Targeting Childhood Obesity in London:

Spatio-Temporal Effects and Socioeconomic Moderators Informing Facility Interventions

---

CASA006 Coursewrok

Preparation

- Github link
- Number of words: 1870
- Runtime: *** hours (*Memory 10 GB, CPU Intel i7-10700 CPU @2.90GHz*)
- Coding environment: SDS Docker
- License: this notebook is made available under the Creative Commons Attribution license.
- Additional library *[libraries not included in SDS Docker or not used in this module]*:
  - **watermark**: A Jupyter Notebook extension for printing timestamps, version numbers, and hardware information.
  - To install required package:

```
!pip install linearmodels
!pip install nbconvert weasyprint
```

Table of contents

1.0 Introduction

[go back to the top]

Childhood obesity remains a serious public health issue in London, harming children's health and imposing substantial economic costs. In response, the UK government's *Childhood Obesity: A Plan for Action* (Department of Health and Social Care, 2018 (Department of Health and Social Care, 2018)) significantly boosted investment in physical activity infrastructure, including schools and community sports facilities. Public Health England's *Local Health Profiles* highlight increased access to parks, playgrounds, and sports centres across London wards (Public Health England, 2023). However, reductions in childhood obesity have been limited and uneven, particularly among younger children, reflecting obesity's complex causes and the limited impact of infrastructure expansion alone (Wyszyńska et al., 2020; Lieb, 2009).

This study quantitatively assesses how various facility types affect childhood obesity rates by age group in London wards from 2009/10 to 2013/14. It also explores the moderating effects of socioeconomic status (SES), greenspace, and public transport accessibility (PTAL). Employing spatial and temporal fixed effects, it aims to reveal spatial heterogeneity and develop an innovative risk-intervention quadrant framework combining spatial effects and residuals to support targeted, equitable public health strategies.

1.1 Requirements to Run the Analysis

An overview of packages used to run the analysis with brief explanation of their role.

44]:
```
!pip install linearmodels
!pip install nbconvert weasyprint
```

45]:
```
import os
import requests
import zipfile
from functools import reduce

import pandas as pd
import geopandas as gpd
from shapely.geometry import Point

import numpy as np
import scipy.stats as stats

import matplotlib.pyplot as plt
from matplotlib.gridspec import GridSpec

import seaborn as sns

from sklearn.ensemble import RandomForestRegressor
from sklearn.preprocessing import StandardScaler

import statsmodels.api as sm
import statsmodels.formula.api as smf
from statsmodels.iolib.summary2 import summary_col
from statsmodels.stats.outliers_influence import variance_inflation_factor
```

```python
from linearmodels.panel import PanelOLS

import nbformat
import nbconvert
from nbconvert.exporters import HTMLExporter
from nbconvert.preprocessors import Preprocessor
from weasyprint import HTML, CSS
```

2.0 Literature Review

[go back to the top]

Increasing evidence challenges the notion that expanding physical activity facilities alone suffices to reduce childhood obesity rates. While infrastructure is necessary, obesity results from complex interactions among behaviours, socioeconomic factors, and environmental contexts (Danielli et al., 2021). Research shows age influences intervention responses: older children engage more independently with their surroundings, whereas younger children's behaviours are more shaped by familial and cultural factors (Wyszyńska et al., 2020; Ayala et al., 2021), indicating the need for age-specific approaches.

Socioeconomic disparities exacerbate risk, as children from lower SES backgrounds often face poor access to quality facilities and supportive environments (Lieb et al., 2009; Yuan, 2024). Environmental factors such as green space and public transport accessibility further modify physical activity opportunities and obesity outcomes (Danielli et al., 2021). Additionally, spatial heterogeneity within London wards manifests uneven obesity prevalence and resources, warranting geographically targeted interventions (Titis et al., 2023).

Notably, Nau et al. (2019) used panel data integrating socioeconomic and built environment factors to study youth BMI trajectories across diverse communities, emphasizing the importance of accounting for spatial and temporal heterogeneity and dynamic interactions over time in obesity research.

Building on these foundations, the present study integrates facility types, age groups, SES, and environmental factors within a spatial-temporal fixed effects framework. This approach aims to better capture how facility impacts vary by age and SES while precisely modeling spatial heterogeneity in childhood obesity patterns, thereby informing more nuanced, targeted interventions.

3.0 Research questions

[go back to the top]

Building on the above, and given that facility effects vary by children’s development and socioeconomic access, while environmental factors (SES, greenspace, PTAL) may modify these effects through space and time, this study addresses:

- **RQ1 |** How do different facility types affect childhood obesity across age and socioeconomic groups, after accounting for spatial and temporal factors?
- **RQ2 |** How do greenspace and public transport accessibility modify the effects of facilities on obesity?

Recognizing that obesity prevalence and environmental influences vary across geographic areas and over time, a comprehensive spatial-temporal analysis is necessary. Consequently:

- **RQ3 |** How can spatial and temporal fixed effects reveal heterogeneous patterns to inform targeted public health interventions?

4.0 Data Collection

[go back to the top]

This study focuses on London wards at the 2011 level (London Data Store, 2011), offering the finest publicly available spatial resolution. The research covers the academic years from 2010/2011 to 2013/2014, spanning four academic years in total.

Table 3.1: Variable Descriptions and Data Sources for Urban Childhood Obesity Analysis

Variable Category	Variable Name	Description	Time Coverage	Data Source
Obesity Rates	`obese_recp`	Reception children obesity (ages 4-5)	2010/2011–2013/2014	Ward Atlas-NCMP
	`obese_y6`	Year 6 children obesity (ages 10-11)	2010/2011–2013/2014	Ward Atlas-NCMP
Facility Types	`facility_edu`	Education facilities per 10,000 pop	2010–2014	Active Places, ONS
	`facility_comm`	Commercial facilities per 10,000 pop	2010–2014	Active Places, ONS
	`facility_local`	Local authority facilities per 10,000 pop	2010–2014	Active Places, ONS
Socioeconomic	`unemp`	Unemployment rate	2010–2014	Ward Profiles
Environmental	`greenspace_pct`	Percentage of greenspace	2010, 2012*	Ward Profiles
	`ptal`	Public Transport Accessibility Level	2010, 2011, 2012, 2014*	Ward Profiles

[Note: Missing years for greenspace and PTAL were estimated via interpolation as described below.]

The key independent variables are densities of physical activity facilities, with the proportion of children from unemployed households and children's age groups used as heterogeneity stratifiers; greenspace coverage and public transport accessibility (PTAL) are included as control variables.

- Facilities were filtered to include only those operational before 2014 by considering their opening and closing dates. Population estimates for wards each year were taken from the Office for National Statistics (ONS, 2023). Facility density per 10,000 population was calculated using the formula:

$$\textbf{Facility density} = \textbf{(Number of operational facilities in ward / Ward population)} \times \textbf{10,000}$$

This calculation standardises facility counts by population size, enabling meaningful comparison across wards with differing population numbers.

48]:
```
# Count active facilities per ward per year
def count_facilities_by_ward(gdf_fac_london, years, facility_types):
    wards_fac_count = wards[ward_cols].copy()
    for yr in years:
        # Filter active facilities for the year
        mask_active = (
            (gdf_fac_london['year_built'] <= yr) &
            (gdf_fac_london['closed_date'].isna() |
(gdf_fac_london['closed_date'].dt.year >= yr))
        )
        active = gdf_fac_london[mask_active]

        # Count facilities by type for each ward
        for ftype in facility_types:
            temp = active[active['mgmt_group'] == ftype]
            by_ward =
temp.groupby('GSS_CODE').size().rename(f'facility_{ftype.lower()}_{yr}').reset_index()
            wards_fac_count = wards_fac_count.merge(by_ward, on='GSS_CODE',
how='left')

    # Fill missing values with 0
    for ftype in facility_types:
        for yr in years:
            col = f'facility_{ftype.lower()}_{yr}'
            if col in wards_fac_count.columns:
                wards_fac_count[col] = wards_fac_count[col].fillna(0)

    return wards_fac_count
```

50]:
```
# Compute facility density per 10,000 population
def compute_facility_density(wards_fac_count, df_pop, pop_years, facility_types,
ward_cols):
    wards_fac_pop = wards_fac_count.merge(df_pop, on='GSS_CODE', how='left')

    for yr in pop_years:
        for ftype in facility_types:
            fac_col = f'facility_{ftype.lower()}_{yr}'
            wards_fac_pop[f'facility_{ftype.lower()}_per_10000pop_{yr}'] = (
                wards_fac_pop[fac_col] / wards_fac_pop[f'pop_{yr}'] * 10000
            )

    to_keep = ward_cols.copy()
    for yr in pop_years:
        for ftype in facility_types:
            to_keep.append(f'facility_{ftype.lower()}_per_10000pop_{yr}')

    return wards_fac_pop[to_keep]

# Main execution
wards_fac_count = count_facilities_by_ward(gdf_fac_london, years, facility_types)
df_pop = load_population_data(pop_years)
wards_fac_pop = compute_facility_density(wards_fac_count, df_pop, pop_years,
facility_types, ward_cols)
```

- Linear interpolation was used to fill missing years in greenspace and PTAL data. The method assumes changes between observed values occur steadily over time, estimating missing data via weighted averages of neighbouring points.

  This approach is reliable since environmental variables like greenspace and PTAL typically change gradually over short periods, and the one- to two-year interpolation span minimises inaccuracies. Linear interpolation is a well-established technique in spatio-temporal analysis, as noted in comprehensive reviews such as Li et al., 2014.

52]:
```
# Interpolate missing years
df_green["greenspace_pct_2011"] = df_green["greenspace_pct_2012"]
df_green["greenspace_pct_2013"] = (df_green["greenspace_pct_2012"] +
df_green["greenspace_pct_2014"]) / 2
df_ptal["ptal_2013"] = (df_ptal["ptal_2012"] + df_ptal["ptal_2014"]) / 2
```

After cleaning and interpolation, all datasets were merged on the standard geographic identifier (ward_code) to produce a comprehensive dataset for analysis.

55]:
```
# Plotting: scatter with marginal histograms
g = sns.JointGrid(data=plot_df, x='facility_education', y='obesity', height=8)
palette = sns.color_palette("Set1", 2)
groups = plot_df['age_group'].unique()

for i, group in enumerate(groups):
    subset = plot_df[plot_df['age_group'] == group]
    sns.regplot(x='facility_education', y='obesity', data=subset, scatter=True,
fit_reg=True,
                scatter_kws={'alpha':0.4, 's':40, 'color':palette[i]},
                line_kws={'color':palette[i], 'linewidth':2}, ax=g.ax_joint,
label=group)
    sns.histplot(subset['facility_education'], color=palette[i], alpha=0.3, bins=30,
ax=g.ax_marg_x)

# Horizontal bar for obesity distribution on right margin
bin_edges = np.linspace(plot_df['obesity'].min(), plot_df['obesity'].max(), 30)
for i, group in enumerate(groups):
    subset = plot_df[plot_df['age_group'] == group]
    counts, _ = np.histogram(subset['obesity'], bins=bin_edges)
    bin_centers = 0.5 * (bin_edges[:-1] + bin_edges[1:])
    g.ax_marg_y.barh(bin_centers, counts, color=palette[i], alpha=0.3,
edgecolor='black')

g.ax_joint.set_xlim(0, 30)
g.ax_joint.set_ylim(0, 25)
g.ax_marg_y.set_ylim(g.ax_joint.get_ylim())
g.ax_joint.legend(title='Age Group')
g.set_axis_labels('Facility Education per 10,000 Population', 'Obesity Rate')

plt.tight_layout()
plt.show()
```

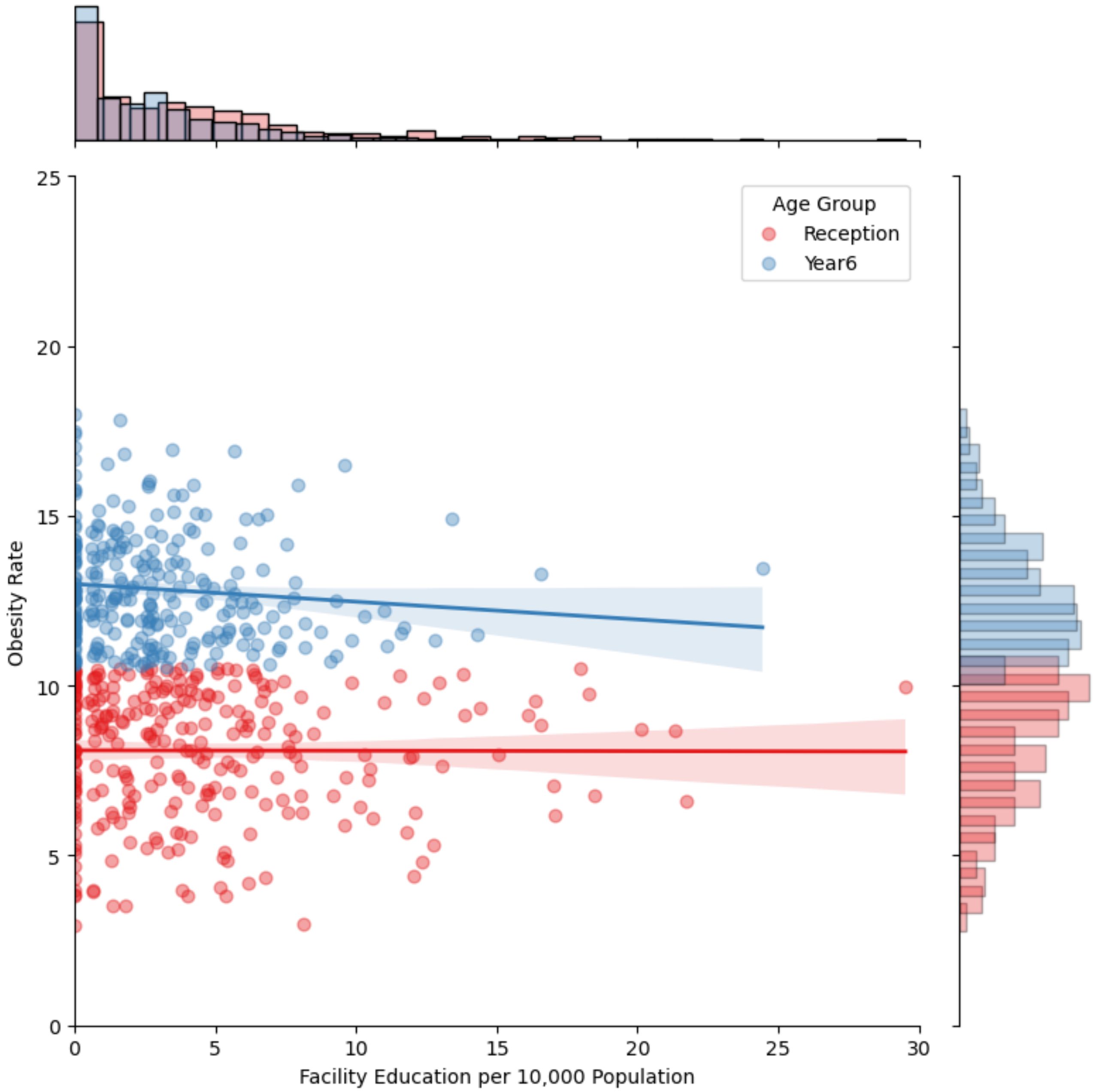


Figure 1: Distribution and relationship between education facility density and childhood obesity rates by age group.

Figure 1 preliminarily visualizes the relationship between education facility density and childhood obesity rates by age group, showing higher obesity rates in Year 6 children and considerable variation in facility density across wards.

5. Methodology

[go back to the top]

Following Nau et al. (2019), this study analyses childhood obesity in London wards from 2010 to 2014 using panel data integrating socioeconomic and built environment variables. We develop age-specific fixed effects regression models at the ward level, incorporating detailed facility types, age groups (Reception and Year 6), and their interactions with socioeconomic status (SES). This approach controls spatial and temporal heterogeneity while accounting for residual variation and spatial risk (see Figure 2).

5.1 Multicollinearity Testing

Multicollinearity is first assessed using Variance Inflation Factor (VIF) and Pearson correlation matrix to avoid high correlations among independent variables that may bias model estimation (Hair et al., 2019): - Variables with VIF > 5 are removed or transformed. - Variables with absolute Pearson correlation > 0.7 are excluded.

This ensures stable and interpretable panel regression models.

5.2 Cross-sectional OLS Regression

To assess baseline spatial associations between childhood obesity and variables, I conduct annual cross-sectional Ordinary Least Squares (OLS) regressions—a standard method estimating linear relationships. Treating years independently captures spatial variation but omits time-invariant unobserved factors.

This approach informs variable selection and model setup for panel fixed effects models addressing temporal dynamics and heterogeneity.

5.3 Panel Fixed Effects Regression Model

Using the screened variables, a panel fixed effects regression on ward-level data (2010–2014) analyses spatial and temporal childhood obesity variations. This model controls for time-invariant regional characteristics and common year effects, reducing omitted variable bias and enhancing causal inference.

The baseline model is specified as:

$$\mathbf{Obesity_{i,t} = \alpha + \sum_k \beta_k\, Facility_{k,i,t} + \gamma_1\, Unemp_{i,t} + \gamma_2\, Green_{i,t} + \gamma_3\, PTAL_{i,t} + \mu_i + \tau_t + \varepsilon_{i,t}}$$

where:

- $\mathrm{Obesity}_{i,t}$: childhood obesity rate in ward i at year t
- $\mathrm{Facility}_{k,i,t}$: density of facility type k in ward i at year t
- $\mathrm{Unemp}_{i,t}$, $\mathrm{Green}_{i,t}$, $\mathrm{PTAL}_{i,t}$: controls for unemployment, greenspace, and public transport accessibility
- μ_i: ward fixed effects (time-invariant factors)
- τ_t: year fixed effects (common temporal shocks)
- $\varepsilon_{i,t}$: error term

Models are fitted separately for Reception (4–5 years) and Year 6 (10–11 years) to capture developmental differences.

Interaction terms between facilities and socioeconomic status (SES) are added:

$$\mathbf{Obesity_{i,t} = \alpha + \sum_k \beta_k\, Facility_{k,i,t} + \sum_k \delta_k (Facility_{k,i,t} \times SES_{i,t}) + \gamma_1\, Unemp_{i,t} + \gamma_2\, Greeni{,}t + \gamma_3\, PTAL_{i,t} + \mu_i + \tau_t + \varepsilon_{i,t}}$$

which explores how SES moderates the impact of different facility types on childhood obesity, enhancing the policy relevance and scientific rigour of recommendations.

5.4 Residuals and Spatial Heterogeneity Analysis

Inspired by approaches combining spatial risk assessment with local variation analysis (Walker et al., 2016), I leverage spatial fixed effects and residuals to capture inherent ward-level risk and unexplained deviations amid current interventions. I then classify wards in a two-dimensional quadrant plot by spatial risk and residual sign:

This classification identifies:

- **High spatial risk & positive residuals**: Persistent hotspots where obesity exceeds expectations, indicating insufficient intervention.
- **High spatial risk & negative residuals**: High-risk wards with better-than-expected outcomes, reflecting effective local measures.
- **Low spatial risk & positive residuals**: Typically low-risk wards showing rising obesity, signaling emerging issues.
- **Low spatial risk & negative residuals**: Wards with low risk and favourable outcomes, representing stable, well-managed areas.

By combining spatial risk and residual deviations, this approach highlights where current strategies succeed or fail, guiding more precise resource allocation and multi-level public health actions.

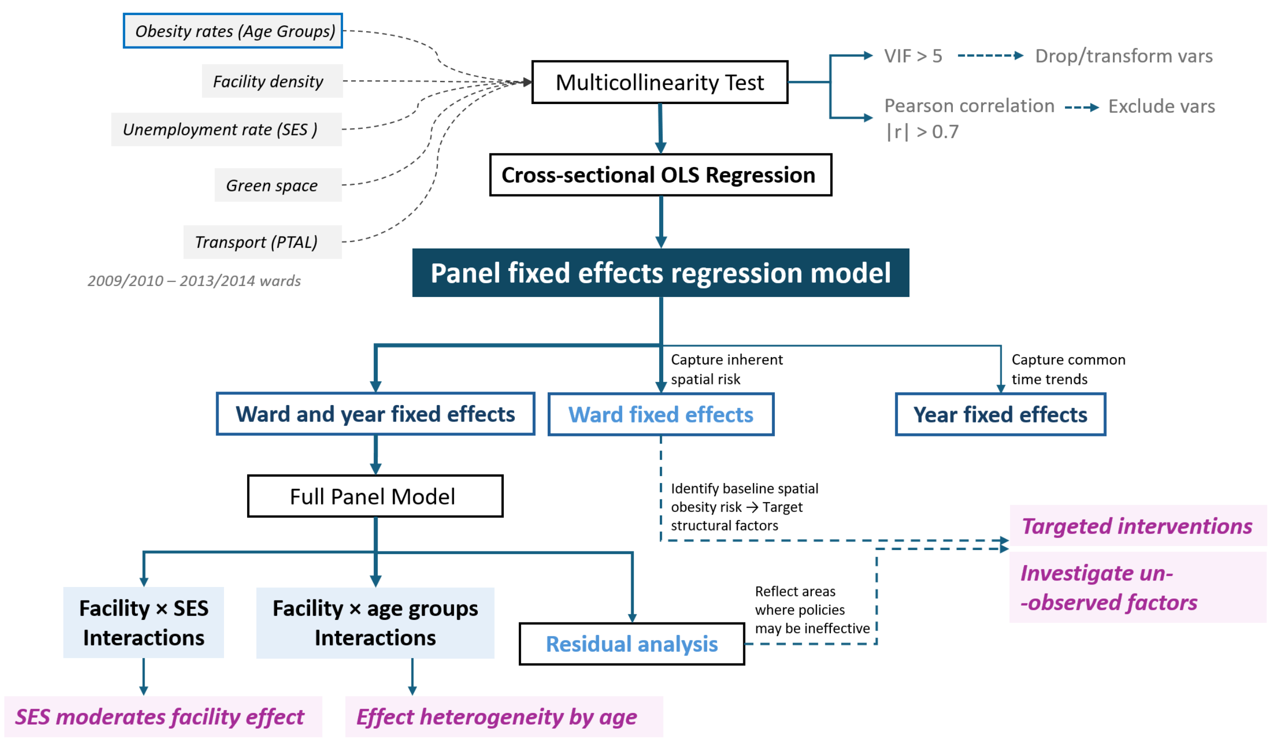


Figure 2: Methodology flow chart.

6.0 Results Analysis

[go back to the top]

6.1 Multicollinearity Check

Table 2 and Figure 3 indicate all explanatory variables have acceptable multicollinearity (VIF < 5). Facility variables show moderate correlations (up to 0.35), while socioeconomic and environmental variables display expected patterns, justifying their inclusion in regression models.

Table 2. Variance Inflation Factor (VIF) for Explanatory Variables

76]:
```
# Define features for VIF and correlation
features = ['facility_education', 'facility_commercial', 'facility_local_authority',
'unemployment', 'greenspace', 'ptal']

# Calculate VIF for multicollinearity check
X = data_clean[features].values
vif_data = pd.DataFrame({'Variable': features, 'VIF': [variance_inflation_factor(X, i)
for i in range(X.shape[1])]})
print(vif_data)
```

```
                   Variable       VIF
0        facility_education  2.037713
1       facility_commercial  1.573460
2  facility_local_authority  1.449113
3              unemployment  3.592629
4                greenspace  3.324651
5                      ptal  4.200670
```

77]:
```
# Compute and plot correlation matrix including obesity
corr_matrix = data_clean[features + ['obesity']].corr()

plt.figure(figsize=(4, 3))
sns.heatmap(
    corr_matrix, annot=True, fmt='.2f', cmap='coolwarm',
    vmin=-1, vmax=1, linewidths=0.5, annot_kws={"size": 6}
)
plt.xticks(fontsize=6, rotation=45)
plt.yticks(fontsize=6)

colorbar = plt.gca().collections[0].colorbar
colorbar.ax.tick_params(labelsize=6)

plt.title("Correlation Matrix of Features and Obesity", fontsize=10)
plt.show()
```

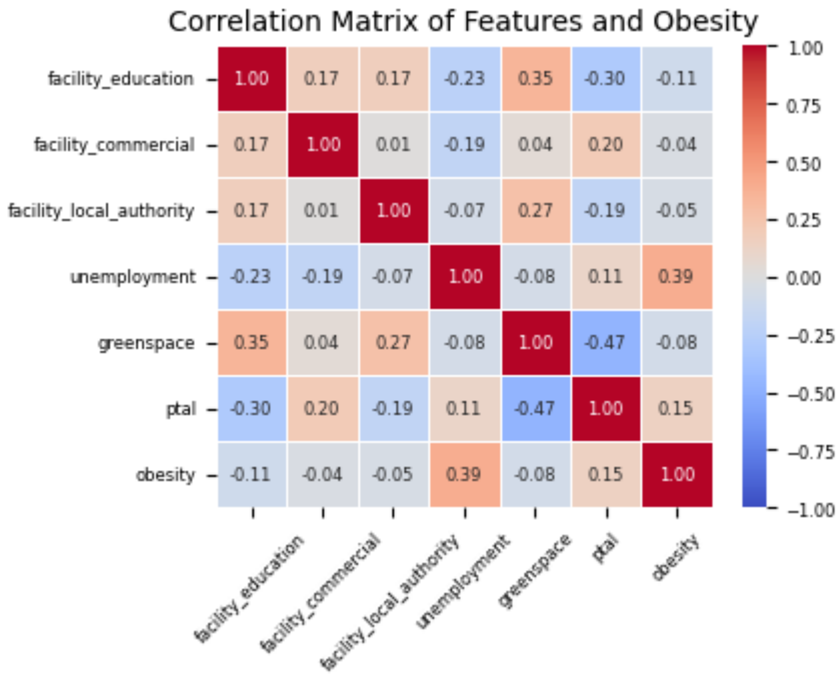


Figure 3. Correlation Matrix of Features and Obesity

Next, the dataset were reorganised into long format by concatenation and set with ward and year as panel indices for regression analysis.

60]:
```
# Concatenate data for all years
panel_df = pd.concat(panel_list, axis=0).reset_index(drop=True)

# Set panel index for panel regression (ward_code and year)
panel_df = panel_df.set_index(['ward_code', 'year'])
```

6.2 OLS Regression Results

Table 3 presents cross-sectional OLS results for 2011–2014. Facility variables exhibit heterogeneity, with coefficients fluctuating between significant and nonsignificant across years; education-related facilities are borderline significant (~0.04). Public transport accessibility (ptal) consistently shows a positive and highly significant association with childhood obesity. Unemployment positively, and greenspace negatively, affect obesity rates significantly.

Figure 4 highlights varied ward-level slopes between facility_education and Year 6 obesity, confirming heterogeneous facility effects.

Table 3. Cross-Sectional OLS Regression Results (2011–2014)

61]:
```
panel_df_reset = panel_df.reset_index()
models = []
model_names = []

for year in sorted(panel_df_reset['year'].unique()):
```

```python
    sub = panel_df_reset[panel_df_reset['year'] == year]
    model = smf.ols("obesity_year6 ~ facility_education + facility_commercial +
facility_local_authority + unemployment + greenspace_pct + ptal", data=sub).fit()
    models.append(model)
    model_names.append(str(year))

table = summary_col(models, stars=True, model_names=model_names,
                    info_dict={"N": lambda x: f"{int(x.nobs)}"},
                    float_format="%.3f")
print(table)
```

```
=================================================================
                          2011      2012      2013      2014
-----------------------------------------------------------------
Intercept                12.759*** 13.216*** 13.480*** 14.293***
                         (0.729)   (0.754)   (0.723)   (0.739)
facility_education       0.020     0.031     0.034     0.022
                         (0.040)   (0.042)   (0.041)   (0.043)
facility_commercial      0.046     0.054     0.058     0.034
                         (0.045)   (0.046)   (0.044)   (0.046)
facility_local_authority -0.056    -0.060    0.034     0.007
                         (0.109)   (0.111)   (0.108)   (0.113)
unemployment             0.007***  0.008***  0.009***  0.011***
                         (0.000)   (0.000)   (0.000)   (0.001)
greenspace_pct           -0.006    -0.013    -0.010    -0.027**
                         (0.010)   (0.010)   (0.010)   (0.011)
ptal                     0.830***  0.770***  0.601***  0.546***
                         (0.142)   (0.146)   (0.138)   (0.142)
R-squared                0.481     0.487     0.508     0.506
R-squared Adj.           0.475     0.481     0.503     0.501
N                        567       567       567       561
=================================================================
Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01
```

74]:

```python
rec_vis = panel_df.reset_index().dropna(subset=['facility_education',
'obesity_year6'])

sns.lmplot(data=rec_vis,
           x='facility_education',
           y='obesity_year6',
           hue='ward_code',
           ci=None,
           legend=False,
           scatter_kws={'alpha':.3},
           height=8,
           palette='husl')
plt.title("Individual Ward Trends: Facility Education vs. obesity Year 6 (10-11)")
plt.show()
```

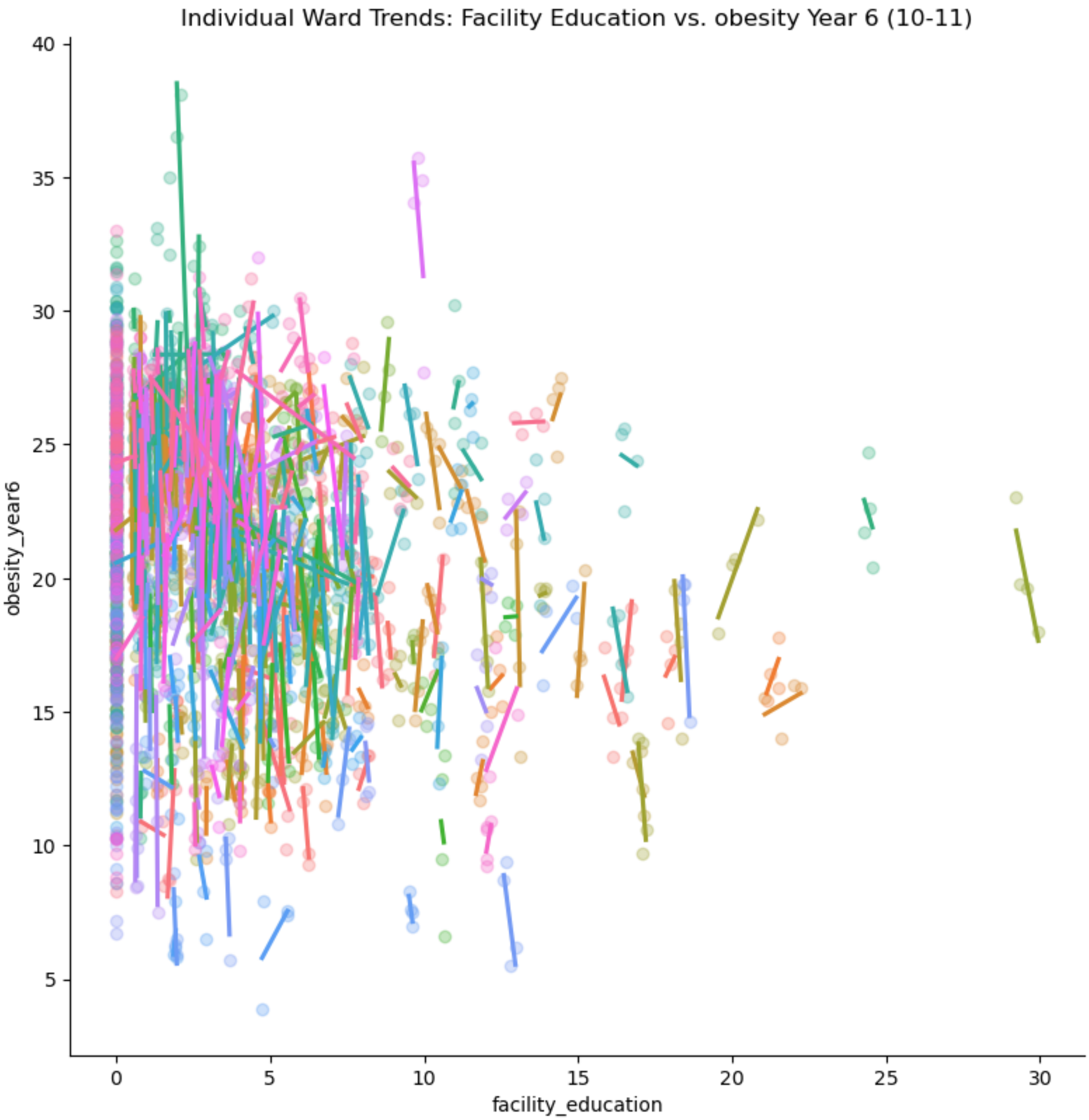


Figure 4. Ward-Level Slopes of Facility Education vs. Year 6 (10-11) Obesity

6.3 Panel Regression Results

The fixed effects panel model (Table 4) shows local authority facilities (facility_local_authority) significantly increase childhood obesity rates ($p = 0.0215$), while other facility types are not significant. This positive association may reflect reverse causality, with higher obesity areas receiving more local authority investments (see code below).

Unemployment shows a negative but insignificant association ($p = 0.123$). Greenspace coverage (greenspace_pct) consistently exhibits a significant protective effect ($p = 0.013$). The model's F-test (12.956, $p < 0.001$) supports the panel approach.

Time fixed effects reveal a gradual increase in obesity rates from 2011 to 2014.

Table 4. Panel Regression Results (Fixed Effects Model)

63]:
```python
panel_data = panel_df_reset.set_index(['ward_code', 'year'])

# Prepare dependent and independent variables
y = panel_data['obesity_year6']
X = panel_data[['facility_education', 'facility_commercial',
'facility_local_authority',
                'unemployment', 'greenspace_pct', 'ptal']]
X = sm.add_constant(X)  # Add intercept term

# Specify and fit the fixed effects panel model with entity and time effects
model_fe = PanelOLS(y, X, entity_effects=True, time_effects=True)
result_fe = model_fe.fit(cov_type='clustered', cluster_entity=True)

# Output model summary results
print(result_fe)
```

```
                          PanelOLS Estimation Summary
================================================================================
Dep. Variable:          obesity_year6   R-squared:                        0.0074
Estimator:                   PanelOLS   R-squared (Between):             -0.2555
No. Observations:                2262   R-squared (Within):               0.0194
Date:                Tue, Apr 29 2025   R-squared (Overall):             -0.2269
Time:                        10:35:47   Log-likelihood                   -4276.2
Cov. Estimator:             Clustered
                                        F-statistic:                      2.1064
Entities:                         567   P-value                           0.0497
Avg Obs:                       3.9894   Distribution:                  F(6,1686)
Min Obs:                       3.0000
Max Obs:                       4.0000   F-statistic (robust):             2.7466
                                        P-value                           0.0117
Time periods:                       4   Distribution:                  F(6,1686)
Avg Obs:                       565.50
Min Obs:                       561.00
Max Obs:                       567.00

                                   Parameter Estimates
============================================================================================
                          Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
--------------------------------------------------------------------------------------------
const                        25.620     3.4231     7.4845     0.0000      18.906      32.334
facility_education          -0.2060     0.2595    -0.7938     0.4274     -0.7149      0.3029
facility_commercial          0.1139     0.4124     0.2763     0.7824     -0.6949      0.9228
facility_local_authority     0.8620     0.3747     2.3004     0.0215      0.1270      1.5969
unemployment                -0.0015     0.0010    -1.5421     0.1232     -0.0033      0.0004
greenspace_pct              -0.1010     0.0406    -2.4875     0.0130     -0.1806     -0.0214
ptal                        -0.2622     0.8717    -0.3008     0.7636     -1.9718      1.4474
============================================================================================

F-test for Poolability: 12.956
P-value: 0.0000
Distribution: F(569,1686)

Included effects: Entity, Time
```

Figure 5. Year Fixed Effects of Obesity

6.4 Age-Group Interaction Models

Sample stratification into Reception (4–5 years) and Year 6 (10–11 years) groups (Table 5) shows:

- No facility or socioeconomic variables are significant in Reception.
- In Year 6, facility_local_authority retains a significant positive effect (p = 0.0215), and greenspace coverage remains a significant negative influence (p = 0.013). Unemployment remains insignificant.

Older children's obesity rates appear more sensitive to environmental facilities and socioeconomic factors.

Table 5. Panel Model Results by Age Group (Reception and Year 6 Combined)

65]:
```
# Define outcome variables by age group
age_groups = {
    'Reception': 'obesity_reception',  # 4-5 years old
    'Year6': 'obesity_year6'           # 10-11 years old
}

results_by_age = {}

for group, outcome in age_groups.items():
    # Subset data and drop missing values
    data_sub = panel_data.dropna(subset=[outcome, 'facility_education',
'facility_commercial',
                                         'facility_local_authority', 'unemployment',
'greenspace_pct', 'ptal'])
    y = data_sub[outcome]
    X = sm.add_constant(data_sub[['facility_education', 'facility_commercial',
'facility_local_authority',
                                  'unemployment', 'greenspace_pct', 'ptal']])

    # Fit fixed effects panel model
    model = PanelOLS(y, X, entity_effects=True, time_effects=True)
    result = model.fit(cov_type='clustered', cluster_entity=True)
    results_by_age[group] = result

    print(f"\n===== Panel Results for {group} group =====")
    print(result.summary.tables[1])
```

```
===== Panel Results for Reception group =====
                              Parameter Estimates
==============================================================================================
                          Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
----------------------------------------------------------------------------------------------
const                        14.242     2.0518     6.9412     0.0000      10.218      18.266
facility_education          -0.0620     0.1456    -0.4260     0.6702     -0.3477      0.2236
facility_commercial          0.0834     0.1963     0.4248     0.6710     -0.3016      0.4683
facility_local_authority    -0.3135     0.3575    -0.8767     0.3808     -1.0147      0.3878
unemployment                 0.0004     0.0008     0.5545     0.5793     -0.0011      0.0020
greenspace_pct              -0.0214     0.0243    -0.8791     0.3795     -0.0691      0.0263
ptal                        -0.8847     0.5338    -1.6573     0.0977     -1.9316      0.1623
==============================================================================================

===== Panel Results for Year6 group =====
                              Parameter Estimates
==============================================================================================
                          Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
----------------------------------------------------------------------------------------------
const                        25.620     3.4231     7.4845     0.0000      18.906      32.334
facility_education          -0.2060     0.2595    -0.7938     0.4274     -0.7149      0.3029
facility_commercial          0.1139     0.4124     0.2763     0.7824     -0.6949      0.9228
facility_local_authority     0.8620     0.3747     2.3004     0.0215      0.1270      1.5969
unemployment                -0.0015     0.0010    -1.5421     0.1232     -0.0033      0.0004
greenspace_pct              -0.1010     0.0406    -2.4875     0.0130     -0.1806     -0.0214
ptal                        -0.2622     0.8717    -0.3008     0.7636     -1.9718      1.4474
==============================================================================================
```

6.5 SES Interaction Models

Introducing interactions between facility variables and unemployment (SES) (Table 6) shows no significant interaction effects in either age group, indicating SES does not substantially moderate facility impacts on obesity.

Table 6. SES Interaction Models Results (Both Age Groups Combined)

66]:

```
results_by_age_ses = {}

for group_name, outcome_var in age_groups.items():
    # Prepare data and create SES dummy and interaction terms
    data = panel_data.dropna(subset=[
        outcome_var, 'facility_education', 'facility_commercial',
'facility_local_authority',
        'unemployment', 'greenspace_pct', 'ptal'
    ]).copy()

    data['high_ses'] = (data['unemployment'] <
data['unemployment'].median()).astype(int)
    data['edu_x_ses'] = data['facility_education'] * data['high_ses']
    data['com_x_ses'] = data['facility_commercial'] * data['high_ses']
    data['local_x_ses'] = data['facility_local_authority'] * data['high_ses']

    y = data[outcome_var]
    X = sm.add_constant(data[[
        'facility_education', 'facility_commercial', 'facility_local_authority',
        'unemployment', 'greenspace_pct', 'ptal',
```

```
        'edu_x_ses', 'com_x_ses', 'local_x_ses'
    ]])

    # Fit fixed effects panel model with clustered SE
    model = PanelOLS(y, X, entity_effects=True, time_effects=True)
    result = model.fit(cov_type='clustered', cluster_entity=True)
    results_by_age_ses[group_name] = result

    print(f"\n===== SES Interaction Results for {group_name} Group =====")
    print(result.summary.tables[1])  # Print coefficients table
```

```
===== SES Interaction Results for Reception Group =====
                                    Parameter Estimates
==========================================================================================
                          Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
------------------------------------------------------------------------------------------
const                        13.984     2.0549     6.8054     0.0000      9.9537      18.014
facility_education          -0.0632     0.1455    -0.4341     0.6643     -0.3485      0.2222
facility_commercial          0.1040     0.2005     0.5186     0.6041     -0.2893      0.4972
facility_local_authority    -0.3305     0.3902    -0.8471     0.3971     -1.0959      0.4348
unemployment                 0.0005     0.0008     0.6045     0.5456     -0.0011      0.0021
greenspace_pct              -0.0197     0.0243    -0.8092     0.4185     -0.0673      0.0280
ptal                        -0.8701     0.5327    -1.6335     0.1026     -1.9148      0.1747
edu_x_ses                    0.0463     0.0245     1.8889     0.0591     -0.0018      0.0944
com_x_ses                   -0.0323     0.0621    -0.5191     0.6038     -0.1541      0.0896
local_x_ses                  0.1009     0.1088     0.9273     0.3539     -0.1125      0.3142
==========================================================================================

===== SES Interaction Results for Year6 Group =====
                                    Parameter Estimates
==========================================================================================
                          Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
------------------------------------------------------------------------------------------
const                        25.782     3.3853     7.6157     0.0000      19.142      32.422
facility_education          -0.2027     0.2563    -0.7906     0.4293     -0.7054      0.3001
facility_commercial          0.1466     0.4439     0.3302     0.7413     -0.7240      1.0171
facility_local_authority     0.8485     0.3752     2.2613     0.0239      0.1125      1.5845
unemployment                -0.0015     0.0010    -1.5632     0.1182     -0.0034      0.0004
greenspace_pct              -0.1033     0.0402    -2.5680     0.0103     -0.1822     -0.0244
ptal                        -0.2659     0.8688    -0.3061     0.7596     -1.9701      1.4382
edu_x_ses                   -0.0094     0.0557    -0.1687     0.8661     -0.1186      0.0998
com_x_ses                   -0.0697     0.1395    -0.4998     0.6173     -0.3433      0.2039
local_x_ses                 -0.0591     0.1921    -0.3076     0.7584     -0.4360      0.3177
==========================================================================================
```

Overall, facility effects depend on age and type, with local authority facilities notably affecting older children. Socioeconomic and environmental variables, especially greenspace, emerge as robust determinants. Potential reverse causality regarding facilities warrants caution in policy interpretation.

6.6 Residuals and Spatial Fixed Effects Analysis

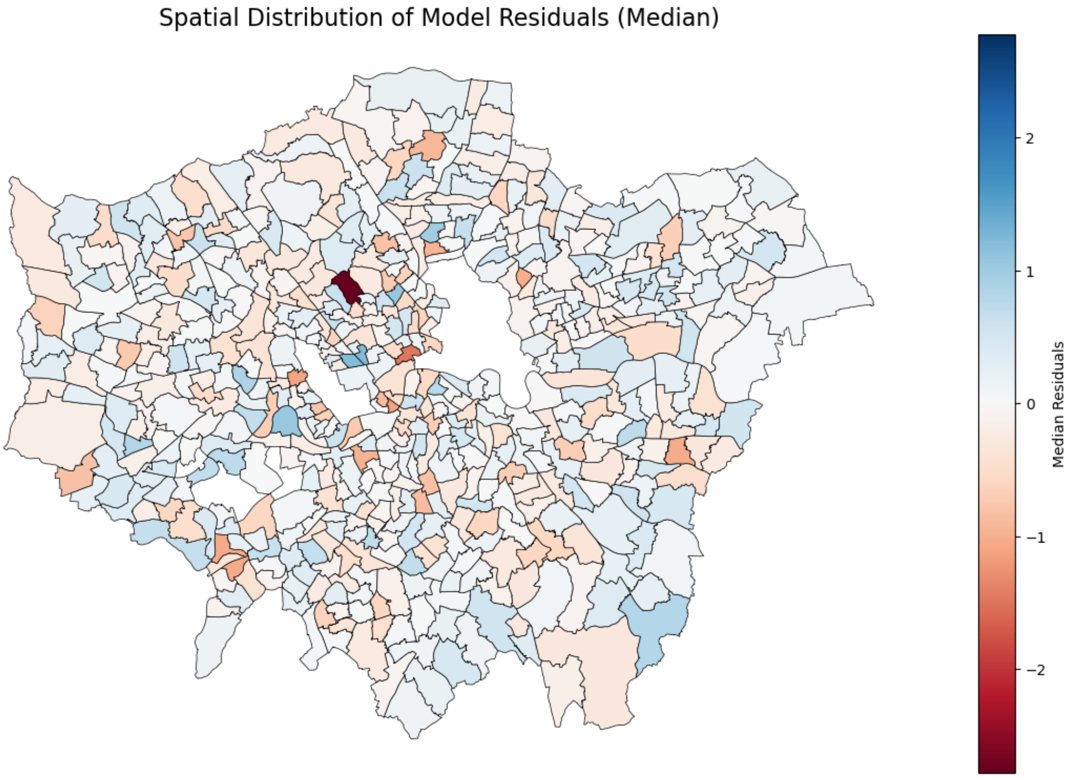
Residuals from the panel model are approximately symmetrically distributed around zero, with slight positive skewness (0.087) and moderate kurtosis (1.14) (Figure 5). The Q-Q plot shows near-normality with minor tail deviations, and residuals versus fitted values reveal no heteroscedasticity, supporting model validity.

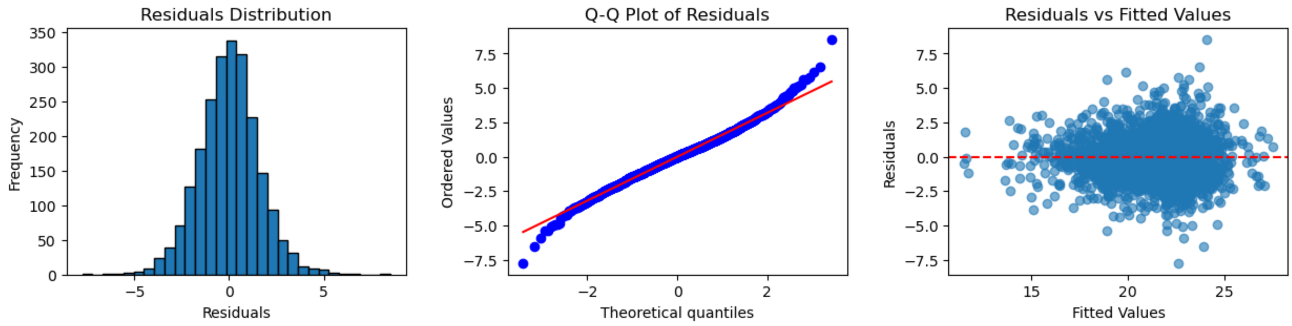
Spatial mapping of median residuals at ward level indicates geographic clustering, suggesting spatial heterogeneity remains despite covariate and fixed effect controls.

67]:
```
# Calculate median residuals by ward for spatial plot
residuals = result_fe.resids
median_residuals = result.resids.groupby(level='ward_code').median()
wards_resid_gdf = wards[['GSS_CODE', 'geometry']].rename(columns={'GSS_CODE':
'ward_code'}).merge(
    pd.DataFrame({'ward_code': median_residuals.index, 'median_residual':
median_residuals.values}),
    on='ward_code')

# Plot spatial map of median residuals
wards_gem = wards[['GSS_CODE', 'geometry']]
wards_gem = wards_gem.rename(columns={'GSS_CODE': 'ward_code'})

vmax = max(abs(wards_resid_gdf['median_residual'].min()),
wards_resid_gdf['median_residual'].max())
```

68]:
```
if vmax < 1e-6:
    vmax = 1e-6  # Set lower bound threshold for visualization

# Use GridSpec to control height ratios: map is taller, bottom row shorter
fig = plt.figure(figsize=(15, 14))
gs = GridSpec(2, 3, height_ratios=[3, 1], hspace=0.3, wspace=0.3)

# Spatial residuals map (top, spans all columns)
ax_map = fig.add_subplot(gs[0, :])
wards_resid_gdf.plot(
    column='median_residual',
    cmap='RdBu',
    linewidth=0.5,
    edgecolor='k',
    legend=True,
    legend_kwds={'label': 'Median Residuals'},
    vmin=-vmax,
    vmax=vmax,
    ax=ax_map
)
ax_map.set_title('Spatial Distribution of Model Residuals (Median)', fontsize=16)
ax_map.axis('off')
ax_map.set_aspect('equal')

# Residual diagnostics plots - three smaller plots in bottom row
ax_hist = fig.add_subplot(gs[1, 0])
ax_qq = fig.add_subplot(gs[1, 1])
ax_scatter = fig.add_subplot(gs[1, 2])

# 1. Histogram of residuals
ax_hist.hist(residuals, bins=30, edgecolor='black')
ax_hist.set_title('Residuals Distribution')
ax_hist.set_xlabel('Residuals')
ax_hist.set_ylabel('Frequency')

# 2. Q-Q plot of residuals
stats.probplot(residuals, plot=ax_qq)
ax_qq.set_title('Q-Q Plot of Residuals')
```

```python
# 3. Residuals vs fitted values scatter plot
ax_scatter.scatter(result.fitted_values, residuals, alpha=0.6)
ax_scatter.axhline(y=0, color='red', linestyle='--')
ax_scatter.set_title('Residuals vs Fitted Values')
ax_scatter.set_xlabel('Fitted Values')
ax_scatter.set_ylabel('Residuals')

plt.show()

# Print summary statistics of residuals
print("Residuals Summary Statistics:")
print(f"Mean: {np.mean(residuals):.6f}")
print(f"Standard Deviation: {np.std(residuals):.6f}")
print(f"Skewness: {stats.skew(residuals):.6f}")
print(f"Kurtosis: {stats.kurtosis(residuals):.6f}")
```






```
Residuals Summary Statistics:
Mean: 0.000000
Standard Deviation: 1.602410
Skewness: 0.075269
Kurtosis: 1.213215
```

Figure 5. Residuals Distribution, Q-Q Plot, and Spatial Mapping of Median Residuals (Combined)

After accounting for spatial heterogeneity via ward-level fixed effects, the model (Figure 6) reveal substantial unexplained spatial variation in obesity rates, with notably higher effects in northern wards and lower effects in some southern areas.

69]:
```
# Extract ward-level fixed effects estimates from model results
entity_effects = result_fe.estimated_effects

# Convert MultiIndex to columns for easier processing
entity_effects_df = entity_effects.reset_index()

# Aggregate by ward_code to calculate average fixed effect per ward
entity_effects_grouped = entity_effects_df.groupby('ward_code')
['estimated_effects'].mean().reset_index()

# Merge aggregated fixed effects with ward geometries for mapping
wards_fe_gdf = wards_gem.merge(entity_effects_grouped, on='ward_code')

# Plot the spatial distribution of ward-level fixed effects
plt.figure(figsize=(12, 10))
wards_fe_gdf.plot(
    column='estimated_effects',
    cmap='coolwarm',
    linewidth=0.5,
    edgecolor='k',
    legend=True,
    legend_kwds={'label': 'Entity Fixed Effects'}
)
plt.title('Ward-level Fixed Effects Distribution')
plt.axis('off')
plt.tight_layout()
plt.show()
```

```
<Figure size 1200x1000 with 0 Axes>
```

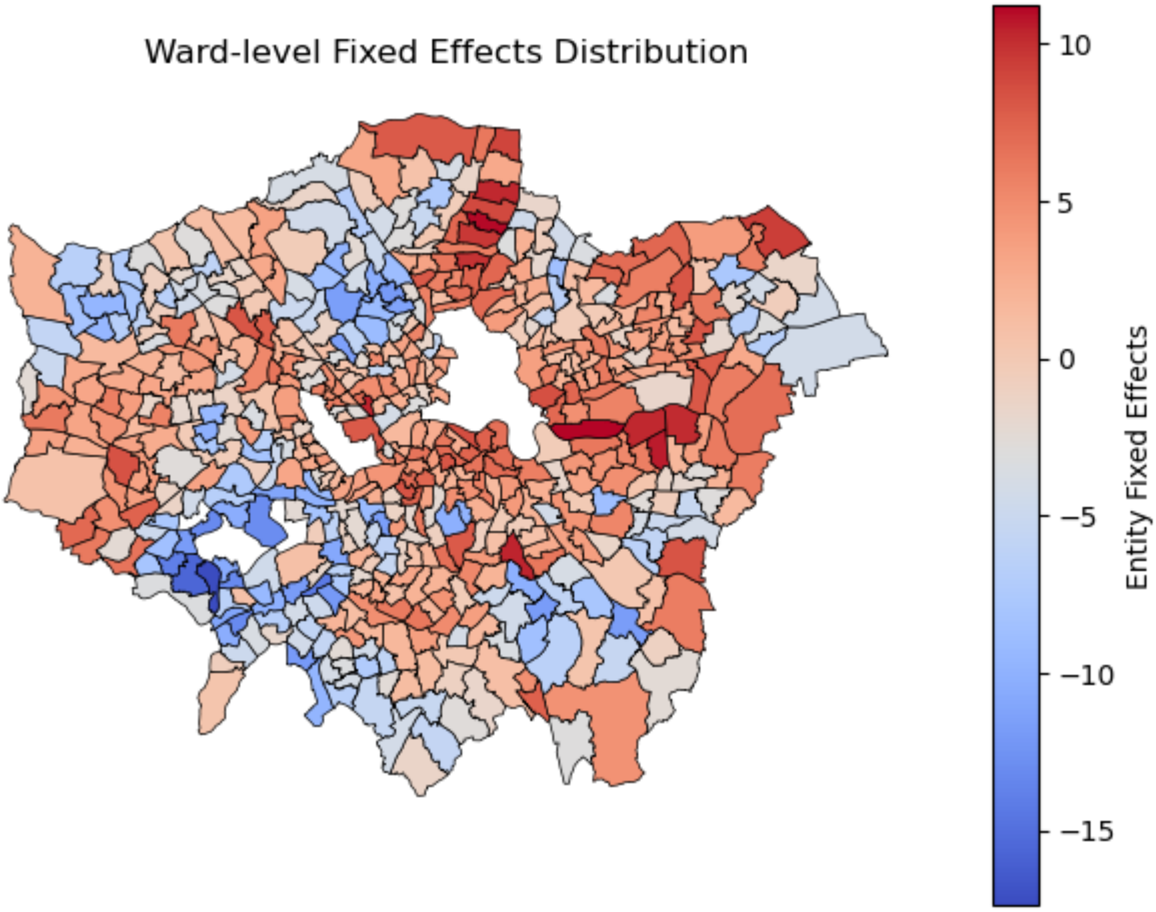


Figure 6. Ward-Level Fixed Effects from Panel Model

Subsequently, a quadrant plot (Figure 7) combining spatial fixed effects and mean residuals with ward labels demonstrates wards with high spatial effects correspond to higher obesity rates, reflecting strong localized influences captured by the model.

Residuals are less systematically linked to obesity, as wards with similar residuals show diverse obesity levels, implying other unobserved, possibly non-spatial or softer policy factors. While spatial fixed effects explain much variation, further research into these complex determinants is recommended.

70]:
```
# Aggregate spatial fixed effects and mean residuals by ward_code
spatial_effects =
result_fe.estimated_effects.groupby(level='ward_code').mean().squeeze()
median_residuals = result_fe.resids.groupby(level='ward_code').median()

# Calculate mean obesity rate and get ward names
obesity_rates = panel_data.groupby('ward_code')['obesity_year6'].mean()
ward_names = panel_data.groupby('ward_code')['ward_name'].first()

# Prepare dataframe combining all key variables
plot_df = pd.DataFrame({
    'ward_code': spatial_effects.index,
    'ward_name': ward_names.reindex(spatial_effects.index),
    'spatial_effects': spatial_effects.values,
    'median_residuals': median_residuals.reindex(spatial_effects.index).values,
    'obesity_rate': obesity_rates.reindex(spatial_effects.index).values
})
```

```
# Define median thresholds to classify quadrants
med_spatial = plot_df['spatial_effects'].median()
median_residuals = plot_df['median_residuals'].median()

def assign_quadrant(row):
    if row['spatial_effects'] >= med_spatial and row['median_residuals'] >=
median_residuals:
        return 'Q1: High Effect, High Residual'
    elif row['spatial_effects'] < med_spatial and row['median_residuals'] >=
median_residuals:
        return 'Q2: Low Effect, High Residual'
    elif row['spatial_effects'] < med_spatial and row['median_residuals'] <
median_residuals:
        return 'Q3: Low Effect, Low Residual'
    else:
        return 'Q4: High Effect, Low Residual'

plot_df['quadrant'] = plot_df.apply(assign_quadrant, axis=1)
```

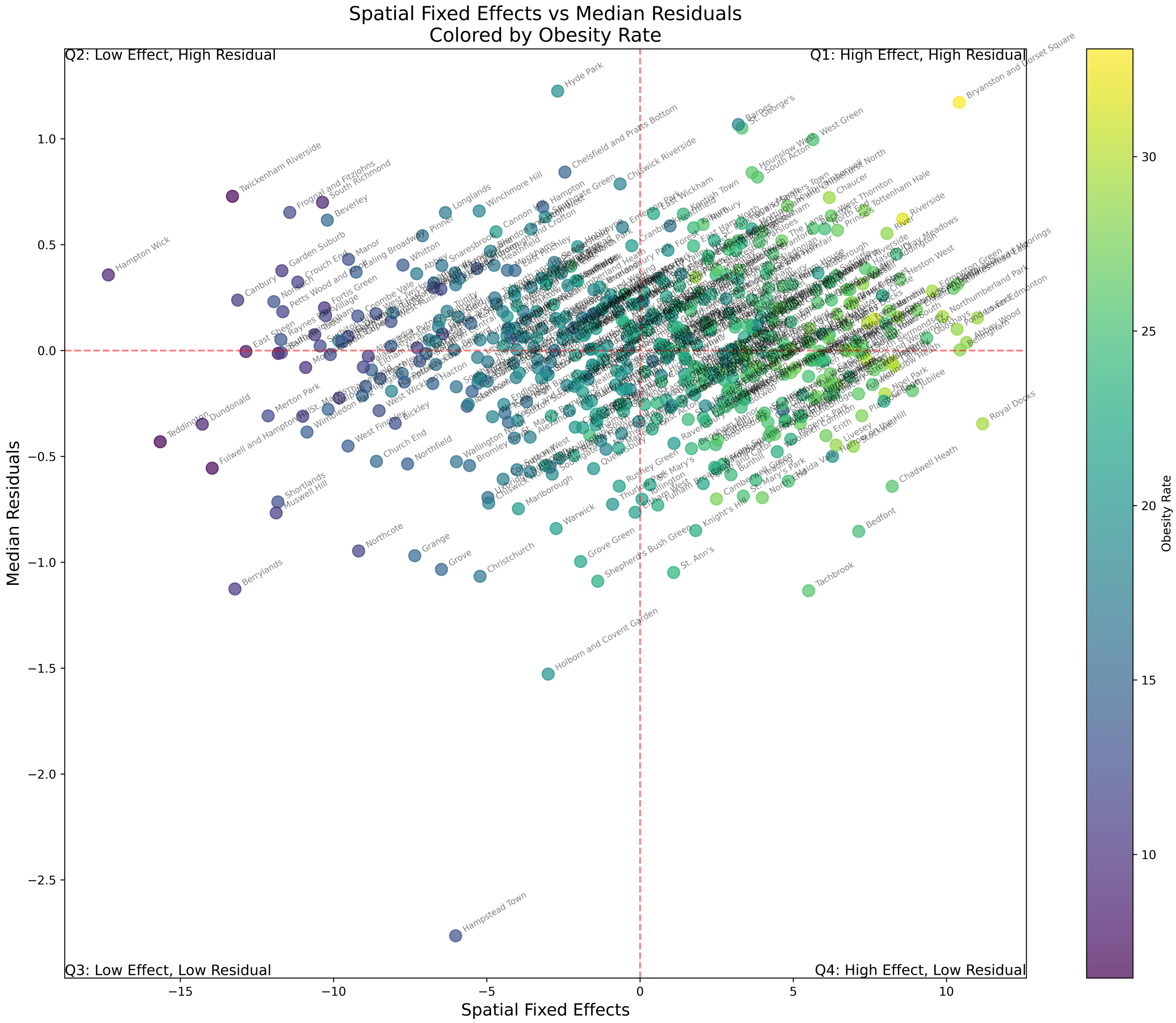


Figure 7. Quadrant Plot: Spatial Fixed Effects and Mean Residuals by Ward

7.0 Discussion and Conclusion

[go back to the top]

By leveraging differnet fixed effects alongside residual analysis, the impact of facilities on childhood obesity shows strong spatial heterogeneity across London wards and varies by age, while socioeconomic status exerts limited influence on this heterogeneity, regarding RQ1.

Consistent with Wyszyńska et al., 2020, older children exhibit heightened sensitivity to local facility availability, emphasising the need for age-tailored policies. Although socioeconomic indicators—particularly unemployment—remain key determinants of obesity prevalence, they do not significantly moderate the impact of facility provision. Greenspace emerges as a robust protective factor citywide, whereas the relationship between PTAL and obesity proves inconclusive, indicating complex underlying mechanisms that merit further investigation.

Spatial analysis, as illustrated in the quadrant scatter plot of spatial fixed effects versus median residuals, identifies wards with high obesity risk and limited intervention effects mainly in East London (notably Newham and Tower Hamlets) and parts of South West London. These areas demand integrated strategies extending beyond infrastructure enhancement to encompass behavioural interventions and community engagement - such as school-based nutrition education, family support services, and community. In contrast, Central and West London wards tend to show more favourable outcomes, reflecting the efficacy of current policy measures. Such spatial differentiation underscores the necessity for bespoke, ward-level approaches to optimise resource allocation and promote health equity.

Limitations

- The dataset spans 2009/10 to 2013/14, thus lacking insight into recent social and urban developments.
- Potential reverse causality between obesity rates and facility investment complicates causal inference.
- Absence of individual-level behavioural data limits elucidation of underlying mechanisms.
- Residual spatial autocorrelation remains unmodelled; future studies should incorporate spatial lag or error models to enhance estimation precision.

Building upon these findings, policymakers should prioritise facility investments particularly for older children and economically deprived wards, alongside intensifying greenspace provision across London. Employing spatial-temporal risk mapping enables tailored, locality-specific interventions. Future research should extend the temporal horizon, integrate behavioural determinants, and apply advanced spatial econometric techniques to refine causal understanding and optimise intervention outcomes.

8.0 References

[go back to the top]

Department of Health and Social Care. (2018). *Childhood Obesity: A Plan for Action*. [online] GOV.UK.

Available at: https://www.gov.uk/government/publications/childhood-obesity-a-plan-for-action [Accessed 29 Apr. 2025].

Ayala, G.X., et al., 2021. The social environment and childhood obesity. *International Journal of Environmental Research and Public Health*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8365653/ [Accessed 20 Apr 2025].

Danielli, A., Parker, L. and Swinburn, B., 2021. Systematic review into city interventions to address obesity. *International Journal of Environmental Research and Public Health*, 18(12). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7910670/ [Accessed 20 Apr 2025].

Lieb, D.C., 2009. Socioeconomic Factors in the Development of Childhood Obesity. *Current Obesity Reports*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5596457/ [Accessed 20 Apr 2025].

Public Health England, 2023. London's Local Health Profiles. London: Public Health England. Available at: https://fingertips.phe.org.uk/profile/health-profiles/data#page/4/gid/8000073/pat/15/par/E92000001/ati/6/are/E12000007/iid/90323/age/201/sex/4/cat/-1/ctp/-1/yrr/1/cid/4/tbm/1/page-options/car-do-0_tre-ao-0 [Accessed 20 Apr 2025].

Titis, E., 2023. Socio-economic correlates of childhood obesity in urban and rural England. *Public Health Nutrition*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10478054/ [Accessed 20 Apr 2025].

Wyszyńska, J., et al., 2020. Physical Activity in the Prevention of Childhood Obesity. *Frontiers in Pediatrics*. Available at: https://www.frontiersin.org/articles/10.3389/fped.2020.535705/full [Accessed 20 Apr 2025].

Yuan, Y., 2024. Public health interventions against childhood obesity in China and globally. *Preventive Medicine Reports*, 29. Available at: https://www.sciencedirect.com/science/article/pii/S2468266724002457 [Accessed 20 Apr 2025].

London Data Store (2011). *Statistical GIS Boundary Files for London*. Available at: https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london (Accessed: 16 April 2025).

London Data Store (2023). *Ward Profiles and Atlas*. Available at: https://data.london.gov.uk/dataset/ward-profiles-and-atlas (Accessed: 16 April 2025).

Office for National Statistics (ONS) (2023). *Ward Level Mid Year Population Estimates (Experimental)*. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/wardlevelmidyearpopulationestimatesexperimental (Accessed: 16 April 2025).

Active Places Power (2023). *Active Places Dataset*. Available at: https://www.activeplacespower.com/pages/downloads (Accessed: 16 April 2025).

Li, J., Heap, A. D., Potter, A., & Daniell, J. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, 53, pp.182–190. https://doi.org/10.1016/j.envsoft.2013.12.008

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (8th Edition). Cengage Learning.

Nau, C., Ellis, H., Huang, H., et al. (2019). Associations of multidimensional socioeconomic and built environment factors with body mass index trajectories among youth in geographically heterogeneous communities. *Preventive Medicine Reports*, 15, 100939. https://doi.org/10.1016/j.pmedr.2019.100939

Walker, P. G. T., et al. (2016). Estimating the most efficient allocation of interventions to reduce malaria burden and transmission. *The Lancet Global Health*, 4(7), e474-e484. Available at: https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(16)30073-0/fulltext30073-0/fulltext) [Accessed 27 April 2025].