## Correlates of mobility trajectory cluster membership

**<Contents>**

In the following analysis, we deployed the classification model to examine the significant factors that help us to interpret the clusters identified through the time-series clustering analysis. MLR LASSO is applied to assess the effects of socioeconomic characteristics between clusters, which grouped into similar trajectories of mobility reduction during the lockdown. MLR LASSO can take advantage of reducing the dimension of input features or variable selection. Some socioeconomic characteristics might be irrelevant to explain that interpreting trajectories of mobility over time. It also outperformed to eliminate the redundant collinear features to avoid multicollinearity (Gao et al., 2020). In MLR LASSO, estimated coefficients yielded the penalised coefficients for the standardised variables. A positive or negative sign is indicated as the direction of the relationship.

The model estimation results provided the estimated coefficients for selected factors between clusters. All explanatory variables were normalised to the range of [0,1] prior to model fitting. The best λ, penalty parameter, was selected by 10-fold cross-validation (CV). Seventeen factors remain through the embedded feature selection process in MLR LASSO. The share of high-income households, self-employed workers, and households with one vehicle were the most effective variables for the classification, yet the estimated coefficients varied between clusters (see Figure 1). The positive coefficients for the share of high-income households and self-employed workers were observed in G1 and G2. Based on the sign of coefficients, it can be inferred that those local authorities with particularly high levels of affluent populations and self-employed workers (are more likely to be low earners in London). Conversely, negative coefficients were indicated in G3 and G4 that high-income households and self-employed workers were negligible. In addition, the positive coefficients for the share of households with one vehicle were observed in G2 and G4, while negative coefficients were indicated in G1 and G3.

Table 1 shows the estimated coefficients for selected factors correlated with the classification of generated clusters. It was generally classified as G4, where the share of households with more than three vehicles, self-employed workers, and people in the lower middle class (i.e., social grade C1) were substantially lower. In contrast, G1 was more likely located in the local authorities; the share of high-income households, cumulative COVID-19 infection rates before the lockdown, and minority ethnic groups were comparatively high. Among the remaining clusters, high levels of the share of middle-income households, and more than three bedrooms in the house were more likely classified as G3. At the same time, G2 was more likely located in local authorities; the share of high-income households, lower middle class, and Black African were relatively higher.

* G1: Share of self-employed workers (0.71), households with one vehicle (0.35), cumulative COVID-19 infection rates before lockdown (0.24), high-income households (0.19), and percentage Other Black (0.12).
* G2: Share of high-income households (0.47), percentage Black African (0.43), households with more than three vehicles (0.39), more than three bedrooms in the house (0.30), and Hospital density per 1,000 population (0.15).
* G3: Share of medium-income households (0.38), social grade C1 (0.25), high-income households (0.23), more than three bedrooms in the house (0.25), households with one vehicle (0.04), and self-employed workers (0.02).
* G4: Share of households with more than three vehicles (0.64), self-employed workers (0.26), more than three bedrooms in the house (0.23), cumulative COVID-19 mortality rates before lockdown (0.20), and percentage Black Caribbean (0.19).

Chart, box and whisker chart

Description automatically generated

Figure . Estimated coefficients of explanatory variables to predict the clusters.

Table 1. Estimated coefficients of explanatory variables to classify clusters.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Domains | | Variable | G1 | | G2 | | G3 | | G4 | |
| Constant | | | -0.59 | | 0.41 | | 0.53 | | -0.34 | |
| Socioeconomic | Income | Share of households in the lowest household  income quintile at the national level |  | |  | |  | | 0.03 | |
| Share of households in median household  income quintile at the national level | -0.08 | |  | | 0.38 | |  | |
| Share of households in top household  income quintile at national level | 0.19 | | 0.47 | | -0.23 | | -0.19 | |
| Occupation | Share of Social Grade C1 (lower middle class) |  | | 0.3 | |  | | -0.23 | |
| Housing type | Share of dwellings with ≥3 bedrooms |  | |  | | 0.25 | |  | |
| Residential density | Resident population density  (1,000 inhabitants per km2) |  | | -0.02 | |  | |  | |
| Accessibility | Car availability | Share of households with 1 vehicle | -0.35 | | 0.04 | | -0.04 | | 0.18 | |
| Share of households with ≥3 vehicles |  | | 0.39 | |  | | -0.64 | |
| Clinical capacity and Allowed premises | Hospitals (per 1,000 inhabitants) | 0.07 | | -0.15 | |  | |  | |
| Activity commitment | Economic activity | Share of self-employed workers in the resident  population aged 16-74 | 0.71 | | 0.02 | | -0.02 | | -0.26 | |
| Perceived risk of COVID-19 | Infection rates | Cumulative COVID-19 reported cases per 100,000  resident population before lockdown | 0.24 | |  | |  | | -0.01 | |
| Mortality rates | Cumulative COVID-19 reported deaths per 100,000  resident population before lockdown |  | |  | |  | | -0.2 | |
| Ethnic composition | Bangladeshi | Percentage Bangladeshi | 0.08 | |  | |  | | -0.06 | |
| Other Asian | Percentage Other Asian |  | | 0.06 | | -0.01 | |  | |
| African | Percentage Black African |  | | 0.43 | |  | |  | |
| Caribbean | Percentage Black Caribbean | 0.09 | |  | |  | | -0.19 | |
| Other Black | Percentage Other Black | 0.12 | |  | |  | | -0.03 | |
| Model criteria information | | | | | | | | | | |
| 10-fold cross validation accuracy (%) | |  | | 44.4 | | 15.8 | | 55.2 | | 70.1 |
| AICc (Akaike's Information Corrected Criterion) | | -222.91 | | | | | | | | |
| BIC (Bayesian information criterion) | | -161.18 | | | | | | | | |