Tracking the Transit Divide: A Multilevel Modelling Approach of Urban Inequalities and Train Ridership Disparities in Chicago

Danial Owen *, Daniel Arribas-Bel and Francisco Rowe

Geographic Data Science Lab, Department of Geography and Planning, Roxby Building, University of Liverpool, Liverpool L69 7ZT, UK
* Correspondence: d.w.owen@liverpool.ac.uk

Abstract: Using a multilevel modelling approach, this study investigates the impact of urban inequalities on changes to rail ridership across Chicago’s “L” stations during the pandemic, the mass vaccination rollout, and the full reopening of the city. Initially believed to have an equal impact, COVID-19 disproportionally impacted the ability of lower socioeconomic status (SES) neighbourhoods to adhere to non-pharmaceutical interventions: working-from-home and social distancing. We find that “L” stations in predominately Black or African American and Hispanic or Latino neighbourhoods with high industrial land-use recorded the smallest behavioural change. The maintenance of higher public transport use at these stations is likely to have exacerbated existing health inequalities, worsening disparities in users’ risk of exposure, infection rates, and mortality rates. This study also finds that the vaccination rollout and city reopening did not significantly increase the number of users at stations in higher vaccinated, higher private vehicle ownership neighbourhoods, even after a year into the pandemic. A better understanding of the spatial and socioeconomic determinants of changes in ridership behaviour is crucial for policymakers in adjusting service routes and frequencies that will sustain reliant neighbourhoods’ access to essential services, and to encourage trips at stations which are the most impacted to revert the trend of declining public transport use.

Keywords: urban inequalities; multilevel modelling; spatial patterns; COVID-19; mobility; health inequalities

1. Introduction

Public transport is one of the most important city services, connecting citizens to essential services including their workplaces, education, and healthcare. Already facing declining ridership levels in several major cities since the mid-2010s in the United States (U.S.) [1,2], public transport experienced the greatest direct impact of all modes of transport during the COVID-19 pandemic [3]. This decline in use, particularly during COVID-19, has not been homogenous across social groups. Studies have revealed that lower socioeconomic status (SES) neighbourhoods and minority groups experienced fewer behavioural change in ridership [2,4] due to a greater reliance on the mode of transport and due to the necessity of trips made. Journeys made on public transport during the pandemic were largely driven by necessity and made by “essential workers”, who were required to work in-person and less able to travel by private modes of transport due to reduced car ownership rates [5–8].

Disparities in transit ridership between social groups subsequently exacerbated existing health inequalities. Some of the most vulnerable groups who were originally deemed “essential” were increasingly exposed to COVID-19 by being less able to adhere to non-pharmaceutical interventions (NPIs) such as working from home (WFH) and performing social distancing [9]. These privileges, according to Nanda [9], were reserved only for the well-off. Whilst changes in public transport use across different social groups were measured during the early stages of the pandemic, it is unclear whether, how, and if these...
changes will persist beyond the pandemic as changing travel and working behaviours become more permanent features in the urban landscape [4].

This study applies a multi-level modelling approach to explore the underlying mechanisms of changes in ridership behaviour across stations in Chicago, particularly following the full reopening and mass vaccination rollout in 2021. A two-level multilevel modelling approach is applied, given the nature of the data, where repeated measurements are structured into groups (transit stations), and changes in ridership are expected to vary between groups. Multilevel modelling is applied, rather than GIS analysis, as this study hopes to discover not only where disparities in ridership exist, but to explore and to measure the variation in ridership caused by station-level characteristics, the extent to which disparities in ridership changes occur between transit stations in Chicago, and how these change over time.

The main contribution of this study is how it extends the work of Hu and Chen [2] and Osorio et al. [8], who quantified the influence of sociodemographic and station characteristics such as race and land-use, respectively, on public transport ridership in Chicago during the early stage of COVID-19. This study extends the current research in two domains: firstly, by adding to the understanding of influencing factors on ridership by including additional variables such as access to private vehicles and vaccination data, which is now possible due to data availability. Secondly, we extend the study period, examining the persistence in differences in transit ridership between socio-economic groups beyond the vaccination rollout and the reopening of Chicago.

This will help to examine the differences in transit ridership at different stages of Chicago’s recovery and the extent to which stations’ transit ridership levels have returned to pre-pandemic levels. The expected outcome of this period is unclear: on the one hand, we expect ridership to recover in commercial and institutional areas as non-essential retail and schools reopen. However, we also suspect that ridership in higher SES neighbourhoods may fail to recover as this population becomes more adapted to remote and hybrid working and more able to access private modes of transport. A better understanding of the changes in public transit across socio-economic groups is key to ensure appropriate transit provision and functioning of the urban transport network.

This study measures the influences of race, land-use, and other contextual factors including the ability to work from home and vaccination status on changes in Chicago’s elevated railway, coined “L” train, ridership levels at 139 stations between February and December 2021. We focus on “L” train ridership behaviour, as it is the fourth-largest rapid-transit system in the U.S. and spans across many of Chicago’s different types of neighbourhoods [2,10]. As the service frequencies and routes did not change during COVID-19, this study can effectively measure the impact of “L” stations’ socioeconomic and land-use characteristics on changes in ridership behaviour.

The rest of the paper is structured as follows: Section 2 will review the literature on the influence of urban inequalities and COVID-19 on mobility behaviour, and the history of residential segregation and “L” train use in Chicago. Section 3 will introduce, explain, and justify the modelling approach, covariates, and data used in this study. Sections 4 and 5 will then describe, interpret, and conclude the results and relate the findings to the literature.

2. Literature Review
2.1. Urban Inequalities, COVID-19, and Mobility Behaviour

Urban inequalities, manifested in residential segregation, are the unequal access to resources and capabilities [11,12]. It determines certain social groups’ power over space and time [13], and ability and ambition to expand and travel for work, healthcare, education, and recreational purposes [14]. Following a new economy of information products rather than manufactured goods in the 1990s, urban inequalities have increased, and urban centres have become increasingly exclusionary as this new economy is suited more towards the highly educated, high-paid white collar worker [12]. Consequently, lower SES neighbourhoods are spatially suppressed and maintain or increase the concentration
of poverty, unemployment, and poor healthcare. Residential segregation also creates an imbalance in the power and ability of individuals to travel over space and time which closely follows social and racial lines [12,13,15].

Lower SES, inner-city neighbourhoods have been disadvantaged in their ability to travel over space and time since the 1950s and the introduction of the federal subsidies for highway construction [14]. The construction of highways and cheap suburban housing subsequently encouraged car-dependent cities in the U.S., which predominately favoured White, suburban neighbourhoods and connected them to downtown city areas [10,13]. The historical favouring of certain groups continues following 2010 and COVID-19, as the use of private vehicles has increased [1,13,16]. The increasing use of private vehicles typically benefits more affluent neighbourhoods, as lower SES neighbourhoods have far fewer car ownership levels [14]. As discussed in Credit et al. [17], private vehicle ownership is found to be correlated with job and economic opportunities, and is one of the biggest non-financial barriers to healthcare. Consequently, lower SES neighbourhoods rely more on public transport, which is described as the mode of transport for the urban poor [15]. This increased reliance on shared modes of transport is a key influencing factor in health inequalities during the COVID-19 pandemic.

In the early stages of the pandemic, COVID-19 was seen to be a ‘great equalizer’ and its transmission and impact would not discriminate [18,19]. However, it is now clear and widely reported in the literature that COVID-19 would exacerbate inequalities and economic and health disparities. Racial and ethnic minorities were disproportionally impacted by COVID-19, with much a higher risk of infection and mortality [18]. Importantly, COVID-19 did not create new health and economic disparities; rather, it accentuated existing underlying weaknesses [6].

In early attempts to suppress the spread of the virus in 2020, two NPIs, social distancing and WFH, uncovered groups of different SES’ abilities to adhere and to limit the spread of COVID-19. Lower SES neighbourhoods were more likely to work in-person and have lower rates of car ownership [7,8,14]. Therefore, despite an initial modal shift towards private modes of transport and active travel, lower SES neighbourhoods were less able to do so [3,20,21]. Consequently, trips made on public transport during the pandemic were largely driven by necessity and the use of public transport in the early stages of the pandemic resulted in significantly longer travel time and increased risk of exposure to the virus [7,15,18].

The literature on COVID-19 grew faster than any previous epidemic or pandemic, and there is an abundance of studies on the impact on mobility behaviour, where researchers originally attempted to use mobility data to monitor and predict the spread of the virus and the public’s compliance to the introduction of NPIs such as travel and social restrictions [22–24]. Early research would focus on measuring the overall decrease in travel with relatively fewer studies specifically researching decreases in travel with regards to socioeconomic factors and urban inequalities [25]. Soon after, it was realised that the decrease in mobility was not homogeneous between social groups and that unpacking differences in changes in mobility behaviour is important as it uncovers the inaccuracy of the initial “great leveller” rhetoric and how urban inequalities determined various groups’ vulnerability and exposure to the virus [26].

By now, studies have explored the underlying mechanisms of disparities in public transport use in cities, particularly during the early stage of the pandemic [2,4,6–8,25,27,28]. Studies consistently report that the smallest changes in public transport ridership are in lower SES neighbourhoods and for minority groups, and that the greatest change during the early stages is in areas with a higher proportion of commercial land-use, and higher-income, White individuals [2].

Largely due to data availability, fewer studies have explored the disparities in ridership during the early stages of recovery from COVID-19 and whether these disparities continued beyond the reopening of urban centres and vaccination schemes. It is important to explore how Chicagoans travel to work, to school, and to shop beyond the first year of the pandemic.
to substantiate the expectation that transit ridership would largely return once the city was fully reopened [8]. This study will contribute to this body of literature, particularly to the two studies in Chicago who use Bayesian structural time series models to infer the impact of COVID-19 on ridership [2,8]. Our study extends the timeline of the research on the subject and investigates if the reopening of Chicago and the vaccination rollout enticed a return to public transport use in more commercial, affluent, White neighbourhoods.

A better understanding of the underlying mechanisms of how the pandemic and post-pandemic impacts certain population groups’ mobility behaviour [26] is particularly important, as it would help to improve future disaster preparedness plans and targeted public transport service times and routes to ensure that the most vulnerable are protected from becoming isolated and immobile [8,26]. Historically, due to a lack of evidence, cuts to public transport services, before and during COVID-19, were made arbitrarily [1,2]; therefore, a better understanding of the spatio-temporal use of public transport systems would allow transport policymakers to more adequately adjust services to reduce the impact on lower SES neighbourhoods.

2.2. Residential Segregation and Chicago’s L-Train

Residential segregation is the spatial separation of social groups [29]. The partitioning of social groups into defined neighbourhoods exacerbates spatial inequalities where economic, racial, and social inequalities are spatially suppressed [10]. In the 1960s in the U.S., segregation and discrimination were supposedly abolished following the Civil Rights Movement. However, the influence of social and racial segregation on public transport remains today [10,13].

Residential segregation influences the spatial and temporal use of public transport, where certain racial and ethnic groups’ public transport use is defined by the day of the week, time of day, and the place of the train [10]. This is particularly apparent in Chicago as it is recognised as one of, if not the most, segregated city in the U.S. [10,30]. The high degree of residential segregation in Chicago makes it possible for this study to explore the differences in changes in public transport use across neighbourhood boundaries and social groups.

With highly racially segregated communities in Chicago, “L” stations distinctly represent neighbourhoods’ race and class [10]. The most prominent segregation in Chicago’s metro area and along the “L” train service route is the Black–White segregation along the north–south red line, the busiest line in the city. Along this line, Black or African Americans are heavily concentrated in the south and more affluent, White neighbourhoods are concentrated in the north (Figures A1 and A2) [10,31]. Historically, segregation between Hispanic or Latinos and White people is less; however, residential segregation is still prominent, where Latino families reside in the southwest and northwest of Chicago [31]. Social groups seldom interact or encounter another due to the high degree of segregation along the “L” train service route [10]. The functionality of each station, particularly passengers’ time and day of travel, purpose of travel, and reliance on the mode of travel is strongly determined by the neighbourhood characteristics and will determine how and when public transport was used during the pandemic. This study hopes to evaluate the influences of these station service area characteristics, residential segregation, and the subsequent inequalities on mobility behaviour, which has generally received less attention in the research literature [17].

3. Materials and Methods

3.1. Modelling Approach and Unit Structure

This study adopts a two-level multilevel modelling (MLM) approach to determine the extent of change in “L” train ridership across 139 stations over an 11-month study period between February and December 2021. The motivations for implementing MLM consist of causal inference, the study of variation, and for future predictions [32]. In this study, we propose a two-level MLM to explore, quantify, and explain the variation in
ridership between stations. Using MLM, we can explore the differences in station-level ridership levels during the COVID-19 pandemic and how each station’s rate of change over time varies. MLM is used when data are structured into groups and the coefficients vary by group: the existing literature justifies the use of MLM, as it outlines heterogeneous ridership loss during the pandemic, which is highly influenced by socio-demographic characteristics [8].

The data for the longitudinal study consist of repeated measurements: monthly percentage changes in ridership at each “L” station over an 11-month period. A two-level MLM is applied, where the Level-1 units (individual monthly station observations) are represented by \(i\) and are nested within Level-2 units (i.e., “L” stations), which are represented by \(j\) (Figure 1).

![Figure 1. Unit diagram of the longitudinal data.](image)

### 3.2. Model Development

Seven models, incrementally increasing in complexity, were implemented before deciding on the Final Model. The significance of the improvement of each model was tested using a Likelihood Ratio (LR) test (a drop in deviance greater than the critical value at the 5% level of 3.84) and only statistically significant covariates were retained in subsequent models. The seven models comprise a mixture of models including unconditional linear regression and variance-component models (null models), random-intercept models, and random-coefficient models (Table 1). Model development and the covariates introduced are discussed in more detail in Sections 3.2 and 3.3, respectively.

**Table 1. Summary of models and model covariates.**

<table>
<thead>
<tr>
<th>Model</th>
<th>Covariates</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>None</td>
<td>Unconditional Linear Regression</td>
</tr>
<tr>
<td>Model 2</td>
<td>None</td>
<td>Variance-Component Model</td>
</tr>
<tr>
<td>Model 3</td>
<td>Time Trend</td>
<td>Random-Intercept</td>
</tr>
<tr>
<td>Model 4</td>
<td>Time Trend</td>
<td>Random-Coefficient</td>
</tr>
<tr>
<td>Model 5</td>
<td>Time Trend; Race and Ethnicity</td>
<td>Random-Coefficient</td>
</tr>
<tr>
<td>Model 6</td>
<td>Time Trend; Race and Ethnicity; Land-Use;</td>
<td>Random-Coefficient</td>
</tr>
<tr>
<td>Model 7</td>
<td>Housing, Health, and Economic Characteristics</td>
<td>Random-Coefficient</td>
</tr>
<tr>
<td>Final Model</td>
<td>Time Trend; Race and Ethnicity; Land-Use; Housing, Health, and Economic Characteristics</td>
<td>Random-Coefficient</td>
</tr>
</tbody>
</table>

#### 3.2.1. Null Models

The first two models implemented, unconditional linear regression (Equation (1)) and variance-component model (Equation (2)), include no covariates and are the simplest forms of regression (Table 1). The intercept, the global mean of the response variable across all stations over the entire study period, is represented by \(\beta_0\), and \(c_{ij}\) and \(u_i\) represent the Level-1 and Level-2 residuals, respectively. Model 2 (Equation (2)) is the simplest form of random-intercept model and the individual station-level residuals \((u_i)\), which represent the difference between the global mean and each station’s mean, can be extracted. These two “null” models are implemented to decompose the total residuals and account for the dependency in the data, which is how the percentage change in ridership is dependent on station-level factors \((u_i)\). The extent of the dependency or clustering in the response
variable can be measured using a variance partition coefficient (VPC). VPC measures the proportion of the total variance that is explained by station-level factors, where a higher VPC indicates a greater degree of clustering in the response variable and justifies the use of MLM.

\[ \text{Percentage Change Ridership}_{ij} = \beta_0 + e_{ij} \] (1)

\[ \text{Percentage Change Ridership}_{ij} = \beta_0 + u_j + e_{ij} \] (2)

\[ \text{VPC}_u = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} \] (3)

3.2.2. Time Trend Models

The first covariate to be introduced in Model 3 (Equation (4)) and 4 (Equation (5)) is a time trend, a sequential variable which starts at 0 and increases monotonically to measure the effect of an additional unit of time on the response variable (Table 1). The intercept in both models, \( \beta_0 \), now represents the average percentage change in ridership across all stations where the time trend equals to 0, which is the first month of the study, February 2021. In Model 3, the time trend is introduced as a fixed effect which assumes that the rate of change in ridership is uniform across all stations over the 11-month period. Subsequently, Model 4 introduces the time trend as a random coefficient, where the rate of change in ridership between stations varies.

\[ \text{Percentage Change Ridership}_{ij} = \beta_0 + \beta_1 n_{\text{month}}+ u_j + e_{ij} \] (4)

\[ \text{Percentage Change Ridership}_{ij} = \beta_0 + \beta_1 n_{\text{month}} + u_{0j} + u_{1j} n_{\text{month}} + e_{ij} \] (5)

3.2.3. Random-Coefficient Model with Covariates Models

The remaining three models are extensions of Model 4 (Equation (5)), incrementally adding Level-2 covariates represented by \( X \) in Equation (6): Model 5 controls for race, Model 6 controls for land-use, and Model 7 controls for all remaining explanatory variables including the percentage of individuals that have received at least one dose of the vaccination, percentage with access to two private vehicles, percentage WFH and with access to facilities such as desktops or laptops, and the percentage of unemployment (Table 1).

Following Model 7, a Final Model is selected by retaining all statistically significant covariates introduced in all previous models. The Final Model does not introduce any further covariates and represents the statistically significant influences on spatial and temporal changes in ridership across L stations in Chicago between February and December 2021.

Covariates such as race and land-use are included, as they have been proven to be an influencing factor on changes in ridership in recent studies in Chicago [2,8]. The other contextual factors are introduced to represent the different forms of urban inequalities: the unequal access to resources, access, and capabilities [11]. Access to private vehicles reflects social groups’ uneven access to different modes of transport, percentage WFH and access to desktops or laptops reflects the disparities in those able to work from home, vaccination data represents the unequal opportunity and access to health facilities, and unemployment data reflects the unequal access to economic opportunities.

These covariates are also introduced based on recommendations from the literature: vaccination status data are now available and will be able to explain whether the rollout encouraged the public to return to shared modes of transport [8,16], and data on access to private vehicles, which are not available in some case studies [25], will measure the impact of car ownership and access to better modal choices on transit ridership. The introduction of these covariates, along with information on those with the facilities to WFH is important.
to attempt answering the open question on whether the significant ridership drop will continue beyond COVID-19 [8].

\[
\text{Percentage Change Ridership}_{ij} = \beta_0 + \beta_1 n_{\text{month}} + \beta_2 X_j + u_{ij} + u_1 n_{\text{month}} + e_{ij}
\]  

(6)

3.3. Data, Study Location, and Study Period

3.3.1. Response Variable

The response variable in this study is calculated using data obtained from the Chicago Data Portal. The data consist of monthly “L” station entries for over 140 “L” stations across Chicago, beginning in 2001 [33]. In this study, 139 stations were used and the percentage change in ridership for a given month is calculated from a five-year, pre-COVID-19 baseline for that respective month. The percentage change in train ridership is studied over an 11-month period from February to December 2021. This period encompasses key events in Chicago’s recovery programme which are expected to influence public transport ridership, such as the vaccination rollout which started in April 2021 and the full reopening of the city on 11 June [34,35].

Chicago was selected for the study as the data which were made openly available traverse many of Chicago’s neighbourhoods and normal service coverage and frequency were maintained throughout the pandemic, ensuring that changes in ridership were not influenced by service changes [8]. We highlight that our study focuses on train ridership; therefore, it does not capture changes in other modes of transport [15].

3.3.2. Supplementary Data and Station Service Areas

The covariates used to capture “L” station characteristics are all openly available and obtained from the American Community Survey (ACS), the Chicago Metropolitan Agency for Planning (CMAP), and the Illinois Department of Public Health (IDPH). These provide information on contextual factors such as race and ethnicity, land-use, and vaccination status in Chicago, respectively.

Demographic, economic, and housing characteristics from the ACS’ 2020 5-year estimates are used to explain changes in train ridership and were obtained, determined by availability, at the census tract level or the zip-code level. At the census tract level, this study uses datasets B03002 and S2801 to measure demographic and housing characteristics with estimates of race and access to computers and internet subscriptions [36,37]. These datasets are used to calculate the proportion of each race or ethnicity and the proportion with access to desktops or laptops in each station service area. At the zip-code level, dataset DP03 is used to capture economic characteristics, including the percentage of those WFH and those unemployed [38].

Chicago’s land-use data are obtained from CMAP’s 2015 land-use inventory, which is openly available [39]. At the time of the study, this inventory was the most up to date information on parcel-level, land-use data, categorizing parcels into nearly 60 categories. In this study, the individual categories were aggregated, according to CMAP metadata, into nine main domains: residential, commercial, institutional, industrial, transportation, agriculture, open-space, vacant, and ‘other’ (Figure A3).

Vaccination status data published by IDPH are updated weekly and report the percentage who have received one dose for each zip code in Illinois. This study was unable to obtain historical vaccination status data for the study period; therefore, the data used in this study include the vaccination status for each zip code in Illinois, updated as of 19 October 2022 [40].

For each “L” station, the average value for each factor within an 800 m walking-distance service area was used as a covariate in our models. The service area is created from a road-network buffer, rather than a circular buffer which is used in most existing studies (Figure A3). The road-network buffer more accurately captures each stations’ service area [41] and is defined as 800 m, as it is typically used as the upper limit of where
most users will travel to access rail stations by foot [42]. Due to the simplicity of buffers, they do induce bias where the influence of neighbourhoods beyond the 800 m buffer is not considered [41]. This could influence the interpretation of the results, particularly with the segregated nature of Chicago’s neighbourhoods, as mainly African American neighbourhoods live near rail stations [10].

4. Results

4.1. Null Models and Time Trend

Models 1 and 2 are the simplest forms of regression models, modelling the percentage change in ridership throughout the study period across all 139 stations, without covariates (Table 2). Both models show that the average percentage change in “L” train ridership over the study period across all stations, relative to a five-year baseline, was 64% below pre-pandemic levels.

Table 2. Model results for both ‘null’ and time trend models.

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est. SE</td>
<td>Est. SE</td>
<td>Est. SE</td>
<td>Est. SE</td>
</tr>
<tr>
<td>Fixed Effects:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-64.12 *** 0.28</td>
<td>-64.12 *** 0.61</td>
<td>-75.49 *** 0.63</td>
<td>-75.49 *** 0.72</td>
</tr>
<tr>
<td>n_month</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2.27 *** 0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.27 *** 0.07</td>
</tr>
<tr>
<td>Random Effects:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Station Level:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between-Station Variance</td>
<td>-</td>
<td>44.59</td>
<td>49.76</td>
<td>68.97</td>
</tr>
<tr>
<td>Intercept-Slope Covariance</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-3.30</td>
</tr>
<tr>
<td>N-month Variance</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.57</td>
</tr>
<tr>
<td>Observational Level:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual Variance</td>
<td>119.37</td>
<td>74.70</td>
<td>17.85</td>
<td>11.53</td>
</tr>
<tr>
<td>Deviance</td>
<td>11,650</td>
<td>11,216</td>
<td>9226</td>
<td>8878</td>
</tr>
</tbody>
</table>

Notes: Statistical significance: *** = \( p < 0.001 \). Est. refers to the estimate and SE indicates the standard error of the estimate.

The intercept does not change between Models 1 and 2; however, the standard error for the intercept in Model 2 does increase slightly. This is because the linear regression model (Model 1) assumes that all observations are independent, resulting in over-confidence and potentially leading to incorrect conclusions, whereas the variance component model (Model 2) accounts for the dependency in the data: repeated measurements are nested within stations. Table 2 indicates that 37% of the total variance is explained by station-level differences (VPC = 0.37) and indicates a notable degree of clustering and dependency in the data. The presence of clustering in the data and a significantly positive LR test justify the use of MLM.

As 37% of the total variance is explained by between-station variation in Model 2, the remaining 63% of the total variance is explained by within-station differences. As a longitudinal study, this within-variation is mostly assumed to be differences between monthly values of percentage change in ridership across the study period. To account for this, Model 3 incorporates the time trend variable, \( n_{month} \), representing each month in the study as a unit of time, increasing sequentially from 0 to 10. With the introduction of the time trend covariate, the intercept now represents the average value for the percentage change in ridership across all stations in February 2021.

Model 3 shows a notable decrease in the intercept relative to Models 1 and 2, suggesting an average percentage change in ridership of \(-75.5\%\) in February 2021, relative to pre-pandemic levels (Table 2). Throughout the study period, ridership increases 2.27% for each passing month (Table 2). However, in Model 3, the introduction of the time trend as a fixed effect assumes that the slope, the rate of change in ridership over time, is uniform.
across all stations. The evidence in the literature provided in Section 2 clearly indicates diverse changes in public transport ridership behaviour between social groups; therefore, assuming a uniform increase in ridership across stations would be naïve.

Consequently, Model 4 introduces \( n_{\text{month}} \) as a random coefficient which allows each station’s slope, the rate of change in ridership over time, to vary between stations (Table 2). Table 2 shows a heterogeneous change in public transport ridership use over the study period, between stations, in Chicago. A negative intercept-slope covariance implies a ‘fanning-in’ pattern in the data, where stations with a higher intercept value typically observe a lower slope value, and vice versa. Therefore, stations which experienced the biggest change in ridership at the beginning of the study period experienced a steeper rate of increase in ridership over time.

The random intercept and random slope values in Model 4 represent the difference between each station’s percentage change in ridership in February 2021 and the rate of change over time, respectively. To explore the spatial relationship between train ridership and the rate of change over time, these values can be extracted and mapped using a bivariate CMAP (Figure 2). To increase the interpretability of Figure 2, the intercept \( (\beta_0 = -75.49\%) \) is added to all random intercept values, and the random coefficient values are added to the mean rate of change over time \( (\beta_1 = 2.27) \). The bivariate CMAP is annotated to indicate the number of stations assigned to each cell (Figure 2).

Figure 2. Spatial distribution of “L” stations’ random effects: intercept and slope. The intercept and slope represent percentage change in ridership at the beginning of the study period and the rate of increase per month, respectively.
In Figure 2, the spatial distribution of Model 4’s random effects uncovers a broadly defined north–south divide in Chicago. “L” stations in the south and southwest of Chicago, predominately orange or green, suggest average or higher than average ridership in the beginning of the study and experience a less than average increase in ridership over the study period. Conversely, several “L” stations in Chicago’s central Loop and northern stations, coloured in pink or light blue, indicate stations with higher-than-average changes in ridership at the beginning of the study and average or above average rate of increase in ridership over the study period. The spatial clustering of the random effects, intercepts, and slopes in Chicago suggests that changes in public transport ridership behaviour during the pandemic and throughout recovery programmes are spatially and socially dependent. The introduction of contextual covariates, population, and land-use characteristics are introduced in the next section to explain and measure the extent of these possible influencing factors on disparities in ridership behaviour.

4.2. Random-Coefficient Models with Covariates

The first contextual factors to be controlled for in this study, in Model 5, are race and ethnicity, where each covariate represents a race or ethnicity’s percentage of the station service area’s total population. The covariates included in Model 5 include all races and ethnicities, other than the percentage of White people which is used as a reference category (Table 3). The intercept for Model 5 represents the response variable, where all covariates are equal to zero. Therefore, the intercept represents the average percentage change in ridership across all stations at the beginning of the study in predominately White neighbourhoods, which is $-81.3\%$, nearly 6% less than in Model 4 (Tables 2 and 3). Supported by findings in the literature, this decrease suggests that race has a notable influence on changes to ridership [2,8]. This can also be measured by the drop of just over 39 points in between-station variance relative to Model 4, which shows that race accounts for approximately 57% of the variance between stations at the beginning of the study period (Tables 2 and 3).

Of the five race and ethnicity groups introduced as covariates, only the effects of three were found to be significant: percentage of Black or African American, Hispanic or Latino, and American Indian or Alaskan Native (Native American). The coefficients for Black or African American and Hispanic or Latino indicate a significant, positive association with ridership and suggest that a 10% increase in each race or ethnicity would see a 1.3% and 1.8% increase in ridership, respectively. Despite a significant coefficient for Native Americans, the standard error is noticeably higher than for the other covariates and both the percentage of Native Americans and Hawaiians have an exceptionally large negative association with the response variable. This may be caused by the smaller sample size, as the maximum percentage of Native Americans in any “L” station buffer is less than 2%. Due to higher standard errors and non-significant coefficients, only the covariates on the percentage of Black or African American and Hispanic or Latino are retained in subsequent models.

In Model 6, station service areas’ land-use is controlled for where each covariate represents the proportion of station service areas covered by that particular land-use. Each land-use consists of two types of covariates: a centred covariate where a value of 0 represents the mean percentage cover for that land-use across all stations and a covariate which is an interaction term with the time trend variable. The interaction term measures the extent of how the effect of the coverage of that particular land-use on the response variable changes for each additional month.

Table 3 shows that, from the centred land-use covariates introduced in Model 6, only the proportion of commercial land-use had a negative association with ridership relative to the intercept at the beginning of the study in February 2021. Although non-significant, the model suggests that a station service area with 10% above average commercial land would experience more than a 1% decline in ridership. Conversely, the proportion of industrial and open-space land-use illustrates a strong, significant, positive association with ridership
at the beginning of the study, where a 10% above average coverage of these land-uses in a station service area would induce increases of 2.8% and 2%, respectively (Table 3).

Table 3. Model results when incorporating contextual covariates.

<table>
<thead>
<tr>
<th>Fixed Effects:</th>
<th>Model 5</th>
<th>Model 6</th>
<th>Model 7</th>
<th>Final Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>−81.29 **</td>
<td>1.53</td>
<td>−82.28 ***</td>
<td>0.98</td>
</tr>
<tr>
<td>n_month</td>
<td>2.27 ***</td>
<td>0.07</td>
<td>2.27 ***</td>
<td>0.06</td>
</tr>
<tr>
<td>Black or African American</td>
<td>0.13 ***</td>
<td>0.02</td>
<td>0.15 ***</td>
<td>0.06</td>
</tr>
<tr>
<td>Asian</td>
<td>0.01</td>
<td>0.90</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>0.18 ***</td>
<td>0.03</td>
<td>0.16 ***</td>
<td>0.02</td>
</tr>
<tr>
<td>Native American</td>
<td>−6.90 **</td>
<td>2.31</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>−7.81</td>
<td>8.92</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Commercial (Centred)</td>
<td>-</td>
<td>-</td>
<td>−0.11</td>
<td>0.08</td>
</tr>
<tr>
<td>Commercial (Interaction)</td>
<td>-</td>
<td>-</td>
<td>0.05 ***</td>
<td>0.01</td>
</tr>
<tr>
<td>Industrial (Centred)</td>
<td>-</td>
<td>-</td>
<td>0.28 **</td>
<td>0.11</td>
</tr>
<tr>
<td>Industrial (Interaction)</td>
<td>-</td>
<td>-</td>
<td>0.03 *</td>
<td>0.01</td>
</tr>
<tr>
<td>Institutional (Centred)</td>
<td>-</td>
<td>-</td>
<td>0.02</td>
<td>0.07</td>
</tr>
<tr>
<td>Institutional (Interaction)</td>
<td>-</td>
<td>-</td>
<td>0.04 ***</td>
<td>0.01</td>
</tr>
<tr>
<td>Open Space (Centred)</td>
<td>-</td>
<td>-</td>
<td>0.20 *</td>
<td>0.09</td>
</tr>
<tr>
<td>Open Space (Interaction)</td>
<td>-</td>
<td>-</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Residential (Centred)</td>
<td>-</td>
<td>-</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>Residential (Interaction)</td>
<td>-</td>
<td>-</td>
<td>0.02 ***</td>
<td>0.01</td>
</tr>
<tr>
<td>Desktop or Laptop (Centred)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Desktop or Laptop (Interaction)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Two Vehicles (Centred)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Two Vehicles (Interaction)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>One Dose (Centred)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>One Dose (Interaction)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Unemployed (Centred)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Unemployed (Interaction)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>WFH (Centred)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>WFH (Interaction)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Random Effects:

<table>
<thead>
<tr>
<th>Est.</th>
<th>Est.</th>
<th>Est.</th>
<th>Est.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between-Station Variance</td>
<td>29.88</td>
<td>26.95</td>
<td>23.34</td>
</tr>
<tr>
<td>Intercept-Slope Covariance</td>
<td>−1.07</td>
<td>−0.75</td>
<td>−0.34</td>
</tr>
<tr>
<td>n_month Variance</td>
<td>0.57</td>
<td>0.41</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Observational Level:

<table>
<thead>
<tr>
<th>Residual Variance</th>
<th>11.53</th>
<th>11.53</th>
<th>11.53</th>
<th>11.53</th>
</tr>
</thead>
</table>

Deviance | 8803 | 8754 | 8683 | 8691 |

Notes: Statistical Significance: * * * = p < 0.001, ** = p < 0.01, * = p < 0.05, and + = p < 0.10. Est. refers to the estimate and SE indicates the standard error of the estimate.

All land-uses’ interaction terms are positive, which indicates that the effect on ridership becomes increasingly positive for each additional month of the study. Despite a negative association with ridership in February 2021, the commercial interaction term, along with institutional land-use, has the strongest, significant, and increasing effect on ridership as the study period progresses. Industrial and residential interaction terms also show significant, positive associations; however, these relationships with the response variable are weaker. The only interaction term in Model 6 to not show a significant increasing association with the response variable over time is the percentage of open space.

In the final models, contextual covariates are introduced to explain the disparities in changes in ridership behaviour which were influenced: the percentage with access to desktops or laptops, access to two vehicles, received the first dose of the vaccination,
unemployed, and WFH (Table 3). Again, each covariate is represented by a centred covariate and an interaction term with the time trend. The Final Model represents the ultimate model where all non-significant covariates are omitted.

In February 2021, three of the five centred covariates have a significant, negative association with the response variable. According to the Final Model, a 10% increase in a station service area’s access to desktops and laptops, percentage of people that received at least one dose of the vaccination, and percentage unemployed would result in a decrease of 3%, 1.7%, and 14.6% in ridership at the beginning of the study, respectively. Despite a substantial impact on the response variable in February 2021, the influence of the percentage of unemployment on ridership behaviour does not significantly change over time. Table 3 shows that, for each additional month, there is an increasingly negative association between the response variable and the percentage with access to two private vehicles. Interestingly, the interaction term for access to desktops and laptops is significantly positive. Therefore, despite an initial negative association, for each additional month, this negative coefficient increases by 0.05, which means that, following the seventh month, there would be a positive association between access to desktops and laptops, and ridership. The interpretation of these findings is discussed in the next section.

5. Discussion and Conclusions

Towards the end of the pandemic’s first year in February 2021, Osorio et al. [8] detailed how the perception of risk was accountable for approximately 1% of rail ridership loss due to “caution fatigue” in Chicago. This suggests that the influences on changes in ridership extend beyond the perception of risk alone. Due to data availability, we can now uncover, in this section, the influences on transit ridership behaviour beyond the first year of the pandemic. Encompassing the reopening of non-essential retail and the vaccination rollout, this section will discuss the results in Section 4 and interpret the spatial distribution and impact of “L” station service area socioeconomic and land-use characteristics on the return to or avoidance of public transport. Our study recognizes that not all of the factors influencing public transport use can be accounted for and that those included in this study are based on suggestions from the literature. Future studies may extend our research by incorporating additional parameters, such as the impact of changing COVID-19 policy stringency on public transport use.

5.1. Race and Ethnicity

The findings of this study extend those in the literature, illustrating that, between February and December 2021, more than a year after the implementation of NPIs and the beginning of recovery programmes, the average ridership across all stations during this study period continued to be substantially lower, 75.5% below pre-COVID-19 levels. Similar to Osorio et al. [8], Brough et al. [4], and Hu and Chen [2], this study finds that this decrease in public transport use is spatially unequal. In this paper, we extend this information and have extracted and mapped each individual station’s change in public transport use and its rate of change over time (Figure 2).

Figure 2 illustrates a defined north–south divide, where lower SES neighbourhoods in the south and southwest with higher percentages of Black and Hispanic population (Figure A1) experience a smaller change in ridership behaviour in comparison to the more commercial and affluent “L’s” stations in Chicago’s Loop and northern stations, respectively. Figure 2 also captures the disparities in rate of changes in station entries over time between stations in Chicago. The negative intercept-slope covariance in Table 2 and Figure 2 indicates a “fanning-in” pattern, where northern and central stations in higher SES neighbourhoods and major commercial areas with lower initial ridership generally experienced an above average rate of increase in ridership over the study period.

The north–south divide illustrated in Figure 2 revealed how disparities in ridership behaviour during the pandemic may have disproportionately impacted minority groups’ infection and mortality rates during the pandemic in Chicago. Smaller changes in initial
ridership in the south and southwest (Table 3) showed predominately Hispanic or Latinos and Black or African American neighbourhoods’ lessened ability to practice social distancing and increased occupation of jobs that cannot be performed from home [2,5,6,9,19,43,44]. Smaller changes in ridership levels at these stations at the beginning of the study in February 2021, during higher stringency and potential infection periods suggest that a large proportion of “L” train journeys in these lower SES neighbourhoods is driven by necessity, perhaps to workplace zones [7].

5.2. Land-Use

A positive association between train ridership and the share of industrial land-use confirms a smaller behavioural change in ridership behaviour at workplaces in February 2021 (Table 3). This positive association may reflect the trips largely driven by necessity, as “essential” work at manufacturing and industrial centres mostly required staff to work in-person [8]. The continuation of station entries at industrial zones accentuates the disparities in certain social groups’ ability to work from home and perform social distancing. Subsequently, minority groups’ reliance on public transport exacerbates existing health disparities in lower SES neighbourhoods, as public transport use increases the exposure to COVID-19 and risk of infection [18,20,26,43].

Despite trips being largely driven by necessity, the Final Model (Table 2) also showed that the proportion of open greenspaces was also associated with limited change in train ridership, pointing to the perceived safety of the use of public transport for recreational visits to parks and open greenspace during the pandemic [45–47]. Consistent with findings in the early stages of the pandemic, the use of public transport for leisure trips to open greenspace differs from those to commercial areas, where, for trips to commercial centres, a larger decline in ridership behaviour was recorded [2]. Although non-significant, Model 6 showed that trips to commercial zones such as the Loop in Chicago were still far below pre-COVID levels in February 2021, nearly a year after the implementation of NPIs, as the reopening of all non-essential retail would not occur until June 2021 [33].

Surprisingly, our model estimates showed an insignificant association between commercial land-use’s interaction term and train ridership. This result may contradict findings in other cities and countries where leisure and social trips, including shopping, returned to more “normal”, pre-pandemic levels [48]. This finding may also reflect a modal shift, with people choosing other modes of transport, particularly private vehicles, and active travel such as walking or cycling, to move within Chicago.

5.3. Urban Inequalities

We now focus on the extent to which urban inequalities influenced changes in ridership behaviour across stations at the beginning of the study, and how this influence changes throughout the study period, following the mass rollout of the vaccination and reopening of Chicago. Overall, the results indicate a global increase in ridership across all stations; however, it is evident that the rate of increase across all stations is not homogenous.

During the early stages of the pandemic, research speculated whether ridership would return to “normal” levels following the introduction of the vaccination programme [8]. With vaccination data now available, this study finds a non-significant association between changes in rail ridership over the study period and vaccination rates (Table 3). Although they were speculated as a catalyst for the return to public transport, these results suggest an unchanging, potentially “new normal” of lessened public transport use in neighbourhoods with greater vaccination rates. Similarly, over the study period, higher access to private vehicles and WFH also showed a relative decrease in the use of public transport.

In the Final Model (Table 3), the negative association between private vehicle ownership and train ridership suggests that the modal shift to private modes of transport in the early stages of the pandemic may have become a more permanent feature of local mobility behaviour [4]. Two stations in Figure 2 exemplify this trend where ridership may not return to “normal” pre-COVID-19 ridership levels due to high ownership of private vehicles at
two “L” stations depicted in turquoise in eastern Chicago. Cumberland and Austin (Forest Park), the former in a predominately White, commercial neighbourhood and the latter in a mixed, residential neighbourhood, have some of the highest proportions of vehicle ownership of all stations. These stations illustrate where ridership was greatly impacted during the earlier stages of the pandemic and have failed to adhere with the general trend of convergence, where the stations which were the most impacted generally observed a stronger return to public transport use over the study period. This finding is important as it identifies where disparities in ridership have persisted and, unlike during other past pandemics and epidemics, such as during Middle East Respiratory System (MERS) where ridership levels returned to normal after six months [8], current ridership levels may not return to pre-COVID-19 levels, as the pandemic has permanently changed travelling and working behaviours.

Not all covariates indicate permanent changes to mobility and working behaviours. The interaction term for the access to desktops or laptops suggests a relative increase in public transport use, becoming positive in October 2021, despite the original negative association (Table 3). This suggests that not all work remained remote or hybrid and that, despite these households having the facilities and capability to WFH, household occupants do not inherently WFH and their occupation or other factors may not allow them to do so. This positive interaction term may capture a glimpse of the beginning of the long-awaited return to public transport services. Only future work will be able to explore and confirm the extent of changing mobility and working behaviours and their influence on public transport use.

5.4. Conclusions

Our results indicate an overall increase in “L” train ridership in Chicago for every month between February and December 2021. Yet, our findings show that rail ridership has remained far below pre-pandemic levels and reveal persisting disparities in ridership between stations following the inauguration of Chicago’s vaccination programme and full reopening. Over a year after the beginning of the pandemic, the recovery programme is yet to see a significant increase in train ridership at some higher-vaccinated, predominately White neighborhoods with greater access to private vehicles and ability to WFH. These findings reflect the persistence of inequalities in public transport use, different social groups’ changing mobility behaviours and new working practices, and the challenge to revert the trend of declining public transport use in the U.S. in recent decades.

These results highlight the need for continued efforts and innovative solutions to support the recovery of public transport in the post-pandemic era: helping transport planners to encourage trips at stations where the biggest change in ridership has persisted, whilst protecting the most vulnerable groups’ potential mobility and reducing the disparities in access to health, economic, and education services. During previous pandemics, incentive programmes have been introduced to revert the modal shift towards private modes of transport. Incentive programmes, such as discount programmes, promotional activities, and improved service quality and frequencies [8] have been introduced to encourage ridership which, as a result, help to mitigate climate change and contribute towards decarbonization [20]. However, incentive programmes following COVID-19 must be carefully considered, as increasing service frequencies may no longer entice those working from home and introducing discount programmes in some of the more affluent neighbourhoods, particularly during a cost-of-living crisis, may seem unfitting. We focused on studying ridership use along Chicago’s “L” train lines. These lines do not extend to all neighbourhoods, most importantly, Chicago’s lower SES neighbourhoods in the far southern or southwestern city limits [10]. Future work should consider the effects of the pandemic and recovery programmes on these neighbourhoods and those omitted from the study by examining the change in mobility behaviour across more or all modes of transport, including populations residing beyond this study’s “L” station service areas. Such analysis would enable a more
comprehensive understanding of overall changes in mobility behaviour and modal shifts after Chicago’s recovery programme across all SES groups and neighbourhoods.

**Author Contributions:** D.O.: Conceptualization, Methodology, Formal Analysis, Software, Writing—original draft, Visualization. D.A.-B.: Conceptualization, Writing—review & editing, Supervision. F.R.: Conceptualization, Writing—review & editing, Supervision. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by UKRI (ESRC), through the Data Analytics and Society CDT with grant number ES/P000401/1.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All datasets used in this study are openly available with links available in [33,36–40].

**Conflicts of Interest:** The authors declare no conflict of interest.

**Appendix A**

![Figure A1. Distribution of racial and ethnic minorities along CTA L stations and lines.](image-url)
Figure A2. Percentage below poverty line along CTA L stations and lines [49].

Figure A3. Example of 800 m, L station service area buffer and aggregated land-use types.
References

1. Dong, X. Investigating changes in longitudinal associations between declining bus ridership, bus service, and neighborhood characteristics. *J. Public Transp.* 2022, 24, 100011. [CrossRef]


10. Swyngedouw, E. The segregation of social interactions in the red line L-train in Chicago. *Symb. Interact.* 2013, 36, 293–313. [CrossRef]


47. Hino, K.; Asami, Y. Change in walking steps and association with built environments during the COVID-19 state of emergency: A longitudinal comparison with the first half of 2019 in Yokohama, Japan. *Health Place* 2021, 69, 102544. [CrossRef] [PubMed]


Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.