Inequalities in Public Transport Accessibility over a Decade: Through the Spatial Distribution of Ethnicity, Age, and Socioeconomic Status

A Case Study in Greater London Area

(Analysis Part Only)

Yuxin Nie

This dissertation is submitted in part requirement for the degree of MSc in the Centre for Advanced Spatial Analysis, Bartlett Faculty of the Built Environment, UCL.

Abstract

Cross-sectional studies have indicated spatial inequalities in public transport accessibility within London, where low-skilled, low-income groups often experience limited accessibility, hindering their access to urban services and opportunities. However, how accessibility to public transport is distributed by demographic groups and how it changed over time have not been studied. This study examined the potential unequal distribution of public transport accessibility with a focus on demographic groups defined by ethnicity, age, and socioeconomic status by cross-sectional models over the past decade, at the LSOA level in the Greater London Area. After accounting for geographical features, car ownership, population density, and spatial autocorrelation in spatial lag models, the disparities for ethnicity were found, as the mixed and other ethnic groups were more disadvantaged both in 2011 and 2021, while the Asian ethnic groups had a more advantaged position. Income also played a role, as wealthier groups tended to have better access to public transport; however, these privileges decreased throughout the decade. The accessibility advantage of the middle-aged and older groups in 2011 diminished significantly by 2021. This was replaced by the median low-level age group, which had the most prominent advantage in tube accessibility. The research aims to inform policymakers on addressing disparities in public transport, optimising accessibility, and developing a fairer and more inclusive urban environment.

Chapter 3: Methodology

This chapter introduces the methodology of this paper based on the review and summary of the literature in chapter 2, including three aspects: 3.1 study area, 3.2 data, 3.3 Methodology Overview. the study area part introduces the spatial scope and spatial units of concern of the study, the data includes Access Index (from PTAL); Neighbourhood demographics; Geographical factors; Potential confounders; Methodology includes spatial dependence analysis, model selection, spatial regression models, and spatial dependence analysis. The methodology includes spatial dependence analysis, model selection, spatial regression models, and model evaluation.

3.1 Study Area

3.1.1 Study Boundary

This study analyses public transport accessibility and demographic characteristics in terms of spatial distribution within the Greater London Area (GLA). Within the GLA, the tube and buses are the most commonly used means of transport by city dwellers to commute within the city, radiating outward from the city's central area.

3.1.2 Spatial Unit

This paper uses the second smallest unit of the census, known as Lower Layer Super Output Areas (LSOAs), which consist of between 400 and 1,200 households or 1,000 to 3,000 individuals. New LSOAs have been established in 2021 census through the merging or division of the 2011 LSOAs due to changes in population (Office for National Statistics, 2021). To ensure consistency, this paper converts the 4994 LSOAs in the 2021 census back to the 4835 LSOAs in 2011.

3.2 Data

To maintain interpretability, this study refrained from applying any transformations to the data. In order to make comparisons across years and different regression models, the demographic features (section 3.2.2) are divided into five quintiles based on a 20% percentage split.

3.2.1 Accessibility

This study use Access Index which was extracted from the Public Transport Accessibility Levels (PTAL) dataset (2010 and 2023) from Transport for London (TFL). Each area is given a PTAL level score that ranges from 0 (very poor access) to 6b (excellent access) based on an underlying Access Index (AI) value, which is continuous measure ranging from 0 to over 100. Here, instead of using PTAL categories, we chose the continuous AI value as our measurement of accessibility of public transportation in Greater London Area. The calculation is based on the service access points and transportation route and service frequency. Below is a summary of formulas from the TFL connectivity guide (Transport for London and Mayor of London, 2015).

$$\text{AWT (average waiting time)}_{\text{(mins)}} = 0.5*(60/\text{frequency}) + \epsilon \qquad (1)$$

$$\text{EDF (equivalent doorstep frequency)} = 0.5*(60/\ (\text{walk time}_{\text{(mins)}} + \text{AWT})) \qquad (2)$$

$$\text{AI (access index)} = \max\text{(EDF)} + 0.5*\text{sum (other EDFs)} \qquad (3)$$

$$\text{AI total} = \text{sum (AI bus + AI rail + AI tube + AI tram)} \qquad (4)$$

This study computed an average AI value for each LSOA based on the original 5m grid published by TFL, in order to combine with other demographic variables. By merging the datasets for 2010 and 2023, covering a span of 12 years, the change of data shows (1) the opening of new stations or the closure of old ones, and (2) the increase or decrease in the frequency of public transport services in terms of overall accessibility.

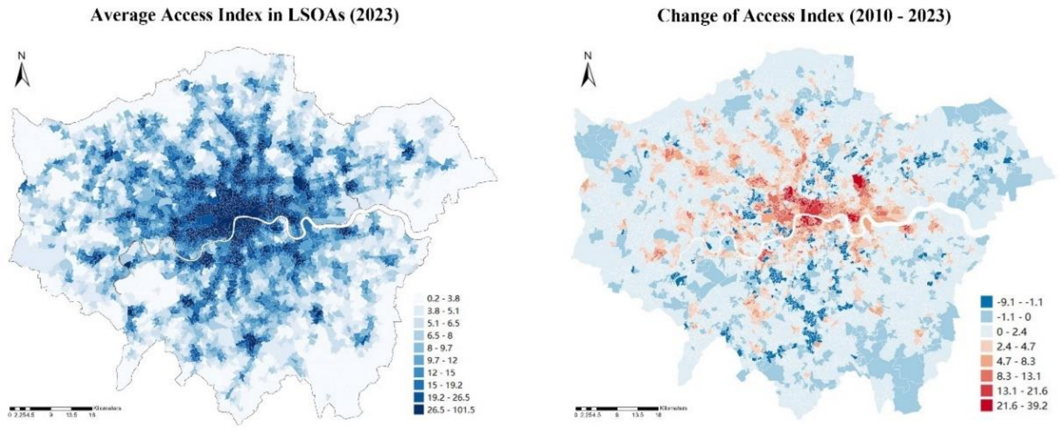


Figure 3.2.1. Spatial distribution of average Access Index at the LSOA level across Greater London Area for 2023; Spatial distribution of change of Access Index from 2010 to 2023 (data source: TFL).

In Figure 2, the access Index shows a decline in some areas due to a number of bus routes, stops and frequency of departures being downgraded or even temporarily cancelled from the start of the 2020 Covid outbreak. Many services have not been fully restored to their original levels until early 2023, when the data is calculated.

3.2.2 Neighbourhood demographics

Ethnicity

Race and ethnicity are from UK census data in 2011 and 2021 (Office for National Statistics, 2022). This study extract LSOAs within GLA and divided 19 types of ethnicity into 5 categories based on the classification of UK government and ethnicity question of The Office for National Statistics (GOV.UK, 2021), and calculate the percentage of each ethnic group within LSOAs. The ethnicity category used in this study is:

- Asian: Bangladeshi, Chinese, Indian, Pakistani, Other Asian;
- Black: African, Caribbean, Other Black;
- Mixed: White and Black Caribbean, White and Black African, White and Asian, Other Mixed;

- White: English/Welsh/Scottish/Northern Irish/British, Irish, Gypsy or Irish Traveller, Roma (only 2021), Other White;

- Other: Arab, Any other ethnic group

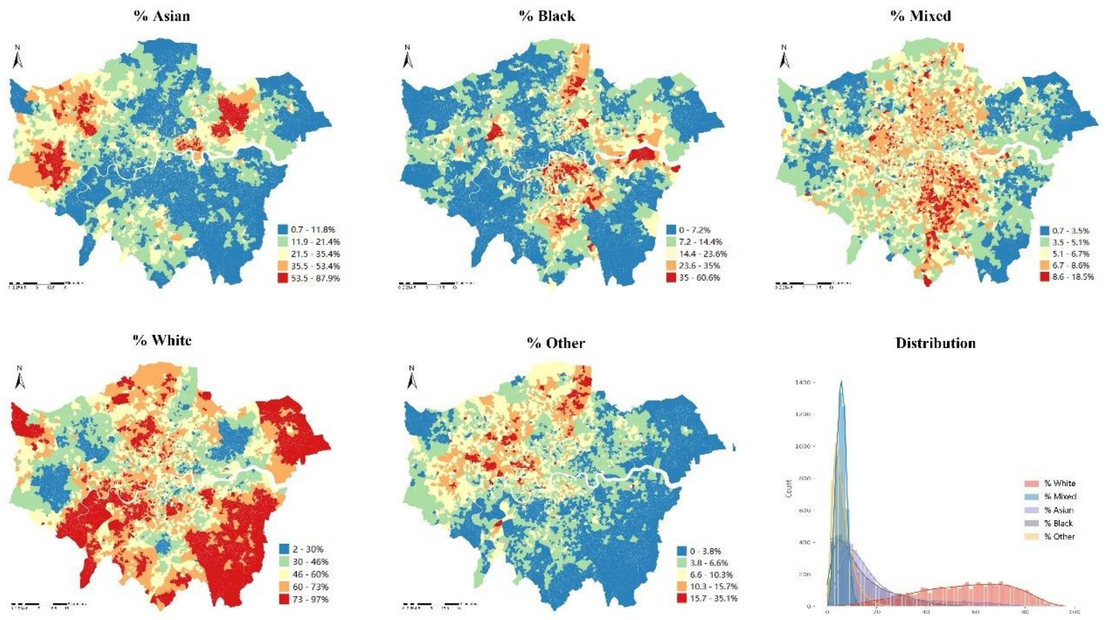


Figure 1.2.2. Spatial distribution of percentages for Asian, Black, Mixed, White, and Other ethnic groups at the LSOA level in 2021. Colour classes have been determined by Fisher-Jenks natural breaks.

Income

To research on different income groups, this study use the rank of income domain in the Index of Multiple Deprivation dataset in 2010 and 2019 (GOV.UK, 2011) (GOV.UK, 2019), and divided London LSOAs into 5 quantiles to represent low (Q1), median low (Q2), median (Q3), median high (Q4) and high income groups (Q4) (Klompmaker *et al.*, 2023).

Age

The age data is also obtained from the census data in terms of LSOA (Office for National Statistics, 2023a). In this study, the LSOAs are sorted by age according to the median age, and the mean age is compared if the median value is the same. In this way, the LSOAs are classified into five quintiles based on aging, ranging from the lowest to the highest: Q1, Q2, Q3, Q4, and Q5, respectively.

3.2.3 Geographical Factors

Opportunity Areas (OA)

Opportunity Areas are significant sites with development potential for new housing, commercial and infrastructure development, which are connected to existing or potential public transport improvements (Greater London Authority, 2021). There are 47 OAs identified by the London Plan including “adopted”, “emerging”, and “boundary to be define” types. This study uses the adopted boundary of OA and extracts LSOAs that intersect with boundary, defining them as "LSOAs benefited by OA."

Inner or Outer London

In academic research on transport accessibility studies, single-centre cities are often studied in different circles from inside to outside. For example, (Kawabata, 2009) divided Boston and San Francisco into "central city, inner suburbs and outer suburbs" to make comparison between these three regions. The study by (Smith *et al.*, 2020) (Smith and Barros, 2021) divides London Metropolitan Region into "inner Greater London Area (GLA), outer GLA and outer metropolitan area", in order to take into account all the traffic volumes and demands of daily commuters travelling in and out of London.

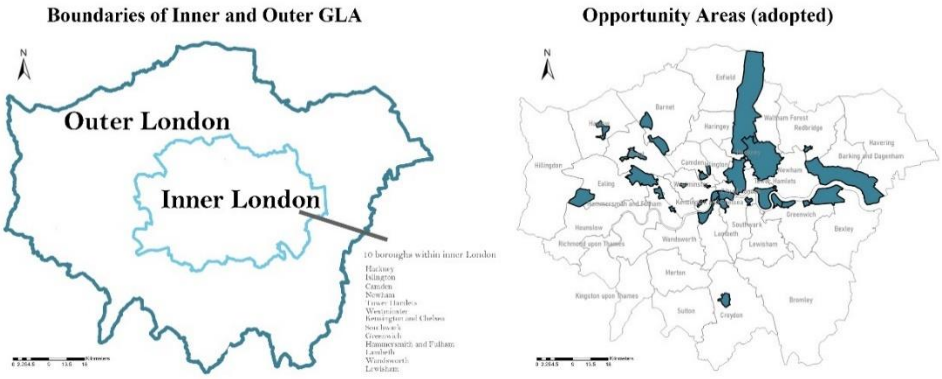


Figure 3.2.3. Boundaries of inner London, outer London of Greater London Area (GLA); Opportunity Areas (adopted ones).

Based on these ideas, this paper divides the GLA into Inner London and Outer London to provide an in-depth comparison to capture transit needs of daily commuters travelling within London.

3.2.4 potential confounders

Population density

Some literatures found a significant correlation between density distribution and accessibility (Gultom *et al.*, 2022). In order to control for the effect of this factor in this study, the number of residents counted in the census data was chosen to calculate the population density of each LSOA (Office for National Statistics, 2022).

Car ownership

Car ownership has been identified as a significant factor influencing the demand for public transport (Holmgren, 2007). On the other hand, research has also shown that affordable and sufficient public transport can deter car ownership (Cullinane, 2002). Temporal analysis reveals a complex interrelationship between the two (Holmgren, 2020). In London, data from the London Travel Demand Survey (LTDS) for the years 2005 to 2011 indicates that people living in areas further from public transport are more likely to own a car (Transport for London, no date).

In this context, if the model is constructed to consider only the relationship between ethnicity and public transport, the accessibility advantage enjoyed by car owners is overlooked. In other words, the fact that public transport accessibility is greater in the city center cannot be solely explained by demographic features such as ethnicity and income. Thus, car ownership (percentage of households with car/van in LSOAs) is considered in this study as the confounding variable, extracted from the 2021 census LSOA housing data (Office for National Statistics, 2023b).

3.2.5 Data summary

Overall, this study divides the data to 2010s variables and 2020s variables for illustration and model construction.

Table 3.2.5. Statistics of continuous variables for both 2010s and 2020s. Values rounded to 2 figures. All variables are calculated at LSOA level (n = 4835).

Variable		Year	Description	Mean	S.D.	Min	Max
Access Index	Total	2010	Average access index at LSOA levels, divided by transport mode	11.39	9.82	0.10	88.74
	Tube			2.23	4.18	0.00	32.53
	Bus			7.62	5.36	0.02	44.77
	Rail			1.54	2.83	0.00	36.26
Access Index	Total	2023		13.24	11.65	0.18	101.51
	Tube			2.98	5.62	0.00	44.86
	Bus			8.24	5.18	0.04	44.64
	Rail			1.94	3.72	0.00	40.74
	Tram			0.08	0.51	0.00	7.27
% Asian		2011	Percentages of ethnic groups at LSOA levels, grouped by sub-categories	17.94	16.17	0.75	86.90
		2021		19.66	16.06	0.65	87.88
% Black		2011		13.07	11.21	0.13	63.65
		2021		13.20	10.70	0.00	60.60
% Mixed		2011		4.92	1.95	0.61	14.39
		2021		5.74	2.03	0.68	18.51
% White		2011		60.69	20.37	3.54	98.16
		2021		55.15	18.75	1.94	96.62
% Other		2011		3.38	2.82	0.00	36.56
		2021		6.25	3.99	0.00	35.14
Income Deprivation Score		2010	Income deprivation score, only meaningful when making rankings	0.19	0.11	0.01	0.58
		2019		0.14	0.08	0.01	0.44
Age		2011	Median age at LSOA level	34.33	5.01	17.00	57.00
		2021		36.45	4.84	17.00	62.00
Population Density		2011	Population per km2 at LSOA levels	9647.63	6652.18	115.89	93957.51
		2021		10002.34	6423.85	119.37	70389.81
Car Ownership		2011		59.96	18.52	13.75	97.32

	2021	Percentage of vhousehold with one or above cars or vans	60.06	18.84	13.27	96.22

To provide a detailed result for each of the ethnic group and prevent omitting one of the whole ethnicity, this study divided each ethnic categories into five equal quintiles (from low to high: Q1, Q2,Q3,Q4,Q5), and created new values based on one-hot encoding. Thus, in the analysis models, every Q1s were removed as the reference group for other higher levels of groups.

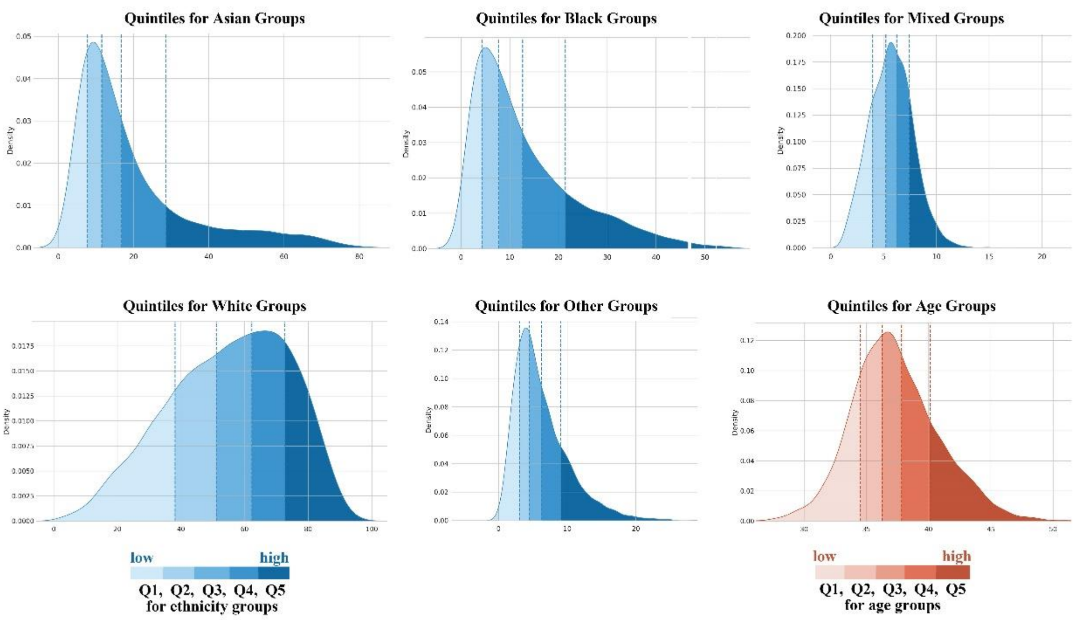


Figure 3.2.5. Distribution of quintiles of five ethnicity groups and age in 2021. All divided by equal percentages (data source: Census 2021).

3.3 Methodology Overview

Main goal of this study is to study at both cross-sectional associations and change. Specifically, our objectives were to estimate the associations between:

(i) 2011 LSOA level demographic characteristics and 2010 LSOA level Access Index;

(ii) 2021 LSOA level demographic characteristics and 2023 LSOA level Access Index;

(iii) LSOA level change in Access Index between 2023 and 2011 and 2011 LSOA level demographic characteristics.

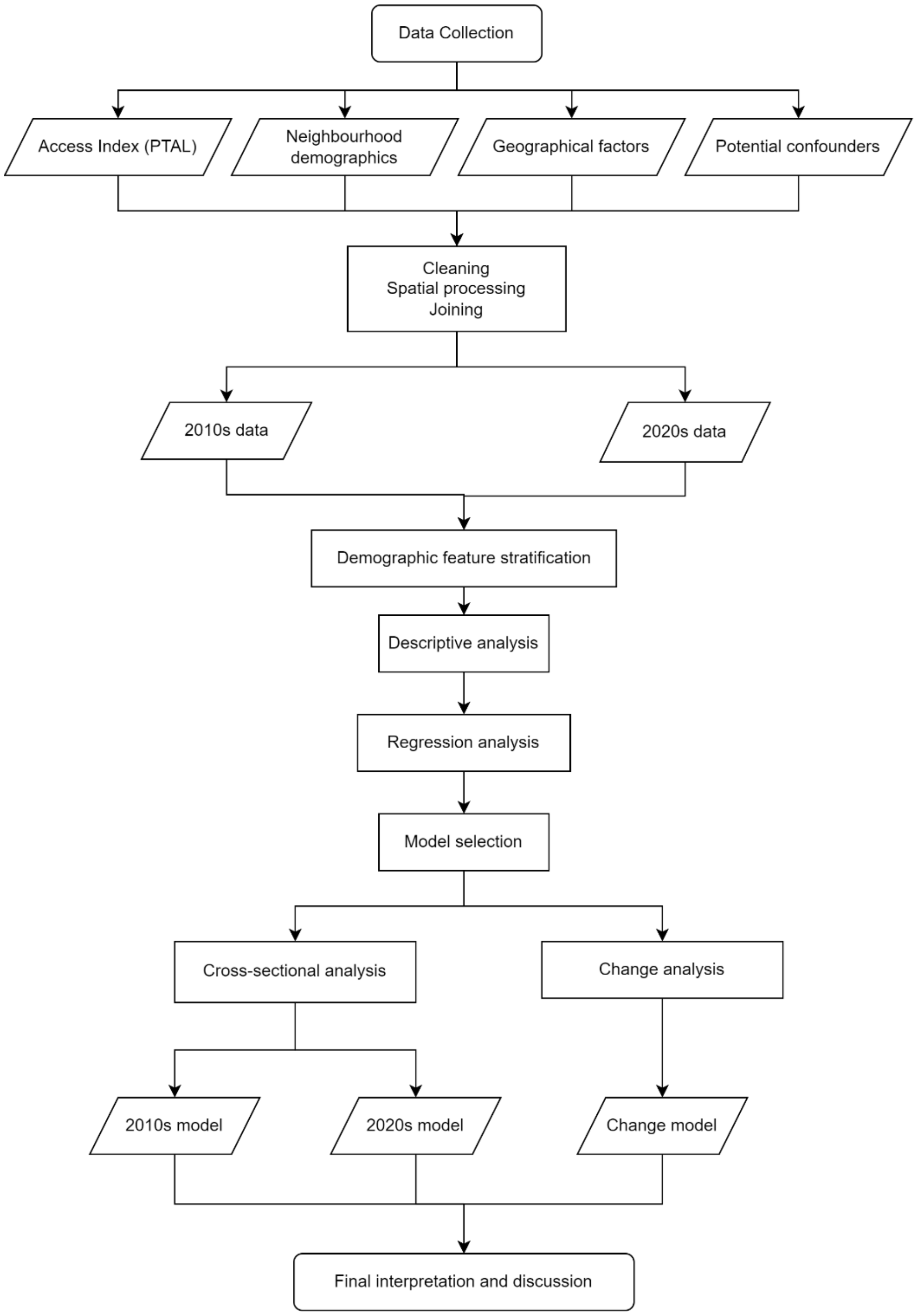


Figure 3.3. Workflow

3.3.1 Spatial dependence

It is reasonable to suspect that there is a high probability of spatial dependence due to the small size of the spatial unit LSOA used in this analysis, as well as the monocentric distribution of public transportation. LSOAs next to areas with stations or stops have higher chance of having greater accessibility, because the public transportation relies on continuous line connections. In other words, the public transport accessibility itself, as a response variable, may have an impact on the public transport accessibility of the surrounding area. Also, the possibility that the demographic features of the neighboring LSOAs have impact on a particular LSOA could not be excluded. Therefore, it is necessary to introduce appropriate spatial terms in regression models.

Before performing spatial regression, it is usually necessary to conduct a full spatial autocorrelation analysis on the dependent variable (Jiang, 2016). Spatial autocorrelation (Moran's I) serves as the primary method for analyzing spatial independence and measuring the degree of interdependence between the data at a given location and those in the surrounding area.

$$I = \frac{n}{\sum_i \sum_j W_{ij}} \frac{\sum_i \sum_j W_{ij}(Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_i (Y_i - \bar{Y})^2} \tag{5}$$

In formula (5) (Moran, 1950), Y_i and Y_j represent the observed values of spatial units i and j, respectively. $\bar{Y}$ denotes the mean value of the observed Y values, n represents the number of spatial units, and W stands for the spatial weight matrix, typically subjected to row normalisation. The value of I falls within the range of [-1, 1]. A significant Moran's I indicates the presence of spatial autocorrelation (Moran, 1950).

3.3.2 Model selection

The significance of Moran’s I does not determine whether a spatial lag model or a spatial error model should be chosen. The empirical analysis process proposed by Anselin et al. (1996) serves as the most commonly used basis for model selection. The overall process is as follows: starting with the OLS regression, followed by the

Lagrange multiplier (LM) test and the robust LM test for the residuals, and then making a decision to use either the spatial error model or the spatial lag model based on whether and which test is significant.

3.3.3 Spatial regression analysis

After the test mentioned above, the initial step in spatial regression involves building the weighted matrix. Given the irregular shape of the London LSOAs, some of the common approaches such as rook or queen rules of selecting spatial neighbors are not meaningful. Therefore, this study chose k nearest neighbors (KNN) and selected the 8 closest neighbors in weighted matrix, to capture the local impact of accessibility (Padeiro, Louro and Da Costa, 2019).

Spatial Lag model (SLM)

$$Y_i = \alpha + \rho \sum_j \omega_{ij} Y_i + \beta X_i + \epsilon_i \tag{6}$$

In formula (6), Y_i represents average Access Index value for each LSOA i. All the independent variables are included in X. The spatial weighted matrix ω_{ij}, based on the k nearest neighbours, captures the impact of Access Index in LSOA i on the neighbouring LSOA j. The symbol ρ denotes the parameter to be estimated for the spatial lag term of the dependent variable, which is also known as the spatial autoregressive coefficient.

Spatial Error Model (SEM)

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

$$\epsilon_i = \lambda \sum_j \omega_{ij} \epsilon_i + \mu_i \tag{7}$$

In formula (7), λ represents the coefficient to be estimated for the spatial autocorrelation error term, which is also known as the spatial autocorrelation coefficient. μ_i denotes the error term. The spatial weighted matrix ω_{ij} is based on the k nearest neighbors, capturing the impact of the error in LSOA i on the neighboring LSOA j. α donates as the constant term, while β stands for the parameter to be estimated, and ϵ_i represents the random error term.

3.3.4 Model Evaluation

In addition to the R-squared and adjusted R-squared, which reflect the overall explanatory power of the fitted model, this study employs three statistics proposed by Anselin (1988) for comparing the performance of spatial models:

- Log Likelihood;
- Akaike Information Criterion (AIC);
- Schwarz Criterion.

Among these, a larger value of the Log Likelihood statistic indicates a better model fit, whereas smaller values of the AIC and Schwarz Criterion statistics indicate better model fit. It's noticeable that the value of R^2 and adjusted R^2 should not be directly compared between OLS model and the spatial lag model as they differ in fit goodness (Anselin, 1988).

Chapter 4: Result

This chapter uses the data and methods introduced in Chapter 3 to present the results of both descriptive and quantitative analyses. The results consist of three main parts: 4.1 Descriptive Results. This part demonstrates the initial bivariate analysis as a starting point; 4.2 Cross-sectional Results. This part is subdivided into 4.2.1 Spatial Independence and Regression Model Selection, and 4.2.2 Analysis of Regression Results; 4.3 Change Data Analysis. This section explores the relationship between changes in accessibility values from 2010 to 2023 and the base-year (2011) demographic characteristics.

Section 4.2 aims to showcase research objectives (1) and (2), while Section 4.3 aims to address research objective (3) as mentioned in Chapter 3.3.

4.1 Descriptive analysis

This study begins with a bivariate analysis of the main focus: ethnicity and accessibility, without considering other variables.

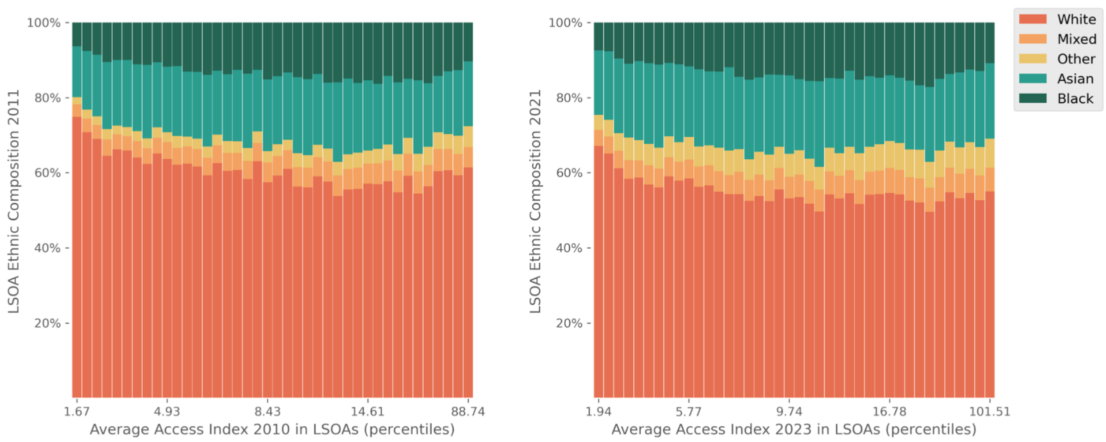


Figure 4.1. Average ethnicity composition by levels of Average Access Index in LSOAs (n = 4,835) in Greater London Area. The x-axis of each plot was divided by 2.5 percentile.

The Figure 4.1 shows that in both 2011 and 2021, the White group has the largest share among all races, with a concentration between 50% and 60%. This is followed by the Asian group, and then the Black group. Comparing the two years, the White racial share declines, while the mixed and other groups, as well as the Asian group, increase; the Black share does not change significantly. Intuitively looking at the figures, from left to right, the higher the accessibility, the lower the proportion of the White groups, while the proportion of the Black group increases slightly. The Asian group, the other group, and the mixed group increase in proportion as accessibility improves, but not significantly. This is due to the fact that the White group is more distributed in the suburban areas of outer London, in line with the distribution of the higher-income, car-owning groups while the public transport is mostly gathered in the centre areas. Therefore, we need to control for this type of potential covariant when analysing accessibility and ethnicity shares.

4.2 Cross-sectional analysis

4.2.1 Spatial dependence and model selection

In this study, the construction of the OLS model is used as a starting point to perform Moran's I test on the residuals of the dependent variable in the model, revealing the degree of spatial autocorrelation existing in the variables. The Lagrange multiplier (LM) test and the robust LM test are also conducted on the OLS model to select the appropriate spatial regression model based on the significance of the test results.

Table 4.2.1. Result of Moran's I, LM test, Robust LM test for OLS models; Statistic result for spatial regression models. All models were built with the Total Access Index as the response variable, and demographic characteristics and potential confounders as independent variables.

	OLS (2011)	OLS (2021)	LAG (2011)	LAG (2021)	ERROR (2011)	ERROR (2021)
Moran's I	0.474	0.567	0.121	0.161	0.643	0.720
P value	0.001	0.001	0.001	0.001	0.001	0.001
Log likelihood	-15793	-16596	-13941	-14140	-13985	-14118
Akaike info criterion (AIC)	31651	33256	27948	28346	28033	28299
Schwarz criterion	31858	33463	28161	28560	28241	28507
LM-Lag	4364 ***	5783 ***				
Robust LM-Lag	465 ***	397 ***				
LM-Error	4663 ***	6674 ***				
Robust LM-Error	764 ***	1288 ***				

The test results illustrate that the original OLS model exhibits significant spatial autocorrelation (significantly with 0.474 and 0.567). Therefore, spatial regression was selected in this study to capture the spatial effects and optimise the model's performance. The results of the LM test indicated significance for both the LM lag and the LM error tests, while in the further investigation with the Robust LM test, the results also show significance for both the lag and error terms, suggesting the potential impact of both on spatial independence. Consequently, this study constructed both models for comparison.

Comparing the Moran's I of the different regression models, it can be concluded that: (1) Only the Spatial Lag models reduces the degree of spatial autocorrelation in the original OLS model, while the Spatial Error model increased this problem. (2) When compared with the Spatial Error model, the Spatial Lag model exhibits the smallest Moran's I value, essentially overcoming the issue of spatial autocorrelation. Referring to Table 4.2.1, the LM lag value is larger than the LM error value in the LM test results, which further confirms that the spatial lag term has a greater impact on this model.

Additionally, the three metrics proposed by Anselin (1988) – Log Likelihood (lower), AIC and Schwarz Criterion (higher) – all satisfy the criteria for model optimisation. This further validates that the spatial model is better tuned to the spatial global autocorrelation problem. Combined with theoretical analyses for the spatial distribution of public transit in Chapter 3.3.1, the spatial lag model is chosen for subsequent regression analyses in this study.

4.2.2 Results of spatial regression

Cross-sectional analysis

The spatial lag models for both 2011 and 2021 are significant, indicating inequalities in access to public transit by ethnicity, income and age after accounting for population density, car ownership, the LSOA's location (inner or outer London, within or out of Opportunity Areas).

The total explained amount of the spatial lag model for 2011 is 0.83, with the spatial lag term "Weighted Access Index" contributing 27% to the explained amount. This indicates that spatial autocorrelation has a strong influence on the model of this study. Similarly, in the model of 2021, the spatial lag term contributes 35% to the total explanation of the model, which was 87%.

All Q1s were removed due to the principle of one-hot encoding. Additionally, in this study, the VIF was set to 5 to test multicollinearity, thus the white group Q5 was also removed.

Table 4.2.2. Result of spatial lag regression between total Access Index 2023 and 2021 demographic variables, adjusting for geographical features, potential confounders, and spatial lag autocorrelation term. Q1s were all removed as reference groups. Figures in the cells were regression coefficients in 2011, the results for 2021 were in the brackets.

Demographic features	Q1	Q2	Q3	Q4	Q5
Race/ethnicity					
% White	ref	-0.294 . (-0.026)	-0.112 . (0.278)	-0.262 . (-0.056)	/
% Mixed	ref	-0.537 ** (-0.371)	-0.873 *** (-0.583 **)	-0.807 *** (-0.521 *)	-0.865 *** (-0.914 ***)
% Asian	ref	0.479 * (0.273 .)	0.594 ** (0.725 ***)	1.113 *** (1.143 ***)	1.376 *** (1.220 ***)
% Black	ref	0.483 * (0.262 .)	-0.105 . (-0.423 .)	-0.089 . (-0.259 .)	-0.508 . (0.001 .)
% Other	ref	-0.430 * (-0.831 ***)	-0.491 * (-0.888 ***)	-0.693 ** (-0.907 ***)	0.911 *** (-1.321 ***)
Income	ref	1.496 *** (0.630 **)	2.365 *** (1.047 ***)	3.079 *** (1.693 ***)	3.651 *** (1.675 ***)
Age	ref	0.339 . (0.210 .)	0.619 ** (0.178 .)	0.526 * (0.022 .)	0.596 * (0.323 .)
Geographical features	0 (outside)	1 (inside)			
Opportunity area (within)	ref	0.256 . (-0.093 .)			
Outer London	1.423 *** (1.713 ***)	ref			
Adjusted variables					
Weighted	0.846 ***				

Access Index	(0.891 ***)
Population density (persons/km^2)	-0.00006 *** (-0.00006 ***)
Car ownership	-0.194 *** (-0.170 ***)
Model summary	
adjusted R2	0.83 (0.87)
p value	0.000 (0.000)

Note: Figures in the cells were regression coefficients in 2011, the results for 2021 were in the brackets. *** stands for statistical significance at a very strong level($p < 0.001$); ** stands for statistical significance at a strong level ($p < 0.01$); * stands for statistical significance ($p < 0.05$); . stands for weak statistical significance ($p < 0.1$).

a) The Association between Ethnicity and Accessibility

In this study, the total access index, Bus Access Index, and Tube Access Index are used as independent variables to construct the spatial lag model. The model's variables include neighbourhood characteristics (ethnicity, income, and age) and adjusted features (including population density, car ownership, and location). The ethnic groups are divided into five quintiles based on the level of percentage in LSOAs, including Q1 (low level), Q2 (median low level), Q3 (median level), Q4 (median high level), Q5 (high level). The coefficients in the result are only comparable to the reference Q1 level. This study compares between the relative relationships of ethnic group levels and the reference level to interpret the temporal differences.

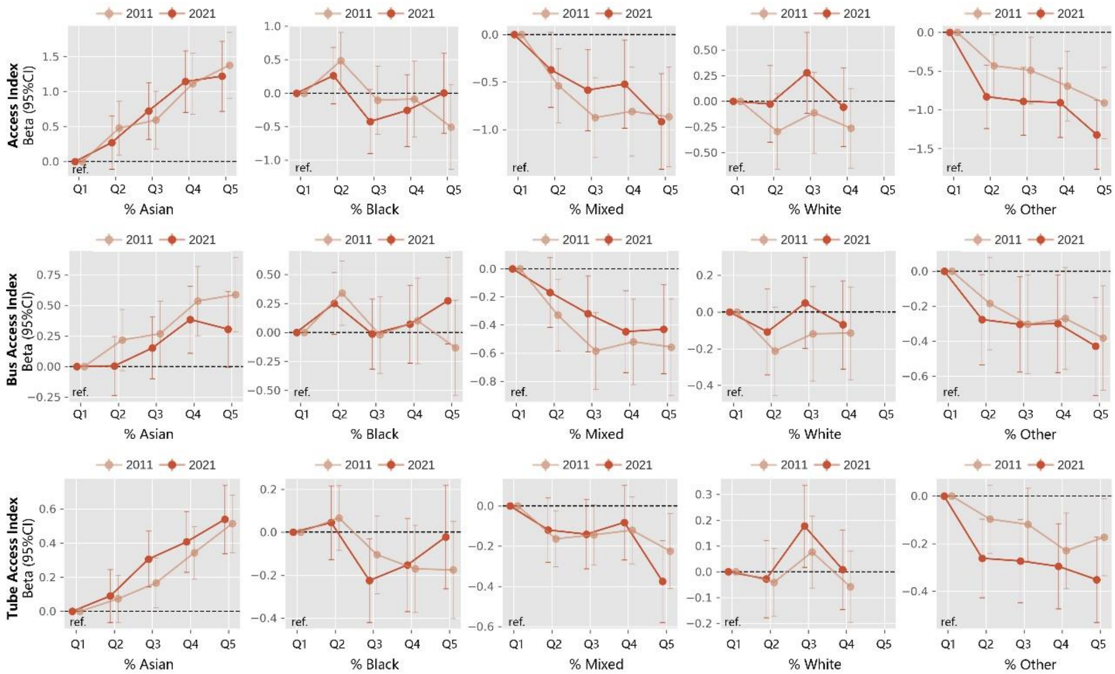


Figure 4.2.2.1. Associations of total Access Index, bus Access Index, tube Access Index in 2010 and 2023 and five quintiles of percentages of ethnic groups at LSOA level in 2011 and 2021 respectively. Percentages rose from Q1 to Q5. All coefficients were within the 95% Confidence Interval. Q1s were all removed as reference groups.

The relationship between each of the five ethnic groups and public transit accessibility varies. While in terms of different transit modes, different levels of ethnic groups demonstrate similar trends of correlation. Specifically, the Asian groups show a consistently significant positive correlation with total public transit access index, indicating the advantaged situation of Asian groups in the spatial distribution of accessibility. Notably, the total accessibility advantage becomes more pronounced as the percentage of Asian groups increases (from Q1 to Q5, with Asian% gradually rising). The LSOAs with the highest percentage of Asian groups have about 1.5 units of access index higher than the LSOAs with the lowest percentage (Q1, the reference group). A comparison between 2011 and 2021 reveals a slight reduction in the dominance of Asian groups in median low (Q2) and hight (Q5) LSOAs. Delving into the breakdown of transport modes, the correlation trend of Asian group is similar for bus and tube, ranging simultaneously from 0 to nearly 0.6. The decline in this advantage in bus accessibility for the Asian race from 2011 to 2021 is evident across all levels. Notably, the high-level (Q5) Asian share of LSOAs experiences the most significant decline in the bus accessibility advantage. This could due to (1) reduced bus service levels, and (2) significant migration within these areas.

On the other hand, the Black group presents a more complex scenario. Firstly, areas with median-low (Q2) values experience a public transport accessibility advantage in both 2011 and 2021, despite a declining one. Secondly, the negative correlations seen in Q3 and Q4 levels are mainly a result of the Black group's tube accessibility disadvantage (all negatively correlated). This disadvantage continues to grow from 2011 to 2021 in Q3. However, in the Q5 level with the highest proportion of Black residents, the enhancements in bus accessibility and tube accessibility are both notable, from -0.1 to 0.25 and -0.2 to 0 respectively. Consequently, in the overall Access Index, the Q5 segment of the Black group reflects a trend of greater benefit. However, this trend of inequality reduction does not extend to other levels of areas.

Counter-intuitively, white groups, often considered advantaged, do not possess an absolute advantage in terms of accessibility to public transport. This is demonstrated by both the total access index and the transit mode breakdown, revealing that only the median level (Q3) LSOAs hold a distinct advantage in public transport accessibility while other levels (Q2 and Q4) experience non-clear or slightly negative relationships. The correlations in 2011 got improved through the ten years.

Finally, the mixed group and the other group are examined together, considering that both have experienced disadvantages over the past decade. In terms of the mixed group, the disadvantage situation in bus accessibility has enhanced between 2011 and 2021, with all levels of districts showing less shortage. However, in the LSOAs with the highest share of the mixed group (Q5), tube accessibility has become significantly more disadvantaged, and the gap with other levels of districts further widens. In terms of other groups, unfortunately, as tube accessibility worsens significantly at all levels while bus accessibility remains more or less the same, the entire group become more disadvantaged in terms of total accessibility.

When comparing among all the ethnic groups, only Asian groups and some levels of black (Q2) and white (Q3) are advantaged with regard to the total accessibility. From 2011 to 2021, tube accessibility shows improvement in all level Asian, white and black areas, except for Q4 black level. Instead, dramatic declines happen in all areas across the other groups and Q5 level. In terms of the bus access index, both Asian, black and white groups have advantages, while this advantage declines in Asian groups and increases in white groups.

b) The Association between Income and Accessibility

Similar to the breakdown of ethnicity, this study also divides the LSOAs into five quintiles based on the income deprivation ranking. The most deprived LSOAs are categorised into Q1, representing the lowest income group. Other groups are median low groups (Q2), median groups (Q3), median high (Q4), high (Q5).

In the spatial lag model, the association between income level and accessibility is significant at all levels. There is a positive correlation between income and the total access index, and the higher the income level, the more clearly the advantage of accessibility is reflected. The LSOAs with the highest income (Q5) exhibits nearly four units of higher accessibility compared to the LSOAs with the lowest income (Q1, reference group) in 2011, however, this significant advantage dramatically went down through the decade (from 3.7 to 1.7), particularly for Q5, the highest income group, levelling to that of Q4 in 2021. The main contributor to the decline in the bus accessibility, where the highest income Q5 areas show a significant decrease (nearly drop to the half). Advantages for bus also drops for all income groups, with the approximately the same amount of decrease.

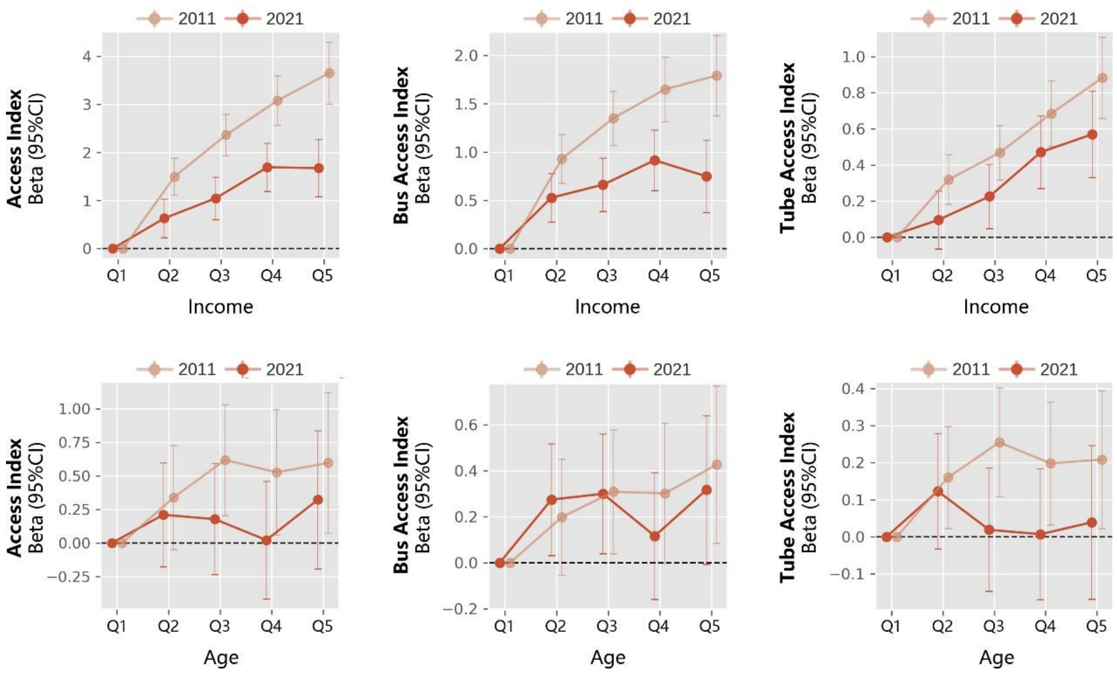


Figure 4.2.2.2. Associations of total Access Index, bus Access Index, tube Access Index in 2010 and 2023 and five quintiles of percentages of income and age groups at LSOA level in 2011 and 2021 respectively. Percentages rose from Q1 to Q5. All coefficients were within the 95% Confidence Interval.

c) The Association between Age and Accessibility

The result shows that the elderly groups are slightly benefiting from public transport accessibility but the median high aged groups do not have obvious advantage.

From 2011 to 2021, the advantage in tube accessibility went down on the aging groups (Q3, Q4 and Q5), leading to the decrease in the total correlation. The trend in bus accessibility is similar to the total value. Only groups in median low age (Q2) have gained more benefits in the past years. Overall, due to the small correlation coefficient between age level and accessibility (less than 0.5), accessibility does not play a strong role in the age distribution.

4.3 Change analysis

In order to quantify the relationship between changes in accessibility and demographic characteristics, this part uses spatial lag model to measure how demographic variables can account for changes in accessibility. Spatial lag models were constructed for the change in the total Access Index and base-year variables (2011) to investigate their relationship.

Table 4.3. Result of spatial lag regression between the change of Access Index from 2010 to2023 and 2011 demographic variables, adjusting for geographical features, potential confounders, and spatial lag autocorrelation term. Q1s were all removed as reference group.

Demographic features	Q1	Q2	Q3	Q4	Q5
Race/ethnicity					
% White	ref	0.038	0.126	-0.032	/
% Mixed	ref	-0.087	0.109	-0.107	-0.071
% Asian	ref	0.068	0.372 ***	0.322 **	0.462 ***
% Black	ref	-0.113	-0.167	-0.237	-0.180
% Other	ref	-0.123	-0.295 **	-0.168 **	-0.323 **
Income	ref	-0.200 *	-0.044	0.086	0.172
Age	ref	-0.172	-0.076	0.096	0.074
Geographical features	0 (outside)	1 (inside)			
Opportunity area (within)	ref	0.247 **			
Outer London	0.062	ref			
Adjusted variables					
Base year (Access Index 2010)	-0.023 ***				
Weighted change of Access Index	0.846 ***				
Population density (persons/km2)	0.000002				
% Car ownership	-0.032 ***				

Model summary	
adjusted R2	0.64
p value	0.000

Note: *** stands for statistical significance at a very strong level(p < 0.001); ** stands for statistical significance at a strong level (p < 0.01); * stands for statistical significance (p < 0.05); . stands for weak statistical significance (p < 0. 1).

Since the objectives of this study—public transport accessibility and demographic characteristics—are both changing during the decade from the 2010s to the 2020s, this study chooses to control for ethnicity distribution in this section. In this case, we assume that the population does not change between 2011 and 2021, and the public transport accessibility is the only variable that changes. Thus, we build a spatial lag model to measure the association between the change of the Access Index from 2010 to 2023 and the base year (2011) neighborhood demographics.

Similar to the cross-sectional model but more significant, the spatial lag term (weighted change of access) contributes to the overall model at 50%. This means that the change of the Access Index itself influences its own spatial changes more than all neighborhood features, geographic features, and neighborhood demographics combined. It explains almost twice as much of the model as all other factors. Consequently, many of the factors in the model are not significant, and their coefficients do not hold strong explanatory value.

Among all the variables, the significance of the factor related to Asian groups is more pronounced. There is more accessibility improvement in areas where Asian groups account for the median, median-high, and high levels (Q3, Q4, Q5), compared to the area with the fewest Asians, Q1. In the previous analyses, we found that in the base year, Asians have a significant advantage in accessibility, and this advantage strengthens as the percentage of Asians increases. However, the results presented in Table 4.3 show that the advantage of accessibility for areas with a large number of

Asians is further expanding. This advantage has the strongest correlation coefficient (0.462) other than the spatial autocorrelation coefficient for the LSOA with the highest level of Asian representation (Q5). Does this mean that public transport is investing further in areas that already have better accessibility? The answer is no. In the model, the effect of the base year's access index on the change in accessibility is significant, but the effect is negative. This implies that new investments in public transport are inclined to target areas that lack adequate accessibility. In this senario, the increase in dominance in areas with a high proportion of Asians can only be explained by the fact that new investment is still going to other areas with a high proportion of Asians, on top of the current dominance of Asians in public transport accessibility.

Another notable ethnic group is the one that stands in contrast to the Asian group. Additionally, other groups, which have faced significant disadvantages, also exhibit a negative correlation with accessibility enhancement. This implies that, instead of ameliorating the situation for these other ethnic groups, new public transport investments further solidify their disadvantaged positions, assuming there is no change in population distribution.

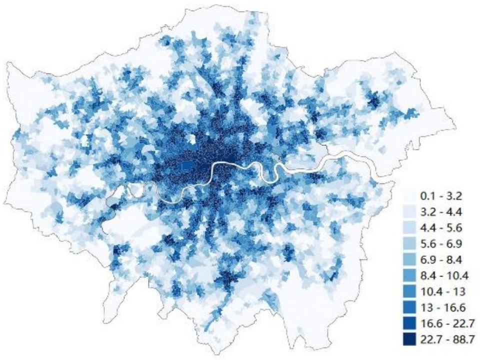
From a geographical perspective, the new public transport investment brings substantial benefits to LSOAs situated in the opportunity area. There is a 0.247-unit advantage in accessibility improvement for LSOAs within the OA, in comparison to those that are located outside it.

Another notable ethnic group is the other ethnic group, which stands in contrast to the Asian group. This group have faced significant disadvantages, and they experience a negative correlation with accessibility improvement. This implies that, under the assumption of an unchanged population distribution, new investments in public transport is potential to widen this existing negative disparity.

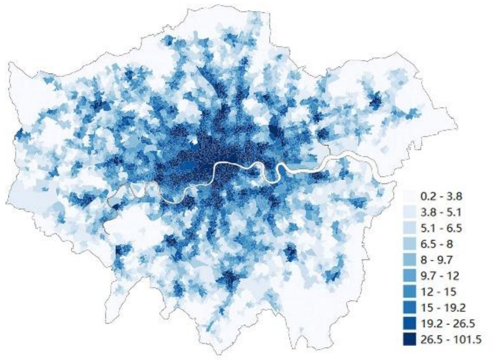
Regarding geographical impact, the new public transport investment brings significant advantages to LSOAs located within the opportunity area. These specific LSOAs witness a notable 0.247-unit improvement in accessibility, contrasting with LSOAs outside the opportunity area.

AppendiX: Spatial distributions of public transport accessibility, population density, percentage of white groups, car ownership levels, and income deprivation.

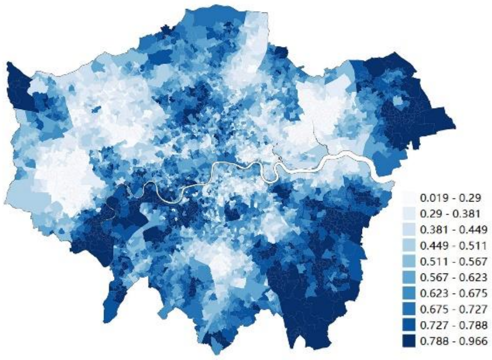
Average Access Index in LSOAs (2010)



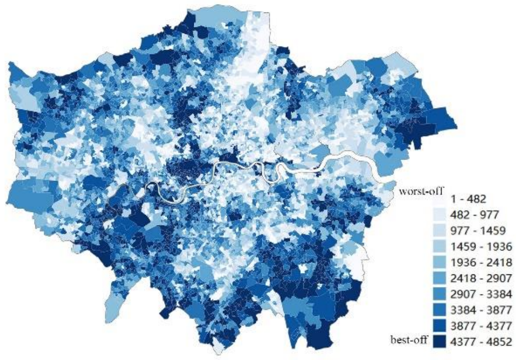
Average Access Index in LSOAs (2023)



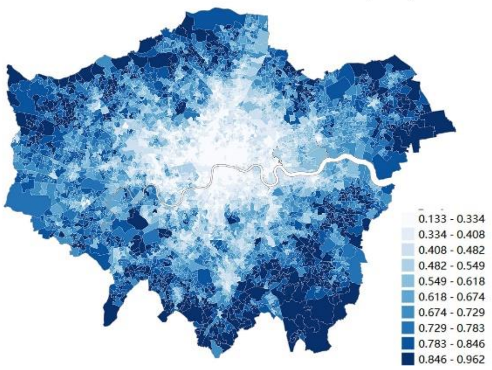
% white groups in LSOAs (2021)



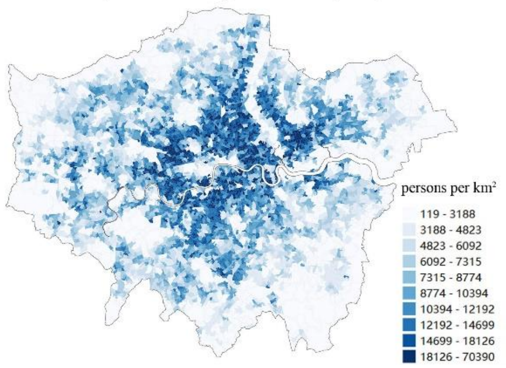
Income deprivation rank in London LSOAs (2019)



% households with car/van in LSOAs (2021)



Population density in LSOAs (2021)



Appendix C: Moran's I plot, Local Moran's I map, and correlation matrix of OLS model with total Access Index 2010 and other 2010s variables.