Neighborhood Mismatch and Visits*

Jing Zhi Lim† Lucas Shen‡

December 2024

Abstract

This study combines GPS traces with census and housing microtransactions to examine how socioeconomic mismatches shape daily movements in Singapore. We measure mismatch as a difference in income levels or ethnic composition, which naturally decomposes into the directional components of mismatch. We find stark asymmetries. Income-based segregation is driven by a lack of visits from richer to poorer neighborhoods, while ethnic-based segregation is driven by a lack of visits from majority- to minority-ethnic neighborhoods. We take our estimates to a counterfactual analysis of an extant ethnic residential integration policy. The findings show that the policy's effect on fostering cross-neighborhood flows would have been three times smaller without accounting for directional asymmetries in ethnic mismatch.

JEL Codes: N35, R23, R28

Keywords: segregation, income, race, GPS, public housing, ethnic integration

*We are especially grateful to CITYDATA.ai (Apurva Kumar and Lee Yew Leong) for an early data grant and advice. The data granters otherwise had no role in the study design, data preparation, analysis, and manuscript preparation. We thank participants at the 16th North American Meeting of the Urban Economics Association, Paul Cheung, Huang Yuting, Gaurav Sood, and colleagues for valuable feedback. We thank Li Changtai for helpful advice on the housing price data and Lee Shu En for research assistance on policies and background. An earlier version of this working paper was circulated under the title, "Segregation Across Neighborhoods in a Small City", which has been briefly discussed at the 2021 UEA Summer School and the Singapore Geospatial Festival 2021. All interpretations in this manuscript are those of the authors and not necessarily those of affiliated institutions. JZL Email: jlim@pardeerand.edu. LS Email: lucas@lucasshen.com (corresponding).

†Pardee RAND Graduate School.

‡National University of Singapore.

1 Introduction

The places people visit daily are more than stops in a trip chain–they are spaces where social exposure and inequalities are experienced (Jones and Pebley 2014; Davis et al. 2019; Dong et al. 2020; Athey et al. 2021; Hilman et al. 2021). While residential markers have long been used as measures of social segregation (Hutchens 2001; Glaeser and Vigdor 2012; Rodriguez-Moral and Vorsatz 2016; Cook et al. 2018; Wang et al. 2018) and to evaluate public policies (Sin 2002a; Loo et al. 2003; Wong 2014; Massey 2015; Choe 2016; Leong et al. 2020; Tan 2023; Leong et al. 2024), they capture only part of the picture. Tracking neighborhood visits, a critical dynamic of social interactions, offers a fuller picture of the realities of social exposure that can often be uneven and asymmetric (Athey et al. 2021; Davis et al. 2019; Dong et al. 2020; Moro et al. 2021; Hilman et al. 2021).

This paper investigates asymmetries in social exposure by combining GPS pings with data on social frictions, spatial frictions, and neighborhood amenities to analyze cross-neighborhood visitation patterns in Singapore. Specifically, we link GPS-derived neighborhood visitation patterns to two primary datasets: (i) housing microtransactions for a measure of neighborhood income and (ii) census data to on neighborhood ethnic composition. Holding spatial frictions and amenities constant, we examine if neighborhood mismatches in income and ethnicity reduce neighborhood visits. We focus on whether directional asymmetries (e.g., poor-to-non-poor vs. non-poor-to-poor) affect visits differently. Finally, we take our estimates on ethnic mismatch to evaluate a key public housing policy that enforces residential ethnic quotas.

Our primary data is GPS records collected over 3 months in 2020, aggregated by neighborhood and day. We link them to neighborhood (subzone) income and ethnic composition. We use geocoded housing resale transactions, weighted by the proportion of residents in public versus private housing within neighborhoods, to proxy for neighborhood income. At the (broader) planning area level, where census income is available, our income proxy

correlates strongly with census income ($\hat{r} = 0.9$). Neighborhood ethnic compositions come directly from the census. We also link to other spatial frictions measures and neighborhood amenities, including census demographic details, places of interest, registered businesses, daily precipitation, and inter-neighborhood spatial distances.

To measure income measure, we start by defining a "poor" neighborhood as one below the 25th percentile.[1] A mismatch between an origin and destination neighborhood occurs when one is poor and the other is not. The mismatch measure is rooted in the exclusive or operator, which assigns 1 when values differ and 0 otherwise. We extend the measure to ethnicity, using ethnic composition to estimate the probability of exposure between a representative minority-ethnic and a majority-ethnic individual from two neighborhoods. To analyze how the direction of mismatch affects visits, we decompose the mismatch measure into its two natural directional components. For example, with income mismatch, this means distinguishing between poor-to-non-poor and non-poor-to-poor directions.

We first examine income mismatch and neighborhood visits. Controlling for spatial frictions, demographics, amenities, rainfall, and places of interest, income mismatch increases neighborhood visits by 2 percent ($\approx 100 \times .021$, as log points, SE 0.9) compared to neighborhood pairs without mismatch. This suggests individuals derive utility from visiting neighborhoods of different income levels. We then decompose the mismatch measure into its directional components to examine asymmetries. We find that the neighborhood flow from income mismatch is driven by visits from poor to non-poor neighborhoods, but not vice versa. The estimate of 0.048 (SE 0.011) implies that visits are 5 percent higher with income mismatch and the origin neighborhood is poor. In contrast, change in visits in the reverse direction—non-poor to poor—is non-existent (estimate of -0.008, SE 0.011). This asymmetry implies that individuals from poorer neighborhoods are more exposed to richer neighborhoods, while the reverse exposure is muted.

Next, we find ethnic mismatch reduces neighborhood visits. The estimate of -0.21 (SE 0.10)

[1] We later systematically test and confirm that our broad findings hold under alternative percentile thresholds.

implies that a one standard deviation increase in ethnic mismatch decreases neighborhood visits by 1.5 percent ($\approx 100 \times 0.21 \times 0.07$, $p < 0.05$). Hence, we detect ethnic preferences in day-today neighborhood visits, consistent with studies showing homophilic preferences in residence location (Wong 2013; Davis et al. 2019; Büchel et al. 2020). As with income mismatch, directional ethnic mismatch reveals stark asymmetries. Our estimates indicate that only one direction—majority- to minority-ethnic neighborhoods—drives all the implied segregated from fewer neighborhood visits.

To confirm our income mismatch extends beyond the 25th percentile thresholds, we re-estimate with different thresholds for "poor" neighborhoods. The broad conclusions hold, including those for including asymmetry in directional mismatch. Using sorted estimates from the full range of thresholds, we further show that the 25th percentile was not chosen to inflate effect sizes. For ethnic mismatch, a constant elasticities framework (closer to a model on utilities from visits) produces similar results. Moreover, ethnic-based segregation reflects deeper behaviors around social frictions than just geographical artifacts.

We build on our estimates and the context of the study to evaluate an extant policy that integrates residents by ethnicity. A decade after the introduction of the Ethnic Integration Policy (EIP), the 2000 official statistics show that neighborhood ethnic compositions moved closer to the city-wide composition compared to counterfactual projections. We use this difference between actual and projected ethnic composition as a counterfactual to compute the implied increase in neighborhood visits resulting from the EIP. Without considering asymmetry, the increase in visits due to the EIP is minimal, no more than eliminating contemporaneous ethnic mismatchin the 1st percentile neighborhood pair. However, when considering asymmetry, the same increase in visits from the EIP is equivalent to eliminating majority-to-minority mismatch for the 74th percentile neighborhood pair. This counterfactual analysis demonstrates that the EIP's value in fostering social exposure is three times higher when accounting for asymmetries, highlighting their importance in shaping social exposure.

A key contribution of this study is to quantify asymmetries in experienced segregation

via day-to-day neighborhood visits using novel data and methods. Holding neighborhood amenities, spatial frictions, and periodicity constant, we find that the segregation effects of income and ethnic mismatches depend on who travels out. We add to a growing set of studies on asymmetries and heterogeneities in segregated interactions (Athey et al. 2021; Dong et al. 2020; Hilman et al. 2021). Dong et al. (2020) find poorer-to-richer neighborhood interactions drive segregation, a result we reproduce with different settings and methods. Applying computational methods to large-scale credit card transaction data, Dong et al. (2020) find poorer-to-richer neighborhood interactions drive segregation, a result we reproduce with different settings and methods. Hilman et al. (2021) find a preference toward visits to affluent places using social location data. Athey et al. (2021) link mobility-based segregation to residential segregation by race, with differences tied to neighborhood traits. Our study targets asymmetries more directly, examining both income- and ethnic-based mismatches, and connects the latter to a counterfactual policy experiment.

Using our model estimates of how ethnic mismatch relates to neighborhood flows, we evaluate an extant residential ethnic integration policy in our setting. Our counterfactual policy experiment emphasizes both the value and complexity of such integration policies, revealing the role of asymmetries in shaping dynamic behavior. Developed cities are often urban, highly mobile, and, at the same time, culturally and ethnically plural. While the housing policy we examine is unique, it is highly salient and frequently referenced as a blueprint for policymakers and international observers (Massey 2015; Fratzke 2017; Czischke and Huisman 2018; di Mauro 2018; Johnson 2019; Lim et al. 2019; Arroyo et al. 2021; Fischer 2021; Tan 2023).

More broadly, we build on a growing set of literature, which typically uses novel data to quantify experienced segregation along different socioeconomic dimensions (e.g., Athey et al. 2021; Davis et al. 2019; Dong et al. 2020; Moro et al. 2021), including segregation in the online space (Bastos et al. 2018; Cinelli et al. 2021; Dong et al. 2020; Eytan et al. 2015; Gentzkow and Shapiro 2011). Even with increasing residential integration (Glaeser

and Vigdor 2012), this literature speaks to the significance of social segregation beyond the confines of residence (Wang et al. 2018), crucial because of polarization and echo chambers (Levy and Razin 2019).[2]

The rest of the paper discusses (i) background and data, (ii) how utility from neighborhood visits depends on neighborhood mismatch, (iii) results, and (vii) the implied value of the ethnic integration policy in increasing visits. The final section concludes.

2 Background and Data

2.1 Geography

Singapore is a metropolitan city-state in Southeast Asia (Figure 1). It spans approximately 725km^2 or 280 square miles.[3] For planning and census taking, Singapore is divided into 5 regions, 55 planning areas, and over 300 subzones (Figure 1). We refer to subzones with census residential records as *neighborhoods* (Table 1). These neighborhoods are the units of analysis. Neighborhoods are small, averaging 1.35km^2 (0.52sq mi).[4] Few areas in Singapore are suburban. Public transit usage is high. According to the transport authority, train and bus trips average over 7 million daily for a population of over 6 million.[5]

2.2 GPS pings

The backbone of our data is the anonymized GPS ping records from CITYDATA.ai. The sample period covers 91 days in Jan–Mar 2020, before the city-wide COVID-19 lockdown on

[2] Our study also contributes to studies highlighting residential segregation and inequality in Singapore (e.g., Choe 2016; Leong et al. 2020; Loo et al. 2003; Sin 2002a; Tan 2023; Wong 2013; Teo 2018). Other studies focusing on segregation in cities include Jones and Pebley 2014; Krivo et al. 2013; Le Roux et al. 2017; Luo et al. 2016 and Wang and Li 2016.

[3] For comparison, about five times as large as San Francisco (121km^2), 1.2 times as large as Madrid city (667km^2), and 0.45 times as large as London city (1,570km^2, Lee et al. 2021).

[4] Comparable to the area of two geohash-6 grids (1.2km $\times$ 0.6km) stacked vertically, and with some neighborhoods as small as 0.05km^2 (0.02sq mi), which is comparable to two geohash-8 grids (38m $\times$ 19m).

In residential population terms, neighborhoods resemble US census tracts but are closer in land area to US census block groups, reflecting Singapore's high density and prevalence of high-rise flats (Lee et al. 2021).

[5] We account for transit-related amenities as part of spatial frictions (Section 3).

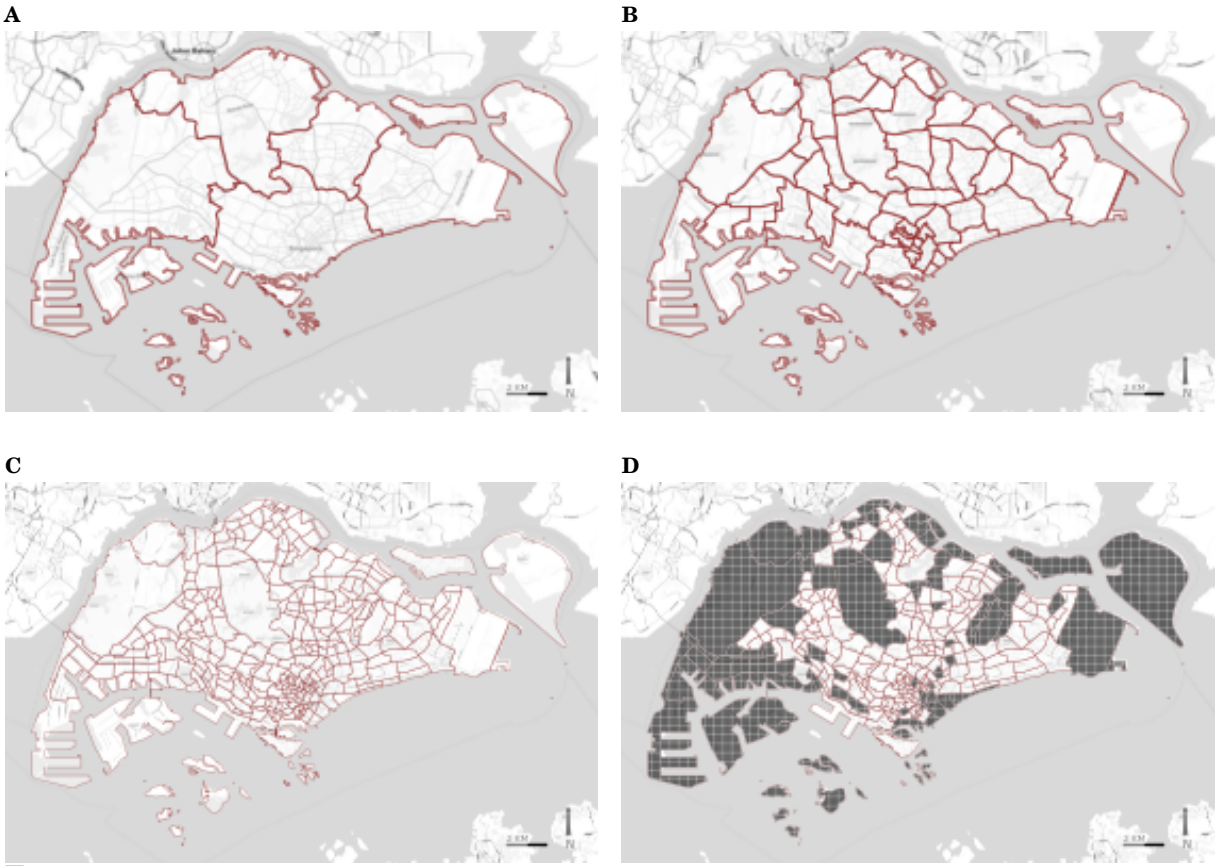

Figure 1. Geographical units. **(A)** Regions. **(B)** Planning areas. **(C)** Subzones. **(D)** Neighborhoods (subzones not shaded, with non-trivial density of residency).

April 7 (Lee et al. 2021). We observe mobile device hashes present in a neighborhood each day. We stitch these observations to construct a panel of daily neighborhood flows, assigning the modal neighborhood as the origin (Chen and Rohla 2018).

We perform a set of tests to check for how representative the GPS pings are for the neighborhoods (please see the supplementary appendix). GPS ping density increases with residential population (Lee et al. 2021), and neighborhood flows peak when destinations are geographically close to origins and on days with moderate precipitation. Overall, neighborhoods with denser populations capture more mobile devices (Figure 2).

Table 1. Descriptive statistics.

	Count/Mean ± s.d.	Min.	Max.
Device hashes	$> 17m$	.	.
Device hashes in sample (> 30 appearances)	$> 125k$	.	.
Census areas (coarser)	55	.	.
Subzones/neighborhoods (finer)	323	.	.
Census areas, with residential records	52	.	.
Subzones/neighborhoods, with residential records	219	.	.
Census areas, with GPS records	52	.	.
Subzones/neighborhoods, with GPS records	301	.	.
Subzone area (km^2)	2.23 ± 5.67	0.04	69.75
Neighborhood area (km^2) (subzones with residential records)	1.35 ± 1.24	0.05	8.45
Mismatch by wealth	0.3422 ± 0.4744	0	1
P $\longrightarrow$ W	0.1724 ± 0.3777	0	1
W $\longrightarrow$ P	0.1698 ± 0.3755	0	1
Mismatch by ethnicity	0.3685 ± 0.0686	0.0494	0.6955
Mnr $\longrightarrow$ Maj (ethnicity)	0.1878 ± 0.0685	0.0074	0.6880
Maj $\longrightarrow$ Mnr (ethnicity)	0.1807 ± 0.0744	0.0074	0.6880
$\lvert \text{wealth}_o - \text{wealth}_d \rvert$	0.0058 ± 0.0061	0	0.0319

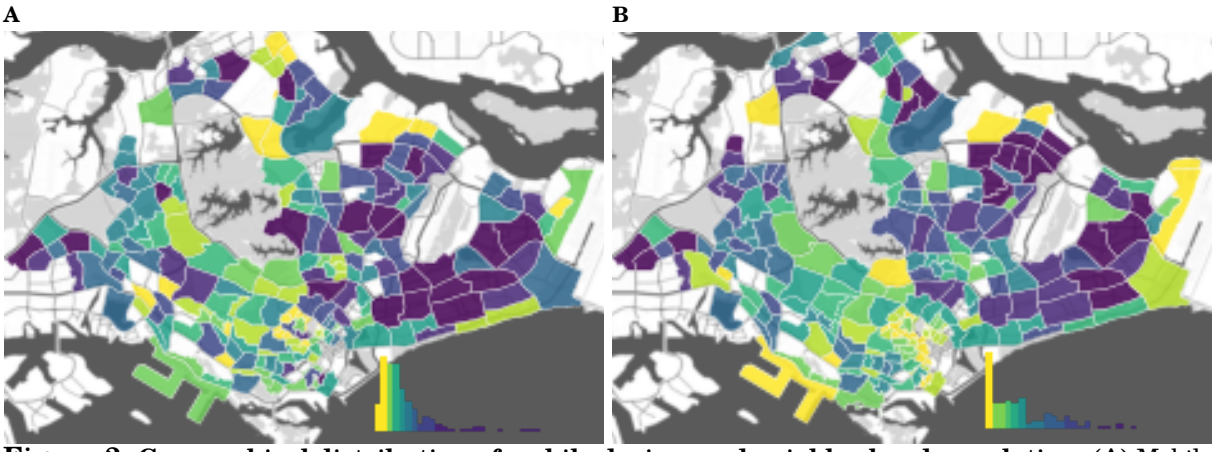

Figure 2. Geographical distribution of mobile devices and neighborhood population. **(A)** Mobile devices. **(B)** Neighborhood resident population.

2.3 Neighborhood income level

Our second key data source is the REALIS database's microtransaction records for housing resale prices. With census neighborhood income unavailable, we use these housing prices (per square meters) to proxy for neighborhood income.[6]

[6] Nighttime lights (e.g., VIIRS) offer high-resolution proxies but face issues like overglow. We find weighted housing prices provide a more direct and validated neighborhood income proxy (Xu et al. 2018).

We use the 133k microtransaction records in the past three years (2017–2019). We geocode street addresses to neighborhoods and compute neighborhood income using price per square meter, weighted by the proportion of private versus public housing residents (Section 2.4). We validate this measure by aggregating to the broader planning areas, where census income data is available. Figure 3 compares the geographical distribution of resale prices to census income across 28 planning areas. Higher housing prices correspond to higher income, with most points near the fitted line. Census income explains 79% of variation in our neighborhood income proxy ($\hat{r} = 0.89$).[7] [8]

2.4 Census demographics

We also link to census data for neighborhood ethnic composition and demographics. Ethnicity groups are Chinese (76%), Indian (7.5%), Malay (15%), and Others (1.5%). We group ethnicities into majority ethnic (Chinese) versus minority ethnic (Indian, Malay, Others), as is common (Leong et al. 2024). Other demographics include age, gender, and population size. Age groups are: below 20, 20–39, 40–64, and 65+.

2.5 Point of interests and businesses

We use data from the official data repository for places of interest (POIs)–including train stations, libraries, supermarkets, parks, schools, sports facilities, and tourist attractions. For businesses, we use corporate entity listings from the Accounting and Corporate Regulatory Authority, filtered to active entities. We focus on three industry divisions based on the Singapore Standard Industrial Classification, which make up the largest share of employment (Ho et al. 2022).

[7] This relationship is similar to Xu et al. (2018), who found a correlation coefficient of 0.89 when correlating housing price with self-reported income from the Household Interview Travel Survey at the same census planning area level. This survey data is not available to us.

[8] Since the $\hat{R}^2$ is itself a sample statistic, we bootstrap ($n = 100,000$) and recompute the $\hat{R}^2$. This yields a left-skewed distribution with a 95% confidence interval of [0.43, 0.91], equivalent to correlation coefficients of 0.66–0.95 (Figure S3).

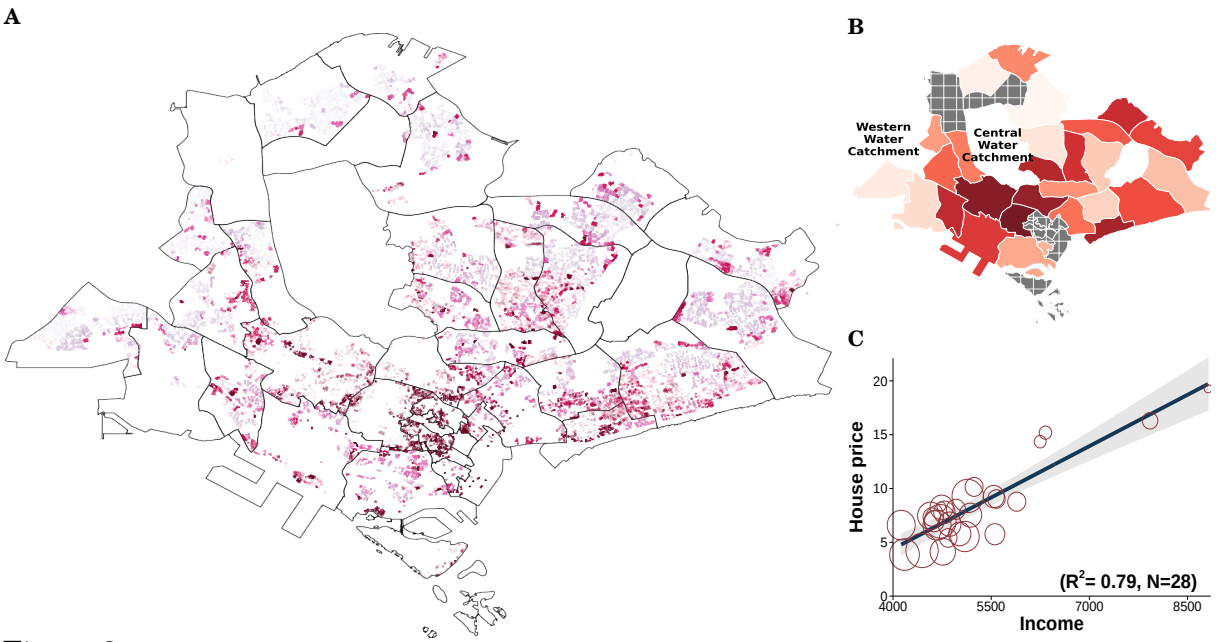

Figure 3. Proxying for neighborhood income levels. **(A)** Geographical distribution of housing prices per square meter (past five years). Each point is a recorded transaction, and darker shades indicate higher prices. **(B)** Geographical distribution of census income by the 28 census planning areas (with census income data). **(C)** House price and census income for the 28 areas. Marker size encodes resident population size.

2.6 Geographical and weather measures

We calculate neighborhood distances using the official shape files, computing contiguity and centroid-based distances. Daily rainfall (mm) records are sourced from the Meteorological Service Singapore. We spatially link rainfall data from 46 weather stations to the nearest neighborhood, applying linear interpolation for days where the measure is missing (Figure S4). Please see the supplementary appendix for all data details.

3 Empirical Approach

3.1 Mismatch and utilities from neighborhood visits

We sketch a simple model based on Kreindler and Miyauchi (2023) where utility arising from visiting neighborhood d from o is:

$$U_{od}(u, m, D) = \frac{u_d^{e_1} m_{od}^{e_2}}{D_{od}^{e_3}}. \tag{1}$$

Utility depends on the destination's intrinsic appeal (u), mismatch as social frictions (m), and spatial frictions (D). The destination's appeal includes wage opportunities and amenities, which can vary by day (Appendix B). For example, neighborhoods near business districts may yield higher utility and vary between weekdays and weekends. Social frictions stem from mismatches in income or ethnicity. Spatial frictions reflect geographical distance or travel inaccessibility (Davis et al. 2019; Kreindler and Miyauchi 2023). While the destination's appeal increases visit propensity, social and spatial frictions reduce it. Equation (1) implies the probability of visiting a neighborhood d depends on its utility relative to the utility from visiting all other neighborhoods (Appendix B). Our objective is to quantify how social frictions—mismatches in income or ethnicity— shape daily mobility through neighborhood visits.

3.2 Mismatch in income

We define mismatch as

$$\textit{mismatch}_{od} = z_o(1 - z_d) + z_d(1 - z_o) \tag{2}$$

where z is an indicator for a poor neighborhood (Chen and Rohla 2018). To operationalize the indicator, we define $z = 1$ when the neighborhood income proxy (Section 2.3) falls below the 25th percentile and $z = 0$ otherwise. We refer to these as poor and non-poor neighborhoods for ease of exposition.[9] Equation (2) is the exclusive or operator, assigning a value one only

[9] We test all income mismatch results for sensitivity to the 25th percentile cutoff below in Section 4.4.

when a neighborhood pair differs—one poor and the other is not—and value zero when both are the same—both are poor, or both non-poor.

3.3 Mismatch in ethnic composition

Equation (2) can also extend to continuous characteristics as an exclusive-or probability (Figure S5), as in Chen and Rohla (2018), which we do for majority and minority ethnic compositions (Section 2.4).[10] When applied to ethnic composition, z is the proportion of minority ethnic residents in a neighborhood, and *mismatch* becomes the imputed probability of an ethnic mismatch between representative individuals from origin and destination neighborhoods.[11]

3.4 Asymmetry in mismatch

To examine asymmetries in mismatch, we extend our analysis by decomposing the *mismatch* measure into the two directional components to have a distinct effect on visits (Chen and Rohla 2018). For example, $z_o(1 - z_d)$ captures mismatch arising from a poor to a non-poor destination neighborhood ($P \rightarrow NP$), while $z_d(1 - z_o)$ captures mismatch in the reverse direction ($NP \rightarrow P$).

With ethnic composition, $z_o(1-z_d)$ captures the imputed probability that a minority-ethnic person from an origin neighborhood encounters a majority-ethnic person in the destination neighborhood (Mnr $\rightarrow$ Maj). This probability is highest when the origin neighborhood has a high minority composition and when the destination neighborhood has a low minority composition. $z_d(1 - z_o)$ captures the reverse. The second component, $z_d(1 - z_o)$, captures the probability that a majority-ethnic person from the origin neighborhood encounters a minority-ethnic person in the destination neighborhood (Mnr $\rightarrow$ Maj). These two directional

[10] Chen and Rohla (2018) uses Equation (2) to quantify how mismatch in US precinct-level two-party vote shares impacts the duration of cross-party Thanksgiving dinners.

[11] The relaxation of the z to a continuous measure of proportion implies that the mismatch measure can be interpreted as the unconditional probability that, for two randomly drawn individuals i and j from neighborhoods o and d, either i is from a minority ethnic, or j is minority ethnic, but not both. The larger this measure is, the larger the probability of mismatch by ethnicity between the two neighborhoods (see Figure S5).

components in ethnic mismatch are central to our counterfactual policy experiment in Section 5.

3.5 Empirical model

To formalize how cross-neighborhoods flows depend on mismatch, we use our novel dataset, which merges GPS-derived neighborhood flows with census demographics, house microtransactions, and measures of amenities and spatial frictions (Section 2), to estimate

$$\log(\textit{visits})_{odt} = \alpha + \beta\, \textit{mismatch}_{od} + \Gamma_t X_{odt} + \varepsilon_{odt}, \tag{3}$$

where $\textit{visits}_{odt}$ is the probability of a representative resident from origin neighborhood o visiting neighborhood d on day t, derived from a model of utility from visits (Section 3.1). β is the coefficient of interest, which measures the effect of neighborhood mismatch on visits.

To account for spatial frictions (Appendix B), Equation (3) includes the origin-to-destination neighborhood distances, a contiguity dummy, and neighborhood sizes. These variables are fully interacted with the day fixed effects, allowing day-specific effects. We also include the census area-by-day fixed effects, subsuming day-of-week fixed effects. This embeds the assumption that utility from neighborhood visits are area- and day-specific. Neighborhood residential density is measured using the census and transient urban population using the density of observed GPS devices. This forms our baseline model.[12]

More demanding specifications of Equation (3) include (i) neighborhood census demographics, such as female proportion (Dong et al. 2017) and age groups, (ii) business composition (services, manufacturing, and construction) from addresses registered in the neighborhoods (Miyauchi et al. 2021; Ho et al. 2022), (iii) neighborhood- and day-specific rainfall (Figure S4), and (iv) POIs (places of interest and amenities such as transit stations, schools, tourist attractions). These variables fully interacted with the day fixed effects. Standard

[12] The area-by-day fixed effects subsumes the area-by-day utility derived from the neighborhood visits (Equation (S1)). While this controls for commuting-related visits, it might exclude some non-commuting visits. One study that disentangles commute from non-commute travels within the data pipeline is the one by Miyauchi et al. 2021. This is not possible with the resolution of the data we have in this study.

errors are clustered at the origin-by-destination planning area level.

4 Results

4.1 Income mismatch

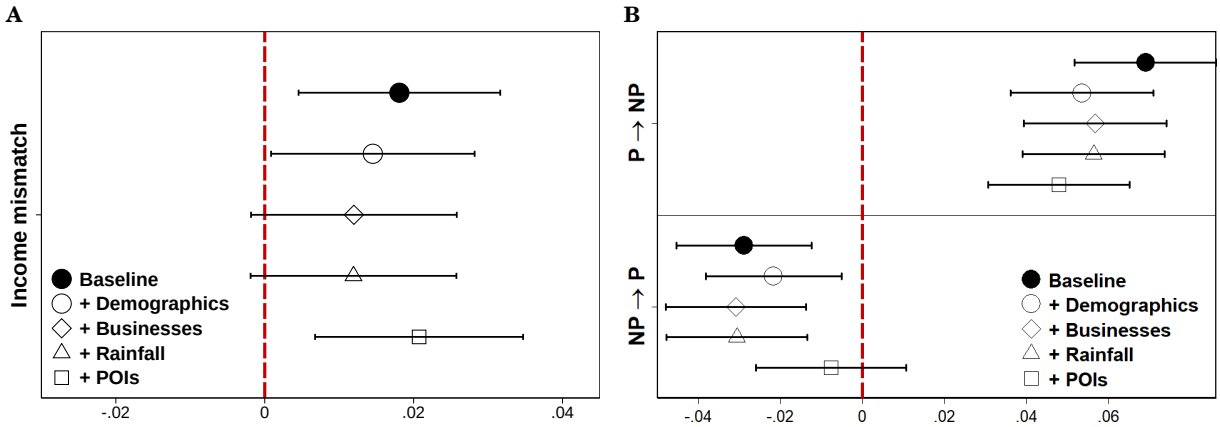

Figure 4. Income mismatch and neighborhood visits. (**A**) Estimated coefficients of mismatch from estimating Equation (3). (**B**) Asymmetry by direction of mismatch. The dependent variable is log visits from origin o to destination d. P $\rightarrow$ NP indicates movement from poor to non-poor neighborhoods ($z_o(1 - z_d)$ in Equation (2)). NP $\rightarrow$ P indicates the reverse ($z_d(1 - z_o)$ in Equation (2)). The baseline model includes neighborhood land area, population density (by both census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Figure 4A corresponds to Table S2. Figure 4B corresponds to Table S3. Capped horizontal lines are 90% CI from standard errors clustered at the origin-by-destination areas.

Figure 4A reports the estimates from estimating Equation (3), where income mismatch is associated with increased neighborhood visits on a daily basis. The solid black circle denotes the baseline specification, and the hollow markers are estimates with additional controls. The point estimate of 0.021 (SE 0.009) from the most demanding specification (hollow square) suggests that income mismatch increases visits by 2.1 percent ($\approx 100 \times .021$, as log points; $p < 0.05$).[13]

This increase in visits due to income mismatch may be asymmetrically driven by trips

[13] Table S7 of supplementary appendix uses absolute Euclidean distance of price per square meter as an alternative measure of income mismatch. The results are similar: a point estimate of 6.45 implies a standard deviation increase in Euclidean distance raises inflow by 3.9 percent ($\approx 100 \times 6.45 \times 0.006$; SE 0.57). However, this method does not allow the decomposition of mismatch.

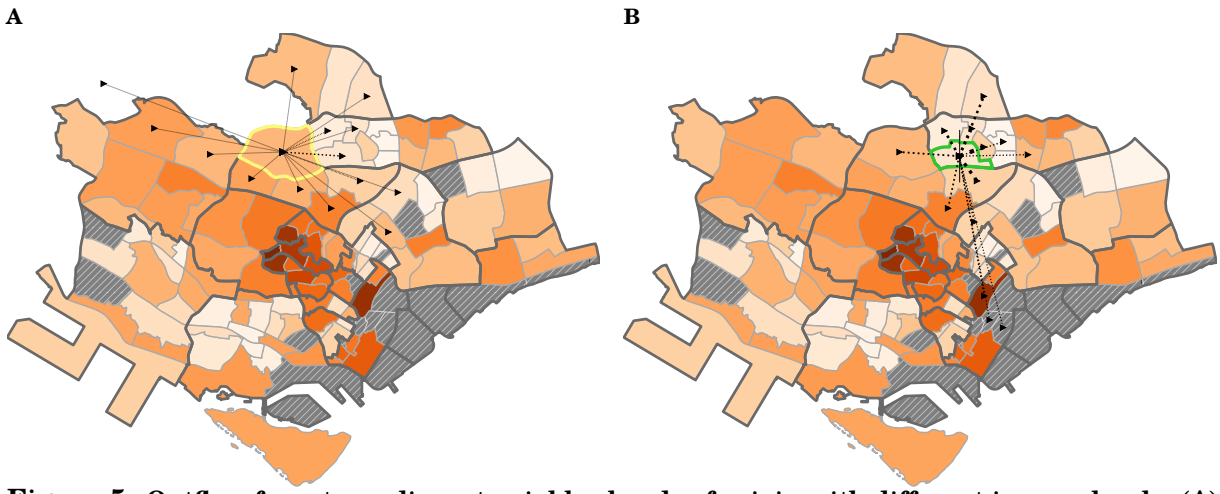

Figure 5. Outflow from two adjacent neighborhoods of origin with different income levels. **(A)** Richer neighborhood outlined in yellow (Mount Pleasant). **(B)** Poorer neighborhood outlined in green (Toa Payoh Central). Black lines show outflow based on GPS pings. The map is of the central region (Figure S1). Darker shades indicate richer areas. Shaded areas indicate non-residential areas.

from poorer to richer neighborhoods rather than the reverse. As illustrative examples, Figure 5 presents two adjacent neighborhoods (outlined in yellow and green) with differing income levels as points of origin. These neighborhoods exhibit different visit patterns. In particular, the farthest neighborhood visited from the richer neighborhood (yellow) is similar in income (left panel). Conversely, the farthest neighborhood visited from the poorer neighborhood (green) has much higher income (right panel).

To formally assess directional asymmetry, Figure 4B decomposes the income mismatch measure into its two directional components (Section 3.4). Specifically, we regress visits on (i) $z_o(1-z_d)$ for poor-to-non-poor ($P \rightarrow NP$) and (ii) $z_d(1-z_o)$ for non-poor-to-poor ($NP \rightarrow P$). The estimates reveal that the result in Figure 4A is driven by poor to non-poor visits. The estimate of 0.048 for $P \rightarrow NP$ implies a 5 percent higher probability of a visit from poor to richer neighborhoods (SE 0.011, p $<$ 0.01). Conversely, the estimate of -0.008 (SE 0.011) for $NP \rightarrow P$ is an order of magnitude lower and not significant at conventional levels.

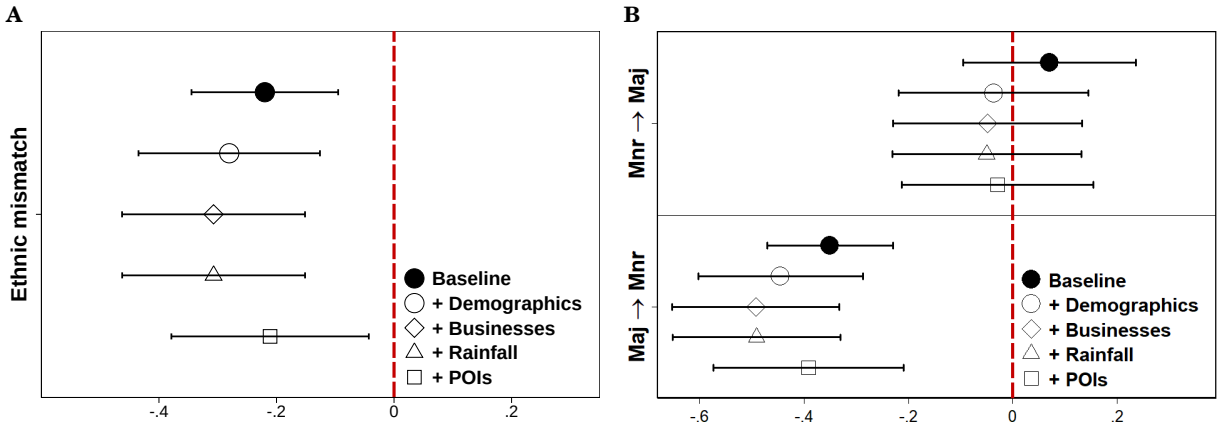

Figure 6. Mismatch in ethnicity and movement. (**A**) Coefficients are for the mismatch measure from Equation (2), with z_ℓ as the proportion of minority-ethnic residents in neighborhood ℓ. (**B**) Asymmetry by direction of mismatch. The dependent variable is log visits from origin o to destination d. Mnr $\rightarrow$ Maj indicates movement from minority-ethnic to majority-ethnic neighborhoods ($z_o(1 - z_d)$. Maj $\rightarrow$ Mnr indicates the reverse ($z_d(1 - z_o)$ in Equation (2)). The baseline model includes neighborhood land area, population density (by both census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Figure 7A corresponds to Table S5. Figure 7B corresponds to Table S6. Capped horizontal lines are 90% CI from standard errors clustered at the origin-by-destination areas.

4.2 Ethnic mismatch

We next examine ethnic mismatch and neighborhood visits. Figure 6A reports the coefficients, where we observe a negative effect. The estimate of -0.21 implies that a one standard deviation increase in ethnic mismatch decreases neighborhood visits by 1.5 percent ($\approx 100 \times 0.21 \times .07$, SE 0.102 and p < 0.05). Alternatively, the estimate implies that going from the median to the 100th percentile neighborhood pair in ethnic mismatch reduces neighborhood visits by 6.9 percent ($\approx 100 \times 0.21 \times [.37 - .7]$).

Figure 6B provides evidence of asymmetry in the mismatch in ethnicity effect. We once again decompose the mismatch measure into the two directional components: minority-ethnic to majority-ethnic neighborhoods (Mnr $\rightarrow$ Maj) and majority-ethnic to minority-ethnic neighborhoods (Maj $\rightarrow$ Mnr). The estimate of -0.39 implies neighborhood visits increase by 2.9 percent ($\approx 100 \times -0.391 \times -.074$) when going from a majority- to a minority-ethnic neighborhood. On the contrary, the estimate of -0.029 (SE 0.12) implies that the direction of minority-ethnic to majority-ethnic neighborhoods is muted. We use these estimates of

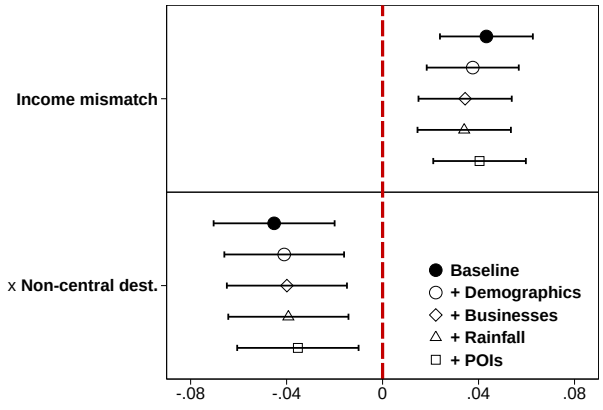

Figure 7. Asymmetry by geography with income mismatch. The top panel is mismatch, and the bottom panel is the mismatch measure interacted with an indicator for destination neighborhoods outside the central region. The baseline model includes neighborhood land area, population density (by both census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. The dependent variable is log inflow from origin o to destination d. The figure corresponds to Table S4. Capped horizontal lines are 90% CI from standard errors clustered at the origin-by-destination areas.

ethnic mismatch in Section 5 below to evaluate how an ethnic quota in public housing fosters neighborhood visits.

4.3 Income mismatch by geography

Another form of heterogeneity we examine is geography. Specifically, we consider whether the destination neighborhood is in the central region, which includes the central business district.[14] Figure 7 reports the results for income mismatch with an interaction term for destination neighborhoods outside the central region. The top panel of Figure 7 reports the baseline estimate of income mismatch, which captures the effect of mismatch on visits to central region neighborhoods, relative to other central neighborhoods without mismatch. The coefficient of 0.04 implies that income mismatch increases neighborhood visits by 4 percent when the destination is in the central region (SE 0.012 and $p < 0.01$).

For peripheral destination neighborhoods, the effect reverses (bottom panel of Figure 7). The estimate of -0.035 implies that the additional effect of income mismatch on peripheral neighborhood visits is approximately -3.5 percent (SE 0.015 and $p < 0.05$) than that for central neighborhood visits, bringing the net effect closer to zero. On net, income mismatch still confers some positive effect on central neighborhood visit (confirmed by an F-test

[14] The central region, including the central business district, features mixed-use neighborhoods, driven by urban planning goals and rental incentives. This means that areas in the central region, including residential ones, are more likely than peripheral neighborhoods to include commercial buildings, hotels, museums, and concert halls, among others, as amenities.

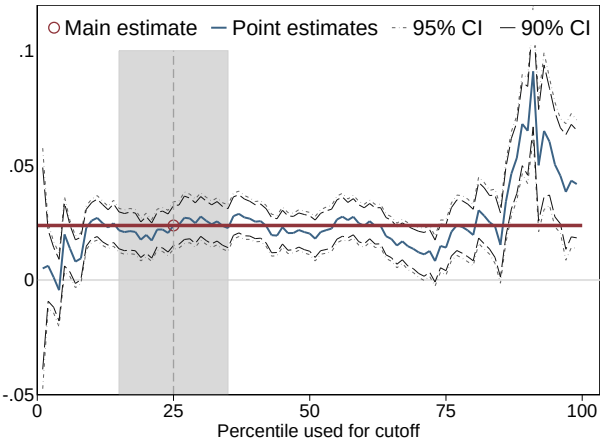

Figure 8. Sensitivity of income mismatch estimates at different thresholds. Each point along the curve is a model estimating Equation (3) with a different percentile that splits neighborhoods into poor versus non-poor. The specification is the one with the full controls. The horizontal red line indicates the main estimate in Figure 4A (25th percentile). Shaded gray areas indicate the 15th to the 35th percentiles (25th percentile $\pm$ 10). See Figure S9 for the same curve sorted by estimate size.

reported in Table S4 that fails to reject the null that the joint additive effect is zero).

Peripheral neighborhoods are as dense as central neighborhoods and are not suburban or rural. The above findings point to the importance of geography in shaping neighborhood flows. Central neighborhoods naturally promote cross-income exposure. In contrast, the negative effect in peripheral neighborhoods highlights the need for targeted policies, such as improved infrastructure and travel accessibility, to promote social exposure akin to those in central neighborhoods.

4.4 Robustness: Sensitivity to income cutoffs

The estimates on income mismatch in Section 4.1 are based on the 25th percentile. Here, we fully test the sensitivity of the estimates to this threshold by repeating the analyses and reporting the coefficients of interest using all possible thresholds.

Figure 8 reports the sensitivity of coefficients from Figure 4A across thresholds from the 1st–99th percentiles. The estimates remain stable near the 25th percentile (gray shaded area and dotted vertical line). This confirms that results are not artifacts of the 25th percentile cutoff. The confidence intervals indicate the estimates are fairly precise. The estimates and standard errors become erratic only at extreme thresholds.[15]

[15] Precision blows up mostly in the tail ends of the distribution, likely because sample imbalance in the poor vs. non-poor neighborhoods is exacerbated.

Figure 9 reports sensitivity for the directional mismatch estimates. Again, the estimates are insensitive to the choice of the threshold. The $NP \rightarrow P$ coefficients are positive and significant only sporadically at certain thresholds. Figure 9 suggests that asymmetry reverses only under an overwhelmingly inclusive definition of "poor." Figure 9A also shows that preferences for visiting more affluent neighborhoods (Davis et al. 2019; Büchel et al. 2020; Hilman et al. 2021) persist across the spectrum and are not unique to the 25th percentile cutoff.[16]

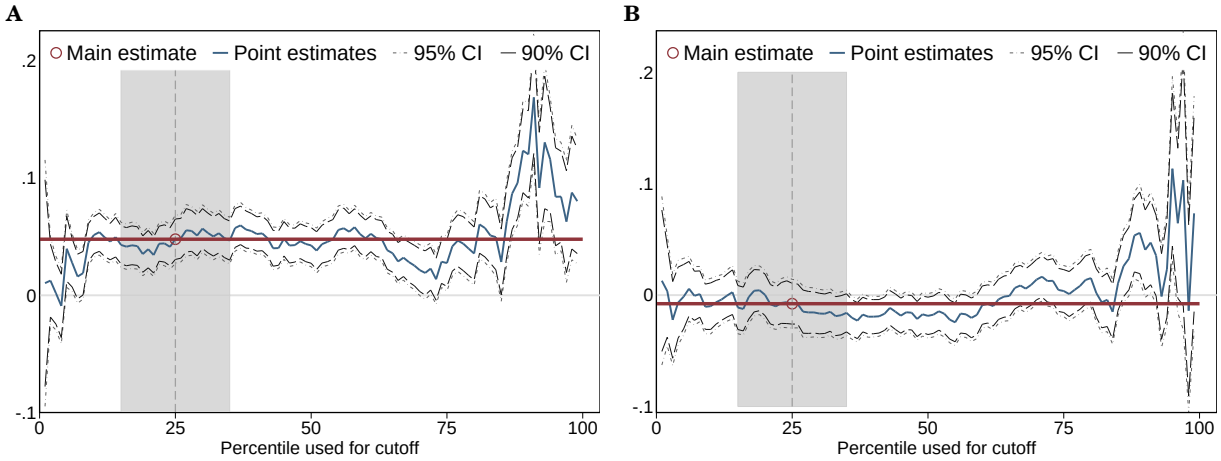

Figure 9. Sensitivity tests of directional mismatch in income. **(A)** Sensitivity of the $(P \rightarrow NP)$ coefficient from the top panel of Figure 4B. **(B)** Sensitivity of the $(NP \rightarrow P)$ coefficient from the bottom panel of Figure 4B. Each point along the curve is a model estimating Equation (3) with a different percentile that splits neighborhoods into poor versus non-poor. The specification is the one with the full controls. Shaded gray areas indicates the 15th to the 35th percentiles (25th percentile $\pm$ 10).66 See Figure S10 for the same curve sorted by estimate size.

In the supplementary appendix, we show that when estimates from Figure 8 and Figure 9 are ordered by magnitude, the estimate from the 25th percentile appears near the middle of the curve. This pattern confirms the threshold was not chosen to inflate effect sizes, which would place our estimates in the right tail.

[16] The supplementary appendix includes a similar sensitivity analysis for income mismatch by geography (Figure 7). The estimates are stable over a considerable range near the 25th percentile cutoff (Figure S6).

4.5 Robustness: Ethnic mismatch

We apply the constant elasticities specification for ethnic mismatch (Appendix B), yielding similar results (Figure S7). The estimate of -0.067 (SE 0.033) for the log-transformed ethnic mismatch implies that a 10 percent increase in ethnic mismatch decreases neighborhood visits by 0.7 percent ($p < 0.05$, Table S11).

We also conduct a form of placebo test, allowing ethnic mismatch to have different effects for central versus peripheral destinations (as with income mismatch in Figure 7). This tests whether ethnic mismatch effects are artificially driven by geography. Unlike income mismatch, ethnic composition should not inherently vary by geography since modern housing policies have dissolved traditional ethnic enclaves (Sin 2002a; Wong 2013, Section 5). The estimates confirm that the ethnic mismatch effect does not differ by geography (Table S12, Table S14). The differential effect of ethnic mismatch for non-central neighborhoods is effectively zero (estimate 0.0008, SE 0.18).

Lastly, including income mismatch and ethnic mismatch in the same model (Figure S2, Tan 2023) shows no substantial changes in estimates—they retain the same signs, magnitude, and statistical significance (Table S16).

5 Ethnic Housing Quotas

We build on our estimates and context of study to perform a counterfactual experiment centered on a key residential integration policy in Singapore–the Ethnic Integration Policy (EIP). The EIP is a cornerstone of Singapore's housing planning. While one of its kind, the EIP serves as a global model for addressing ethnic and immigrant integration in cities worldwide (Massey 2015; Fratzke 2017; Czischke and Huisman 2018; di Mauro 2018; Johnson 2019; Lim et al. 2019; Arroyo et al. 2021; Fischer 2021; Tan 2023).

The counterfactual we examine is the reduction in ethnic mismatch between two extant towns that were ethnic enclaves before the EIP. This counterfactual experiment serves two

purposes: i) to quantify the value of residential ethnic quotas in fostering cross-neighborhood flows and ii) to highlight the value of quantifying asymmetric segregation.[17]

5.1 Background

Before the establishment of the Housing and Development Board (HDB) in 1960, communal enclaves formed as immigrant ethnic groups settled and concentrated in different parts of the city under the previous colonial administration (Choe 2016). The Jackson Plan of the early 1800s segregated the city into ethnic subdivisions (Koh et al. 2006).

By 1989, HDB introduced the EIP to (i) prevent further formation and consolidation of ethnic enclaves and (ii) ensure a balanced ethnic mix (Chinese 76%, Indian 7.5%, Malay 15%, Others 1.5%) in public housing, which accommodates over 80% of the population (Sin 2002a; Leong et al. 2020; Lim et al. 2019; Government of Singapore 2020). The EIP enforces hierarchical ethnic quotas, limiting the percentage of residents from each ethnicity at block and at neighborhood level so that the neighborhood ethnic mix approaches the city-wide ethnic mix (Sin 2002a; Lim et al. 2019).[18] [19]

Table 2 reports pre- and post-EIP ethnic composition for two extant towns with strong legacies as ethnic enclaves, Bukit Merah and Bedok (in the central and east regions, Sin 2002a).[20] Table 2 shows predicted ethnic composition for 1988 and 2000, and actual values for 2000 and 2015. The predictions are based on observed trends in ethnic preferences for housing and movements in the resale market. Table 2 shows that, prior to the introduction of the EIP, both towns had ethnic compositions far from the city-wide average. The predicted 2000 ethnic composition suggests that the ethnic concentrations would have intensified: the

[17] Extant studies on the EIP typically focus on the implied distortions that come from the imposed quotas (Sin 2002b; Wong 2014; Lim et al. 2019).

[18] Once an ethnic limit is reached, no further flat sales to that ethnic group are allowed. The exception is when the buyer and seller are of the same ethnic group. For instance, in a block "over-occupied" by ethnic A, ethnic A owners may sell to ethnic A buyers, but non-ethnic A owners cannot sell to ethnic A buyers. While the quota is binding, some residences have been successful with appeals (Ng 2021; Iau 2023).

[19] Owing to the difference in the distribution of household income among different ethnic groups (Figure S2), the EIP also indirectly integrates residents of varying income groups, although to a smaller extent (see Tan 2023).

[20] Towns are larger administrative units, similar to census areas.

Table 2. Pre- and Post-EIP ethnic composition for selected towns.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ethnic composition				Ethnic mismatch		
1988 (Predicted)							
Town	Chinese	Indian/Others	Malays		Majority	Minority	Mismatch
Bukit Merah	88.4	6.1	5.5		88.4	11.6	0.35
Bedok	69.8	5.7	24.5		69.8	30.2	.
2000 (Predicted)							
Town	Chinese	Indian/Others	Malays		Majority	Minority	Mismatch
Bukit Merah	93.1	5.6	1.3		93.1	6.9	0.48
Bedok	52.0	5.0	43.0		52.0	48.0	.
2000 (Official statistics)							
Town	Chinese	Indian	Malays	Others	Majority	Minority	Mismatch
Bukit Merah	83.8	9.5	5.9	0.8	83.8	16.2	0.34
Bedok	73.1	7.2	17.4	2.4	73.1	26.9	.
2015 (Official statistics)							
Town	Chinese	Indian	Malays	Others	Majority	Minority	Mismatch
Bukit Merah	78.7	9.7	8.6	3.0	78.7	21.3	0.37
Bedok	72.1	8.7	15.2	4.1	72.1	27.9	.

EIP is the ethnic integration policy implemented since 1989. Predicted values of ethnic composition for 1988 and 2000, which comes directly from Sin 2002b who sourced it from Ooi 1993, are estimates of ethnic composition in the absence of the EIP. Mismatch is ethnic mismatch as defined in Equation (2) with z as the proportion of ethnic minorities (column 6). Systematic pre-EIP ethnic records at the neighborhood level is not publicly available.

proportion of Chinese residents in Bukit Merah would have risen to 93.1%, and Malays in Bedok to 43.0%.

Table 2 suggests that the EIP successfully moderated these disparities. By 2015, Bukit Merah's Chinese proportion fell to 78.7%, and Bedok's Malays to 15.2%, nearing city-wide averages. The mismatch values in Table 2 confirm that ethnic mismatch between Bukit Merah and Bedok is now lower than it would have been without the EIP intervention.

5.2 Counterfactuals

How might the EIP-driven reduction in ethnic mismatch affect neighborhood visits between these two towns? To evaluate the impact of the EIP on fostering cross-neighborhood flows, we compare the implied increase in visits based on the post-EIP ethnic mismatch versus its counterfactual at 2000. We evaluate two scenarios: without and with asymmetry in ethnic

mismatch.

Without accounting for asymmetries in ethnic mismatch, the difference in ethnic mismatch from predicted to actual 2000 ethnic composition implies a 2.9 percent ($\approx 100 \times -0.21 \times [.34 - .48]$) increase in visits after the EIP. For context, a one standard deviation decrease in ethnic mismatch today increases visits by 1.5 percent ($\approx 100 \times -0.21 \times .069$).

Accounting for asymmetries, the change in the majority-to-minority ethnic mismatch with and without the EIP implies that visits are 8.7 percent ($\approx 100 \times -0.391 \times [.269(1-.162) - .48(1-.069)]$) higher than it would have been without the EIP. A one standard deviation decrease in today's majority-to-minority ethnic mismatch implies only a 2.9 percent increase in visits ($\approx 100 \times -0.391 \times .074$).[21]

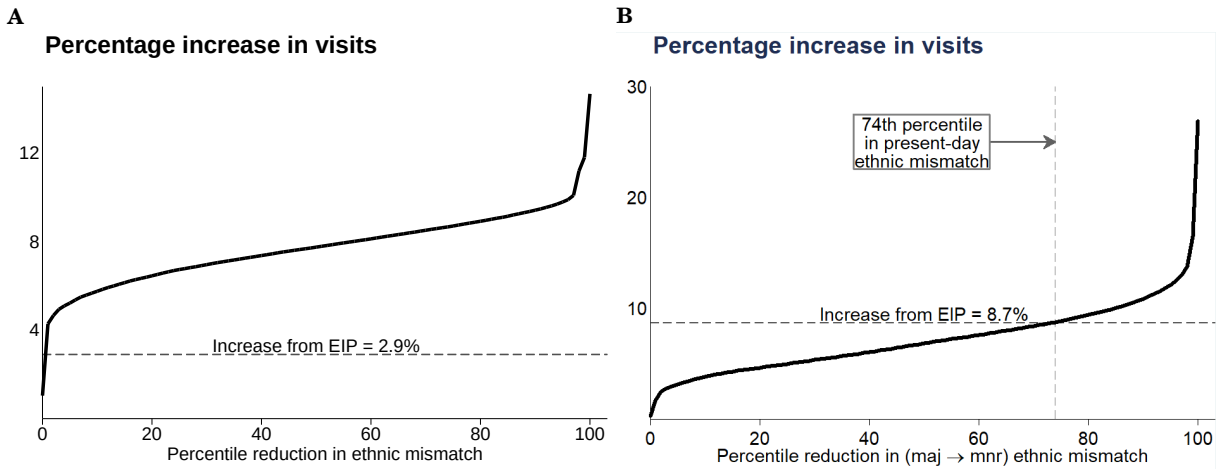

Figure 10. Reduction in ethnic mismatch and predicted increase in visits. **(A)** Without accounting for asymmetry: Reduction in ethnic mismatch and the percentage increase in visits. **(B)** Accounting for asymmetry: Reduction in (Maj → Mnr) ethnic mismatch and the increase in visits. The horizontal axes are ranked by the percentiles of the mismatch measure for the 75,390 neighborhood-pairs. Each point is an increase in visits if mismatch was reduced to zero: we take the mismatch *value* at the centile and compute ($100 \times -.21 \times$ *value*, Figure 6A) for Panel A and ($100 \times -.391 \times$ *value*, Figure 6B) for Panel B.

To systematically contextualize the above counterfactual estimates, we form all possible pairwise combinations of neighborhoods in our sample and rank them by their ethnic mismatch. For each pair, we compute the implied increase in neighborhood visits if we eliminate

[21] Our counterfactual analyses nominally require assumptions about similar distributions in amenities between the two towns for our sample period and in the past. However, since our focus is on the difference with and without accounting for asymmetry, such concerns loom less large.

their ethnic mismatch. We repeat this for majority-to-minority asymmetry. Figure 10 depicts the curves relating to the increase in neighborhood visits from total reduction in ethnic mismatch for the 75,390 neighborhood-pairs, ranked by the percentile in ethnic mismatch (horizontal axis). Each point on the curve corresponds to the increase in visits (vertical axis) from a total elimination of ethnic mismatch between the neighborhood pair. For instance, the 50th percentile value for ethnic mismatch is .37, and eliminating that mismatch—going from a value of .37 to 0—implies an increase in visits of 7.7 percent ($\approx 100 \times -0.21 \times .37$).

Without accounting for asymmetries, the 2.9% increase in neighborhood visits between the two towns after the EIP is equivalent to no more than that from eliminating present-day ethnic mismatch of the 1st percentile neighborhood pair (Figure 10A). However, a very different picture emerges when asymmetries are taken into account. To achieve the 8.7% increase in neighborhood visits implied by the EIP, we would need to eliminate present-day majority-to-minority mismatch from the 74th percentile neighborhood-pair (Figure 10B).

Overall, this counterfactual experiment underscores the critical role of asymmetries–the value of the EIP in fostering cross-neighborhood flows is three times greater than it would appear without accounting for how asymmetries influence visits.[22]

6 Conclusion

This study highlights how socioeconomic mismatch and behavior asymmetrically shape daily neighborhood visits. By integrating GPS-based cross-neighborhood flows with housing microtransactions, demographic records, and measures of spatial frictions and neighborhood amenities, we show that social exposure is nuanced and reinforced by directional asymmetries. Social exposure can have profound consequences along social and economic dimensions (Cutler and Glaeser 1997; Cutler et al. 2008; Ananat 2011; Cook et al. 2018), and asymmetries—more visits from majority- to minority-ethnic neighborhoods, but not

[22] Comparing the 2000 predicted values to the 2015 actual values shows an even larger difference in efficacy measure ($\sim$factor of 4).

the reverse—highlight nuances in how socioeconomic disparities shape exposure in urban societies.

Integration of diverse populations across ethnicities and immigrants remains a staple global concern (Massey 2015; Fratzke 2017; Czischke and Huisman 2018; di Mauro 2018; Johnson 2019; Arroyo et al. 2021; Fischer 2021). We build on estimates of asymmetry in ethnic mismatch for a counterfactual case study of a binding residential ethnic quota. Our counterfactual experiment reveals that the ethnic quota's impact on neighborhood visits, and therefore reduced experienced segregation, would have been severely understated had we not accounted for asymmetries.

GPS data offers studies of segregation beyond residence but has limitations. Notably, we cannot pinpoint individuals' specific reasons for visiting neighborhoods; we only observe presence. Despite this, trip hops within commute activities still potentially constitute meaningful interactions, as individuals share physical spaces during their daily activities (Le Roux et al. 2017; Cagney et al. 2020; Athey et al. 2021).

References

Ananat, Elizabeth Oltmans. 2011. "The wrong side(s) Of the tracks: The causal effects of racial segregation on urban poverty and inequality." *American Economic Journal: Applied Economics* 3 (2): 34–66, http://www.jstor.org/stable/41288628. 23

Arroyo, Ivette, Norma Montesino, Erik Johansson, and Yahia Moohammed Wasim. 2021. "Social integration through social connection in everyday life. Residents' experiences during the COVID-19 pandemic in SällBo collaborative housing, Sweden." *Archnet-IJAR: International Journal of Architectural Research* 15 (1): 79–97. 10.1108/ARCH-10-2020-0236. 4, 19, 24

Athey, Susan, Billy Ferguson, Matthew Gentzkow, and Tobias Schmidt. 2021. "Estimating experienced racial segregation in US cities using large-scale GPS data." *Proceedings of the National Academy of Sciences* 118 (46): e2026160118. 10.1073/pnas.2026160118. 1, 4, 24

Bastos, Marco, Dan Mercea, and Andrea Baronchelli. 2018. "The geographic embedding of online echo chambers: Evidence from the Brexit campaign." *PLoS ONE* 13 (11): 1–16. 10.1371/journal.pone.0206841. 4

Blanchard, Paul, Douglas Gollin, and Martina Kirchberger. 2023. "Perpetual Motion: High-Frequency Human Mobility in Three African Countries." April, https://ideas.repec.org//p/tcd/tcduee/tep0823.html. 36

Büchel, Konstantin, Maximilian V Ehrlich, Diego Puga, and Elisabet Viladecans-Marsal. 2020. "Calling from the outside: The role of networks in residential mobility." *Journal of Urban Economics* 119 103277. https://doi.org/10.1016/j.jue.2020.103277. 3, 18

Cagney, Kathleen A., Erin York Cornwell, Alyssa W. Goldman, and Liang Cai. 2020. "Urban Mobility and Activity Space." *Annual Review of Sociology* 46 (1): 623–648. 10.1146/annurev-soc-121919-054848. 24

Chen, M Keith, and Ryne Rohla. 2018. "The effect of partisanship and political advertising on close family ties." *Science* 360 (6392): 1020–1024. https://doi.org/10.1126/science.aaq1433. 6, 10, 11

Choe, Alan F. C. 2016. *The early years of nation-building: Reflections on Singapore's urban history*. Chap. CHAPTER 1 3–21, World Scientific, . 10.1142/9789814656474_0001. 1, 5, 20

Cinelli, Matteo, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. "The echo chamber effect on social media." *Proceedings of the National Academy of Sciences* 118 (9): e2023301118. 10.1073/pnas.2023301118. 4

Cook, Lisa D, Trevon D Logan, and John M Parman. 2018. "Racial segregation and southern lynching." *Social Science History* 42 (4): 635–675. https://doi.org/10.1017/ssh.2018.21. 1, 23

Cutler, David M, and Edward L Glaeser. 1997. "Are ghettos good or bad?" *The Quarterly Journal of Economics* 112 (3): 827–872. https://doi.org/10.1162/003355397555361. 23

Cutler, David M, Edward L Glaeser, and Jacob L Vigdor. 2008. "When are ghettos bad? Lessons from immigrant segregation in the United States." *Journal of Urban Economics* 63 (3): 759–774. https://doi.org/10.1016/j.jue.2007.08.003. 23

Czischke, Darinka, and Carla J Huisman. 2018. "Integration through Collaborative Housing? Dutch Starters and Refugees Forming Self-Managing Communities in Amsterdam." *Urban Planning* 3 (4): 156–165, https://ideas.repec.org/a/cog/urbpla/v3y2018i4p156-165.html. 4, 19, 24

Davis, Donald R, Jonathan I Dingel, Joan Monras, and Eduardo Morales. 2019. "How segregated is urban consumption?" *Journal of Political Economy* 127 (4): 1684–1738. https://doi.org/10.1086/701680. 1, 3, 4, 10, 18, 37

Dong, Xiaowen, Alfredo J Morales, Eaman Jahani et al. 2020. "Segregated interactions in urban and online space." *EPJ Data Science* 9 (1): 20. 10.1140/epjds/s13688-020-00238-7. 1, 4

Dong, Xiaowen, Yoshihiko Suhara, Burçin Bozkaya, Vivek K. Singh, Bruno Lepri, and Alex Pentland. 2017. "Social bridges in urban purchase behavior." *ACM Transactions on Intelligent Systems and Technology* 9 (3): . 10.1145/3149409. 12

Eytan, Bakshy, Messing Solomon, and Adamic Lada A. 2015. "Exposure to ideologically diverse news and opinion on Facebook." *Science* 348 (6239): 1130–1132. 10.1126/science.aaa1160. 4

Fischer, Robert. 2021. "Singapore Housing Lessons for the Biden Administration." https://www.planetizen.com/blogs/112077-singapore-housing-lessons-biden-administration. 4, 19, 24

Fratzke, Susan. 2017. "Weathering Crisis, Forging Ahead: Swedish Asylum and Integration Policy." https://www.migrationpolicy.org/sites/default/files/publications/TCM-Asylum-Sweden-FINAL.pdf. 4, 19, 24

Gentzkow, Matthew, and Jesse M Shapiro. 2011. "Ideological segregation online and offline." *The Quarterly Journal of Economics* 126 (4): 1799–1839. 10.1093/qje/qjr044. 4

Glaeser, Edward, and Jacob Vigdor. 2012. "The end of the segregated century: Racial separation in America's neighborhoods, 1890-2012." *Civic Report* 66 1–36, https://media4.manhattan-institute.org/pdf/cr_66.pdf. 1, 4

Government of Singapore. 2020. "HDB's Ethnic Integration Policy: Why it still matters." https://www.gov.sg/article/hdbs-ethnic-integration-policy-why-it-still-matters. 20

Hilman, Rafiazka Millanida, Gerardo Iñiguez, and Márton Karsai. 2021. "Socioeconomic biases

in urban mixing patterns of US metropolitan areas." https://doi.org/10.48550/arXiv.2110.04183. 1, 4, 18

Ho, Terence, Jing Zhi Lim, and Lucas Shen. 2022. "Surveying the Singapore Urban Labour Market Using Online Job Postings." https://lkyspp.nus.edu.sg/docs/default-source/aci/acirp202215.pdf. 8, 12

Hutchens, Robert. 2001. "Numerical measures of segregation: desirable properties and their implications." *Mathematical social sciences* 42 (1): 13–29. https://doi.org/10.1016/S0165-4896(00)00070-6. 1

Iau, Jean. 2023. "One-third of ethnic integration housing policy appeals succeeded in 2022, up from 21https://www.channelnewsasia.com/singapore/hdb-ethnic-integration-housing-policy-successful-increase-2022-shanmugam-ips-3555721. 20

Johnson, Olatunde C. 2019. "Unjust Cities? Gentrification, Integration, and the Fair Housing Act." February, https://papers.ssrn.com/abstract=3629763. 4, 19, 24

Jones, Malia, and Anne R Pebley. 2014. "Redefining neighborhoods using common destinations: social characteristics of activity spaces and home census tracts compared." *Demography* 51 (3): 727–752. 10.1007/s13524-014-0283-z. 1, 5

Koh, Tommy, Timothy Auger, Jimmy Yap, and Ng Wei Chian. 2006. *Singapore: The Encyclopedia*. Singapore: Editions Didier Millet. 20

Kreindler, Gabriel E., and Yuhei Miyauchi. 2023. "Measuring Commuting and Economic Activity inside Cities with Cell Phone Records." *The Review of Economics and Statistics* 105 (4): 899–909. 10.1162/rest_a_01085. 10, 36

Krivo, Lauren J, Heather M Washington, Ruth D Peterson, Christopher R Browning, Catherine A Calder, and Mei-Po Kwan. 2013. "Social isolation of disadvantage and advantage: The reproduction of inequality in urban space." *Social Forces* 92 (1): 141–164. 10.1093/sf/sot043. 5

Le Roux, Guillaume, Julie Vallée, and Hadrien Commenges. 2017. "Social segregation around the clock in the Paris region (France)." *Journal of Transport Geography* 59 134–145. https://doi.org/10.1016/j.jtrangeo.2017.02.003. 5, 24

Lee, Shu En, Jing Zhi Lim, and Lucas Shen. 2021. "Segregation Across Neighborhoods in a Small City." *Asia Competitiveness Institute Research Paper Series* Research Paper #07-2021, https://lkyspp.nus.edu.sg/docs/default-source/aci/acirp202107.pdf?sfvrsn=a862240a_2. 5, 6

Leong, Chan-Hoong, Angelica Ting Yi Ang, and Siok Kuan Tambyah. 2024. "Using spatial big data to analyse neighbourhood effects on immigrant inclusion and well-being." *International Journal of Intercultural Relations* 102 102020. https://doi.org/10.1016/j.ijintrel.2024.102020. 1, 8

Leong, Chan-Hoong, Eugene Teng, and William Weiliang Ko. 2020. "The state of ethnic congregation in Singapore today." In *Building Resilient Neighbourhoods in Singapore*, edited by Leong, Chan-Hoong, and Lai-Choo Malone-Lee 29–50, Springer Nature Singapore Pte Ltd, . 10.1007/978-981-13-7048-9. 1, 5, 20

Levy, Gilat, and Ronny Razin. 2019. "Echo chambers and their effects on economic and political outcomes." *Annual Review of Economics* 11 (1): 303–328. 10.1146/annurev-economics-080218-030343. 5

Lim, Travis, Chan-Hoong Leong, and Farzaana Suliman. 2019. "Managing Singapore's residential diversity through Ethnic Integration Policy." *Equality, Diversity and Inclusion: An International Journal* 39 (2): 109–125. 10.1108/EDI-05-2019-0168. 4, 19, 20

Loo, Lee Sim, Ming Yu Shi, and Sheng Han Sun. 2003. "Public housing and ethnic integration in Singapore." *Habitat International* 27 (2): 293–307. https://doi.org/10.1016/S0197-3975(02)00050-4. 1, 5

Luo, Feixiong, Guofeng Cao, Kevin Mulligan, and Xiang Li. 2016. "Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago." *Applied Geography* 70 11–25. https://doi.org/10.1016/j.apgeog.2016.03.001. 5

Massey, Douglas S. 2015. "The Legacy of the 1968 Fair Housing Act." *Sociol. Forum* 30 (Suppl 1): 571–588. 10.1111/socf.12178. 1, 4, 19, 24

di Mauro, Beatrice. 2018. "Building a Cohesive Society: The Case of Singapore's Housing Policies." 4, 19, 24

Miyauchi, Yuhei, Kentaro Nakajima, and Stephen J Redding. 2021. "Consumption access and the spatial concentration of economic activity: Evidence from smartphone data." *National Bureau of Economic Research Working Paper Series* No. 28497. 10.3386/w28497. 12, 36, 37

Moro, Esteban, Dan Calacci, Xiaowen Dong, and Alex Pentland. 2021. "Mobility patterns are associated with experienced income segregation in large US cities." *Nature Communications* 12 (1): . 10.1038/s41467-021-24899-8. 1, 4

Ng, Michelle. 2021. "Successful ethnic quota appeals for HDB flats up to 21https://www.straitstimes.com/singapore/politics/successful-ethnic-quota-appeals-for-hdb-flats-up-to-21-last-year-from-14-in-2018. 20

Ooi, G L. 1993. "The Housing and Development Board's ethnic integration policy." In *The Management of Ethnic Relations in Public Housing Estates*, edited by Ooi, G L, S Siddique, and K C Soh 4–24, Times Academic Press. 21

Rodriguez-Moral, Antonio, and Marc Vorsatz. 2016. *An overview of the measurement of segregation: Classical approaches and social network analysis*. 93–119, Cham: Springer International Publishing, . 10.1007/978-3-319-40803-3_5. 1

Sin, Chih Hoong. 2002a. "Segregation and marginalisation within public housing: The disadvantaged in Bedok New Town, Singapore." *Housing Studies* 17 (2): 267–288. https://doi.org/10.1080/02673030220123225. 1, 5, 19, 20

Sin, Chih Hoong. 2002b. "The Quest for a Balanced Ethnic Mix: Singapore's Ethnic Quota Policy Examined." *Urban Studies* 39 1347–1374. 10.1080/00420980220142673. 20, 21

Tan, Shin Bin. 2023. "Do ethnic integration policies also improve socio-economic integration? A study of residential segregation in Singapore." *Urban Stud.* 60 (4): 696–717. 10.1177/00420980221117918. 1, 4, 5, 19, 20

Teo, You Yenn. 2018. *This is what inequality looks like*. Singapore: Ethos Books. 5

Wang, Donggen, and Fei Li. 2016. "Daily activity space and exposure: A comparative study of Hong Kong's public and private housing residents' segregation in daily life." *Cities* 59 148–155. https://doi.org/10.1016/j.cities.2015.09.010. 5

Wang, Qi, Nolan Edward Phillips, Mario L Small, and Robert J Sampson. 2018. "Urban mobility and neighborhood isolation in America's 50 largest cities." *Proceedings of the National Academy of Sciences* 115 (30): 7735 LP – 7740. 10.1073/pnas.1802537115. 1, 5

Wong, Maisy. 2013. "Estimating ethnic preferences using ethnic housing quotas in Singapore." *The Review of Economic Studies* 80 (3): 1178–1214. 10.1093/restud/rdt002. 3, 5, 19

Wong, Maisy. 2014. "Estimating the distortionary effects of ethnic quotas in Singapore using housing transactions." *Journal of Public Economics* 115 131–145. https://doi.org/10.1016/j.jpubeco.2014.04.006. 1, 20

Xu, Yang, Alexander Belyi, Iva Bojic, and Carlo Ratti. 2018. "Human mobility and socioeconomic status: Analysis of Singapore and Boston." *Computers, Environment and Urban Systems* 72 51–67. 10.1016/j.compenvurbsys.2018.04.001. 7, 8

Supplementary Information

Table of Contents

A. Data Appendix

The primary source of data is CITYDATA.ai, which aggregates anonymized GPS (global positioning system) ping records as a third party via SDKs (Software Development Kits) within applications installed on mobile phones. All ping records are anonymized and include a device hash ID and a day record, where the underlying location signals have horizontal accuracy of up to 25m. See Table S1 for mobile phone GPS ping data sources. We receive these records as flat files of `area-day` for the presence of devices in an area on a given day for the 91 days in Jan–Mar 2020. This period is before the two-month-long city-wide lockdown on 7th April (announced three days prior). The areas are assigned according to the URA (Urban Redevelopment Authority) 2014 Master Plan map, and devices are recorded. We combine these GPS ping records to the official census records (where available) and to public and private housing transaction records as a proxy for neighborhood income levels.

1. To build the gravity panel by `origin-destination-date`, we start by using the device list records from CITYDATA.ai which are stored as flat files. We have a recorded list of captured devices in that subzone using the device hash for each census subzone and each day. To get cross-area movement flows, we treat each device hash as an individual and aggregate the CITYDATA.ai records up to the `origin-destination-date` level. This gives for any `origin-destination-date` the count of individuals going from the `origin` area to the `destination` area on a given `date`. We then divide this inflow count by the number of devices for that `origin-date` to get the origin-to-destination flow measure.
2. To infer the "origin" area of an individual, we aggregate the CITYDATA.ai records up to the `device-area`, and make the simple assumption that the area with the highest appearance count for a given device hash is the origin area. Devices that appear < 30 times in the 91-day sample period are dropped. Certain devices have ties in rank, and we treat them as different individuals. This yields records from approximately 125k devices.

As a form of validation exercise, and to test the extent to how representative the GPS ping record and the constructed neighborhood flows data is of the population, we correlate the GPS data to neighborhood demographics in the supplementary appendix. We find that captured GPS pings increase with resident population size and do not vary too much by neighborhood-level characteristics such as age group, gender, ethnic composition, and house type. We also find that geographical distance and rainfall affect neighborhood visits in an expected manner. More neighborhood visits are observed when the origin-destination pair in neighborhoods are adjacent and less when not. This observation extends to the centroid and edge-based distances between neighborhoods. More neighborhood visits are observed with moderate precipitation, with the dryest and wettest periods having the least amount of neighborhood visits. For the 91-day sample period, we observe progressively fewer GPS pings in the weeks leading up to the lockdown. Captured GPS pings in origin neighborhoods are decreasing with housing prices, and this likely captures the fact that more expensive houses are private properties in less densely populated areas.

3. Census income data at the geographically smaller subzone level is unavailable to the public. To measure wealth/poverty at geographical units (census subzones) smaller than the available census income records (census/planning area), we use the public HDB housing micro-transaction records from the official and public repository (https://data.gov.sg/dataset/resale-flat-prices) and the proprietary private housing micro-transaction records are from REALIS maintained by the Urban Redevelopment Authority (URA).

   Transaction prices for public housing at the unit level are not available, and only the aggregated story range is provided to identify the unit. Further, public housing transaction prices only include street addresses as geographical information. To crosswalk from the public transaction prices to subzones, we first query the OneMap database for postcodes using the street address. We then use the postcodes to obtain latitude and longitude coordinates from which we can perform point-in-polygon analyses with the shape files to assign transaction prices to subzones. If a postal code maps to multiple coordinates, we will take the mean of the coordinates to get a representative coordinate for the postal code. Finally, we do a simple point-in-polygon query to see if the coordinate falls in a subzone to crosswalk from addresses to subzones.

   Transaction prices for private housing are available at the unit level, and postcodes are directly available in the REALIS data set. We crosswalk from the private transaction prices to the subzone using postcodes, which are available in REALIS, in the same way above. REALIS also contains land transaction data, which we omit.

4. We retain all transaction records in the past three years (2017, 2018, 2019) before our sample year 2020, this yields about 133k public and private transactions. We then aggregate the housing price (per square meter) separately for public and private transactions up to the census subzone level.

   For each subzone, we compute neighborhood wealth as the average housing price (per square

meters) weighted by the share of public residence and private residence:

$$\text{share}_i^{\text{public}} \text{PSM}_i^{\text{public}} + \text{share}_i^{\text{private}} \text{PSM}_i^{\text{private}}$$

where share is the proportion (between 0 and 1) of residents in neighborhood i who live in the public HDB residences or private residences; PSM is the housing price per square meter. To validate the use of the weighted housing price per square meters measure, we aggregate the measure from the subzone level up to the census planning area level and then correlate the aggregated wealth proxy with the census income data available at 28 planning areas, which yields a correlation coefficient of approximately 0.89. Figure 3 shows the geographical distribution of the transactions and the correlation between housing prices and census income at the census planning area level.

5. As an alternative measure of wealth/poverty at the smaller subzones using the census records, we use the proportion of residents living in 1–2-room public housing flats—this is the count (rounded off to tens in the official records) of residents living in 1–2-room HDB (Housing & Development Board) flats divided by the total number of residents in the subzone. Records for "HUDC Flats (excluding those privatized)" and "Others" are excluded from both the numerator and denominator. All residence type records are for the year 2019, available at https://storage.data.gov.sg/singapore-residents-by-subzone-and-type-of-dwelling-jun-2018/resources/planning-area-subzone-age-group-sex-and-type-of-dwelling-june-2011-2019-2020-03-06T03-39-39Z.csv. The results using residence type as a proxy for poverty and wealth are available in the supplementary materials.

6. Census variables for basic demographics at the subzone level are derived in the same manner. Demographic data available at the smaller subzone level are limited to age, gender, ethnicity, and population size. Age and gender records are from https://storage.data.gov.sg/resident-population-by-planning-area-subzone-age-group-and-sex-2015/resources/resident-population-by-planning-area-age-group-and-sex-2019-07-30T03-02-18Z.csv. Ethnicity records are from https://storage.data.gov.sg/resident-population-by-planning-area-subzone-ethnic-group-and-sex-2015/resources/resident-population-by-planning-area-ethnic-group-and-sex-2019-08-01T03-23-57Z.csv. In our sample, we group age demographics at the smaller subzone level as: (i) below 20 years old, (ii) 20–39 years old, (iii) 40–64 years old, and (iv) 65 and above. For ethnicity demographics, four main ethnicities are recorded (Chinese, Indians, Malay, and Others with the approximate composition of 76%, 7.5%, 15%, 1.5%), where the first group is the majority ethnicity. In our sample and results, we group the neighborhood ethnicity into a single non-majority ethnic variable as the proportion of residents in a subzone that is from the non-majority ethnic to compute mismatch by ethnicity.

7. To measure poverty/wealth at the larger census planning area unit, we use the census records of both residential house types and (gross) income from work for resident working persons aged 15 years and above. This is available from https://storage.data.gov.sg/resident-working-persons-aged-15-years-over-by-planning-area-gross-monthly-income-from-work-2015/

resources/resident-working-persons-aged-15-yrs-over-by-pa-gross-monthly-income-from-work-2019-08-08T04-36-54Z.csv. Income is a per-population weighted-average: for each income, bin reported we take the midpoint and multiply it by the number of residents in that income bin, then we aggregate up to the area level and divide by the total number of working residents in that area.

8. We use the 2014 masterplan data from https://data.gov.sg/dataset/master-plan-2014-subzone-boundary-web to spatially compute centroid-to-centroid distances. The geographic data is stored in a projected coordinate system - SVY21, allowing a more accurate representation of the Singapore area. For each subzone, we calculate the coordinates of its centroid. We then take the distance between all pairs of centroids to obtain the inter-subzone distance, `subzone-centroid-distance`. Additionally, we obtain the edge-to-edge distance between all pairs of subzones. If the edge-to-edge distance between two subzones equals zero, we consider them contiguous.

9. For the area-specific POIs (places of interest), we mostly default to the official records found in https://data.gov.sg. We do a point-in-polygon query for POI records stored in shapefiles or their equivalent to match POIs to areas. The number of POIs in a subzone includes libraries, supermarkets, parks, preschools, schools (primary and secondary), silver zones, sports facilities, train stations, and tourist attractions.

   For businesses, the listing of corporate entities from the Accounting and Corporate Regulatory Authority comes from https://storage.data.gov.sg/acra-information-on-corporate-entities/resources/acra-information-on-corporate-entities-a-2021-01-15T00-37-18Z.csv. From the entity status description of the official records, we retain those that are "live", and then focus on three main industry divisions, based on the SSIC (Singapore Standard Industrial Classification), that account for a large portion of the employment force: construction (SSIC 41), manufacturing (SSIC 10), and services (SSIC 46, 47, 49). The corporate entities' records only have street addresses as geographical metadata. We geocode the street addresses for coordinates and then map them to areas.

10. Daily rainfall (mm) records come from the MSS (Meteorological Service Singapore) at http://www.weather.gov.sg/climate-historical-daily/. To derive rainfall for each `subzone-date`, we map each subzone to the nearest recorded weather station using the centroid of the subzone. Our data includes 46 weather stations on record (Figure S4). We apply linear interpolation for days where the rainfall record is missing. Temperature and wind records are sporadically available only for certain weather stations and dates and are thus not included in our data sample.

Table S1. Indicative mobile application sources.

App. Type	Share	App. Type	Share
AppGenre\Games	15%	AppGenre\Non-Games Apps\Reference	<1%
AppGenre\Non-Games Apps\Tools	13%	AppGenre\Non-Games Apps\Shopping	<1%
AppGenre\Games\Puzzle	8%	AppGenre\Games\Educational	<1%
AppGenre\Non-Games Apps\Media & Video	6%	AppGenre\Non-Games Apps\Events	<1%
AppGenre\Games\Casual	5%	AppGenre\Non-Games Apps\Autos & Vehicles	<1%
AppGenre\Non-Games Apps\Communication	4%	AppGenre\Non-Games Apps\Food & Drink	<1%
AppGenre\Games\Arcade	4%	AppGenre\Non-Games Apps\Libraries & Demo	<1%
AppGenre\Non-Games Apps\Lifestyle	3%	AppGenre\Games\Role Playing	<1%
AppGenre\Games\Action	3%	AppGenre\Games\Strategy	<1%
AppGenre\Non-Games Apps\Productivity	3%	AppGenre\Games\Adventure	<1%
AppGenre\Games\Simulation	3%	AppGenre\Non-Games Apps\Personalization	<1%
AppGenre\Non-Games Apps\Music & Audio	2%	AppGenre\Non-Games Apps\Comics	<1%
AppGenre\Non-Games Apps\Photography	2%	AppGenre\Non-Games Apps\Navigation	<1%
AppGenre\Non-Games Apps\Sports	2%	AppGenre\Games\Sports Games	<1%
AppGenre\Non-Games Apps\Books & Reference	2%	AppGenre\Non-Games Apps\Education	<1%
AppGenre\Non-Games Apps\Social	2%	AppGenre\Games\Casino	<1%
AppGenre\Games\Word	2%	AppGenre\Non-Games Apps\Weather	<1%
AppGenre\Non-Games Apps\Music	2%	AppGenre\Non-Games Apps\Health & Fitness	<1%
AppGenre\Games\Board	2%	AppGenre\Non-Games Apps\Travel & Local	<1%
AppGenre\Non-Games Apps\Entertainment	2%	AppGenre\Games\Sports	<1%
AppGenre\Non-Games Apps\Travel	2%	AppGenre\Non-Games Apps\Dating	<1%
AppGenre\Non-Games Apps\Social Networking	1%	AppGenre\Non-Games Apps\Business	<1%
AppGenre\Non-Games Apps\Photo & Video	1%	AppGenre\Non-Games Apps\News	<1%
AppGenre\Games\Racing	1%	AppGenre\Non-Games Apps\House & Home	<1%
AppGenre\Games\Trivia	1%	AppGenre\Non-Games Apps\Finance	<1%
AppGenre\Games\Music	1%	AppGenre\Non-Games Apps\Art & Design	<1%
AppGenre\Non-Games Apps\News & Magazines	1%	AppGenre\Non-Games Apps\Transportation	<1%
AppGenre\Games\Card	1%	AppGenre\Non-Games Apps\Medical	<1%
AppGenre\Non-Games Apps\Utilities	1%	AppGenre\Non-Games Apps\Beauty	<1%

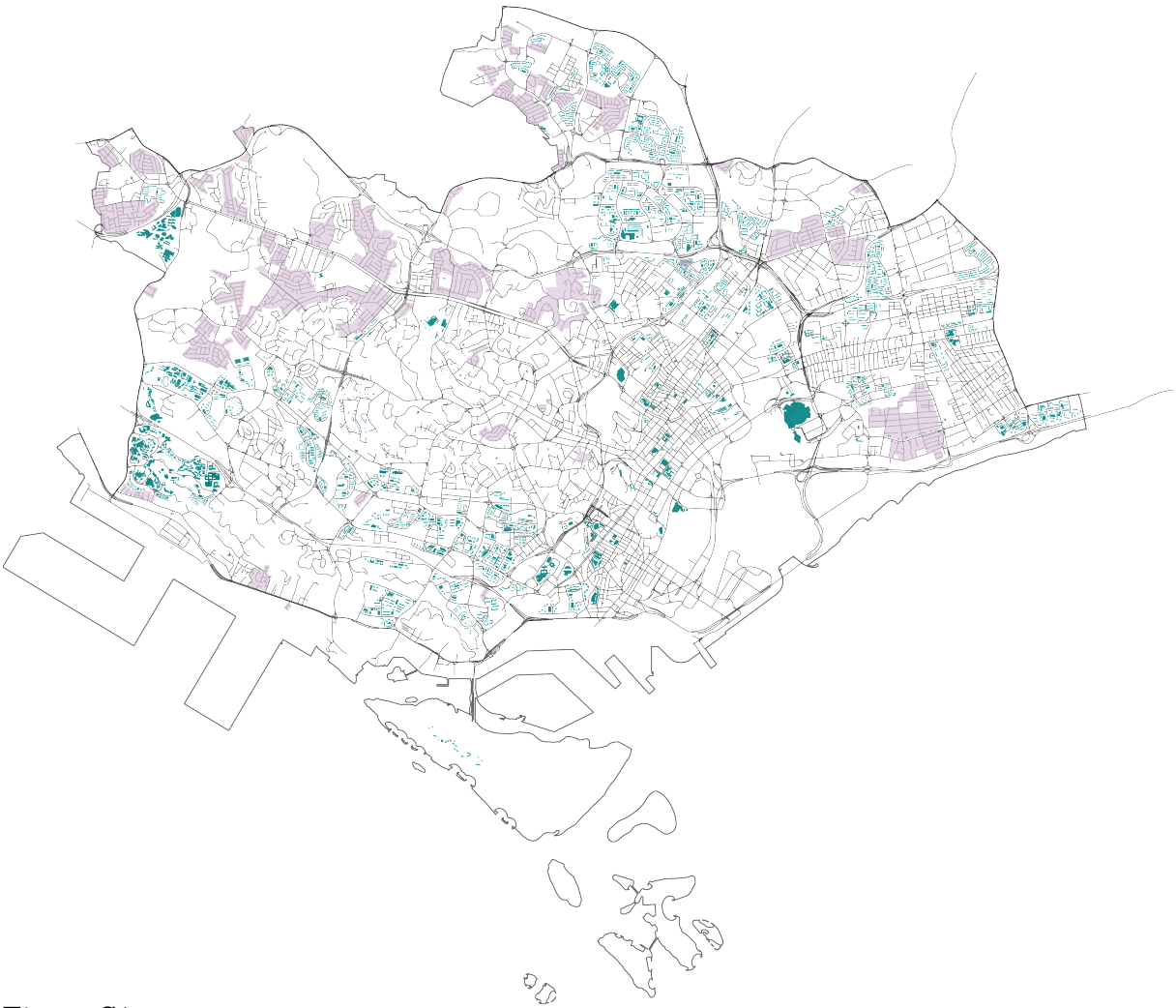

Figure S1. Urban overview of central region. Thin black lines are roads. Thick gray lines are highways. Blue polygons are buildings, which include the public housing flats. Pink polygons are the single-family houses with dedicated land space. Corresponds to Figure 5.

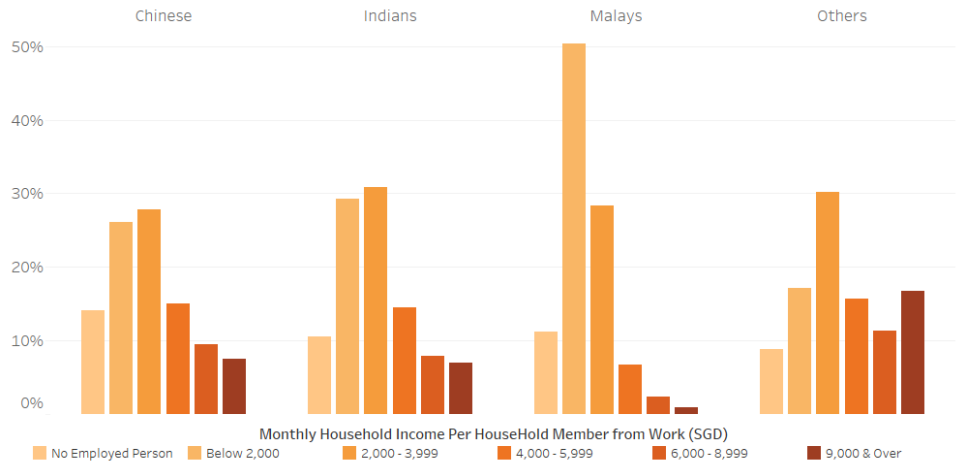

Figure S2. Percentage of Resident Households by Monthly Household Income Per Household Member and Ethnic Group. Each bar is calculated by taking the number of residents in each income group and dividing by the total number of residents across income groups for each ethnic group. The distribution of monthly household income per household member of Malay residents skews largely towards the lower income groups with 50.4 per cent and 1.0 per cent earning below $2,000 (below median income) and earning above $9,000 (highest bracket), respectively. The distributions for Chinese, Indian, and 'Others' Residents are more spread out with 26.0 and 7.5 per cent, 29.3 and 6.9 per cent, and 17.2 and 16.7 per cent earning below $2,000 and earning above $9,000, respectively.

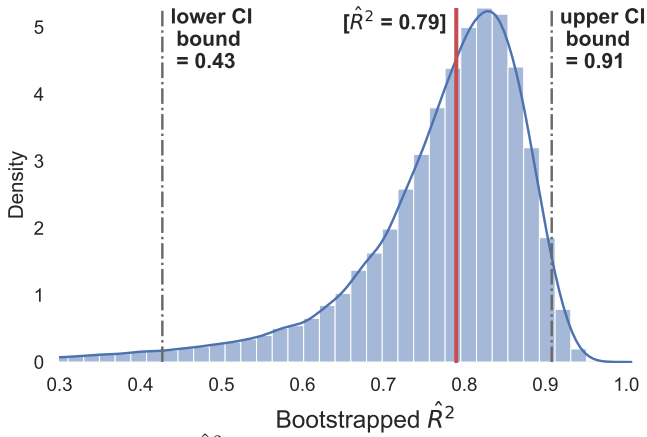

Figure S3. Bootstrapped values of $\hat{R}^2$ from regressing house price on census income. At the 28 census planning areas where census income data is available. Number of bootstrap samples is 100,000. Red vertical line indicates the sample $\hat{R}^2$ value. The gray dotted lines indicate the lower and upper 95% confidence interval bounds. Corresponds to Figure 3.

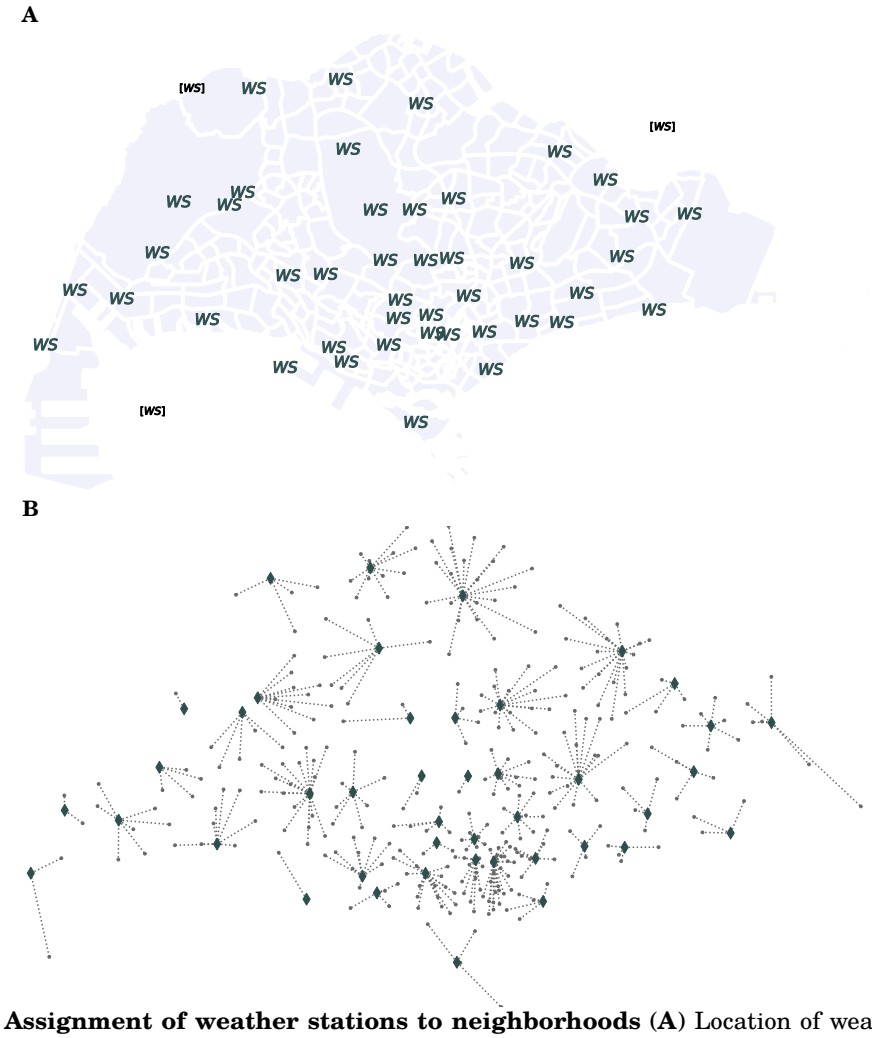

Figure S4. Assignment of weather stations to neighborhoods (**A**) Location of weather stations in Singapore. The three weather stations that end up with no area mapping is shown in brackets. (**B**) Matches encoded by dotted lines, connecting weather stations to neighborhoods.

B. Model

Here, we sketch a simple model where individuals derive utility via inter-neighborhood visits, closely following Kreindler and Miyauchi (2023). An individual i going from neighborhood o to neighborhood d on day t derives total utility

$$U_{i,odt} = \frac{u_{dt}^{e_1} m_{od}^{e_2}}{D_{od}^{e_3}} \epsilon_{i,odt}$$

where u_{dt} is the utility individual i derives from visiting neighborhood d on day t, which subsumes wage-differentials across neighborhoods in a labor-specific context as in Kreindler and Miyauchi 2023 as well as utility from amenities (Miyauchi et al. 2021). For example, certain neighborhoods may coincide or be close to business districts and yield higher utility u_{dt}. Moreover, u_{dt} allows the utility of neighborhood d to differ across time, and in our analyses, it varies by day of the week. This nests the assumption that certain neighborhoods would yield different utility depending on whether it is a weekday or a weekend (e.g., central business districts, Blanchard et al. 2023).

The key factor of our study is m_{od}. This is the utility (disutility if $e_2 < 0$) from visiting neighborhood d because of inherent mismatches (Figure S5) in social dimensions with individual i's neighborhood of origin o. D_{od} is the disutility from travel costs between neighborhoods o and d. $\epsilon_{i,odt}$ is the individual-specific idiosyncratic utility shock.

Given the origin neighborhood o and day t, the probability that the representative individual i from neighborhood o visits d on day t is

$$f_{odt|ot} = \frac{u_{dt}^{e_1} m_{od}^{e_2} / D_{od}^{e_3}}{\sum_{\ell \neq o} \left(u_{\ell t}^{e_1} m_{o\ell}^{e_2} / D_{o\ell}^{e_3} \right)} \in [0, 1]$$

and taking logs gives

$$\log f_{odt|ot} = e_1 \log u_{dt} + e_2 \log m_{od} - e_3 \log D_{od} - \log \left\{ \sum_{\ell \neq o} exp\Big(e_1 \log u_{\ell t} + e_2 \log m_{o\ell} - e_3 \log D_{o\ell} \Big) \right\}$$

which can be estimated using

$$\log(\text{visits})_{odt} = \theta_{dt} + \beta \log m_{od} + \delta \log D_{od} + \varepsilon_{odt}. \tag{S1}$$

β in Equation (S1) is the key parameter of interest and captures the effect of mismatch in social dimensions across neighborhoods (defined in later in Equation (2)) on inter-neighborhood movement. If the estimate of β is negative, then this suggests that mismatch induces disutility from inter-neighborhood travels, and vice versa.

θ_{dt} is the area-by-day fixed effects. Below, we use the planning census area-by-day fixed effects, where a census area is a geographical unit larger than a neighborhood so that the regression analyses

can identify the two components of the mismatch measures.

δ captures the effect of distance between neighborhoods on neighborhood visits. Our analyses include centroid-based distances and contiguity indicators for each origin-to-destination neighborhood pair. More broadly, δ captures the effect of spatial frictions (as in Davis et al. 2019; Miyauchi et al. 2021) while β captures the effect of social frictions on flows between neighborhoods. To mitigate concerns that β approximates the true effect of neighborhood mismatches on neighborhood flows, we extend Equation (S1) to include the date fixed effects, which nests the day-of-week fixed effects, neighborhood-specific demographics, business, rainfall, and places of interest, where the effects are allowed to vary by time. The error term is ε_{odt}.

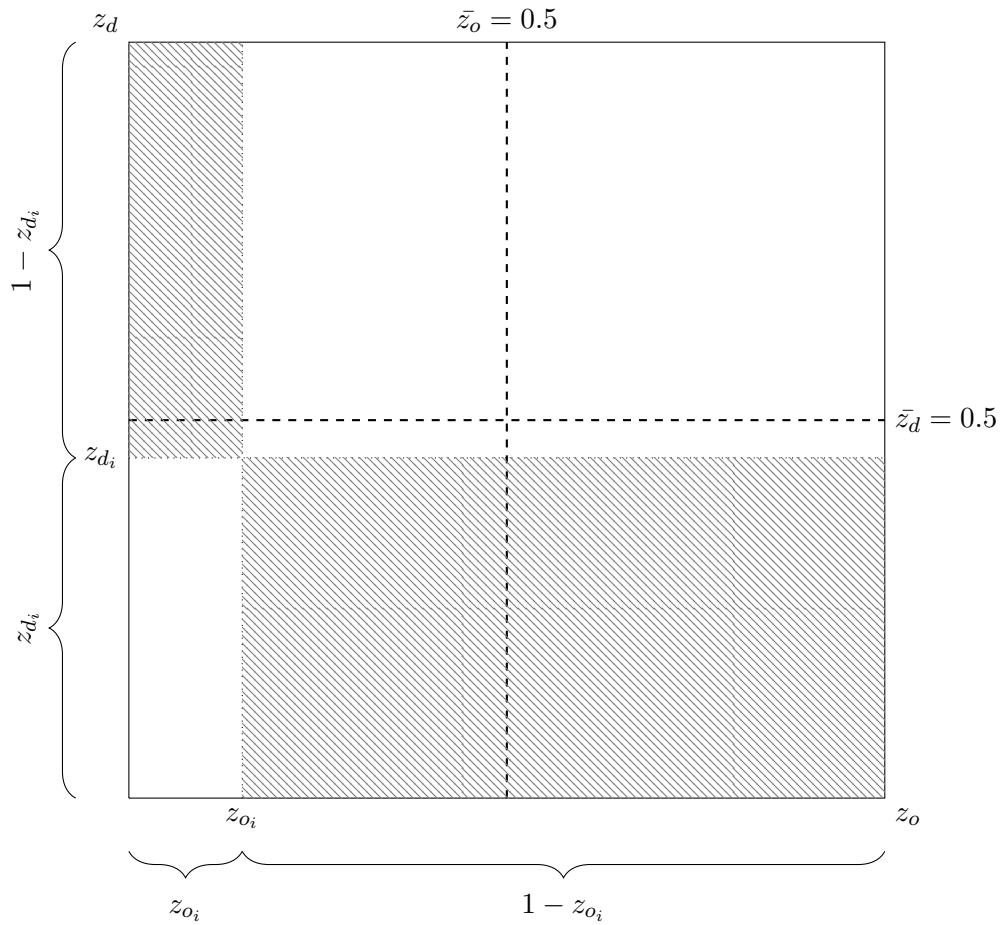

Figure S5. Sample space of mismatch. Unit square with length one on all sides. Shaded area is $mismatch_{od} = z_o(1 - z_d) + z_d(1 - z_o)$ which indicates the magnitude of the level of mismatch between neighborhoods o and d. $\bar{z_o}$ and $\bar{z_d}$ indicated by the dotted lines are thresholds. Without loss of generality, given z_o, an increase in z_d increases or decreases the shaded area depending on whether z_o is more or less than $\frac{1}{2}$:

$$\frac{\partial \{z_o(1 - z_d) + z_d(1 - z_o)\}}{\partial z_d} = 1 - 2z_o \lesseqgtr 0 \text{ if } z_o \lesseqgtr \frac{1}{2}$$

and z is bounded above by $\frac{1}{2}$ in the sample (ignoring extreme values on the right tail), so an increase in z_d increases the mismatch measure in the sample.

C. Supplementary Tables to Main Figures

Table S2. Effect of income mismatch on neighborhood visits.

	Dependent variable: Log neighborhood visits				
	(1)	(2)	(3)	(4)	(5)
Income mismatch	0.0181**	0.0145*	0.0120	0.0119	0.0207**
	(0.0082)	(0.0083)	(0.0084)	(0.0084)	(0.0085)
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
R^2	0.7332	0.7351	0.7353	0.7354	0.7367
Days	91	91	91	91	91
Clusters	1,439	1,434	1,434	1,434	1,434
Observations	1,321,467	1,228,385	1,228,385	1,228,385	1,228,385

Notes: Coefficients are estimated from Equation (3), with mismatch defined in Equation (2). Income is measured as the price per square meter of houses, weighted by the share of public and private residences in the neighborhood population. The model in column (1) includes neighborhood land area, population density (from census and real-time records), neighborhood contiguity and distance, and census area-by-day fixed effects interactions. Columns (2)–(5) progressively add more controls as indicated. The estimates correspond to Figure 4A. Standard errors are clustered at the origin-by-destination census planning area level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table S3. Asymmetry in income mismatch and neighborhood visits.

	Dependent variable: Log neighborhood visits				
	(1)	(2)	(3)	(4)	(5)
P $\rightarrow$ NP	0.0690^{***}	0.0536^{***}	0.0567^{***}	0.0564^{***}	0.0479^{***}
	(0.0105)	(0.0106)	(0.0106)	(0.0106)	(0.0105)
NP $\rightarrow$ P	-0.0289^{***}	-0.0217^{**}	-0.0309^{***}	-0.0307^{***}	-0.0077
	(0.0100)	(0.0101)	(0.0104)	(0.0104)	(0.0111)
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
p-val: (P $\rightarrow$ NP) $\neq$ (NP $\rightarrow$ P)	.015	.054	.123	.125	.018
R^2	0.7336	0.7353	0.7357	0.7357	0.7368
Days	91	91	91	91	91
Clusters	1,439	1,434	1,434	1,434	1,434
Observations	1,321,467	1,228,385	1,228,385	1,228,385	1,228,385

Notes: Coefficients are estimated from Equation (3). P $\rightarrow$ NP indicates movement from poor to non-poor neighborhoods (or, $z_o(1 - z_d)$ in Equation (2)). NP $\rightarrow$ P indicates the reverse (or, $z_d(1 - z_o)$ in Equation (2)). Income is measured as the price per square meter of houses, weighted by the share of public and private residences in the neighborhood population. The model in column (1) includes neighborhood land area, population density (from census and real-time records), neighborhood contiguity and distance, and census area-by-day fixed effects interactions. Columns (2)–(5) progressively add more controls as indicated. The table also reports p-values for the null that (P$\rightarrow$NP) + (NP$\rightarrow$P) = 0. The estimates correspond to Figure 4B. Standard errors are clustered at the origin-by-destination census planning area level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table S4. Asymmetry in income mismatch by geography.

	Dependent variable: Log neighborhood visits				
	(1)	(2)	(3)	(4)	(5)
Income mismatch	0.0433^{***}	0.0376^{***}	0.0344^{***}	0.0340^{***}	0.0404^{***}
	(0.0118)	(0.0117)	(0.0118)	(0.0118)	(0.0117)
Income mismatch × Non-central destination	-0.0452^{***}	-0.0410^{***}	-0.0399^{***}	-0.0393^{***}	-0.0353^{**}
	(0.0153)	(0.0152)	(0.0152)	(0.0152)	(0.0154)
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
p-val: $\hat{\beta} + \hat{\gamma} = 0$	.859	.755	.62	.634	.65
R^2	0.7333	0.7352	0.7354	0.7355	0.7368
Days	91	91	91	91	91
Clusters	1,439	1,434	1,434	1,434	1,434
Observations	1,321,467	1,228,385	1,228,385	1,228,385	1,228,385

Notes: Coefficients estimated from fully interacting income mismatch with an indicator for destinations outside the central region:

$$\log(\textit{visits})_{odt} = \alpha + \beta \textit{mismatch} + \gamma(\textit{mismatch} \times \mathbb{1}^{\textit{Non-central dest.}}) + \Gamma_t X_{odt} + \varepsilon_{odt},$$

where $\mathbb{1}^{\text{Non-central dest.}}$ is the non-central region indicator. P → NP indicates movement from poor to non-poor neighborhoods (or, $z_o(1 - z_d)$ in Equation (2)). NP → P indicates the reverse (or, $z_d(1 - z_o)$ in Equation (2)). Income is measured as the price per square meter of houses, weighted by the share of public and private residences in the neighborhood population. The model in column (1) includes neighborhood land area, population density (from census and real-time records), neighborhood contiguity and distance, and census area-by-day fixed effects interactions. Columns (2)–(5) progressively add more controls as indicated. The estimates correspond to Figure 7. Standard errors are clustered at the origin-by-destination census planning area level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table S5. Effect of ethnic mismatch on neighborhood visits.

	Dependent variable: Log neighborhood visits				
	(1)	(2)	(3)	(4)	(5)
Ethnic mismatch	-0.2198^{***}	-0.2804^{***}	-0.3071^{***}	-0.3070^{***}	-0.2109^{**}
	(0.0758)	(0.0938)	(0.0946)	(0.0945)	(0.1020)
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
R^2	0.7338	0.7345	0.7347	0.7348	0.7360
Days	91	91	91	91	91
Clusters	1,592	1,510	1,510	1,510	1,510
Observations	1,394,127	1,255,583	1,255,583	1,255,583	1,255,583

Notes: Coefficients are estimated from Equation (3), with mismatch defined in Equation (2). Mismatch is as defined in Equation (2), computed using the proportion of minority ethnic residents (z in Equation (2)) in the neighborhoods. The model in column (1) includes neighborhood land area, population density (from census and real-time records), neighborhood contiguity and distance, and census area-by-day fixed effects interactions. Columns (2)–(5) progressively add more controls as indicated. The estimates correspond to Figure 6A. Standard errors are clustered at the origin-by-destination census planning area level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table S6. Asymmetry in ethnic mismatch and neighborhood visits.

	Dependent variable: Log neighborhood visits				
	(1)	(2)	(3)	(4)	(5)
Mnr $\rightarrow$ Maj	0.0703	-0.0370	-0.0486	-0.0500	-0.0292
	(0.1004)	(0.1100)	(0.1101)	(0.1098)	(0.1115)
Maj $\rightarrow$ Mnr	-0.3498^{***}	-0.4445^{***}	-0.4922^{***}	-0.4907^{***}	-0.3911^{***}
	(0.0733)	(0.0959)	(0.0974)	(0.0974)	(0.1106)
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
p-val: (Mnr $\rightarrow$ Maj) $\neq$ (Maj $\rightarrow$ Mnr)	.081	.011	.005	.005	.04
R^2	0.7341	0.7347	0.7349	0.7350	0.7362
Days	91	91	91	91	91
Clusters	1,592	1,510	1,510	1,510	1,510
Observations	1,394,127	1,255,583	1,255,583	1,255,583	1,255,583

Notes: The two key independent variables, (Mnr $\rightarrow$ Maj) and (Maj $\rightarrow$ Mnr), are the first and second term in the mismatch measure defined in Equation (2). Income is measured as the price per square meter of houses, weighted by the share of public and private residences in the neighborhood population. The model in column (1) includes neighborhood land area, population density (from census and real-time records), neighborhood contiguity and distance, and census area-by-day fixed effects interactions. Columns (2)–(5) progressively add more controls as indicated. The table also reports the p-value for the null that (Mnr $\rightarrow$ Maj) + (Maj $\rightarrow$ Mnr) = 0. The estimates correspond to Figure 6B. Standard errors are clustered at the origin-by-destination census planning area level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

D. Additional Robustness Tests

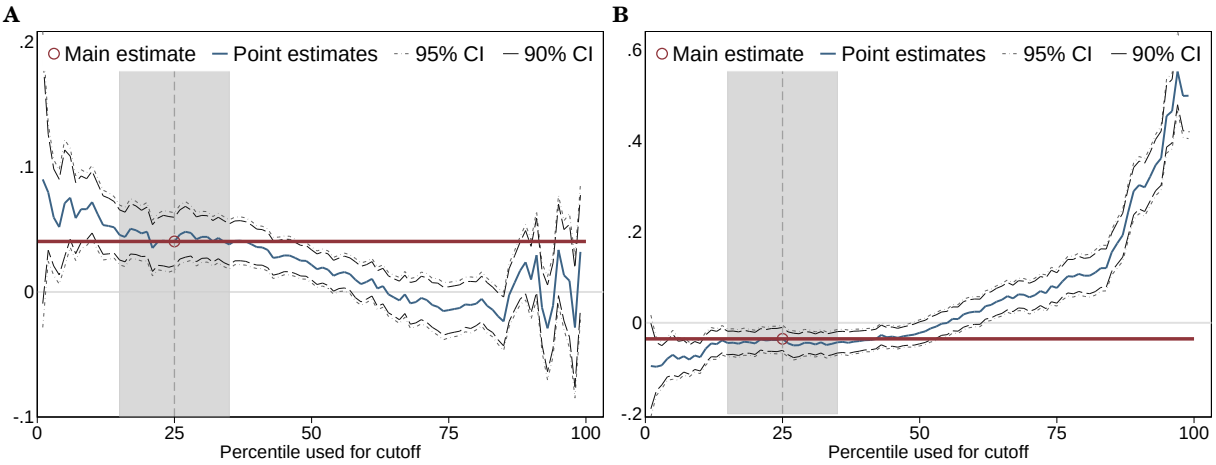

Figure S6. Sensitivity tests for wealth mismatch by geography. **(A)** Sensitivity of the mismatch by wealth coefficient in the top panel of Figure 7. **(B)** Sensitivity of the mismatch by wealth interacted with non-central destination coefficient in the bottom panel of Figure 7. Shaded gray areas indicate the 15th to the 35th percentiles for a $\pm$ 10 range around the 25th percentile. The specification is the one with the full controls. Figure corresponds also to Table S4. Figure S11 reports the results after sorting by estimate size.

Table S7. Effect of absolute distance in wealth on neighborhood visits.

	Dependent variable: Log neighborhood visits				
	(1)	(2)	(3)	(4)	(5)
$\lvert\text{income}_o - \text{income}_d\rvert$	8.6386^{***}	6.0795^{***}	6.0222^{***}	6.0103^{***}	6.4518^{***}
	(1.0152)	(0.9503)	(0.9528)	(0.9539)	(0.9471)
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
R^2	0.7344	0.7356	0.7358	0.7359	0.7372
Days	91	91	91	91	91
Clusters	1,439	1,434	1,434	1,434	1,434
Observations	1,321,467	1,228,385	1,228,385	1,228,385	1,228,385

Notes: Income is measured as the price per square meter of houses, weighted by the share of public and private residences in the neighborhood population. The model in column (1) includes neighborhood land area, population density (from census and real-time records), neighborhood contiguity and distance, and census area-by-day fixed effects interactions. Columns (2)–(5) progressively add more controls as indicated. Standard errors are clustered at the origin-by-destination census planning area level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table S8. Effect of income mismatch on neighborhood visits (neighborhood-by-day fixed effects).

	Dependent variable: Log neighborhood visits				
	(1)	(2)	(3)	(4)	(5)
Income mismatch	0.0345^{***}	0.0268^{***}	0.0268^{***}	0.0268^{***}	0.0268^{***}
	(0.0100)	(0.0098)	(0.0098)	(0.0098)	(0.0098)
Day fixed effects	✓	✓	✓	✓	✓
Neighborhood-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
R^2	0.7485	0.7484	0.7484	0.7484	0.7484
Days	91	91	91	91	91
Clusters	1,439	1,434	1,434	1,434	1,434
Observations	1,321,408	1,228,357	1,228,357	1,228,357	1,228,357

Notes: Coefficients are estimated from Equation (3), with mismatch defined in Equation (2). Corresponds to Table S2, but with neighborhood-by-day fixed effects instead of census area-by-day fixed effects. The model in column (1) includes neighborhood land area, population density (from census and real-time records), neighborhood contiguity and distance, and census area-by-day fixed effects interactions. Columns (2)–(5) progressively add more controls as indicated. The estimates correspond to Figure 4A. Standard errors are clustered at the origin-by-destination census planning area level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table S9. Asymmetry in wealth mismatch by geography (neighborhood-by-day fixed effects).

	Dependent variable: Log neighborhood visits				
	(1)	(2)	(3)	(4)	(5)
Income mismatch	0.0595^{***}	0.0512^{***}	0.0512^{***}	0.0512^{***}	0.0512^{***}
	(0.0162)	(0.0155)	(0.0155)	(0.0155)	(0.0155)
Income mismatch × Non-central destination	-0.0407^{**}	-0.0396^{**}	-0.0396^{**}	-0.0396^{**}	-0.0396^{**}
	(0.0191)	(0.0185)	(0.0185)	(0.0185)	(0.0185)
Day fixed effects	✓	✓	✓	✓	✓
Neighborhood-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
p-val: $\hat{\beta} + \hat{\gamma} = 0$	.111	.32	.32	.32	.32
R^2	0.7486	0.7485	0.7485	0.7485	0.7485
Days	91	91	91	91	91
Clusters	1,439	1,434	1,434	1,434	1,434
Observations	1,321,408	1,228,357	1,228,357	1,228,357	1,228,357

Notes: Coefficients estimated from fully interacting income mismatch with an indicator variable for the destinations outside the central region:

$$\log(\text{inflow})_{odt} = \alpha + \beta \text{mismatch} + \gamma(\text{mismatch} \times \mathbb{1}^{\text{Non-central dest.}}) + \Gamma_t X_{odt} + \varepsilon_{odt},$$

where $\mathbb{1}^{\text{Non-central dest.}}$ is an indicator for destinations outside the census central region. Mismatch measure is as defined in Equation (2). Income is the price per square meters of houses weighted by the share of public and private residence in the neighborhood population. Corresponds to Table S4, but with neighborhood-by-day fixed effects instead of census area-by-day fixed effects. The model in column (1) includes neighborhood land area, population density (from census and real-time records), neighborhood contiguity and distance, and census area-by-day fixed effects interactions. Columns (2)–(5) progressively add more controls as indicated. Standard errors are clustered at the origin-by-destination census planning area level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table S10. Effect of ethnic mismatch on neighborhood visits (neighborhood-by-day fixed effects).

	Dependent variable: Log neighborhood visits				
	(1)	(2)	(3)	(4)	(5)
Ethnic mismatch	-0.7238^{***}	-1.0694^{***}	-1.0694^{***}	-1.0694^{***}	-1.0694^{***}
	(0.2678)	(0.2560)	(0.2560)	(0.2560)	(0.2561)
Day fixed effects	✓	✓	✓	✓	✓
Neighborhood-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
R^2	0.7484	0.7478	0.7478	0.7478	0.7478
Days	91	91	91	91	91
Clusters	1,592	1,510	1,510	1,510	1,510
Observations	1,394,088	1,255,555	1,255,555	1,255,555	1,255,555

Notes: Coefficients are estimated from Equation (3), with mismatch defined in Equation (2). Mismatch is as defined in Equation (2), computed using the proportion of minority ethnic residents (z in Equation (2)) in the neighborhoods. Corresponds to Table S5, but with neighborhood-by-day fixed effects instead of census area-by-day fixed effects. The model in column (1) includes neighborhood land area, population density (from census and real-time records), neighborhood contiguity and distance, and census area-by-day fixed effects interactions. Columns (2)–(5) progressively add more controls as indicated. The estimates correspond to Figure 6A. Standard errors are clustered at the origin-by-destination census planning area level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table S11. Effect of ethnic mismatch on movement (log-log specification).

	Dependent variable: Log neighborhood visits				
	(1)	(2)	(3)	(4)	(5)
log(Ethnic mismatch)	-0.0819^{***}	-0.0868^{***}	-0.0944^{***}	-0.0943^{***}	-0.0667^{**}
	(0.0271)	(0.0308)	(0.0310)	(0.0310)	(0.0331)
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
R^2	0.7338	0.7345	0.7347	0.7348	0.7360
Days	91	91	91	91	91
Clusters	1,592	1,510	1,510	1,510	1,510
Observations	1,394,127	1,255,583	1,255,583	1,255,583	1,255,583

Coefficients estimated from Equation (3). Mismatch is as defined in Equation (2), computed using the proportion of minority ethnic residents (z in Equation (2)) in the neighborhoods. The model in column (1) includes neighborhood land area, population density (from census and real-time records), neighborhood contiguity and distance, and census area-by-day fixed effects interactions. Columns (2)–(5) progressively add more controls as indicated. Corresponds to Figure 6A. Standard errors are clustered at the origin-by-destination census planning area level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

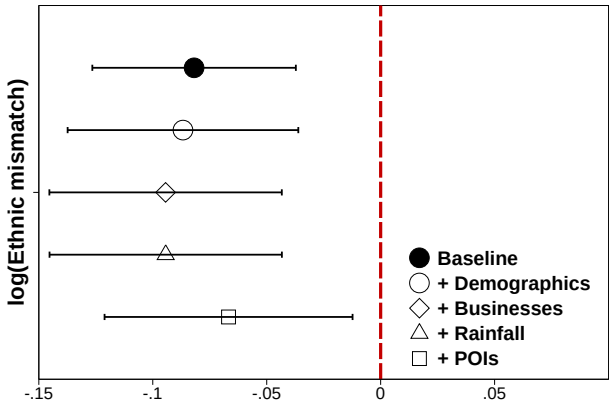

Figure S7. Ethnic mismatch and neighborhood visits—log-log specification. The dependent variable is log visits from origin o to destination d. The ethnic mismatch measure is logged (Appendix B), unlike in Figure 6. Baseline model includes neighborhood land area, population density (by both census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Figure corresponds to Figure 6 and Table S11. Capped horizontal lines are 90% CI from standard errors clustered at the origin-by-destination areas.

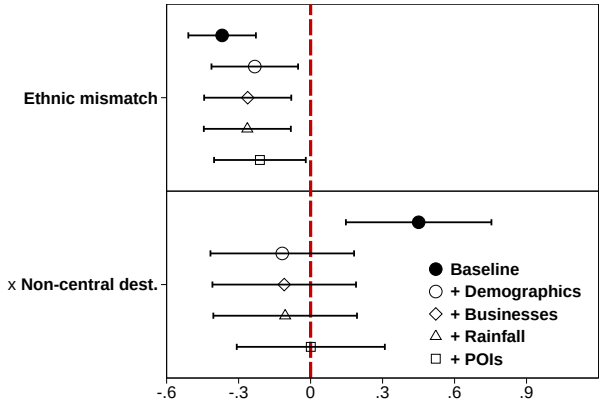

Figure S8. Asymmetry in ethnic mismatch by geography. Top panel is mismatch, and bottom panel is the mismatch measure interacted with an indicator for destination neighborhoods outside the central region. Baseline model includes neighborhood land area, population density (by both census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Dependent variable is log inflow from origin o to destination d. Figure corresponds to Table S12. Capped horizontal lines are 90% CI from standard errors clustered at the origin-by-destination areas.

Table S12. Asymmetry in ethnic mismatch by geography.

	Dependent variable: Log neighborhood visits				
	(1)	(2)	(3)	(4)	(5)
Ethnic mismatch	-0.3687^{***}	-0.2327^{**}	-0.2621^{**}	-0.2637^{**}	-0.2112^{*}
	(0.0856)	(0.1096)	(0.1103)	(0.1102)	(0.1162)
Ethnic mismatch × Non-central destination	0.4504^{**}	-0.1181	-0.1099	-0.1060	0.0008
	(0.1843)	(0.1819)	(0.1818)	(0.1819)	(0.1875)
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
p-val: $\hat{\beta} + \hat{\gamma} = 0$	.606	.023	.016	.017	.2
R^2	0.7340	0.7345	0.7347	0.7348	0.7360
Days	91	91	91	91	91
Clusters	1,592	1,510	1,510	1,510	1,510
Observations	1,394,127	1,255,583	1,255,583	1,255,583	1,255,583

Table S13. Coefficients estimated from fully interacting the wealth mismatch measure with an indicator variable for the destinations outside the central region:

$$\log(\text{inflow})_{odt} = \alpha + \beta \text{mismatch} + \gamma(\text{mismatch} \times \mathbb{1}^{\text{Non-central dest.}}) + \Gamma_t X_{odt} + \varepsilon_{odt},$$

where $\mathbb{1}^{\text{Non-central dest.}}$ is an indicator for destinations outside the census central region. The model in column (1) includes neighborhood land area, population density (from census and real-time records), neighborhood contiguity and distance, and census area-by-day fixed effects interactions. Columns (2)–(5) progressively add more controls as indicated. Table also reports p-values for the null that $\beta + \gamma = 0$. Standard errors are clustered at the origin-by-destination census planning area level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table S14. Asymmetry in ethnic mismatch by geography (neighborhood-by-day fixed effects).

	Dependent variable: Log neighborhood visits				
	(1)	(2)	(3)	(4)	(5)
Ethnic mismatch	-0.9852^{***}	-1.1058^{***}	-1.1058^{***}	-1.1058^{***}	-1.1058^{***}
	(0.2692)	(0.2626)	(0.2626)	(0.2626)	(0.2627)
Ethnic mismatch × Non-central destination	0.7485^{***}	0.1036	0.1036	0.1036	0.1036
	(0.2307)	(0.2155)	(0.2155)	(0.2155)	(0.2156)
Day fixed effects	✓	✓	✓	✓	✓
Neighborhood-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
p-val: $\hat{\beta} + \hat{\gamma} = 0$	.46	.001	.001	.001	.001
R^2	0.7487	0.7478	0.7478	0.7478	0.7478
Days	91	91	91	91	91
Clusters	1,592	1,510	1,510	1,510	1,510
Observations	1,394,088	1,255,555	1,255,555	1,255,555	1,255,555

Table S15. Coefficients estimated from fully interacting the wealth mismatch measure with an indicator variable for the destinations outside the central region:

$$\log(\text{inflow})_{odt} = \alpha + \beta \text{mismatch} + \gamma(\text{mismatch} \times \mathbb{1}^{\text{Non-central dest.}}) + \Gamma_t X_{odt} + \varepsilon_{odt},$$

where $\mathbb{1}^{\text{Non-central dest.}}$ is an indicator for destinations outside the census central region. The model in column (1) includes neighborhood land area, population density (from census and real-time records), neighborhood contiguity and distance, and census area-by-day fixed effects interactions. Columns (2)–(5) progressively add more controls as indicated. Corresponds to Table S2, but with neighborhood-by-day fixed effects instead of census area-by-day fixed effects. Table also reports p-values for the null that $\beta + \gamma = 0$. Corresponds to Table S12. Standard errors are clustered at the origin-by-destination census planning area level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table S16. Effect of income and ethnic mismatch on neighborhood visits.

	Dependent variable: Log neighborhood visits				
	(1)	(2)	(3)	(4)	(5)
Income mismatch	0.0214**	0.0182**	0.0159*	0.0158*	0.0233***
	(0.0085)	(0.0084)	(0.0084)	(0.0084)	(0.0085)
Ethnic mismatch	−0.2873***	−0.2713***	−0.2945***	−0.2934***	−0.2223**
	(0.0924)	(0.0959)	(0.0964)	(0.0962)	(0.1034)
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
R^2	0.7336	0.7352	0.7354	0.7355	0.7368
Days	91	91	91	91	91
Clusters	1,439	1,434	1,434	1,434	1,434
Observations	1,308,012	1,228,385	1,228,385	1,228,385	1,228,385

Table S17. Same as Table S2 and Table S5, but with income and ethnic mismatch considered together. The model in column (1) includes neighborhood land area, population density (from census and real-time records), neighborhood contiguity and distance, and census area-by-day fixed effects interactions. Columns (2)–(5) progressively add more controls as indicated. Standard errors are clustered at the origin-by-destination census planning area level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

E. Sensitivity Tests Ordered by Effect Size

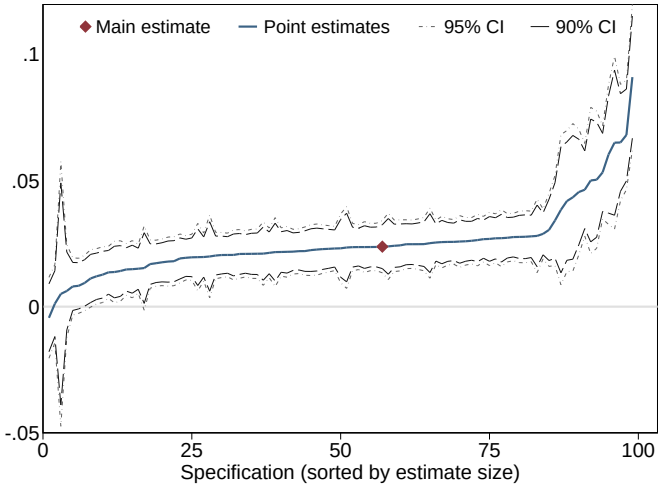

Figure S9. Sensitivity tests of income mismatch (sorted). Corresponds to the unsorted plot in Figure 8 and Table S2. The main estimate is the red diamond. Coefficients are sorted by effect size instead of percentiles as in Figure 8.

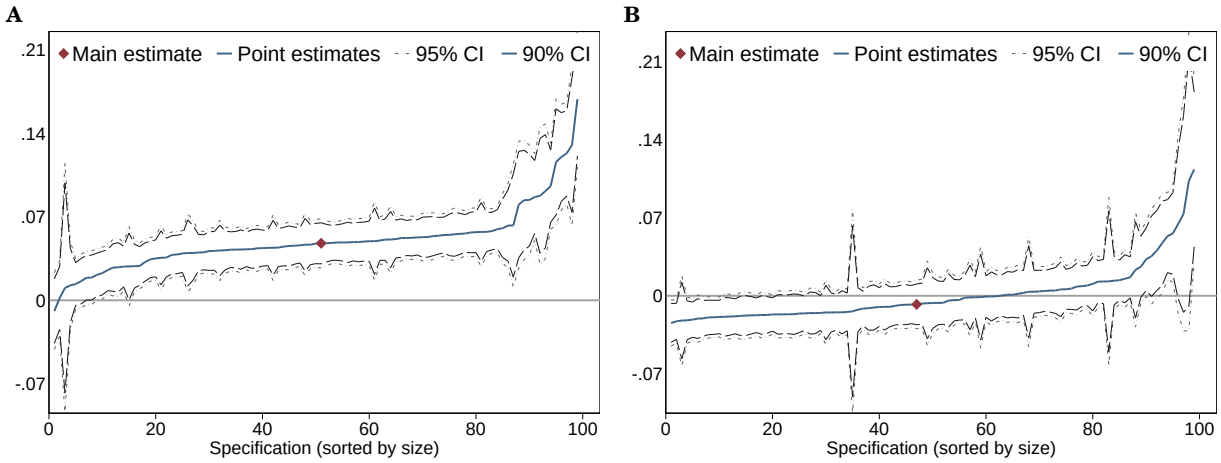

Figure S10. Sensitivity tests of income mismatch by direction of mismatch (sorted). **(A)** Sensitivity of the ($P \rightarrow NP$) coefficient from the top panel of Figure 4B. **(B)** Sensitivity of the ($NP \rightarrow P$) coefficient from the bottom panel of Figure 4B. Corresponds to the unsorted plot in Figure 9 and Table S3. The main estimate is the red diamond. Coefficients are sorted by effect size instead of percentiles as in Figure 9.

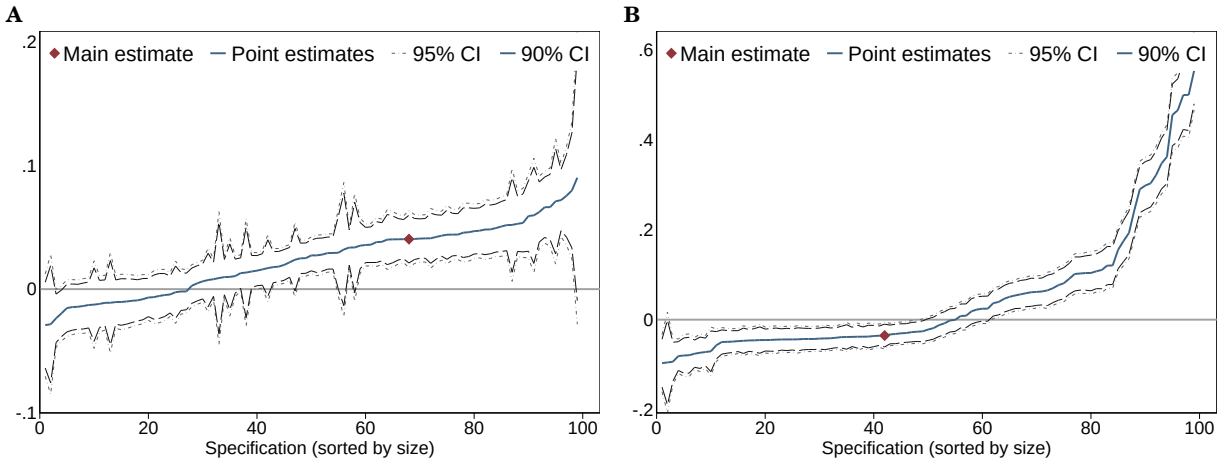

Figure S11. Sensitivity tests for wealth mismatch by geography (sorted). **(A)** Sensitivity of the mismatch by wealth coefficient in the top panel of Figure 7. **(B)** Sensitivity of the mismatch by wealth interacted with non-central destination coefficient in the bottom panel of Figure 7. Corresponds to the unsorted plot in Figure S6 and Table S4. The main estimate is the red diamond. Coefficients are sorted by effect size instead of percentiles as in Figure S6.

F. Representativeness of GPS pings by neighborhood demographics

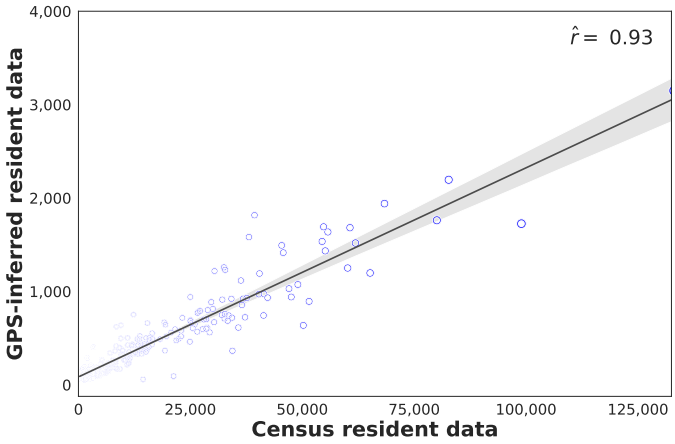

Figure S12. Inference of residence location. GPS-inferred resident data is the average number of captured devices across neighborhood-days inferred as residing in that neighborhood. Supplementary materials section A. Data Appendix provides more details on inferring residence. Figure reports the correlation between the GPS-inferred resident size and the census resident size.

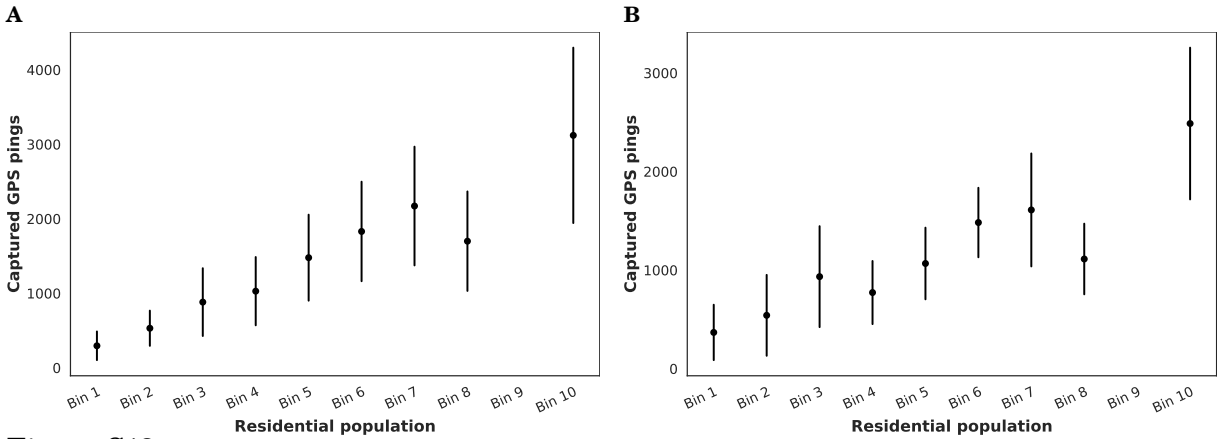

Figure S13. Variation in GPS pings by neighborhood population size. **(A)** Origin neighborhoods. **(B)** Destination neighborhoods. Horizontal axis is bins for neighborhood populaton size. Vertical axis is the number of captured GPS pings by neighborhood-date. Vertical bars are standard deviations.

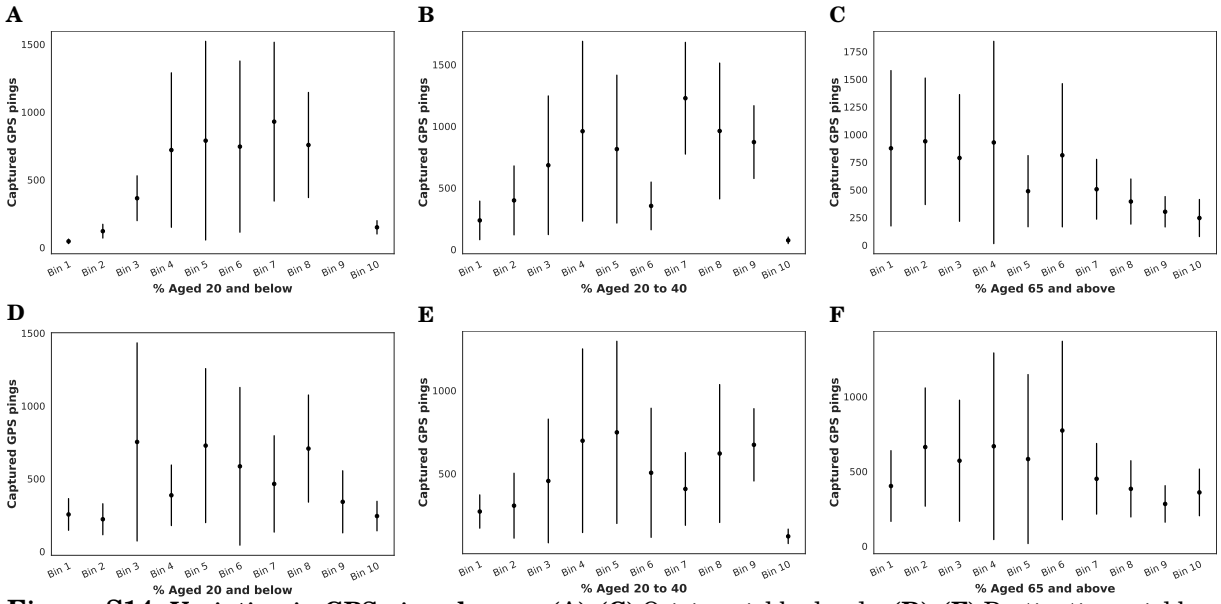

Figure S14. Variation in GPS pings by age. **(A)–(C)** Origin neighborhoods. **(D)–(F)** Destination neighborhoods. Horizontal axis is day of the week. Vertical axis is the number of captured GPS pings by neighborhood-date. Vertical bars are standard deviations.

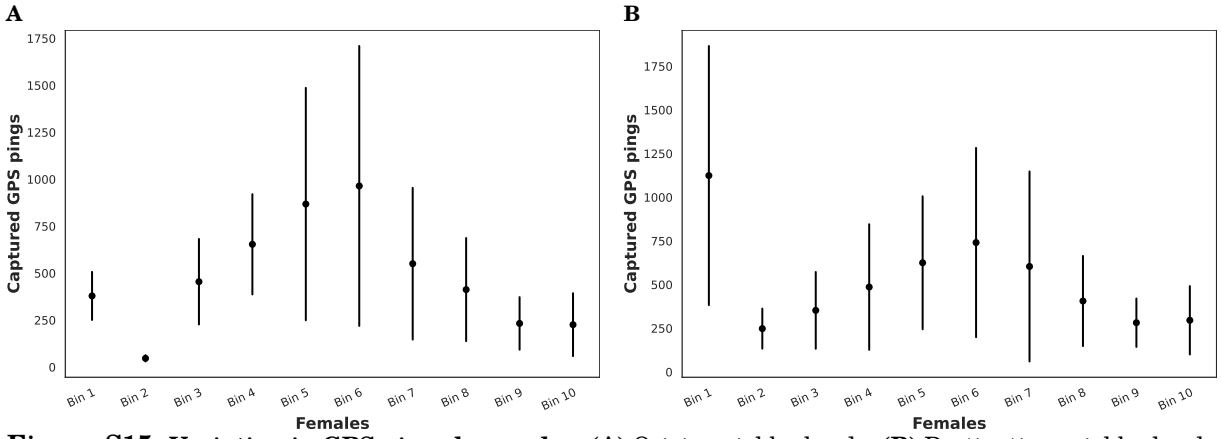

Figure S15. Variation in GPS pings by gender. **(A)** Origin neighborhoods. **(B)** Destination neighborhoods. Horizontal axis is female proportion in a neighborhood. Vertical axis is the number of captured GPS pings by neighborhood-date. Vertical bars are standard deviations.

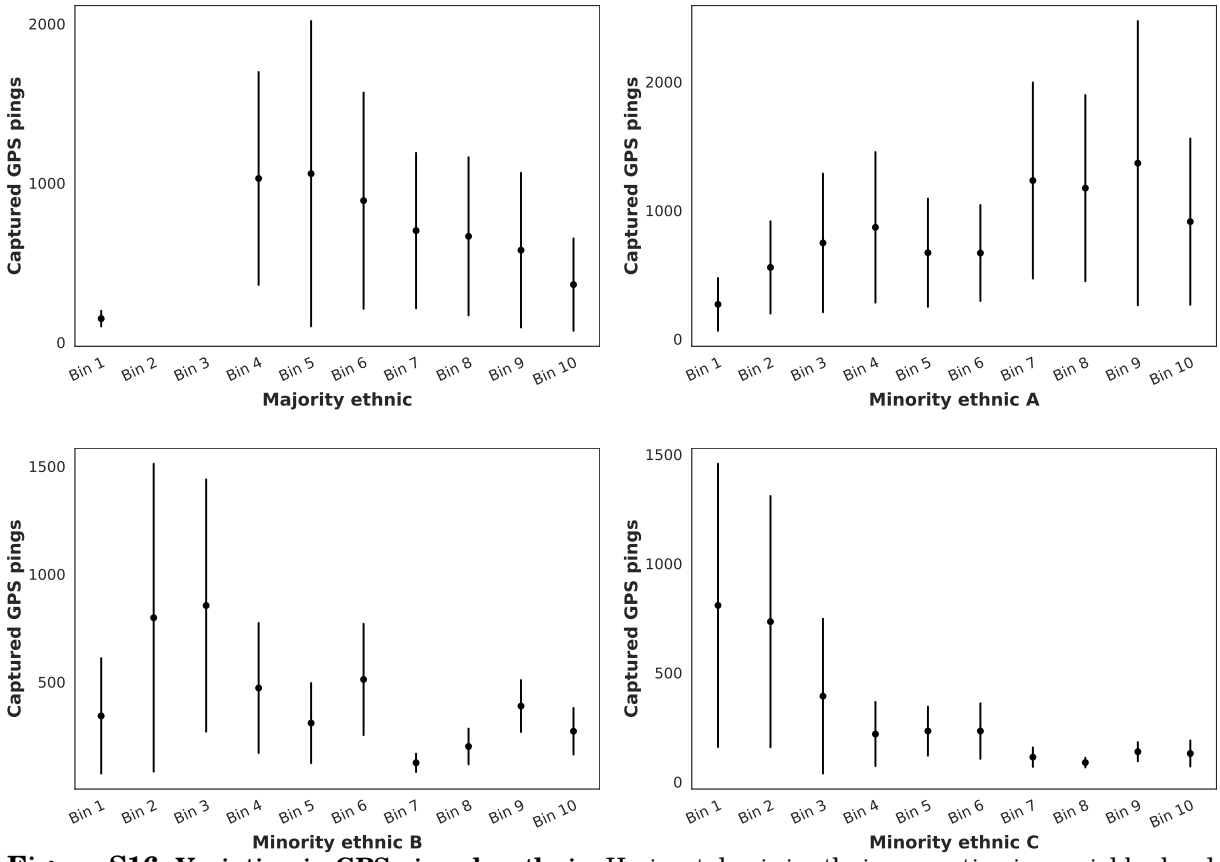

Figure S16. Variation in GPS pings by ethnic. Horizontal axis is ethnic proportion in a neighborhood. Vertical axis is the number of captured GPS pings by neighborhood-date. Vertical bars are standard deviations.

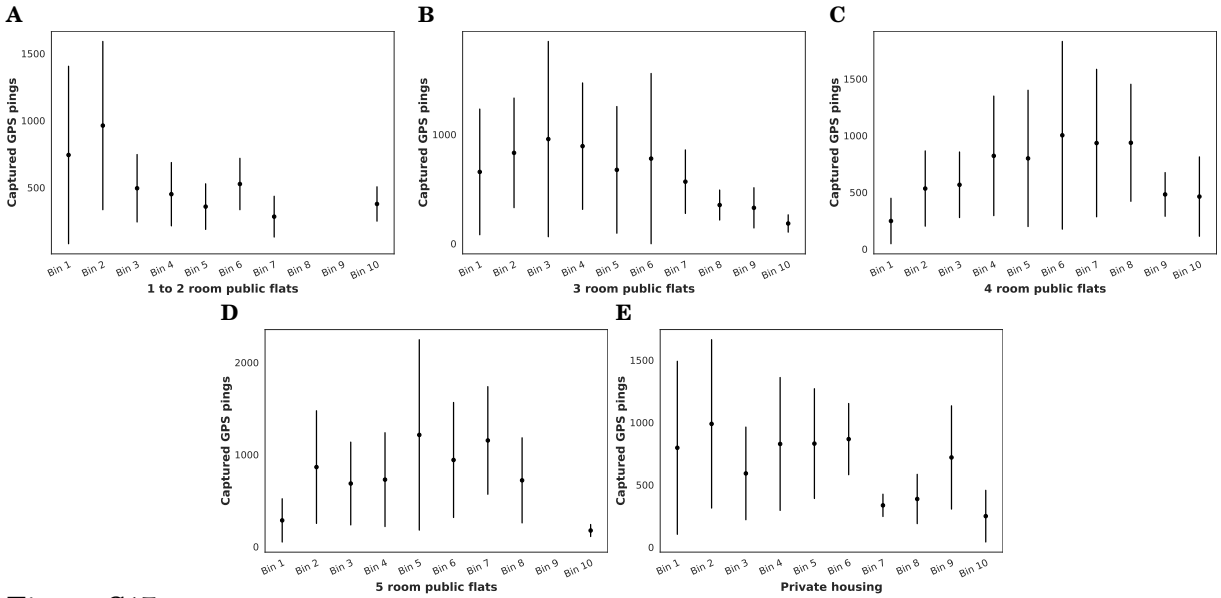

Figure S17. Variation in GPS pings by house type (origin neighborhoods). Horizontal axis is day of the week. Vertical axis is the number of captured GPS pings by neighborhood-date. Vertical bars are standard deviations.

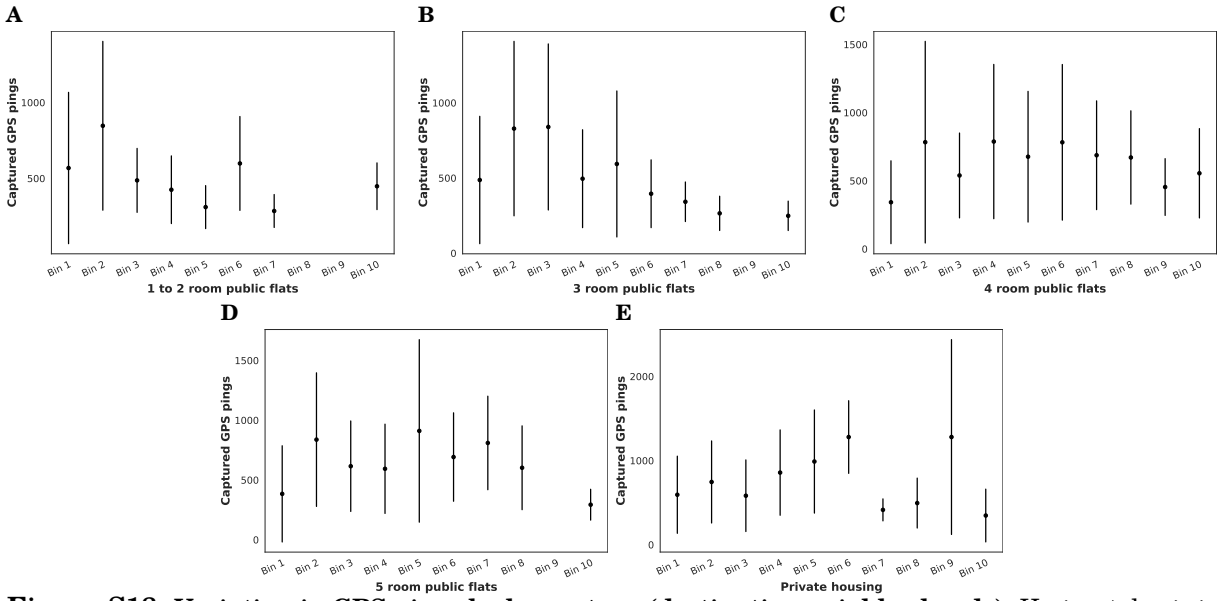

Figure S18. Variation in GPS pings by house type (destination neighborhoods). Horizontal axis is day of the week. Vertical axis is the number of captured GPS pings by neighborhood-date.

G. Representativeness of GPS pings by other characteristics

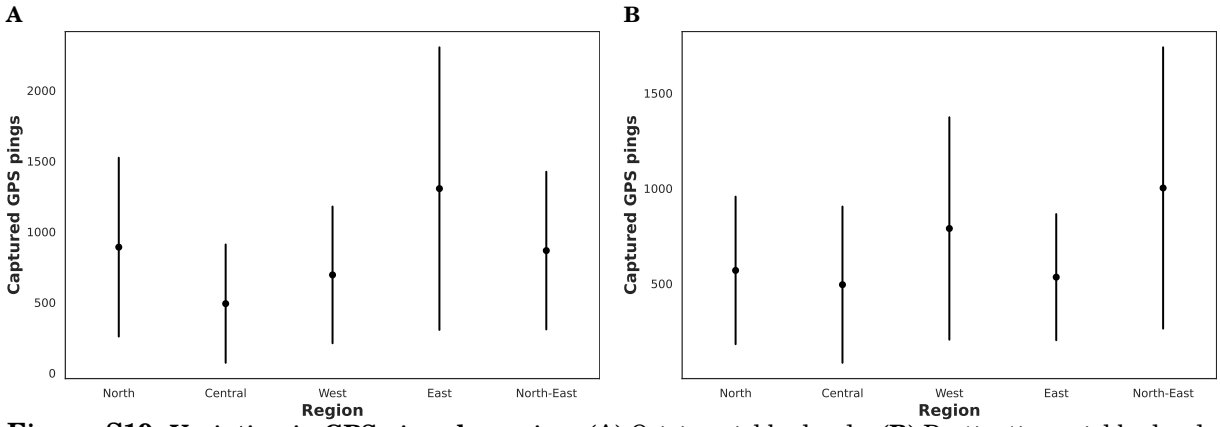

Figure S19. Variation in GPS pings by region. **(A)** Origin neighborhoods. **(B)** Destination neighborhoods. Horizontal axis is for the five regions of administrative boundaries. Vertical axis is the number of captured GPS pings by neighborhood-date. Vertical bars are standard deviations.

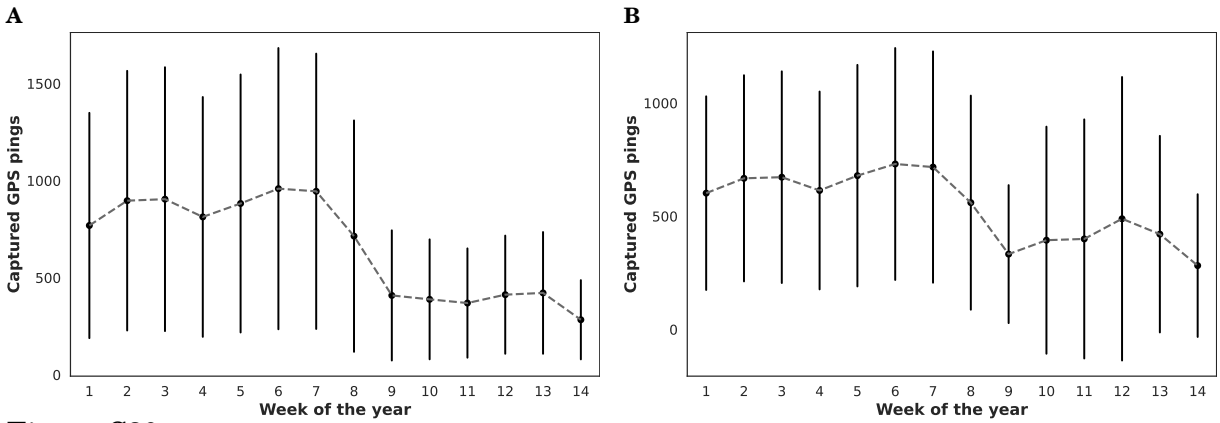

Figure S20. Variation in GPS pings by week of the year. **(A)** Origin neighborhoods. **(B)** Destination neighborhoods. Horizontal axis is the week of the year. Vertical axis is the number of captured GPS pings by neighborhood-date. Vertical bars are standard deviations.

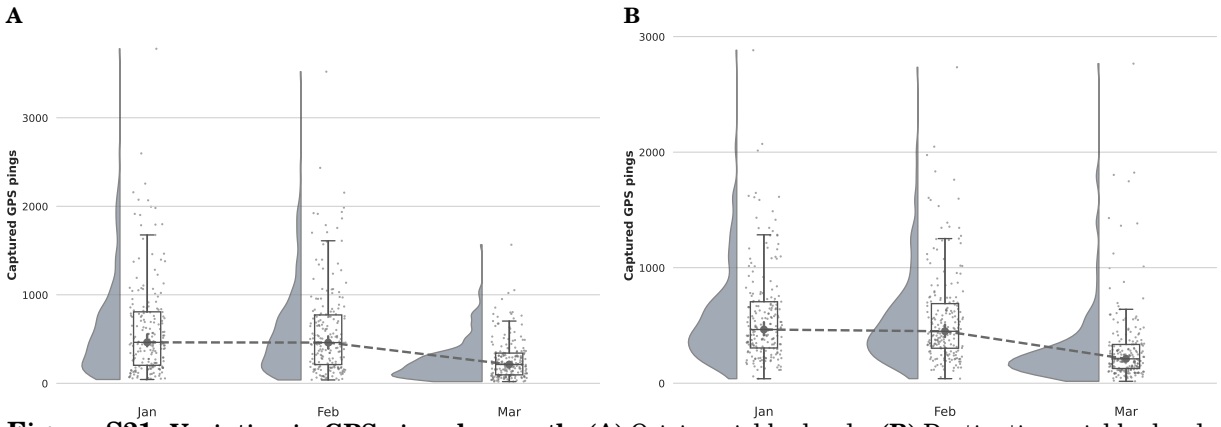

Figure S21. Variation in GPS pings by month. **(A)** Origin neighborhoods. **(B)** Destination neighborhoods. Horizontal axis is month. Vertical axis is the number of captured GPS pings by neighborhood-date.

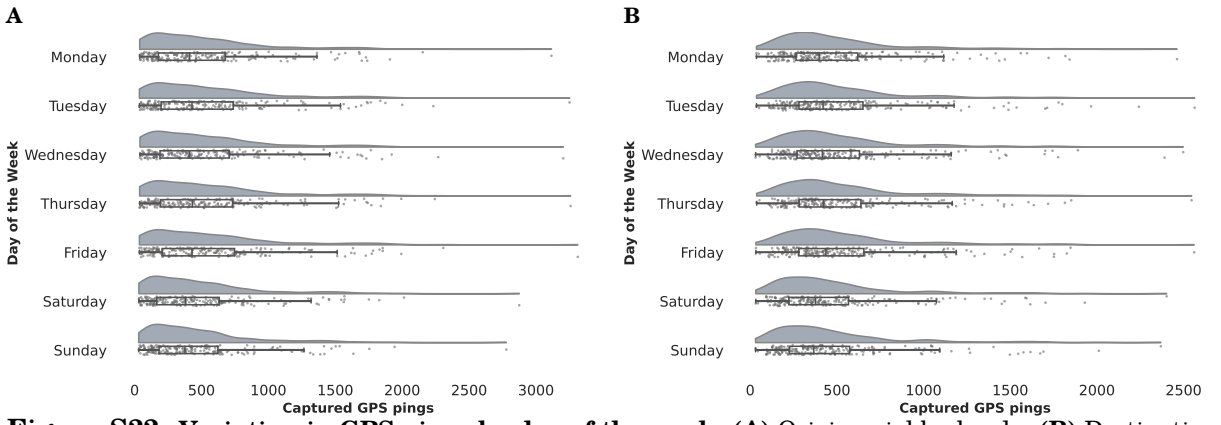

Figure S22. Variation in GPS pings by day of the week. **(A)** Origin neighborhoods. **(B)** Destination neighborhoods. Horizontal axis is day of the week. Vertical axis is the number of captured GPS pings by neighborhood-date.

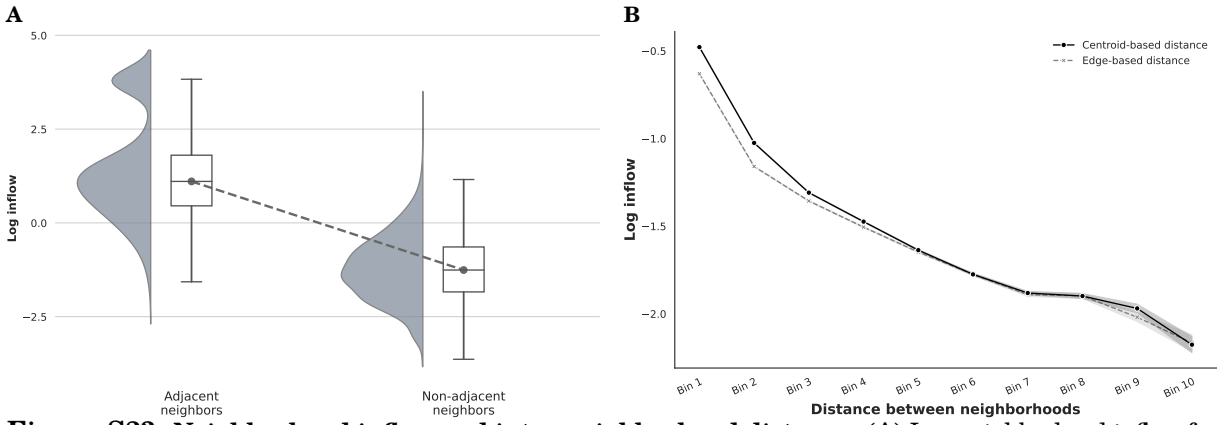

Figure S23. Neighborhood inflow and inter-neighborhood distance. **(A)** Log neighborhood inflow for adjacent vs non-adjacent neighborhoods. **(B)** Log neighborhood inflow and distance for non-adjacent neighborhoods. Shaded areas are the 95% confidence intervals constructed using 10,000 bootstraps.

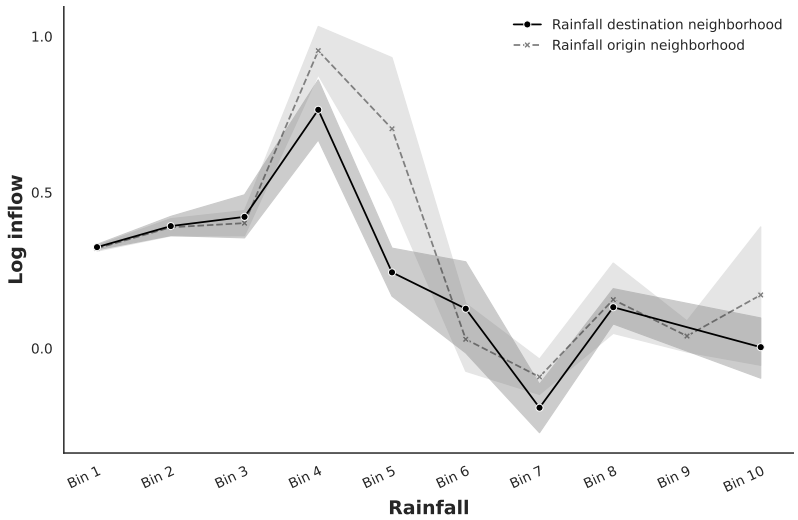

Figure S24. Neighborhood inflow and precipitation. Solid black line is the level of precipitation in the destination neighborhood. Dashed gray line is the level of precipitation in the origin neighborhood. Corresponding shaded areas are the 95% confidence intervals constructed using 10,000 bootstraps.

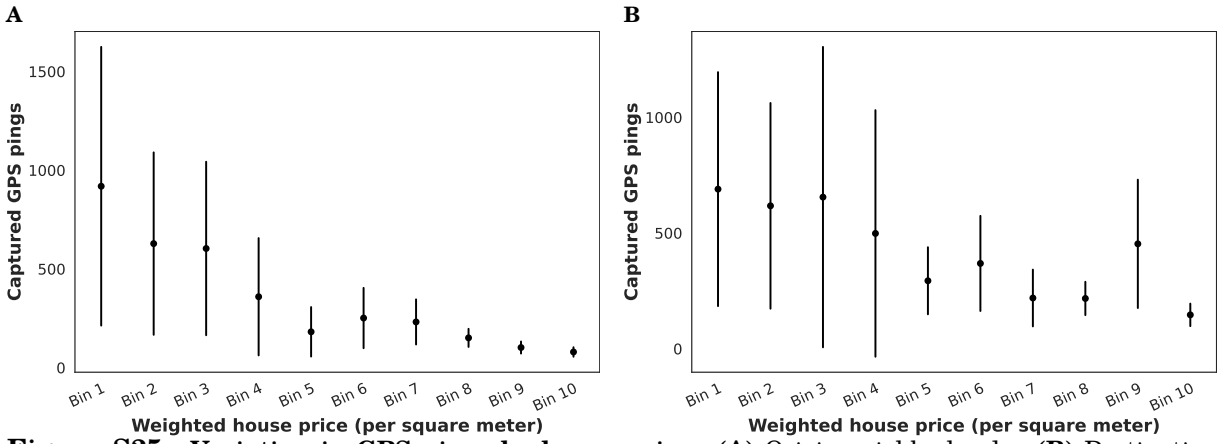

Figure S25. Variation in GPS pings by house price. **(A)** Origin neighborhoods. **(B)** Destination neighborhoods. House price is the weighted price per square meters determined by the proportion of neighborhood residents who reside in public vs private housing. Vertical bars are standard deviations.

H. Neighborhood Flows During Chinese New Year

This analysis tests whether intergenerational mobility can be detected, to the extent that the neighborhood visiting patterns reverses during a key period where younger individuals and families disproportionately travel to visit their elders, including parents. Specifically, we pivot the analysis on the two days from 25–26 Jan for the Chinese New Year holidays where Chinese individuals visit their elders.

Table S18. Effect of wealth levels on movement.

	Dependent variable: Log neighborhood visits				
Independent variable	**1**	**2**	**3**	**4**	**5**
Wealth in origin	-0.0026^{*}	-0.0028^{*}	-0.0029^{*}	-0.0029^{*}	-0.0029^{*}
	(0.0014)	(0.0016)	(0.0016)	(0.0016)	(0.0016)
Wealth in destination	0.0046^{***}	0.0054^{***}	0.0056^{***}	0.0056^{***}	0.0058^{***}
	(0.0010)	(0.0012)	(0.0013)	(0.0013)	(0.0013)
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
R^2	0.7334	0.7353	0.7355	0.7356	0.7368
Days	91	91	91	91	91
Clusters	1,439	1,434	1,434	1,434	1,434
Observations	1,321,467	1,228,385	1,228,385	1,228,385	1,228,385

Wealth levels are measured by the price per square meters (in thousands) of houses weighted by the share of public and private residence in the neighborhood population. The model in column (1) includes neighborhood land area, population density (both by census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Standard errors are clustered at the origin-by-destination census planning area level. *** Significant at the 1 per cent level. ** Significant at the 5 per cent level. * Significant at the 10 per cent level.

In Table S18, we find that neighborhood visits is highly correlated with wealth levels. In origin neighborhoods, higher wealth is correlated with lower outgoing flows, while in destination neighborhoods, higher wealth is‘ correlated with higher incoming flows.

If there is intergenerational mobility, then the estimated effects from Table S18 should reverse during the Chinese New Year period where the young visit their elder relatives. Specifically, during the Chinese New Year period, origin neighborhoods with higher wealth should be linked to higher outflows while destination neighborhoods with higher wealth should be linked to lower inflows. We model this reversal by interacting a Chinese New Year indicator variable with the origin and destination wealth levels, with the interactions allowing for the differential effects during Chinese New Year discussed above.

From Table S19, the coefficients of the interactions are all statistically indistinguishable from zero, suggesting that our sample is unable to capture the hypothesized movement patterns during the Chinese New Year period. This could be because of a lack of statistical power or because of non-linearities.

Table S19. Movement patterns during CNY.

	Dependent variable: Log neighborhood visits				
Independent variable	**1**	**2**	**3**	**4**	**5**
Wealth in origin	-0.0027^{*}	-0.0028^{*}	-0.0029^{*}	-0.0029^{*}	-0.0029^{*}
	(0.0014)	(0.0016)	(0.0016)	(0.0016)	(0.0016)
Wealth in destination	0.0047^{***}	0.0054^{***}	0.0056^{***}	0.0056^{***}	0.0057^{***}
	(0.0010)	(0.0012)	(0.0013)	(0.0013)	(0.0013)
Wealth in origin $\times \mathbb{1}^{\text{CNY}}$	0.0005	0.0001	0.0001	0.0003	0.0003
	(0.0012)	(0.0014)	(0.0014)	(0.0014)	(0.0014)
Wealth in destination $\times \mathbb{1}^{\text{CNY}}$	-0.0015	0.0006	0.0007	0.0007	0.0012
	(0.0009)	(0.0011)	(0.0011)	(0.0011)	(0.0012)
H_a : (Wealth in origin $\times \mathbb{1}^{\text{CNY}}$) < 0, p-val	.67	.519	.516	.581	.576
H_a : (Wealth in dest. $\times \mathbb{1}^{\text{CNY}}$) > 0, p-val	.946	.31	.268	.264	.165
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
R^2	0.7334	0.7353	0.7355	0.7356	0.7368
Days	91	91	91	91	91
Clusters	1,439	1,434	1,434	1,434	1,434
Observations	1,321,467	1,228,385	1,228,385	1,228,385	1,228,385

Table S20. Coefficients estimated from Equation (3), with the poverty measures fully interacted with an indicator variable for the Chinese New Year period (24–27 Jan 2020):

$$\log(\text{inflow})_{odt} = \alpha + \beta_d P_d + \beta_o P_o + \gamma_d (P_d \times \mathbb{1}^{\text{CNY}}) + \gamma_o (P_o \times \mathbb{1}^{\text{CNY}}) + \Gamma_t X_{odt} + \varepsilon_{odt},$$

where $\mathbb{1}^{\text{CNY}}$ is an indicator variable for the Chinese New Year period (25–26 Jan 2020). Wealth levels is the price per square meters of houses weighted by the share of public and private residence in the neighborhood population. The model in column (1) includes neighborhood land area, population density (both by census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Table also reports the one-sided tests for the hypothesized interaction effect. Standard errors are clustered at the origin-by-destination census planning area level. *** Significant at the 1 per cent level. ** Significant at the 5 per cent level. * Significant at the 10 per cent level.