A Appendix

A.1 Spatial Analysis: Labor Supply Models with Discrimination

A.1.1 Taste-Based Discrimination

Let $j = 0, 1$ serve to index demographic group concentration, where 0 denotes a low minority share and 1 indicates a high minority share, $p = 0, 1$ indexes large geographic places, and $i = 1, ..., N$ indicates cab drivers. Places have no demographic profile unless paired with groups to form "areas" a (group-place pairings). I assume: 1) driver utility is intertemporally separable, and 2) driver period-specific utility functions are identical. Assumption 2 simplifies the setup but could be relaxed if desired. Assumption 1, while strong, allows for two stage budgeting, where in stage 1, a driver allocates total per period consumption and leisure/labor across periods, and in stage 2, a driver allocates total within-period consumption and leisure/labor between those two choice variables.

I focus on the stage 2 problem for a given period t, thereby suppressing time notation. In choosing total leisure and labor within a period, a driver also simultaneously decides across which areas to allocate labor. The driver i problem is:

$$\max_{C_i, H_{i1}, ..., H_{iA}} U = U(C_i, H_{i1}, ..., H_{iA}),$$

where C is consumption and H_a is hours driven in area a. If T is defined as the total available hours within a period (for example, 24 hours in a day), H is total hours, and L is leisure, $T = H_i + L_i = \sum_a H_{ia} + L_i$ and is fixed. In words, driver i's choice regarding the amount of leisure time pins down the amount of total hours driven, which is jointly determined with how those labor hours are allocated across areas, so in the driver problem, labor hours can be used as the choice variables instead of leisure.

For simplicity but without loss of generality, assume that there are only two areas: $a = 0 \equiv \{j = 0, p = 0\}$ (a low minority share-place 0 pairing), and $a = 1 \equiv \{j = 1, p = 1\}$ (a high minority share-place 1 pairing).[38] The cost to driver i of not supplying labor to a given area a is $H_{ia}(W_a - d_{ia})$, where W_a is the area wage, common to *all* drivers (that is, the area wage is demand-driven, so it varies by area a but not driver by i), and d_{ia} is area-specific distaste on the part of driver i, where $\forall i$, $d_{i0} = 0$ and $d_{i1} \neq 0$. In words, d_{ia} measures the strength of prejudice for area a by driver i due to area a's minority composition. This area-specific distaste lowers the cost of *not* working in that area.[39]

For a high-minority area (here, area 1), there is a distribution of distaste parameters d_{i1} across drivers, $f_1(d)$. In a low-minority area (here, area 0), the distribution of distaste parameters across drivers is degenerate at 0. Also, for a given driver (for example, driver 1), there is a distribution of distaste parameters d_{1a} across areas, $g_1(d)$. Thus, the heterogeneity of discrimination across drivers or areas, respectively, is captured by the f and g distributions.

[38] In other words, $\{j = 0, p = 1\}$ and $\{j = 1, p = 0\}$ pairings do not exist.

[39] Even without time notation suppressed, d_{ia} is assumed to be time-invariant. While discrimination preferences might change over time for some individuals, it seems plausible to assume that they are stable for Boston cab drivers given the older age of this population (49 years old on average, as compared to a 35-year-old average age in the Boston population, according to 2010–2015 American Community Survey data).

Within the driver-area dimension, there is no variation in distaste.

The extensive margin choice to supply any labor in an area will be dependent on the area-specific wage net of the distaste ($W_a - d_{ia}$) being greater than some driver-specific reservation wage, r_i, constant across areas. Assuming an interior solution, the within-period tradeoff between hours worked in a low minority area and a high minority area is:

$$\frac{\partial U(C_i, H_{i1}, ..., H_{iA})/\partial H_{i0}}{\partial U(C_i, H_{i1}, ..., H_{iA})/\partial H_{i1}} = \frac{W_0}{W_1 - d_{i1}}.$$

If I assume a functional form for U, using MaCurdy (1981) as a guide, I can specify $U(C_i, H_{i1}, ..., H_{iA}) = \gamma_i C_i^{\omega_c} - (\sum_a \phi_{ia} H_{ia}^{\omega_a})$. Here, ω is related to how consumption and hours are substituted across periods, where ω_a specifically relates to the intertemporal wage elasticity of substitution for area a, $\beta_a = 1/(\omega_a - 1)$. Additionally, γ and ϕ represent taste shifters that vary across individuals or individual-area pairings, respectively, and are unrelated to area-specific discrimination. The within-period hours tradeoff across areas is now:

$$\frac{\omega_0 \phi_{i0} H_{i0}^{\omega_0 - 1}}{\omega_1 \phi_{i1} H_{i1}^{\omega_1 - 1}} = \frac{W_0}{W_1 - d_{i1}} \Leftrightarrow \frac{H_{i0}^{\omega_0 - 1}}{H_{i1}^{\omega_1 - 1}} = \frac{\omega_1 \phi_{i1} W_0}{\omega_0 \phi_{i0}(W_1 - d_{i1})}.$$

Without trying to simplify this expression further, it is already apparent that the hours differential across low and high minority areas for driver i will be related to the wage difference across areas, the distaste for the high minority area by driver i, the intertemporal elasticity, as well as any driver-specific tastes that are not related to discrimination. I can impose the simplifying assumptions that non-discriminatory, driver-specific tastes do not differ across areas ($\phi_{i0} = \phi_{i1}$), and that the parameter related to the intertemporal elasticity also does not differ across areas ($\omega_0 = \omega_1 = \omega$), resulting in:

$$\ln\left(\frac{H_{i0}}{H_{i1}}\right) = \beta \ln\left(\frac{W_0}{W_1 - d_{i1}}\right),$$

where the log hours differential across areas for driver i is a function of the log differential in wages across areas net of any driver-area-specific discriminatory tastes, scaled by the intertemporal, net-of-discrimination wage elasticity, $\beta = 1/(\omega - 1)$. In the absence of discrimination ($d_{i1} = 0$), differences in hours across areas are fully explained by differences in wages, and $\beta = \beta_0$ is the observed wage elasticity. In this case, while there may be an intercept difference in driver labor supply across areas, there will not be a slope difference, as β_0 will be the same regardless of minority representation in an area.[40]

However, with discrimination present ($d_{i1} \neq 0$), now a slope difference across areas is expected, as higher values of wages in the minority area, W_1, are now required to obtain the same work hours as before, H_{i1}. In other words, β_0 is no longer the observed wage elasticity in this case since β now incorporates the unobserved wage net of discrimination

[40] While technically not identifying a disparity in labor supply at the x-axis, I nevertheless refer to a "level" disparity in hours at a given set of wages as the intercept difference. Similarly, while specifically identifying an elasticity difference in labor supply, I refer to a "change" disparity in hours across various sets of wages as the slope difference. The signs of the slope and elasticity parameters will be the same.

in the minority area, $\widetilde{W}_{i1} = W_1 - d_{i1}$. The observed elasticity, β_1, which summarizes the relationship between observed wages (W_0, W_1) and hours (H_{i0}, H_{i1}) with discrimination present, is smaller than β_0, reflecting diminished wage sensitivity in the minority area due to discrimination.

A.1.2 Statistical Discrimination

Let $j = 0, 1$ serve to index demographic group concentration, where 0 is a low minority share and 1 is a high minority share, $p = 0, 1$ indexes large geographic places, and $i = 1, ..., N$ indicates cab drivers. Once again, places have no demographic profile unless paired with groups to form "areas" (group-place pairings).

Define a_{pj} as the component of the log wage anticipated by *all* drivers (the log wage is demand-driven and varies by place p and group j, not driver i), w_{ipj} is the realized ("measured") log wage faced by driver i (an imperfect indicator of a_{pj}), and u_{ipj} is the component of the log wage unanticipated by driver i ("errors," driven by both demand and supply). Let $w_{ipj} = a_{pj} + u_{ipj}$, with $E(u_{ipj}|a_{pj}, j, p) = 0$. Thus, u_{ipj} is uncorrelated with: 1) anticipated wage a_{pj} (classical measurement error), 2) minority share j ("errors" are unbiased, that is, no distaste is present), and 3) place p (no spatial component to "errors").[41]

Regarding the extensive margin hours choice, I can define r_i as the reservation log wage, constant across place-groups, while H_{ipj} is hours driven by driver i in place-group pj. Let $H_{ipj} = 0$ if $E(a_{pj}|w_{ipj}) < r_i$, while $H_{ipj} > 0$ if $E(a_{pj}|w_{ipj}) \geq r_i$. Focusing on the intensive margin hours choice, h_{ipj} is log hours driven, defined for $H > 0$. Thus, $h_{ipj} = ln(H_{ipj}) = \beta E(a_{pj}|w_{ipj})$, where β is the "expected wage" elasticity of labor supply.

Assuming the minority share j is observable in each area (that is, each pj pairing), I can define the expected conditional anticipated wage as:

$$E(a_{pj}|w_{ipj}, j) = (1 - \psi_{ij})\overline{a}_j + \psi_{ij}(w_{ipj}),$$

where $\psi_{ij} = \frac{\sigma^2_{a,j}}{\sigma^2_{w,ij}} = \frac{\sigma^2_{a,j}}{\sigma^2_{a,j}+\sigma^2_{u,ij}} \in [0, 1]$. Define ψ_{ij} as the "reliability ratio" (that is, the signal to total variance ratio) displaying the reliability of driver i's place-specific realized wage, in terms of indicating the anticipated component of wages, in areas with a given minority share j (and, thus, the weight placed on those observations).

Examining extreme cases offers some intuition on how the reliability ratio affects labor supply. If $\psi_{ij} = 1$, then $h_{ipj} = \beta(w_{ipj})$, showing that the minority share does not affect driver i's hours decision in any place(s) p with minority share j (that is, there are no variables indexed only by j). Conversely, if $\psi_{ij} = 0$, then $h_{ipj} = \beta\overline{a}_j$, showing that the minority share is the only factor that matters for driver i's hours decision in any place(s) p with minority share j (that is, there are only variables that are solely indexed by j). Here, a driver's hours decision for a place-minority share pairing will be identical for all places with a given

[41] Alternatively, one could model how well the anticipated wage proxies for the realized wage given potential errors from the unanticipated wage. I use the given approach instead because: a) the classical measurement error assumption would not hold with the alternative approach (unanticipated wages are correlated with realized wages by definition); and b) for both approaches, the focus is on how closely the variation in realized wages corresponds to the variation in anticipated wages.

minority share. Note that $\partial\psi_{ij}/\partial\sigma^2_{a,j} > 0$, pushing results closer to the $\psi_{ij} = 1$ case, while $\partial\psi_{ij}/\partial\sigma^2_{u,ij} < 0$, pushing results closer to the $\psi_{ij} = 0$ case.

To derive more general results than these extreme cases, assume that by minority share, there are: i) different anticipated wage means $\overline{a}_j$, and ii) the same reliability of ψ_{ij}. More specifically, I assume $\overline{a}_0 > \overline{a}_1$ and $\psi_{i0} = \psi_{i1} = \psi_i < 1$. Given this, I can write down the expectation for the low minority place anticipated wage:

$$E(a_{p0}|w_{ip0}, j = 0) = (1 - \psi_{i0})\overline{a}_0 + \psi_{i0}(w_{ip0}),$$

as well as the expectation for the high minority place anticipated wage:

$$E(a_{p1}|w_{ip1}, j = 1) = (1 - \psi_{i1})\overline{a}_1 + \psi_{i1}(w_{ip1}).$$

The log hours differential between low and high minority share places with the same realized wage for driver i ($w_{ip0} = w_{ip1}$) is $\beta(1 - \psi_i)(\overline{a}_0 - \overline{a}_1) > 0$. Thus, for an equivalent set of realized wages, a driver will still spend more hours driving in the low minority place than the high minority place. Here, by construction, the wage elasticity β does not differ by minority group, and so the hours disparity reflects an intercept difference in driver labor supply across areas. However, if the elasticity is allowed to vary by minority group, the log hours differential for driver i is $(1 - \psi_i)(\beta_0\overline{a}_0 - \beta_1\overline{a}_1) > 0$ given a sufficient assumption that $\beta_0 \geq \beta_1$ (labor supply is weakly more elastic in the low minority place). In this case, the log hours differential gets larger as β_0 gets more elastic or as β_1 gets more inelastic, introducing the possibility of a slope difference in driver labor supply across areas.

Lastly, it can also be shown that at each level of the *anticipated* wage, low minority places are serviced more compared to high minority places. The *expected* log hours differential for driver i between low and high minority share places with the same anticipated wage ($a_{p0} = a_{p1}$) is $\beta(1 - \psi_i)(\overline{a}_0 - \overline{a}_1) > 0$, which is the same expression obtained earlier when conditioning on realized wages. In summary, the result regarding the log hours differential across low and high minority places, conditional on driver realized wages, as well as the result regarding the expected log hours differential across low and high minority places, conditional on anticipated wages, both reflect supply-side discrimination.

A.2 Spatial Analysis: Additional Empirics

A.2.1 Alternative Estimation Strategy

To examine the impact of area demographic composition on cab driver labor supply while incorporating local earnings opportunities, rather than estimating the main text equation with area fixed effects, I could alternatively estimate the following equation:

$$lnH_{kidcta} = \mu + \mathbf{M}'_a\zeta + \beta lnW_{kidcta} + (lnW_{kidcta} \times \mathbf{M}_a)'\eta + \phi_d + \gamma_c + \theta_t + \pi_{dct} + \mathbf{X}'_a\lambda + \varepsilon_{kidcta}, \quad (3)$$

where, for shift k, driver i, day of the week d, calendar week of the year c, year t, and area a, H is the area-specific duration of a shift in hours, $\mathbf{M}$ is a vector of "minority"/demographic population shares (that is, black, Asian, Hispanic, female, and 65 years of age and older, all as measured in the 2010 Census), W is the area-specific average hourly earnings on a shift, and ε is an error term, with standard errors clustered at the driver level.

This specification also includes controls that either help to account for supply differences within areas, non-discriminatory supply differences across areas, or differences in demand across areas. First, $\mathbf{X}$ is a vector of area-specific characteristics potentially relevant to demand by individuals or non-discriminatory supply by drivers across areas, all as measured in the 2006–2010 five-year American Community Survey (Minnesota Population Center 2010). Namely, $\mathbf{X}$ includes the area share of workers 16 years of age and older who use a taxicab for transportation to work, the area share of workers 16 years of age and older who use a motorized vehicle for transportation to work, the log of area median household income, the log of area median gross rent (all intended to capture resident taxi demand), and three area shares of the population 25 years of age and older whose educational attainment is less than a high school diploma or GED, a high school diploma or GED, or else some college or an associate's degree (intended to capture non-discriminatory driver supply, if resident education is correlated with area amenities that drivers might care about).

Also, ϕ controls for day-of-week fixed effects, γ controls for week-of-year fixed effects, θ controls for year fixed effects, and π controls for major holidays. Similar to Farber (2015) and the main text equation, these additional controls help account for the anticipated variation in wages, which likely contributes to driver supply differences within areas, as well as passenger demand and non-discriminatory driver supply across areas. The ζ coefficients in equation (3) cannot be identified in the main text equation with area fixed effects, as these coefficients reflect differences in shift hours by the demographic population shares of an area *conditional* on the market wage (that is, these are the intercept differences at a given wage). Meanwhile, as in the main text equation, the η coefficients reflect differences in wage elasticities by area demographic shares. If there is discrimination, I expect $\zeta < 0$ and/or $\eta < 0$.

While equation (3) accounts for local earnings opportunities via area-specific wages, I may nevertheless remain concerned that the limited controls in $\mathbf{X}$ do not sufficiently account for demand-relevant or non-discriminatory supply-relevant characteristics across areas that are correlated with $\mathbf{M}$, thus resulting in inconsistent estimates of the ζ and η coefficients. Indeed, in estimating equation (3), I observe signs on some of the controls that do not align with a priori reasoning (for example, a negative sign on the area share of workers 16 years of age and older who use a taxicab for transportation to work), perhaps indicating biased estimation.

One possible solution to this concern is the inclusion of fixed effects at the region level or the region-year level (for example, large neighborhoods) if the correlation of $\mathbf{M}_a$ with ε_{kidcta} occurs at this larger geographic level rather than at the area a level. For instance, perhaps non-discriminatory driver supply decisions are affected by criminal activity at a larger geographic boundary but not at the block group boundary, and such crime is correlated with block group demographics. Thus, with region or region $\times$ year effects, the estimation occurs only within regions or region-years, respectively, rather than across them.[42] However, the

[42]Regions are large Boston neighborhoods or Massachusetts counties. They are specified as 25 Boston neighborhoods inside of Boston (largely following neighborhood boundaries and underlying 2010 Census tracts from the Boston Planning & Development Agency (2010), except that "Downtown" is split into Chinatown (tract 702) and the remainder of Downtown, and "Dorchester" is split into North Dorchester (tracts 914, 915, 910.01, and all other Dorchester tracts located north of those) and South Dorchester (tracts 903, 916, 918, 921.01, and all other Dorchester tracts located south of those)), with another region being the

signs on control variables from such specifications still raise doubts about whether consistent estimation has been achieved, as region or region-year effects may not fully account for all differences in demand or non-discriminatory supply across areas. Thus, rather than pursue equation (3) and estimating both ζ and η, I focus on the main text equation and estimating η only.

balance of Suffolk County outside of Boston, and the remaining 13 regions being the rest of the 13 counties in Massachusetts apart from Suffolk County, for a total of 39 regions. Thus, the implication for identification is that demand-relevant or non-discriminatory supply-relevant characteristics vary at a smaller geographic boundary within Boston than in the rest of Suffolk County or Massachusetts. For instance, in the case of non-discriminatory supply choices motivated by area crime rates, this assumption is consistent with cab drivers from the Boston area having more detailed knowledge of how crime varies across areas within Boston than outside of Boston.

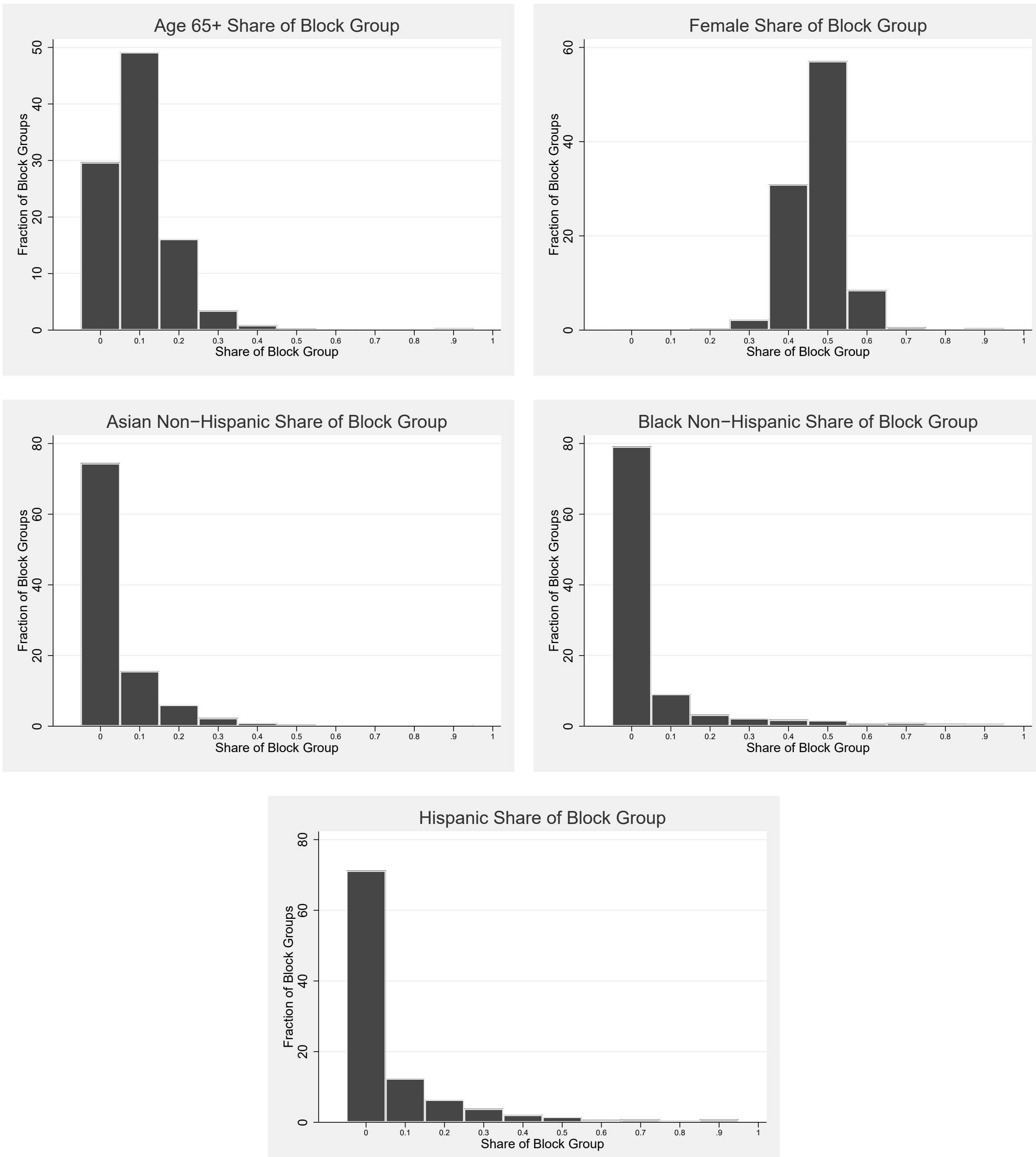

Figure A1: Boston Block Group 2010 Population Shares, by Demographic Group and Decile
Source: 2010 U.S. Census and author's calculations.

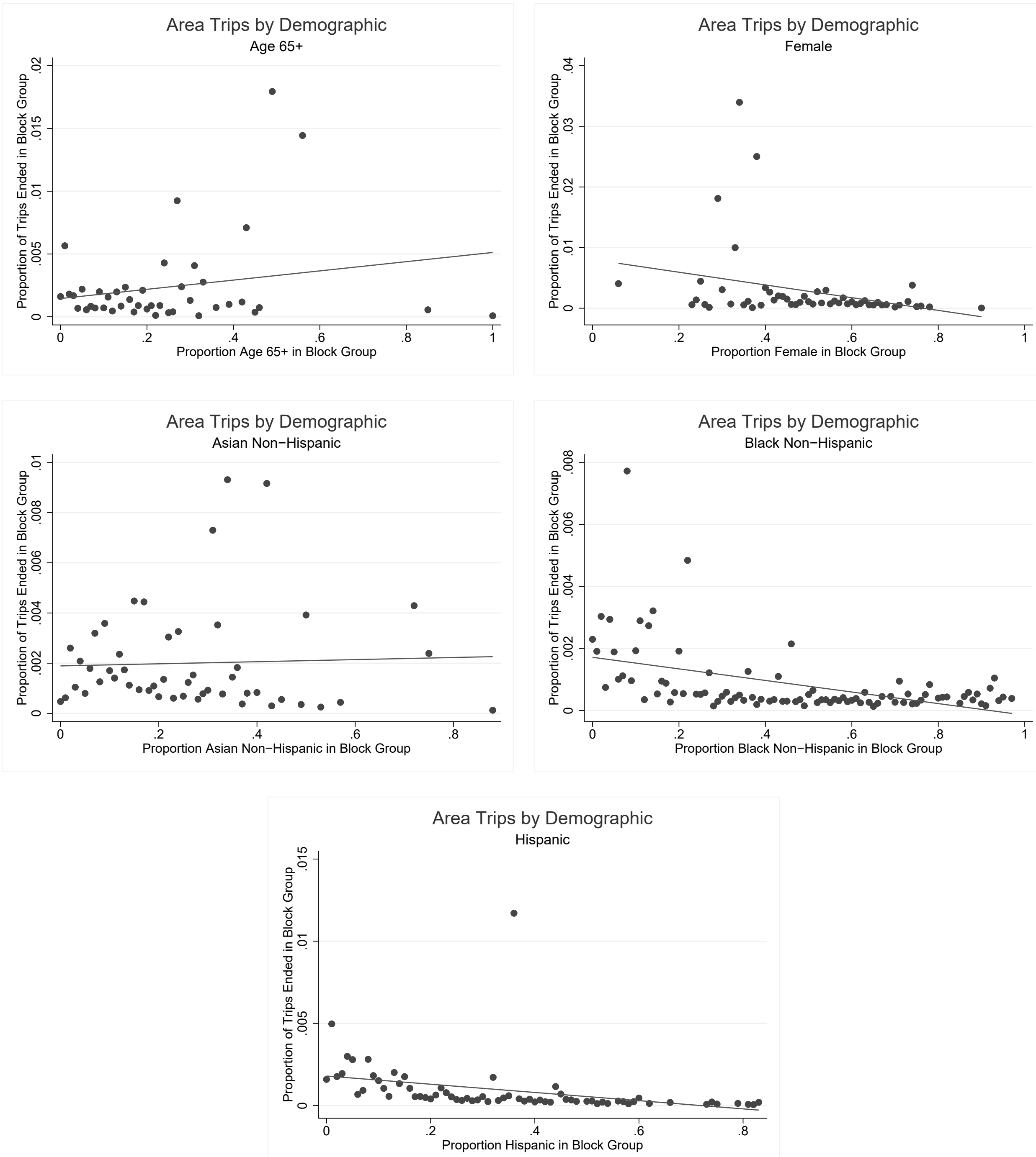

Figure A2: Trip Drop-off Areas and Area Population, by Demographic Group
Source: Boston taxi data, 2010 U.S. Census, and author's calculations.

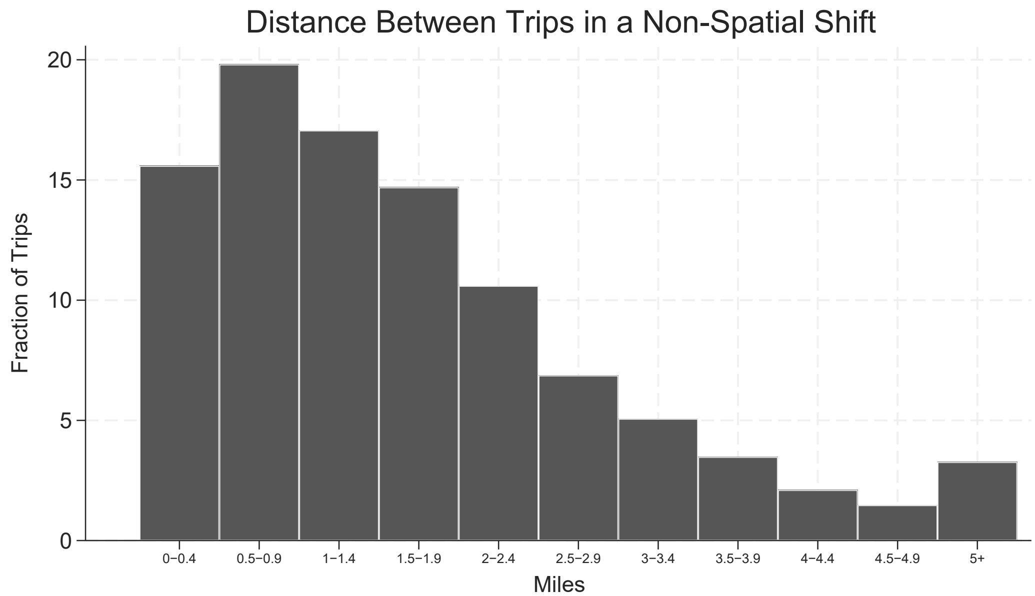

Figure A3: Distribution of Intervening Trip Distance
Source: Boston taxi data and author's calculations.

Table A1: Characteristics of Taxi Drivers and Population

	Boston (Means)		United States (Means)	
Variable	Drivers	Population	Drivers	Population
Wage Income (2010 USD)	20,298	48,233	19,267	40,961
Usual Hours Worked (Weekly)	42.0	38.4	40.7	39.0
Share, Population (%)	0.3	100.0	0.1	100.0
Age (Years)	48.7	35.4	49.2	37.8
Share, Age 65+ (%)	12.7	10.3	15.6	13.9
Share, Female (%)	8.4	52.2	14.6	50.8
Share, White Non-Hispanic (%)	22.1	46.0	45.4	62.6
Share, Black Non-Hispanic (%)	61.5	22.7	23.8	12.3
Share, Asian Non-Hispanic (%)	1.8	9.2	11.4	5.1
Share, Hispanic (%)	10.7	18.5	16.5	17.0
Share, Other Non-Hispanic (%)	3.9	3.7	2.9	3.0
Share, Foreign-Born (%)	75.6	30.5	42.8	14.5
Person Count	108	38,174	23,816	18,699,149

Notes: 2010–2015 American Community Survey and author's calculations. Person weights applied to means. Wage income and usual hours worked restricted to employed persons only. "Drivers" are from occupation code 9140: taxi drivers and chauffeurs.

Table A2: More Robustness Checks, IV Regressions of Log Area Shift Duration in Hours

Model Description	(1) Day Shifts	(2) Other Shifts	(3) Elasticity, Wage	(3) *continued* Elasticity, Cost (Distance)
Elasticity	0.050	0.025	0.096	–0.050*
	(0.191)	(0.302)	(0.150)	(0.029)
× Female	–1.030***	0.071	–0.625**	0.120**
	(0.350)	(0.622)	(0.270)	(0.051)
× Black	–0.910**	–1.019	–0.702**	0.153***
	(0.408)	(1.112)	(0.297)	(0.048)
× Asian	–0.523*	–0.654*	–0.746***	0.127***
	(0.306)	(0.340)	(0.234)	(0.044)
× Hispanic	0.915*	0.864	0.612	–0.031
	(0.507)	(0.863)	(0.487)	(0.079)
× Age 65+	0.586**	0.234	0.354*	–0.093**
	(0.234)	(0.358)	(0.188)	(0.037)
Area Shifts	1,640,790	625,651		3,584,793
Drivers	2,353	2,398		2,962

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Author's calculations using Boston taxi data from 2010–2015. Each model displays a set of estimated elasticities from a single IV regression of log area shift duration, as noted. "Elasticity" is the estimated coefficient of log area average hourly earnings, where an "area" is a 2010 U.S. Census block group. "× 'Group' " is the estimated coefficient of log area average hourly earnings interacted with a vector of demographic group area population shares, where "Group" is either female, black non-Hispanic, Asian non-Hispanic, Hispanic, or 65 years of age and older. The instrument for log area average hourly earnings is the log average across drivers of area average hourly earnings for a non-overlapping sample of drivers on the same day and in the same area, with additional instruments also interacted with "Group." All regressions include as controls indicators for major holiday, day of week × year, calendar week × year, area × year, and driver × area. Singleton observations (area shifts) within a given fixed effect indicator are dropped. Model (1) restricts to area shifts occurring during a non-spatial shift that starts between 4AM and 9:59AM. Model (2) restricts to area shifts occurring during a non-spatial shift that starts between 10AM and 1:59PM or between 8PM and 3:59AM. Model (3) includes an estimate of log average area hourly trip distance (using trip start and end location pairings when both are available, at four-decimal point latitude-longitude coordinates, and estimating trip routes via an Open Source Routing Machine at `http://project-osrm.org`) as a proxy for area hourly costs. Standard errors clustered by driver are in parentheses.

Table A3: Alternative Area Wage Elasticity and Alternative Area Wage Elasticity × Area Population Shares, Trip-Level IV Regressions of Log Intervening Trip Distance in Miles (v.2)

Model	(1)	(2)	(3)	(4)
Key F.E.'s	Area	Area × Year	Area × Year, Driver	Area × Year, Driver × Area
End Elasticity	–1.649	–1.560	–1.306	–1.347
	(1.123)	(1.347)	(1.133)	(0.944)
× Female	–1.931	–1.559	–1.295	–0.591
	(1.487)	(1.326)	(1.189)	(1.052)
× Black	1.444	1.972**	1.584**	0.525
	(0.926)	(0.869)	(0.803)	(0.680)
× Asian	0.130	0.333	–0.066	0.405
	(1.203)	(1.067)	(0.965)	(0.779)
× Hispanic	0.742	0.511	0.238	0.693
	(1.561)	(1.696)	(1.446)	(1.229)
× Age 65+	1.611	1.342	0.973	0.755
	(0.993)	(1.063)	(0.923)	(0.882)
Start Elasticity	3.261***	3.597**	2.941**	2.315*
	(1.026)	(1.511)	(1.275)	(1.231)
× Female	–1.887*	–3.461***	–2.627**	–1.613*
	(1.062)	(1.307)	(1.095)	(0.963)
× Black	–1.325	–1.580	–1.387	–0.262
	(1.034)	(0.994)	(0.854)	(0.872)
× Asian	–0.645	–0.106	–0.187	–1.333
	(1.206)	(1.215)	(1.061)	(1.062)
× Hispanic	1.867	3.132**	2.579**	0.071
	(1.256)	(1.324)	(1.120)	(0.951)
× Age 65+	–0.771	–0.604	–0.869	–0.475
	(0.686)	(0.743)	(0.630)	(0.532)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Author's calculations using Boston taxi data. Each column displays a set of estimated elasticities from a single IV regression of log intervening trip distance, as noted. For v.2, trip distance is determined using trip start and end location pairings when both are available, at four-decimal point latitude-longitude coordinates, and estimating trip routes via an Open Source Routing Machine at `http://project-osrm.org`. " 'Location' Elasticity" is the estimated coefficient of log area average hourly earnings, where an "area" is a 2010 U.S. Census block group, and "Location" is either the current trip ending block group ("End") or the next trip starting block group ("Start"). "× 'Group' " is the estimated coefficient of log area average hourly earnings interacted with a vector of demographic group area population shares, where "Group" is either female, black non-Hispanic, Asian non-Hispanic, Hispanic, or 65 years of age and older, and "area" corresponds to "End" or "Start" as indicated. The instrument for log area average hourly earnings is the log average across drivers of area average hourly earnings for a non-overlapping sample of drivers on the same day and in the same area ("end" or "start"), with additional instruments also interacted with "Group." All regressions include an indicator for major holiday as a control. Model 1's additional controls are indicators for day of week, calendar week, year, and area (three area types: current trip start [origin], current trip end, and next trip start). Model 2 replaces Model 1's additional controls with indicators for day of week × year, calendar week × year, and area × year (for each of the three area types). Model 3 has all of Model 2's additional controls plus indicators for driver. Model 4 has all of Model 2's additional controls plus indicators for driver × area (for each of the three area types). The sample is estimated using 1,097,819 trips for 2,386 drivers from 2010–2015. Standard errors clustered by driver are in parentheses.

Table A4: Correlations of Pick-up Area Demographics and Drop-off Area Demographics

Measure	Pick-up Female Share	Pick-up Black Share	Pick-up Asian Share	Pick-up Hispanic Share	Pick-up Age 65+ Share
Drop-off Female Share	0.0694	0.0794	0.0068	0.0584	–0.0341
Drop-off Black Share	0.0891	0.4610	–0.0582	0.0983	–0.0301
Drop-off Asian Share	0.0082	–0.0658	0.0787	–0.0113	–0.0005
Drop-off Hispanic Share	0.0419	0.1160	–0.0060	0.0828	–0.0065
Drop-off Age 65+ Share	–0.0317	–0.0160	–0.0139	0.0150	0.0352

Notes: Author's calculations using Boston taxi data from 2010–2015. Each cell displays the correlation between the "Group" area population share in a trip's pick-up area and the "Group" area population share in a trip's drop-off area, where an "area" is a 2010 U.S. Census block group, and a "Group" is one of the following demographic groups: female, black non-Hispanic, Asian non-Hispanic, Hispanic, or 65 years of age and older. Estimated using a sample of 21,635,727 trips from 2010–2015. All correlations are statistically significant with $p < 0.01$.