)	(0.106)	(0.091)	(0.113)	(0.100)	(0.102)
Own-group, Fare	0.164	0.224**	0.294***	0.330***	0.447***	0.006
	(0.114)	(0.099)	(0.087)	(0.107)	(0.096)	(0.096)
Exp. match, Time	0.263*	0.374***	0.284***	0.500***	0.451***	0.030
	(0.156)	(0.127)	(0.099)	(0.135)	(0.118)	(0.114)
Own-group, Time	0.268*	0.321***	0.309***	0.558***	0.419***	-0.026
	(0.151)	(0.121)	(0.095)	(0.128)	(0.116)	(0.108)

Notes: This table reports the estimates plotted in Figure 6, adding the trip-time outcome and exact standard errors. Each cell reports γ_3 (Sim×Metered×Post) from the constrained triple-difference (Eq. 5) estimated on the indicated tercile subsample, with the standard error in parentheses. Green-zone exposure is the driver's pre-entry share of dropoffs in the contestable green zone; pre-entry income is the driver's average pre-entry shift earnings (total fare). Both are predetermined (Jan–Jul 2013). The Goldberg mechanism predicts the narrowing concentrates among the most-exposed and lowest-income drivers. Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.12: Ruling Out Local Knowledge: Own-Group Exposure

	(1) Fare ($)	(2) Time (min)
Panel A: Favoritism gradient γ_1 (Sim×Metered)		
Full sample	-0.391***	-0.318***
	(0.061)	(0.056)
Excluding home dropoffs	-0.412***	-0.336***
	(0.062)	(0.057)
Panel B: Home-interaction specification (full sample)		
Sim×Metered (non-home trips)	-0.407***	-0.337***
	(0.061)	(0.057)
Sim×Metered×Home	0.588***	0.636***
	(0.182)	(0.221)
Driver, block, hour, day-of-week FE	Yes	Yes
Home-dropoff share of trips	3.8%	
Observations, full sample	5,206,073	
Observations, excluding home dropoffs	5,009,566	

Notes: As Table 6 with own-group exposure in place of expected match. The implied home-trip gradient (the sum of the Panel B coefficients) is +0.18 dollars and +0.30 minutes. Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.13: Ruling Out Rational Tip Extraction: All Similarity Measures

	Own-Group (1)	Exp. Match (2)	Cosine (3)
Dependent variable: similarity slope on each outcome			
Fare ($)	-0.377***	-0.402***	-0.219***
	(0.100)	(0.111)	(0.070)
Tip ($)	-0.035	-0.033	-0.011
	(0.039)	(0.043)	(0.027)
Driver, block, hour, day-of-week FE	Yes	Yes	Yes
Ratio, p10–p90 fare gap / tip gap	10.9	12.0	19.7
Observations	573,224	573,224	573,224

Notes: Table 7 reproduced for all three similarity measures (dropoff Census-block level). Each cell reports the coefficient from regressing the row's outcome on the column's similarity measure within driver and destination block, with hour and day-of-week fixed effects, on airport metered credit-card trips. Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure C.1: Favoritism and the Flat-Fare Placebo: Own-Group Exposure

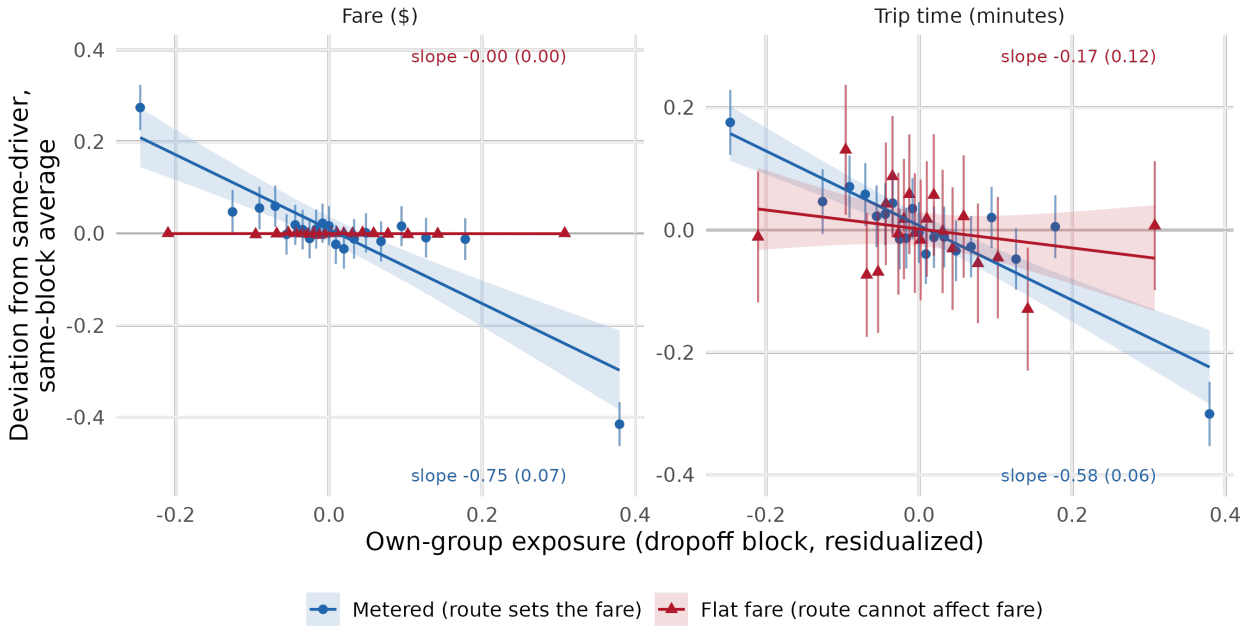


Table C.14: Non-Airport Route Efficiency: Favoritism and Its Change After Entry

	Fare ($)		Time (min)	
	Block FE	Tract-pair FE	Block FE	Tract-pair FE
	(1)	(2)	(3)	(4)
Similarity (expected match)	-0.4959***	-0.0367***	-0.5601***	-0.0847***
	(0.0200)	(0.0086)	(0.0231)	(0.0133)
Similarity × Post	0.0682***	0.0452***	0.1582***	0.1332***
	(0.0181)	(0.0104)	(0.0243)	(0.0180)
Post	0.2005***	0.1655***	0.4608***	0.4244***
	(0.0046)	(0.0027)	(0.0061)	(0.0046)
Driver, hour, day-of-week FE	Yes	Yes	Yes	Yes
Dropoff-block FE	Yes	No	Yes	No
Pickup × dropoff tract-pair FE	No	Yes	No	Yes
Observations	20,000,000			

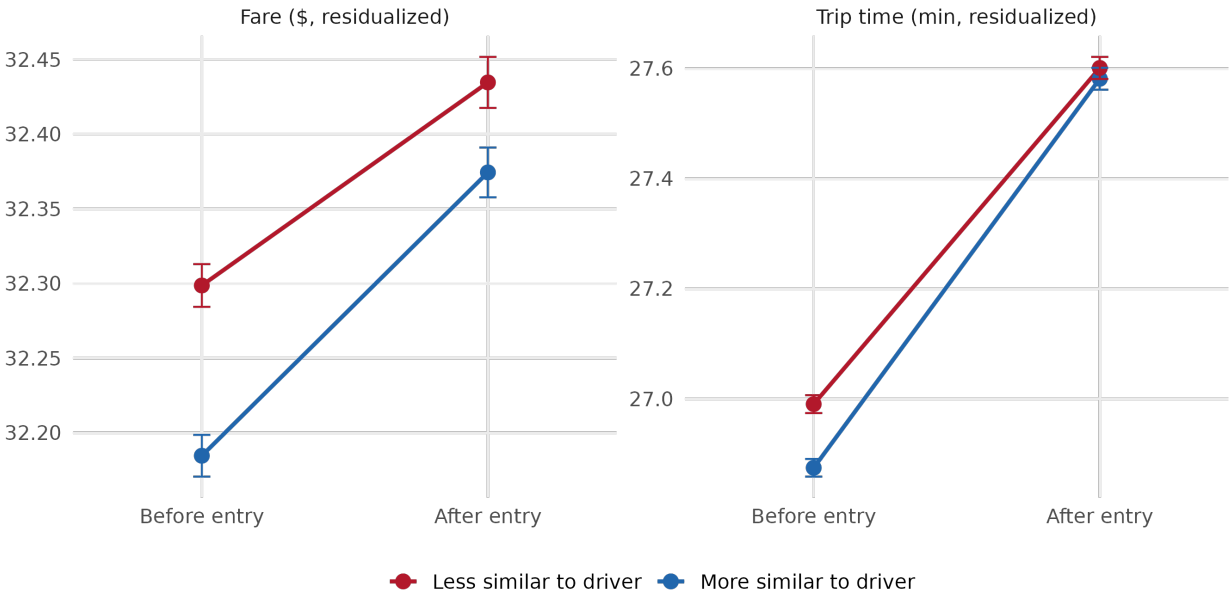
Notes: Estimates from a 20-million-trip random sample of non-airport (street-hail) trips, using expected match, the paper's primary similarity measure; own-group exposure and cosine similarity give the same pattern (Appendix Table C.15). *Post* indicates trips after July 2013; similarity is measured at the dropoff Census block. Columns (1) and (3) control for the dropoff block, the same specification as the airport design (Table 5), and yield a favoritism gradient of the same magnitude as at airports. That comparison is in specification, not identification: off-airport, drivers choose where to cruise, so a driver dropping off in an own-group block likely picked up nearby, and the block-level gradient mixes favoritism with pickup geography. Columns (2) and (4) add pickup×dropoff tract-pair fixed effects, which hold the route fixed and remove that confound, leaving a smaller lower bound. The two columns bracket the discrimination gap the calibration extrapolates (Table 9): the airport-identified gap, scaled to non-airport trip sizes, falls between the block-level gradient above and the route-pinned gradient below. The Post coefficient, the market-wide leveling-down input, is stable across the two designs and is not extrapolated. Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.15: Non-Airport Route Efficiency: All Similarity Measures and Both Designs

	Dropoff-block FE		Pickup × dropoff tract-pair FE		
	Fare ($)	Time (min)	Fare ($)	Distance (mi)	Time (min)
	(1)	(2)	(3)	(4)	(5)
Panel: Own-Group Exposure					
Similarity	-0.4375***	-0.4756***	-0.0406***	-0.0015	-0.0548***
	(0.0181)	(0.0210)	(0.0078)	(0.0023)	(0.0122)
Similarity × Post	0.0645***	0.1071***	0.0250**	-0.0029	0.0627***
	(0.0174)	(0.0230)	(0.0098)	(0.0029)	(0.0169)
Post	0.2133***	0.4909***	0.1742***	-0.0099***	0.4500***
	(0.0029)	(0.0040)	(0.0018)	(0.0007)	(0.0031)
Panel: Expected Match					
Similarity	-0.4959***	-0.5601***	-0.0367***	0.0009	-0.0847***
	(0.0200)	(0.0231)	(0.0086)	(0.0026)	(0.0133)
Similarity × Post	0.0682***	0.1582***	0.0452***	-0.0057*	0.1332***
	(0.0181)	(0.0243)	(0.0104)	(0.0033)	(0.0180)
Post	0.2005***	0.4608***	0.1655***	-0.0088***	0.4244***
	(0.0046)	(0.0061)	(0.0027)	(0.0009)	(0.0046)
Panel: Cosine Similarity					
Similarity	-0.3381***	-0.3779***	-0.0018	0.0037**	-0.0215***
	(0.0129)	(0.0146)	(0.0053)	(0.0016)	(0.0080)
Similarity × Post	0.0307***	0.0809***	0.0215***	-0.0032*	0.0705***
	(0.0086)	(0.0118)	(0.0052)	(0.0018)	(0.0090)
Post	0.2015***	0.4593***	0.1658***	-0.0086***	0.4222***
	(0.0045)	(0.0060)	(0.0027)	(0.0010)	(0.0046)
Observations			20,000,000		
Driver, hour, day-of-week FE	Yes	Yes	Yes	Yes	Yes

Notes: Table C.14 reproduced for all three similarity measures, with the distance outcome added under the route-pinned design. Columns (1)–(2) use the airport design's dropoff-block fixed effects; columns (3)–(5) hold the route fixed with pickup×dropoff tract-pair fixed effects. Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure C.2: Leveling Down: Own-Group Exposure



D Theoretical Derivations

D.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 $m \to 0$. □

D.2 Sorting

This section formalizes the sorting prediction discussed informally in Section 4. 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 eroding the surplus 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 loss of surplus, not a change in preferences, drives the reduction. □

D.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 zero-profit constraint binds ($\phi > m$). A competition shock that lowers m is therefore informative about the underlying mechanism.

D.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 zero-profit 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.

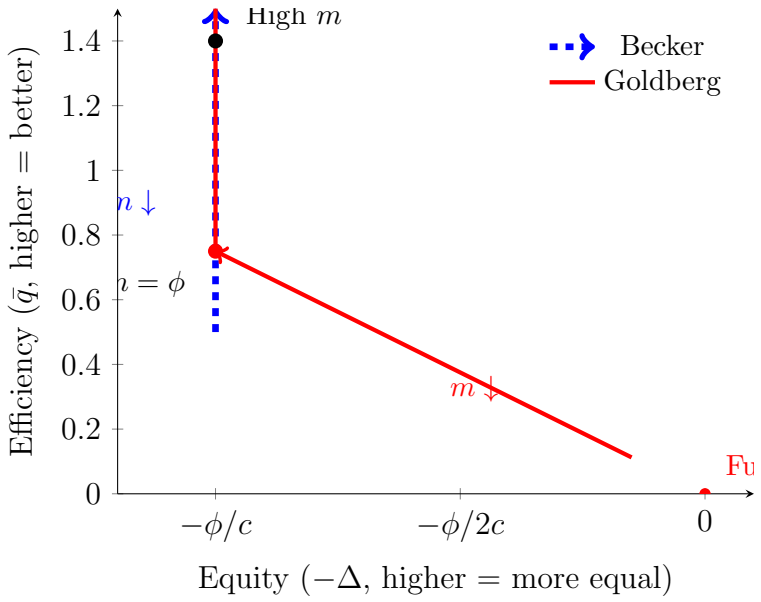
The *leveling-down ratio*, total quality degradation per unit of gap reduction, has a closed form. For a marginal reduction in m:

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

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, so the path is vertical. Under Goldberg with a binding constraint, efficiency falls *and* equity increases, so the path is downward-sloping. Figure D.1 illustrates.

Figure D.1: Equity-Efficiency Frontier Under Competition



As margins m fall, both models reduce efficiency (average quality). Under Becker (blue, dashed), equity is unchanged, so the path is vertical. Under Goldberg (red, solid), the paths coincide while unconstrained; once the zero-profit 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: $c = 2$, $d = \phi = 1$, $\mu = 0.5$.

D.5 Robustness to Cost Function Specification

This section shows that the leveling-down prediction is robust to alternative cost specifications, while the Becker constant-gap prediction is not. Consider a provider with cost function $C(q)$ satisfying $C(0) = C'(0) = 0$ and $C''(q) > 0$ for all $q > 0$. Results are stated for the power-cost family $C(q) = (c/n)\, q^n$ with $n \geq 2$, which nests the baseline quadratic ($n = 2$).

Proposition 4 (Leveling down under power costs)**.** *Under the Goldberg mechanism with cost $C(q) = (c/n)\, q^n$ ($n \geq 2$), unconstrained optimal qualities are $q_F = [(m + \phi)/c]^{1/(n-1)}$ and $q_D = (m/c)^{1/(n-1)}$. The zero-profit constraint $mq \geq C(q)$ yields a maximum feasible quality $\bar{q}(m) = (nm/c)^{1/(n-1)}$. When the constraint binds, the discrimination gap is*

$$\Delta^*(m) = \left(\frac{m}{c}\right)^{1/(n-1)} \left(n^{1/(n-1)} - 1\right),$$

which is strictly increasing in m and converges to zero as $m \to 0$. Favored quality falls at $n^{1/(n-1)}$ times the rate of disfavored quality.

Proof. The provider's objective for group g is $\Pi_g(q; m) = mq - (c/n)\, q^n + \phi \cdot \mathbf{1}[g = F] \cdot q$. The first-order condition for favored customers, $m + \phi - cq^{n-1} = 0$, yields $q_F = [(m+\phi)/c]^{1/(n-1)}$.

For disfavored customers, $m - cq^{n-1} = 0$ yields $q_D = (m/c)^{1/(n-1)}$.

The zero-profit constraint requires $mq - (c/n)\, q^n \geq 0$, i.e., $q^{n-1} \leq nm/c$, so $q \leq (nm/c)^{1/(n-1)} \equiv \bar{q}(m)$. When the constraint binds:

$$\Delta^*(m) = \bar{q}(m) - q_D(m) = (nm/c)^{1/(n-1)} - (m/c)^{1/(n-1)} = (m/c)^{1/(n-1)}(n^{1/(n-1)} - 1).$$

Since $n^{1/(n-1)} > 1$ for all $n \geq 2$ and $(m/c)^{1/(n-1)}$ is strictly increasing in m, $\Delta^*(m)$ is strictly increasing in m and converges to zero as $m \to 0$.

Differentiating, $\partial\bar{q}/\partial m = n^{1/(n-1)}/[(n-1)(m/c)^{(n-2)/(n-1)}c]$ and $\partial q_D/\partial m = 1/[(n-1)(m/c)^{(n-2)/(n-1)}c]$. The ratio is $n^{1/(n-1)}$. For $n = 2$, this ratio is 2, recovering the baseline. For $n = 3$, the ratio is $\sqrt{3} \approx 1.73$. □

Proposition 5 (Becker gap under general costs)**.** *Under the Becker mechanism with cost $C(q)$ satisfying $C''(q) > 0$, the discrimination gap $\Delta^B(m) = (C')^{-1}(m) - (C')^{-1}(m-d)$ satisfies*

$$\frac{\partial \Delta^B}{\partial m} = \frac{1}{C''(q_F^B)} - \frac{1}{C''(q_D^B)},$$

whose sign equals $-\text{sgn}(C''')$:

- *$C''' = 0$ (quadratic): $\partial\Delta^B/\partial m = 0$. The gap is constant in margins.*
- *$C''' > 0$ (e.g., $n > 2$): $\partial\Delta^B/\partial m < 0$. The gap* widens *as margins fall.*
- *$C''' < 0$: $\partial\Delta^B/\partial m > 0$. The gap narrows as margins fall.*

Proof. Under Becker, the first-order conditions $C'(q_F^B) = m$ and $C'(q_D^B) = m - d$ yield $q_F^B = (C')^{-1}(m)$ and $q_D^B = (C')^{-1}(m-d)$. Differentiating:

$$\frac{\partial q_F^B}{\partial m} = \frac{1}{C''(q_F^B)}, \qquad \frac{\partial q_D^B}{\partial m} = \frac{1}{C''(q_D^B)}.$$

Since $q_F^B > q_D^B$, the sign of $\partial\Delta^B/\partial m = 1/C''(q_F^B) - 1/C''(q_D^B)$ depends on whether C'' is increasing or decreasing. When $C''' > 0$, $C''(q_F^B) > C''(q_D^B)$, so $1/C''(q_F^B) < 1/C''(q_D^B)$ and $\partial\Delta^B/\partial m < 0$: the gap widens as margins fall. When $C''' = 0$, C'' is constant and $\partial\Delta^B/\partial m = 0$. When $C''' < 0$, the inequality reverses.

For the power-cost family, $C'''(q) = c(n-1)(n-2)\, q^{n-3}$, which is positive for $n > 2$, zero for $n = 2$, and negative for $1 < n < 2$ (at $q > 0$). □

Discussion. The leveling-down prediction under Goldberg is qualitatively robust across the power-cost family: for all $n \geq 2$, the zero-profit constraint caps favored-customer quality

and the gap narrows as margins fall. The key mechanism, that preferential treatment requires surplus and competition erodes surplus, depends on the zero-profit constraint, not on the specific curvature of costs. The Becker constant-gap prediction, by contrast, is a knife-edge of the quadratic specification. Under the empirically natural case $C''' > 0$ (the marginal cost of quality improvements is itself increasing), the Becker model predicts that competition *widens* the discrimination gap. This prediction runs opposite to the empirical finding, strengthening the case for the Goldberg mechanism independently of functional form.

E Spatial Sorting (Extensive Margin)

This appendix examines the extensive margin: where drivers position themselves between trips, a choice over neighborhoods rather than routes. Similarity is therefore measured at the Community District Tabulation Area (CDTA) level, New York City's 68 community districts, using the construction of Section 2.3 with CDTA population shares in place of block shares. The location-choice sample comprises 4.88 million airport trips with an observed next pickup in the same shift; requiring an observed next pickup excludes shift-ending trips.

Drivers position their next pickup toward destinations that resemble them. Table E.1 regresses an indicator for whether a driver's next pickup falls in the same community district as the current dropoff on driver–destination similarity, within driver and dropoff-district fixed effects. The similarity coefficient is positive and significant for every measure: a driver who drops off in a demographically similar district is more likely to seek the next fare there. Figure E.1 shows the same gradient in binned form. The interaction with the post-entry indicator is small and, for the two primary measures, statistically insignificant, so competition did not clearly erode this sorting the way it compressed the route-efficiency gap. The extensive margin is a weaker and less cleanly identified channel than the intensive margin, which is why it sits in the appendix rather than the main text.

The pattern extends to non-airport trips. Table E.2 estimates the same specification on street hails: drivers again steer the next pickup toward own-group districts, and again the post-entry interaction is indistinguishable from zero. Sorting is a robust feature of how drivers move through the city, but its response to competition is not.

Figure E.1: Spatial Sorting: Location Choice and Own-Group Exposure

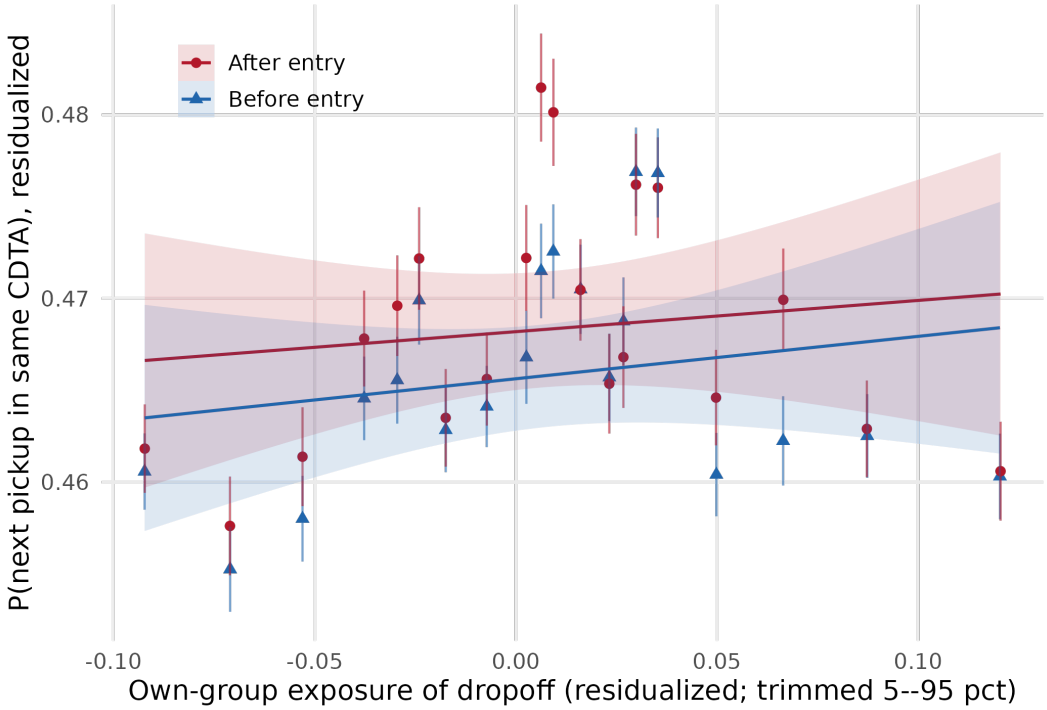


Table E.1: Location Choice: Spatial Sorting by Racial Similarity

	Own-Group (1)	Expected Match (2)	Cosine Sim. (3)
Similarity	0.0173***	0.0219***	0.0125***
	(0.0043)	(0.0045)	(0.0024)
Similarity × Post	-0.0016	-0.0054	-0.0036**
	(0.0037)	(0.0037)	(0.0015)
Post	0.0023***	0.0033***	0.0038***
	(0.0005)	(0.0009)	(0.0008)
Driver FE	Yes	Yes	Yes
Dropoff-CDTA FE	Yes	Yes	Yes
Hour, day-of-week, JFK FE	Yes	Yes	Yes
Mean dep. var.	0.467	0.467	0.467
Observations	4,884,367	4,884,367	4,884,367

Notes: Extensive-margin spatial sorting, reported in the appendix because it is not a headline mechanism in this version. The dependent variable is an indicator for whether the driver's next pickup occurs in the same community district (CDTA) as the current dropoff. *Similarity* is the CDTA-level driver–destination racial similarity; *Post* indicates trips after July 2013 (green-cab entry). A positive Similarity coefficient means drivers sort their next pickup toward destinations demographically like themselves; Similarity × Post measures how competition changes that sorting. All columns include driver, dropoff-CDTA, hour-of-day, day-of-week, and JFK-pickup fixed effects. Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table E.2: Non-Airport Spatial Sorting Before and After Entry

	Own-Group	Expected Match	Cosine
Similarity	0.0362***	0.0471***	0.0285***
	(0.0026)	(0.0030)	(0.0017)
Similarity × Post	0.0029	-0.0012	-0.0006
	(0.0020)	(0.0020)	(0.0009)
Observations		19,155,441	
Driver, hour, day-of-week, dropoff-CDTA FE	Yes	Yes	Yes

Notes: Outcome is an indicator that the driver's next pickup is in the same CDTA as the current dropoff. Estimated on the non-airport sample; *Post* indicates trips after July 2013. Standard errors clustered by driver. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.