**University of Glasgow**
**School of Social and Political Sciences**
**Urban Studies**

# Public Transport Job Accessibility in Greater Manchester

**(Michael) Sam Allwood**
**2938711**

**Presented in partial fulfilment of the requirements for the degree of M.Sc. in Urban Transport**

**Word Count:** ***[14,658]***

**Awarded Grade B1: Very Good**

**August 2024**

## Abstract

Transport infrastructure is pivotal in shaping human societal interactions and has long been understood to facilitate economic development in urban centres. Some of the productivity benefits of larger cities have been attributed to the expedience with which people can move around, as access to job opportunities lubricates the labour market. For a long time, this meant building urban motorways, before their disbenefits were properly understood, but urban mobility advancements now focus primarily on public transport and active travel. This study explores the intersection of urban economics and public transport analysis to understand the relationship between public transport accessibility and employment outcomes as a proxy to economic prosperity in Greater Manchester Combined Authority. A novel public transport job accessibility index is developed using commuting travel data to establish contextualised decay constants and fixed effects regression utilised to derive a relationship with employment outcomes. Instrumental variables are deployed to suppress the pervasive endogeneity in employment and transport studies. The relationship is statistically significant and implies that improved expeditious public transport access to jobs reduces unemployment when controlling for ethnicity, qualification levels and access to private vehicles.

Keywords: public transport job accessibility, unemployment, instrumental variables, fixed effects regression, Greater Manchester Combined Authority

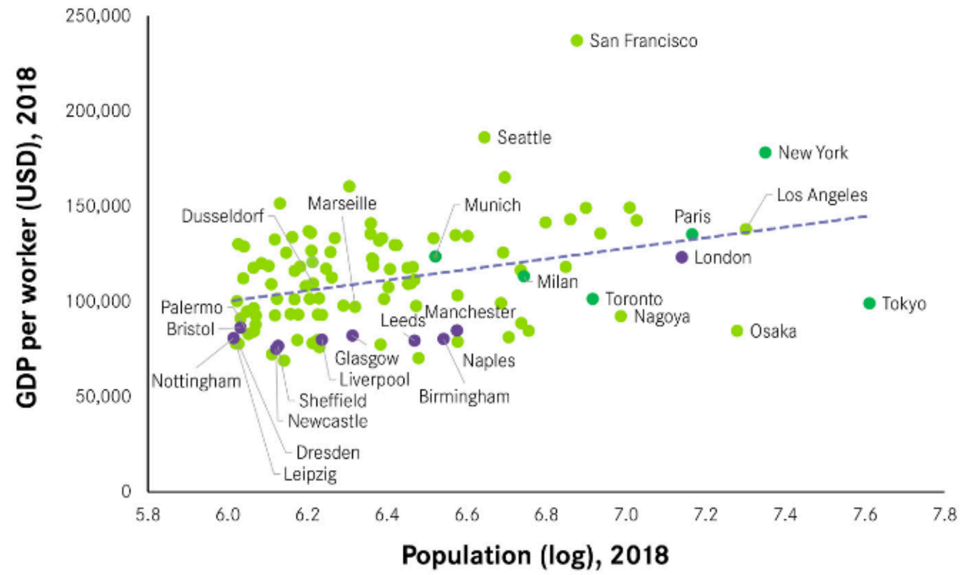
# Table of Contents

## List of Tables

## List of Figures

# 1 Introduction

Urban transport networks are capable of supporting economic growth of local economies by several mechanisms including lubrication of the labour market (NBER, 2022). A more fluid labour market has a positive effect on nominal wages (Galí, 2011) through expanded employment options for people and businesses. This results in higher productivity (measured as out per worker or hour worked), which is where the UK urban environment outside London lags behind the rest of the G7 group of nations. Of the 112 large cities across the G7, UK cities outside of London make up 9 of the 20 with lowest productivities, as shown in Figure 1-1. This productivity lag is a key economic indicator measured by the UK government (2024). Measuring levels of unemployment is an expeditious way to monitor the labour market and wider economy (UK Government, 2024). It is transparent and objective, and data are widely available at a granular level.




Source: OECD City Statistics 2018; ONS BRES 2018; NISRA Labour Market Report 2018; Eurostat City Statistics 2018; US Census Bureau, American Community Survey 2018; Statistics Canada, Census 2021; Japan Economic Census Table 9-1B, 2021.

*Figure 1-1. Productivity and Population from 2018 for the 112 large cities in the G7 group of countries.* (Breach and Swinney, 2024)

## 1.1 Private vs Public Transport Networks

Cities of the Global North have a long history of urban public transport networks. The industrial revolution replaced horse-drawn vehicles with steam powered ferries and underground trains and later electric streetcars moved through the pedestrian-friendly urban landscape. Access to public transportation had a strong influence on the expanding urban form through the 19th and early 20th centuries, and dense residential growth was guided along transit lines in what might be described today as Transit-Oriented Development (TOD) (Carr and Whitehand, 2001).

In the mid-20th century, the anthropological disbenefits of the private vehicle were not publicised, even if they were understood by some (Scientific American, 2015). This led to planning authorities prioritising the freedoms afforded by the private automobile in spatial planning strategies, allowing developers to build high-return, single-family suburban homes and sprawl into green belt land (Carr and Whitehand, 2001). Consequently, population densities in UK cities fell, which increased dependence on the automobile for mobility as public transit struggled to serve the less dense neighbourhoods, establishing a negative feedback loop as less-used transit became more expensive to run, warranting cuts to services.

US cities have achieved high productivities with very low public transport ridership, as is demonstrated in Figures 1 and 2. However, their urban population densities are much lower than European cities (Qiang, Xu and Zhang, 2020), in part due to their interstate programme in the mid-20th century to demolish neighbourhoods to build urban freeways (DiMento and Ellis, 2012, Taylor, Morris and Brown, 2023). The urban form required to allow private vehicle mobility to support a high-productivity local economy requires significant urban land dedicated

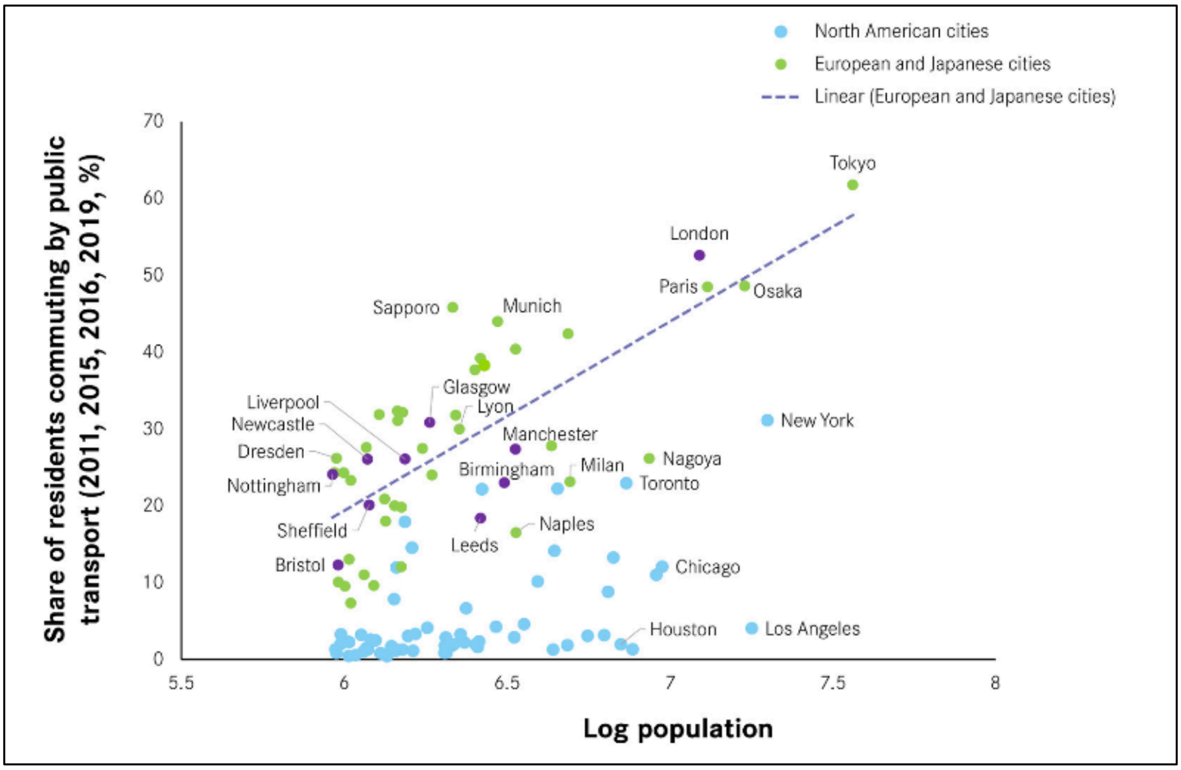



*Figure 1-2. Relationship between urban population and public transport commuting mode share for G7 large cities* (Breach and Swinney, 2024)

to roads and car parks, suppressing urban residential density, which in turn reduces the viability of public transport networks (Cooke and Behrens, 2017). US cities also benefit from unprotected green-belt land into which their cities can expand, unlike their European counterparts which have been densifying urban conurbations since the turn of the century to facilitate public transport and protect the surrounding countryside (Livingstone, Fiorentino and Short, 2021; Pont et al., 2021).

Other factors which should lead to a strategy favouring public transport ahead of the private vehicle include climate change, particulate emissions from internal combustion engines, particulates from brake pads and tire wear causing poor air quality, the deleterious effects of noise pollution, public safety implications, sedentariness of driving culture worsening public health and promotion of equality through wealth redistribution (Miner et al., 2024). There will likely always be a place for private vehicles in urban centres, but most major conurbations' transport strategies indicate a move away from prioritising them in favour of public transport and active travel (Transport for Greater Manchester, 2021; Liverpool City Region, 2022; Glasgow City Council, 2022).

## 1.2 Public Transport Accessibility

Public transport describes all travel modes which are readily available to the public; bus, coach, train (above ground, underground, light, heavy, metro), tram, ferry, and walking, and advanced transport models account for all available modes. The urban bus is by far the most utilised mode in the UK, with 3.7billion journeys by bus in 2022-2023 representing 60% of all journeys by public transport, compared with 36% taken by rail (GOV.UK, 2024a). Those percentages are reversed when considering passenger-km (25% bus vs 70% rail) as a predominance of intercity public transport is by rail. This emphasises the importance of the bus networks in urban environments (Abrantes, Fuller and Bray, 2013). Transport accessibility studies through the 20$^{th}$ century had focussed on private vehicles and road/highway networks, building from the seminal works on accessibility by Hansen (1959). More contemporary studies model urban public transport networks to understand their relationships with the wider economy (Naro, Biraghi and Lenzi, 2024). Public transport accessibility is defined inconsistently across different regions and institutions, but Curl, Nelson and Anable (2011) categorise the discipline into affordability, physical accessibility, availability/reliability and safety/security aspects. They suggest that public authorities should consider all aspects of accessibility in planning and optimising public transport networks, but that academic studies would typically be focussed on one or two of the categories to isolate and understand the underlying effects. This research will consider only the availability aspect of accessibility and focus on the timetabled service provision, and the mobility benefits this provides to the local population. The other aspects of accessibility are subjective, for example affordability is observed differently depending on wealth, but availability affects the population equally.

The literature base covering public transport accessibility consistently highlights the importance of context and that recommendations are often only applicable to the specific contextual environment of the study. There is an ever-growing empirical knowledgebase for different geographies and systematic reviews of the literature can infer generalisations, though they are applied to other geographies with caution (Bastiaanssen, Johnson and Lucas, 2020).

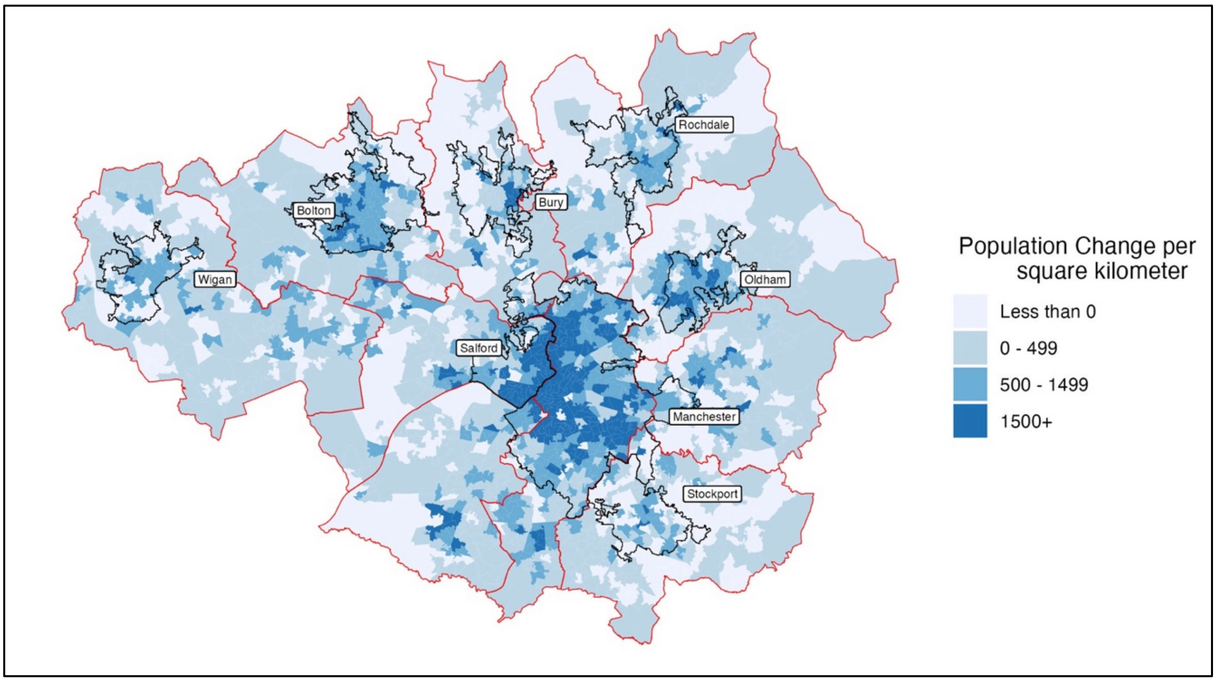



*Figure 1-3. Population density change across Greater Manchester Combined Authority (GMCA) by LSOA, 2001-2021. Data from Census 2001 and 2021* (ONS, 2024c). *The red lines demarcate the Local Authority Districts and the black lines the major town and city boundaries from the ONS* (ONS, 2018).

## 1.3 Greater Manchester Context

One conurbation which has recognised the spatial planning issues of the 20th century and has been working to reverse the trend of de-densification is Manchester in the Northwest of England. This study contributes to the growing empirical literature base in the UK, specifically addressing Greater Manchester Combined Authority (GMCA) as a case study. GMCA was founded in 2011 and has devolved powers over housing, public transport, regeneration, skills, waste management, carbon neutrality and planning permission (GMCA, 2024). GMCA housed a population of 2.87 million people in 2021; an increase of 15.6% since 2001 against a national average increase of 14.6%. Figure 1-3 shows the distribution of population changes across GMCA from 2001 to 2021, with many peripheral areas declining in population and the city centre and immediate suburbs showing significant increases as the city densifies it's housing stock. The city region incorporates 10 councils, 8 of which have a significant town or city as their administrative centre. These are shown on Figure 1-3 and make the case study particularly interesting due to the polycentric nature of the city region.

## 1.4 Metrolink and the Bee Network

Manchester's second age of tram-based mobility, the Metrolink, comprises 103km of tramway with 99 stations across 8 lines and carries 40 million trips each year (GMCA, 2021). It was constructed in several phases from 1992, the most recent addition completed in 2020 and a plan for 3 new lines has been released in July 2024 (Carey, 2024). The system forms the backbone of the Bee Network, Greater Manchester's integrated transport system, which is in the process of enveloping the local bus networks (the final tranche of services will be brought under public control in January 2025) and will also include all local rail services by 2028. This will create the second fully integrated transport network in UK, after London.


*Figure 1-4. Metrolink Tram*
(Page, 2011)

Despite the population densification and public transport infrastructure, Figure 1-1 shows Manchester lying significantly below the line of average G7 city productivity. If this productivity gap were closed, an additional £8.8billion would be added to the UK GDP (Centre for Cities, 2022). Improved public transport accessibility will not completely close the productivity gap, but it is likely to improve labour market outcomes which will contribute towards increasing productivity by improving wages (Galí, 2011).

## 1.5 Research Questions

This research aims to supplement the empirical research base by evaluating unemployment as an indicator of the economic benefits invoked by improved urban mobility, in a context (GMCA) where the local authority has the power to make changes to improve the provision of public transport.

RQ-1) How does access to jobs by public transport vary across the Greater Manchester Combined Authority (GMCA) area?

RQ-2) Is fixed effects regression a suitable approach and what geographical level should it be set at?

RQ-3) How does public transport job accessibility relate to unemployment rate across GMCA?

Delivering a case-study and transport model for GMCA at this time will provide usable information to Transport for Greater Manchester (TfGM) at a time when they are taking control of the full public transport network. The wider academic literature on the subject will be reviewed to ensure appropriate methods are deployed, followed by a detailed description of the data used and methods adopted. The results are then presented and discussed before summarising the findings and drawing conclusions.

# 2 Literature Review

The relevant literature on transport accessibility and unemployment has been reviewed to understand the approaches, shortcomings and findings with which to compare. The summary below describes the review but is in no way exhaustive.

## 2.1 Job Accessibility and Employment Outcomes

Since Kain (1968) proposed a spatial mismatch hypothesis, which focused primarily on racial segregation of housing in US downtowns, there has been extensive debate and empirical studies of the relationship between spatial access to job opportunities and the relationship with employment outcomes. The Kain study is referenced in almost all papers on job accessibility as the initial recognition of impacts of spatial disparities, though it's findings and theoretical underpinning are weak, and conclusions drawn are highly tentative. It gave rise to a great debate amongst scholars, though the resultant US commentary ascribes differences in outcomes to ethnicity, making it difficult to distinguish the truth from the racial motivations of the authors (Holzer, 1991; Ihlanfeldt and Sjoquist, 1991).

Korsu and Wenglenski (2010) described that 20th century studies of employment outcomes and accessibility were often methodologically flawed in either poor indicators of accessibility, small sample sizes, endogeneity or use of spatially aggregated macro-data as an inference to lower-level outcomes. They also make a case for the strongly contextualised nature of the topic, contrasting the conclusions drawn from US studies with those in their own Parisian context. They highlight specifically the spatial distribution of jobs and residences, the different ubiquity of public transit and levels of peak time congestion, all of which will have significant impacts on their model and would be highly geographically contextual. Each could introduce endogeneity in different circumstances, thus making model comparisons between different contexts challenging and potentially misleading. This challenging analysis ecosystem led to several re-assessments of Kain's data drawing conflicting conclusions (Jin and Paulsen, 2018) as scholars tried to verify or contradict his conclusions. Kain highlighted the highly contextual nature of the topic and his recommendations, only ever claiming they would be applicable to Chicago.

Bastiaanssen, Johnson and Lucas's (2020) meta-regression and systematic review of the transport factors affecting employment status of individuals studied 4 key effects; car ownership, public transport access, commuting times and job accessibility. Car ownership was indicated to have a strong effect, though it is highlighted that 10 of the 12 studies in the analysis were US-based where only 8% of households are without a private vehicle (United States Census Bureau, 2021), compared with 22% in the UK (GOV.UK, 2021a). Similarly, commute time studies were largely US metropolitan and demonstrated a negative relationship with probability of employment across public transport and private vehicle commuting, indicating longer commutes to employment centres reduce probability of employment. Using the commuting time to the nearest employment centre as a proxy for generalised commuting time simplifies the complex urban landscape to assume all jobs are located in employment centres and is likely to underestimate the relationship between employment outcomes and travel time, as there is a large portion of the population who do not commute to the local employment centre for work and jobs are typically distributed across conurbations, away from the nominal centre, particularly since the decentralisation of the late 20th century (Internet Geography, 2024). Some studies (see table 2-1) look alternatively at the individual level commuting time measures, though they cannot take account of the missing commuting durations for the unemployed residents. Only 4 of the 22 studies measuring commuting time used a gravity model calculation for the mean commuting time impact on employment outcome, even though the gravity model has been widely adopted by transport planners to describe decreasing attractiveness as travel time or distance increases (Cervero, Rood and Appleyard, 1995). Job accessibility was the fourth effect-group in the study, which was typically measured using a fixed threshold of 45- or 60-minutes travel time, removing any benefit from jobs further than 60 minutes away or any advantage of being closer. The meta-regression did not show a relationship between public transport accessibility and employment, which was attributed to the variety of methods of measurement and that the relationship is small enough not to be discernible across different study areas. However, their systematic review of the wider literature shows a generally positive correlation between transport accessibility and employment outcomes (Bastiaanssen, Johnson and Lucas, 2020).

The quantitative studies of job accessibility and employment frequently use binomial logit (Hu, 2017) or probit (Matas, Raymond and Roig, 2010; Bastiaanssen, Johnson and Lucas, 2022) modelling on individual-level survey datasets to establish factors effecting probabilities of employment. The meta-regression analysis (Bastiaanssen, Johnson and Lucas, 2020) only selected the studies using binomial logit regression, of which there were 20. There are far fewer studies using other regression models, though Johnson, Ercolani and Mackie (2017) used fixed effects ordinary least squares linear regression and found a relationship with employment levels, discussed later. Korsu and Wenglenski (2010) described studying aggregate-level employment statistics as one of the methodological flaws of the early works on job accessibility which led to the discrepancies described above in the late 20th century, though doesn't add detail as to what the methodological flaws are.

The most relevant published study is that of Johnson, Ercolani and Mackie (2017) who used the Middle Super Output Area (MSOA) data from the 2011 census to assess the relationship between employment and public transport accessibility across Great Britain. The assertion from Korsu and Wenglenski (2010) that using aggregate-level data would introduce methodological flaws was addressed in part by their focus on inclusion of several district-level fixed effects on employment, reducing the effect of omitted variable bias. The key dependent variables only measured travel time to the nearest employment centre by car or by public transport, which has several limitations compared with a gravity-based index, as discussed above. Their study concluded the relationship between public transport travel times and employment is statistically significant, and that shorter travel times to employment centres do correlate with employment outcomes.

It should be noted that Public Transport Job Accessibility (PTJA) is not a widely adopted term, and several of the studies use different terminology to describe the same concept, though it has been normalised herein for consistency.

As the lowest income groups are more likely to have no access to a car and utilise public transport the most, studies have often focussed on the access to the lowest paid or unskilled jobs and lowest income segments of the population (Shen, 1998; Hu and Giuliano, 2017). Alternatively, Hu (2017) assessed how the modelled advantage inferred by job accessibility was distributed across different income groups and found the middle-income groups to benefit the most from access to jobs by car. He also published in the same year (Hu and Giuliano, 2017) that the poor are not provided any benefit from job accessibility with regards to employment outcomes, though it did have a significant impact on poor people's commuting distance, suggesting that it had lubricated the labour market, but not in order to match jobs with job seekers.

| Reference | Analysis Method | Level of analysis | Dependent Variable | Context | Transport time measure | Conclusion |
|---|---|---|---|---|---|---|
| Bastiaanssen, Johnson and Lucas, (2022) | Binomial probit regression with (IV) | Individual | Employed (1/0) | Great Britain | PTJA index with decay function derived from labour force survey | Higher PTJA improves employment probability |
| Matas, Raymond and Roig (2010) | Binary probit regression with latent variable (tried IV but to no avail) | Individual | Employed (1/0) | Barcelona, Madrid | PTJA index as sum of number of jobs in location divided by travel time to location by public transport | Higher PTJA increases employment probability in women |
| Johnson, Ercolani, Mackie (2017) | Multiple linear regression with Fixed effects for Local Authority Districts and IV | Mid-layer Super Output Area (MSOA) (Aggregated) | Employment level (n) | Great Britain | Public Transport and private vehicle travel time to employment centre | Shorter public transport times correlate with higher employment outcomes |
| Hernandez, Hansz and Massobrio (2020) | Multilevel regression with fixed effects | Individual | Employed (1/0) | Montevideo, Uruguay | Cumulative opportunity index for jobs within 40min travel threshold by public transport | Improved PTJA reduces probability of unemployment |
| Rogers (1997) | Multiple regression | Individual, with zip-code level transport access index | Unemployment duration (weeks) | Pittsburg, USA | Driving time, zip code to zip code | Increased job accessibility reduces expected unemployment duration |
| Johnson (2006) | Multiple regression | Individual | Unemployment duration (weeks) | Atlanta, USA | Driving and public transit time published by municipality | |
| Hu (2017) | Binomial logit regression with IV by income quintile | Individual | Employed (1/0) | Los Angeles region, USA | PTJA and driving accessibility, adjusted for demand and income level | Job accessibility improves employment but not for all income groups. |
| Di Paolo, Matas and Raymond (2017) | Simultaneous 3-equation regression. Job-education mismatch, employment-selection binomial probit, car ownership by ordered probit | Individual, by gender | 1) Years of education above peers'<br>2) Employed (1/0)<br>3) Number of cars (0, 1, 2+) | Barcelona, Spain | PTJA estimated as number of jobs in district divided by time by public transport to district | Public transport connectivity to job locations improves matching between education level and occupation. |
| Jin and Paulsen (2018) | Fixed effects with IVs on Generalised Spatial 2-Stage Least Squares regression. | Block-groupings (aggregated) | Unemployment rates and income levels in block group | Chicago, USA | Gravity-decayed distance-based job accessibility index, adjusted for demand potential | Improved accessibility reduces unemployment and enhances income levels |
| Hu and Giuliano (2017) | Simultaneous OLS regression<br><br>(IV made no significant difference from OLS) | Census tract | i) Workers-in-poverty to population ratio<br><br>ii) Commute distance | Los Angeles region, USA | Job Access Index is sum of jobs for poor job seekers decayed by travel time (driving) impedance function and local competitive demand | Job accessibility had no relation with employment outcomes |

*Table 2-1. Transport Accessibility and Employment Studies*

## 2.2 Dependent Variables

Johnson, Ercolani and Mackie (2017) used the absolute employment level as their dependent variable, allowing for population by including it as one of the covariates. This does not take account of the portion of the population not pursuing employment (retired, in education or long-term illness) though they used it to reflect more readily to policy-makers the advantage of transport investments in numbers of employments filled. Other linear regression models have used job search duration (Rogers, 1997; Johnson, 2006) or workers-population ratio (Hu and Giuliano, 2017) as dependent variables. Using census data provides a more accurate reflection of the status (i.e. does not rely upon survey extrapolations) but does not provide the data for job search durations. The worker-population ratio is an alternate formulation of the absolute employment level described above, and similarly takes no account of the people not in the job search market.

## 2.3 Independent variables

The spatiotemporal relationship between job opportunities and residential location is only one aspect of the employment marketplace. Another important factor is qualification to job-level matching. This results in a sub-model for each qualification-job level match (Cervero, Rood and Appleyard, 1995; Di Paolo, Matas and Raymond, 2017), significantly increasing the complexity of the model and reducing communicability of the results. Alternatively, the studies can focus on one particular qualification level (Johnson, 2006) to understand the mechanisms for a particular segment of society. Also, as job location data are derived from surveys, separating those approximations into several job levels per geographic zone firstly reduces the sample size from which it is derived (assuming the survey size doesn't increase proportionally) and secondly increases the relative error of the estimated number of jobs. Also, there are many people who work in job different to their qualification level for a number of reasons, such as being unable to obtain a suitably skilled job or having climbed the corporate ladder from their initial qualification level.

Controlling for qualification level can be achieved using dummy variables ((Johnson, 2006; Matas, Raymond and Roig, 2010; Hu, 2017; Bastiaanssen, Johnson and Lucas, 2022). However, Shen (1998) and Hu and Giuliano (2017) modelled the low wage / low skilled workforce only, on the premise that the relationships between job accessibility and employment would be stronger amongst lower skilled employees (as they would be unable to work from home).

Several studies control for ethnicity (Di Paolo, Matas and Raymond, 2017; Dixon, Johnson and Batley, 2019; Hernandez, Hansz and Massobrio, 2020; Bastiaanssen, Johnson and Lucas, 2022), not least the 20th century US studies designed to model black unemployment in downtowns (Ihlanfeldt and Sjoquist, 1991; Cervero, Rood and Appleyard, 1995; Rogers, 1997; Shen, 1998). All demonstrate a statistical significance in their at least some of their models.

## 2.4 Job Opportunity Distribution

A significant majority of studies on job accessibility use the numbers of employed positions within a geographic area as a proxy for number of job vacancies (Matas, Raymond and Roig, 2010; Bastiaanssen, Johnson and Lucas, 2022). This has the disadvantage that held positions don't really reflect opportunities; there are many companies and sectors which have much higher or lower turnover than average, and those companies may tend to cluster together spatially. There are novel datasets which do represent job opportunities derived from online job advertisements (Darby, McIntyre and Roy, 2022). However, the current opportunities are only a snapshot in time, and a time-series only captures a subset of the actual job opportunities which are advertised on this particular online portal which would introduce selection endogeneity into the model. Alternatively, Rogers (1997) used local employment tax receipt data and estimated company turnover percentages to derive an estimate of numbers of job opportunities in Pittsburg, USA. Johnson (2006) also used detailed multi-city urban inequality data to estimate the number of non-college job vacancies at each employer in Atlanta, Boston and LA, calculating the vacancies from annual employee turnover plus those estimated from company growth. Both of these methods require survey data which is unavailable in the UK.

## 2.5 Instrumental Variables, Endogeneity

Endogeneity is an underlying relationship between one of the variables and the unexplained (or residual) error term, and can be caused by simultaneity, selection, measurement error or omitted variable bias. Any of these elements will cause bias in the model and is likely to increase or decrease estimates of the coefficients and standard errors, thereby giving confidence where it is not due, and particularly concerning is that it cannot (always) be tested for or addressed (Hill *et al.*, 2021). It is of particular concern in transport and employment outcomes for a number of reasons (outlined later).

A common approach to addressing endogeneity between unemployment and transport accessibility in cross-sectional studies is to deploy Instrumental Variables (IVs) in 2-stage least squares regression which is described in section 3.3.4. The important requirements of IVs are that they are strongly correlated with the endogenous independent variable, but also exogenous (i.e. minimal correlation with the residual of the model). This second criterion cannot be tested for empirically and must be determined theoretically (Ullah, Zaefarian and Ullah, 2021).

In approaching the endogeneity between transport accessibility and employment outcomes, Bastiaanssen, Johnson and Lucas (2022) and Hu and Giuliano (2017) utilised an IV derived from the population density. Bastiaanssen, Johnson and Lucas (2022) reported a positive effect from the IV for lower incomes, but a detrimental effect for higher income groups. Hu and Giuliano (2017) reported no significant difference from the OLS models due to their IVs. Johnson, Ercolani and Mackie (2017) also utilised instrumented variables to control for endogeneity between public transport accessibility and employment, but their model showed the benefits of the IV insignificant, most likely due to their choice of IV. They used variables lagged 2 or 10 years earlier than the variable the endogenous variable, which according to (Ullah, Zaefarian and Ullah, 2021) would be a poor choice of instrument. Johnson, Ercolani and Mackie (2017) used the bus travel time from 2009 as an IV to estimate bus travel time from 2011. As bus timetables change very slowly over time, the travel times from 2 years previous would be expected to be almost identical, therefore have a very similar relationship with the error term of the regression as the current travel times.

To address the endogeneity between employment status and private vehicle accessibility, Hu (2017) utilised the rate of car non-ownership as their IV to estimate PTJA. Non-ownership of cars is often included as a model covariate, (Matas, Raymond and Roig, 2010; Johnson, Ercolani and Mackie, 2017; Bastiaanssen, Johnson and Lucas, 2020) indicating a belief that there is believed an underlying relationship with employment levels thus is not exogenous. They have also not included all the exogenous variables in their first stage regression model, as required by instrument variable estimation (Ullah, Zaefarian and Ullah, 2021; Bastardoz *et al.*, 2023). They also followed the conclusions of another study (Matas, Raymond and Roig, 2010) whose arguments were contextual, and applying conclusions to an entirely different context is contrary to good practice (Hill *et al.*, 2021).

The estimation and application of IVs is trialled and determined within each study and is contextual and dependent on availability of suitable data for instruments (Bastiaanssen, Johnson and Lucas, 2020). Accepting weak IVs can make the results less reliable and invalid instruments affect the results even where there is no endogeneity present (Ullah, Zaefarian and Ullah, 2021). Very few studies of transport and employment deal satisfactorily with endogeneity (Bastiaanssen, Johnson and Lucas, 2020) and they present challenges in a number of areas of academia (Ullah, Zaefarian and Ullah, 2021; Bastardoz *et al.*, 2023).

# 3 Data and Methods

This study will supplement the empirical evidence on the relationship between access to jobs and unemployment levels, using the Greater Manchester Combined Authority (GMCA) as a case study. This section describes the approach adopted and the data utilised.

## 3.1 Methodology

In order to explore the relationship between employment levels and public transport job accessibility, a quantitative, secondary-data, cross-sectional approach is developed using open data, primarily from the 2021 census, the Business Register and Employment Survey and public transport agencies. This has been adopted over a qualitative approach as the results can be fed back to cost benefit analyses or policy reviews. It is also possible to obtain results spanning the whole city region, whereas quantitative or primary data studies in this timescale would only be able to assess a smaller area. All data used is currently and widely available and open source and as such there was no application for ethical approval required (Clark, Foster and Sloan, 2021).

The level of measurement throughout is the Lower Super Output Area (LSOA) layer as per the 2021 census (ONS, 2024e). These geographical areas accommodate between 400 and 1200 households representing a population of between 1,000 and 3,000 persons and are slightly adjusted each census to ensure the enclosed populations remain approximately consistent. Therefore, any datasets noted below as downloaded from the 2011 census have been adjusted using the lookup tables available from the 2021 census website (ONS, 2024c).

In the 2021 census there are 1,702 LSOAs within the GMCA boundary, representing a population of 2,867,719.

## 3.2 Datasets

The data sources combined in this study and their purpose are described in Table 3-1 and diagrammatised in Figure 3-1.

| **Dataset** | **Source** | **Purpose** |
|---|---|---|
| ***Geographic Data*** | | |
| LSOA Boundaries | Office for National Statistics (ONS) Open Geography | Identify area of analysis for graphical presentation |
| LSOA Population-weighted centroids | ONS Open Geography | LSOA origin point for transport modelling |
| Combined authority boundaries | ONS Open Geography | Filtering study area |
| Local Authority District (LAD) Boundaries | ONS Open Geography | Geographical presentation |
| Metrolink shapefile | Data.gov | Geographical presentation |
| LSOA 2011 – 2021 Lookup Table | Data.gov | Convert 2011 LSOA 2021 LSOA geographies and lookup LADs |
| Job centre locations | Gov.uk | Calculate time to closest job centre from LSOA using public transport model |
| Major Towns and Cities boundaries | ONS Open Geography | Calculate travel time to closest employment centre.<br>Add Urban/Suburban attribute to each LSOA as a covariate |
| ***Unemployment Data and Neighbourhood Characteristics*** | | |
| Households with no car available, ethnicity, highest household qualification, unemployment rate | Census 2021 | Covariates in regression analysis |
| Rural Urban Classification | Census 2011 | Determine number of rural LSOAs in GMCA |
| ***Job Opportunity Data*** | | |
| Business Register and Employment Survey | Nomisweb.co.uk (ONS online service) | Employment levels in LSOAs as proxy for employment opportunities |
| ***Data for Public Transport Job Accessibility Index*** | | |
| Bus and tram timetables | Bus Open Data Service (BODS) | Transport model to determine travel times |
| Rail timetables | Rail Delivery Group | Transport model to determine travel times |
| OpenStreetMap | Geofabrik | Transport modelling of walking and bus routes along roadways |

*Table 3-1. Summary of Datasets*

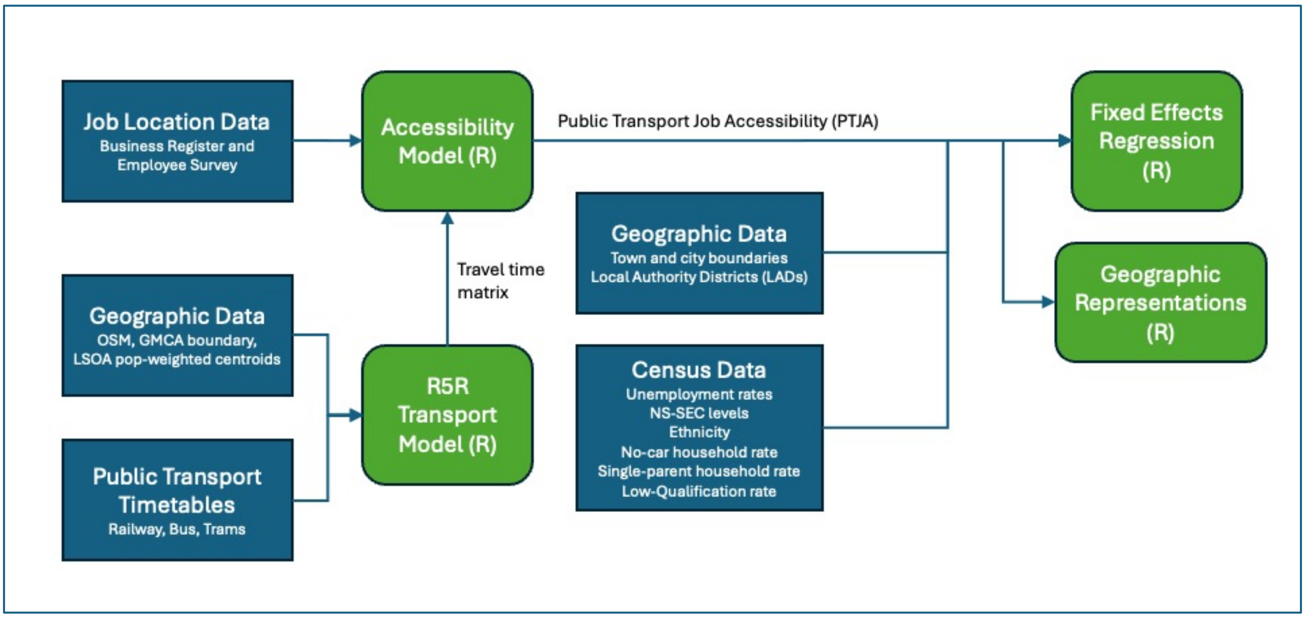



*Figure 3-1. Dataset Flow Diagram*

### 3.2.1 Geographic Datasets

The 2021 LSOA boundaries and population-weighted centroids, Combined Authority boundaries (ONS, 2018) and Major Town and City Boundaries (ONS, 2024d) were downloaded from the ONS Open Geography Portal. All geographical datasets are downloaded for the whole of England and trimmed to a 20km buffer around the boundary of GMCA in order to minimise boundary effects.

### 3.2.2 Census Data

Census 2021 datasets have been obtained from nomisweb.co.uk (ONS, 2024c). On census day, 21st March 2021, there was a nationwide COVID-19 lockdown in place and accordingly the workday population and travel to work data are not considered highly accurate (ONS, 2023). As such, the data obtained from the census has been limited to the number of households with no car available, the households with no/low qualifications, the number of people who identified as white ethnicity, socio-economic classification statuses and unemployment rate. All variables used are confirmed by the Census organisation to be an accurate reflection of society (ONS, 2023). With particular reference to unemployment rate during the COVID-19 pandemic, the data from the census are considered a good representation of actual unemployment rates due to the UK government’s Coronavirus Job Retention Scheme and the Self-Employment Income Support Scheme, which largely prevented additional unemployment during the lockdown periods. The unemployment rate is calculated as the percentage of economically active people 16 years and over who are not in full-time education who are not currently employed. Using the economically active population as the denominator means those who would not be helped into work by improved public transport links (students, full time carers, long term sick) are not included and hence will not dilute the modelled relationship.

### 3.2.3 Job Location Data

Employment opportunities are assumed to be proportional to the numbers of employed positions, the locations of which are recorded in the Business Register and Employment Survey (BRES) from 2021. The data are obtained and compiled by the ONS from a survey of 85,000 of the 2,600,000 VAT- or PAYE-registered entities in Great Britain. It is regarded as the official source of employment and employee data by geographical area and is used by local authorities to forecast trends in employment (ONS, 2024a). The BRES does have some rounding applied to the numbers of jobs in each LSOA in order to protect businesses' privacy, particularly where there is one significantly larger employer present. This does not present a significant loss of accuracy owing to the relatively large numbers involved, and the level of rounding being proportional to the number of jobs in each (ONS, 2024b). This job location data assumes that all employment needs to take place at the place of employment, i.e. does not account for working from home. This is a significant simplification in post-COVID era, as there has been a significant move amongst many jobs to remote working (Naseer, 2023).

The survey reports self-employed persons separately, and they will not be incorporated as they do not represent the job marketplace; it is assumed that unemployed people cannot easily join the self-employed workforce. Using the numbers of current jobs as a proxy for job opportunities is common practice in the literature from all regions (Dixon, Johnson and Batley, 2019; Bastiaanssen, Johnson and Lucas, 2022; Verduzco Torres and McArthur, 2022). The alternative strategies for estimating job opportunities in the literature review were declined due to BRES being surveyed and provided by the government as the source of employment truth and is deemed the most accurate source available.

### 3.2.4 Public Transport Data

The current Openstreetmap (OSM) dataset for England was downloaded from Geofabrik (OpenStreetMap, 2018) and filtered to focus on the GMCA area according to the combined authority boundary provided by the ONS (2024d).

The General Transit Feed Specification (GTFS) was developed by Google in 2005 to provide public transit travel recommendations similar to those already available for private vehicles. It is an open standard data structure and is used consistently by thousands of transit agencies around the world. This allows developers to structure analysis apps to digest information from a global basis and has given rise to many excellent platforms such as OpenTripPlanner (OTP), Rapid Realistic Routing in R (Pereira et al., 2021) and City Transport Analyser (Naro, Biraghi and Lenzi, 2024).

As a result of improved algorithms (particularly R5R), the ability of researchers today to utilise computing power to analyse transport networks far exceeds that of just 5 years ago. A transport model which would have taken a mainframe computer 6 hours to run can now be analysed in a few minutes on a laptop (Pereira et al., 2021). This is a direct result of the wide implementation of the General Transit Feed Specification (GTFS) protocol and of open sourcing and standards, and global indirect collaboration (GTFS, 2024).

To model the public transport network, the bus and tram timetable data was obtained from the Bus Open Data Service (BODS)(GOV.UK, 2024b). The BODS does not provide any historic datasets, only the current timetable, and as such the May 2024 timetable will be used. The timetabled bus services in 2021 were reduced due to the Coronavirus lockdowns, though the objective here is to understand how historic accessibility to public transport has affected employment levels. The current level of unemployment may be affected by the current level of transit accessibility, but there will likely be inertia to the relationship as well, as change the accessibility to an area will gradually take effect as people realise the service, start to utilise it and change their lives around it. The change will not happen as quickly as the change in the service itself. Therefore, using a timetable from a slightly different time period only has a minor effect, as the timetables change very slowly as well.

The train timetable data was obtained from the Rail Delivery Group website (RDG, 2024) in Common Interface Format (CIF) and converted to the commonly used (GTFS) format using a converter program from Leeds university (Morgan, 2024). The Rail Delivery Group also do not offer historic timetabling datasets, so the contemporary timetables are used (from May 2024).

### 3.2.5 Public Transport Job Accessibility (PTJA) Index

There are 2 principal quantitative methods for modelling job accessibility; absolute numbers of jobs within an isochrone or a non-dimensional index representing a function of all jobs and the travel cost to their location (Verduzco Torres and McArthur, 2024). Either method requires a cost applied to the number of jobs at the destination, typically a travel time derived from a transport model. Some models use monetary cost, though they are in the significant minority due to the availability of data and the complexity of analysing the effects of monetary costs on behaviours. The first method sums the number of jobs within a specific threshold cost, but this approach infers no advantage to any jobs outwith the defined threshold, and also assigns equal weighting to all jobs regardless of their access cost (Dixon, Johnson and Batley, 2019).

This study utilises a PTJA index. Following a methodology initially developed by Hansen (1959), this will result in a dimensionless index which is more useful for comparisons across geographies and accounts for all jobs, with their attractiveness decayed according to the travel time to reach. This addresses the issues associated with an absolute threshold and equality of appeal and provides a relative measure of job accessibility between the geographical areas. However, the results are abstract and unintuitive as any relationship established will be to a dimensionless index, rather than in terms of absolute numbers of jobs. Equation (1) below defines the accessibility index, PTJA.

$$PTJA_i = \sum_{j=1}^{n} f\left(t_{ij}, EMP_j\right) \tag{1}$$

*Where;*

*$PTJA_i$ is the public transport job accessibility index of LSOA i.*

*$t_{ij}$ is the travel time by public transport from i to j*

*$EMP_j$ is the number of employees based in LSOA j*

*n is the number of LSOAs in the study area*

Other studies (Shen, 1998) have expanded the job accessibility index to include a demand-side function, representing the competition from the local population for jobs, which could be explored as an extension to these works.

The bus, tram and train timetables and OpenStreetMap were loaded to Rapid Realistic Routing package R5R which was used to calculate the travel times matrix by public transport between each pair of LSOAs, in each direction. As timetable data is loaded for May 2024, the journey departure time is set to 8:00am on Tuesday 21st May 2024. The sensitivity to changing this time to 7:00am or 8:30am was checked and the difference is negligible for a majority of locations. Only some more rural areas which are served by an hourly bus route have significant changes in accessibility. The R5R package calculates the duration of the fastest route for a departure each minute within a 10-minute window and provides the median trip duration (and appropriate departure time) as the single trip time for that origin-destination pair. The sensitivity to changing this departure window to 30-minutes and 60-minutes was tested and noted to make a negligible difference, as per the authors of the accessibility R package (Pinna and Murrau, 2018; Pereira et al., 2021). The walking speed has been set to 1.2m/s or 4.32k/h as the mean speed for ages 20-50 (Pinna and Murrau, 2018). The maximum walking time is 60 minutes which has been derived from the National Travel Survey (Department for Transport, 2024) as the 99% quantile of commuting trips in Northwest England by foot. The maximum trip duration has been left as infinite in order to avoid any gaps in the dataset. The gravity decay function will reduce the attractiveness of jobs at the end of longer journeys.

### 3.2.6 Gravity Decay Function

A key advantage of the accessibility index method compared with a fixed threshold is the ability to reduce the attractiveness of the job opportunities according to their travel-time cost. There are several decay functions applied to travel cost in the literature, including lognormal (Bastiaanssen, Johnson and Lucas, 2022), and negative exponential (Johnson, 2006; Dixon, Johnson and Batley, 2019) but the distribution of the National Travel Survey (Department for Transport, 2024) travel times for the Northwest region by public transport visually approximates more closely to a logistic distribution. This agrees with the findings of de Vries, Nijkamp and Rietveld (2009) who modelled commuting behaviour and follows the method used by Bauer and Groneberg (2016) in modelling floating catchment areas. The logistic decay function is implemented as a standard function in the Accessibility package in R (Pereira et al., 2021) and is defined as;

$$f(t_{ij}) = \frac{1 + e^{\frac{-\bar{x}*\pi}{SD*\sqrt{3}}}}{1 + e^{\frac{(t_{ij}-\bar{x})*\pi}{SD*\sqrt{3}}}} \tag{2}$$

Where;

$t_{ij}$ is the travel time from LSOA i to LSOA j

$\bar{x}$ is the mean of the distribution of commuting travel times

SD is the standard deviation of the distribution of travel times

The decay function parameters, $\bar{x}$ and SD, have been estimated using the National Travel Survey dataset covering the period 2002-2022 (Department for Transport, 2024). The trip-level data was filtered for commuting by public transport in the Northwest region of England in order to account for regional variations in commuting behaviour, and total travel time less than 120 minutes to remove the 1% of trips in the right-hand tail of the distribution. Remaining was a sample of 11,568 observations with which to approximate the decay constants. Bastiaanssen, Johnson and Lucas (2022) noted that rural residents will typically travel further for work, but as GMCA has limited rural-classified areas there is only a single set of decay parameters calculated for the study. The DEFRA Urban Rural classifications were obtained for the 2011 census and converted to the 2021 census geographies using the lookup table referenced above, confirming that out of the 1,702 LSOAs in GMCA, only 13 were classified as rural. No account shall be taken for the urban-rural classification differences due to the low quantity of rural LSOAs in the city region.

## 3.3 Analysis

### 3.3.1 Fixed Effects

The objective is to specify a robust model of the relationship between unemployment rate and public transport job accessibility, controlling for as many variables as practical. It is expected that unemployment (and other variables) will be highly geographically clustered, thus the residuals will be autocorrelated and contravene the assumption for linear regression modelling i.e. there will be unmeasured, intragroup correlations (Schork, 2024). There are several options for modelling datasets which lack observation independence. Many machine learning algorithms require observation independence, and spatial regression models are highly complex, resulting in complex conclusions. There is an option to use time-series data to observe chronological changes, but as the public transport networks change very slowly and the effect on unemployment rates is expected to be gradual, a long time period would be required, and public data collections often change gradually over time. Due to the expected small magnitude of the changes to be detected, minor adjustments to datasets may distort the picture.

Following the process implemented by Johnson, Ercolani and Mackie (2017), I will use a fixed effects regression model. Fixed effect modelling allows for inter-group correlations by subtracting the group-mean from the observations within the group ('demeaning') (Mouw, 2000). It offers the advantage that the underlying intra-group effects do not need to be understood nor quantified, providing they remain fixed within the group (Breuer and Dehaan, 2024). This emphasises the importance of setting the geographical fixed-effect area at the right level, as if the geography is very small the fixed effects will serve to reduce the variation in the public transport accessibility. It also offers some protection from the effects of omitted variable, but only the components of the effects which are constant within the groups (Breuer and

Dehaan, 2024). The areal fixed effects are estimated to describe the contributions to unemployment from geographic influences due to large historical employers, industrial mix and skills within the labour pool (Johnson, Ercolani and Mackie, 2017).

To approach RQ-1, a comparison is made between the 'baseline' model (associating unemployment rates and public transport accessibility only, without controlling for any covariates), and models allowing the intercept to vary by i) Local Authority District (LAD), ii) Town and City Boundaries iii) a combination of the 2. This third model applies LAD boundaries to the large area which surrounds all of the town and city boundaries and is termed the urban/suburban split in this study. The comparison uses the Akaike Information Criterion (AIC) to assess the model fit, and an F-test to determine if the improvement over the Baseline model is statistically significant.

The fixed effects linear regression model is represented by equation (3);

$$U_i - \overline{U}_k = \beta_1(PTJA_i - \overline{PTJA}_k) + \beta_2(x_{2i} - \bar{x}_{2k}) + \cdots \beta_n(x_{ni} - \bar{x}_{nk}) + \varepsilon \tag{3}$$

where;

*$U_i$ is the unemployment rate of LSOA i, in fixed effect area k,*

*$\overline{U}_k$ is the average unemployment rate across fixed effect area k,*

*$PTJA_i$ is the Public Transport Job Accessibility (PTJA) Index for LSOA i,*

*$\overline{PTJA}_k$ is the average PTJA across fixed effect area k,*

*$x_{2i} \ldots x_{ni}$ are the other independent variables in LSOA i,*

*$\bar{x}_{2i} \ldots \bar{x}_{nk}$ are the means of the respective variables across fixed effect area k,*

*$\beta_1 \ldots \beta_n$ are the coefficients to be estimated by the model*

*$\varepsilon$ is the error term*

(Econometrics with R, 2024)

### 3.3.2 Dependent Variable – Unemployment Rate

The dependent variable in the analysis is the unemployment rate from the 2021 census. This is defined as the proportion of economically active people (i.e. excluding students, retirees, long term ill-health and carers) 16 years old and above who are currently unemployed. Using the unemployment rate rather than the level avoids needing to include the population as a covariate (Jin and Paulsen, 2018), which in turn simplifies the results slightly. Excluding those who are not actively seeking work should improve the model fit, as transport accessibility to employment will not compel people into the job market.

### 3.3.3 Independent Variables

The key independent variable is the PTJA index as described above for each LSOA. This index value is intended to quantify accessibility to jobs by public transport, which become decreasingly attractive as travel time increases. There is no allowance made for monetary cost of travel as the complexities surrounding equity and affordability would require an entire study to perform a sufficient analysis.

The other independent variables in the regression model, as observed from various studies in the wider literature (noted in the table below) and where data were publicly available, are;

| Variable description | Reference |
|---|---|
| Proportion of the LSOA households which do not have access to a private vehicle | (Hu and Giuliano, 2017) |
| Proportion of the LSOA population of white ethnicity | (Patacchini and Zenou, 2005) |
| Proportion of LSOA households which are designated as single-parent | (Bastiaanssen, Johnson and Lucas, 2022) |
| Proportion of the LSOA population who have either no qualifications, apprentice level or level 1 or 2 qualifications (collectively termed 'low qualified') | (Matas, Raymond and Roig, 2010) |

*Table 3-2. Independent Variables*

Several other analyses have used travel time to closest employment centre as the key independent variable, to reflect the transport accessibility of the location. The employment centres typically have a significant concentration of the job opportunities (see Figure 4-1 for job distribution) thus representing accessibility in this way is a rational simplification, but owing to advances in computational software it is now possible to model access to all jobs. The travel time variable is not used in regression analysis here but is modelled and discussed in the results.

### 3.3.4 Instrumental Variables – Endogeneity

The relationship between unemployment and public transport job accessibility is often endogenous (Johnson, Ercolani and Mackie, 2017; Bastiaanssen, Johnson and Lucas, 2022). This comes from several sources including reverse causality through residential location choice; did people with jobs move to areas with sufficient transport access to serve their needs, or did they already live there and gain advantage in job searching due to the beneficial transport access? Residential prices are also affected by local public transport access (Costa, Ramos and Zheng, 2022), resulting in another bidirectional causal link to employment outcomes. There are numerous reasons for reverse causality in public transport accessibility models, owing to the complex urban interrelatedness in which the services operate. This is one of the primary reasons that studies are highly context-specific and very difficult to generalise conclusions, making case-studies so important.

Also present in this model, unemployed people can often not afford to own a car due to lack of resources. But having no car also influences the numbers of jobs available, both through job requirements and access to opportunities, and hence probability of unemployment. Therefore, no-car rate is endogenous with unemployment rate as well. Having multiple endogenous variables in a study on unemployment is not uncommon, as the topic is highly interrelated and understanding causality is therefore complex. Reducing the endogeneity throughout the overall model will increase the overall robustness, even though the relationship between no-car rate and unemployment is not of particular interest here.

The common solution to address endogeneity in cross-sectional transport accessibility and employment studies has been instrumental variables (IV). The IV is applied in a first stage regression to estimate the exogeneous component of the endogenous variable, then the estimated values for the endogenous variable are used in the second stage regression model with the dependent variable. This is intended to remove the component of the original endogenous variable which is associated with the reverse-causal relationship, so the second stage regression should estimate only the causal effect on unemployment rate. The standard test for exogeneity of an IV-regression model is the Wu-Hausman algorithm (Waddell, 2019; Fox, Kleiber and Zeileis, 2023). The algorithm compares the second stage regression model with the OLS equivalent to establish any improvement of consistency, and therefore cannot be conducted on single stage regression models – it only measures the improvement from the 2SLS algorithm.

Population density has been used successfully in previous employment and transport accessibility studies (Johnson, Ercolani and Mackie, 2017; Bastiaanssen, Johnson and Lucas, 2022) and will be used again here. The strength can be measured statistically but there is no formal test for exogeneity. We must rely upon the theoretical separation of population density and unemployment rate. There are factors which will indirectly relate them such as social housing typically being built in more dense neighbourhoods and used to house some unemployed people, but there is not a causal relationship between a densely populated area and the rate of unemployment. Population density may have a causal effect on PTJA as it can be used to plan for public transport developments, in order that the service is sufficiently well utilised (Cooke and Behrens, 2017). Also, areas around transit stations and interchanges have been targeted for increased residential density (on the TOD philosophy), so there is potentially reverse causality between population density and PTJA as well.

No-car rate will be instrumented by the proportion of people in management roles in the LSOA. This is formally surveyed in the census and issued by the ONS and is being used here as an indicator of the affluence of the LSOA. There is expected a strong negative correlation between no car rate and percentage of people in management roles as affluence will provide resources to purchase a car. The causal relationship between number of management-types and unemployment is distant and indirect. It could be said that as the rate of managers increases, the rate of unemployed persons must decrease as they are mutually exclusive, but the percentages are sufficiently small that other categories in the variable allow changes in one not having a significant relationship with another. Also, the category in the Socio-Economic Classification (SEC) variable is 'long-term unemployed or never worked', which is not identical to unemployed as measured in the employment variable. I.e. if you are temporarily unemployed your SEC will remain as per the job you most recently occupied. There are no perfect instrumental variables as described consistently throughout the literature and having some endogeneity with other variables in a complex system such as this is not uncommon because it is so difficult to avoid (Johnson, Ercolani and Mackie, 2017; Ullah, Zaefarian and Ullah, 2021; Bastiaanssen, Johnson and Lucas, 2022).

The importance of implementing strong instruments cannot be overstated. Weak instruments can increase bias in the model and give confidence where none should exist. The strength is tested using correlation analysis and a F-test for a regression between only the instrument and the endogenous variable (Fox, Kleiber and Zeileis, 2023) and also in a 'Weak Instrument' test on the final model.

The 2-stage least squares regression equations are shown below. Equation (4) is the 1$^{st}$ stage regression for PTJA, equation (5) is the first stage for no-car rate, and equation (6) is the second stage regression. Note each equation is also deploying fixed effects as shown in equation (3) but is shown as a standard OLS regression model below for simplicity.

$$PTJA_i = \pi_1 PD_i + \pi_2 x_{2i} + \cdots \pi_n x_{ni} \quad (4)$$

$$NCR_i = \gamma_1 M_i + \gamma_2 x_{2i} + \cdots \gamma_{ni} x_{ni} \quad (5)$$

$$U_i = \beta_1 \widehat{PTJA_i} + \beta_2 \widehat{NCR_i} + \beta_3 x_{3i} \ldots + \beta_n x_{ni} \quad (6)$$

where;

*$PTJA_i$ is the PTJA for LSOA i,*

*$NCR_i$ is the no-car rate for LSOA i,*

*$\widehat{PTJA_i}$ is the fitted variable estimated for PTJA in LSOA i from eq. (4)*

*$\widehat{NCR_i}$ is the fitted variable estimated for no-car rate in LSOA i from eq. (5)*

*$\pi_1 \ldots \pi_n$ are the regression coefficients for 1st stage regression for PTJA,*

*$\gamma_1 \ldots \gamma_n$ are the regression coefficients for 1st stage regression for NCR,*

*$PD_i$ is the population density in LSOA i (the instrument for PTJA),*

*$M_i$ is the proportion of management in LSOA i (the instrument for NCR),*

*$x_{2i} \ldots x_{ni}$ are the exogenous variables in the regression model for LSOA i,*

*$\beta_1 \ldots \beta_n$ are the variable coefficients to be estimated from second stage regression,*

*$U_i$ is the unemployment rate in LSOA i.*

### 3.3.5 Regression Modelling and Assumption Testing

The baseline fixed effects model is linear and estimated using least squares, so the Best Linear Unbiased Estimated (BLUE) (Stock and Watson, 2019) assumptions will be applied to assess model fit. These are i) exogeneity (no correlation between covariates and the error term), ii) correct functional form, iii) homoskedasticity (constant variance of errors), iv) no autocorrelation (errors are independent of each other). The output will also be checked for normal distribution of errors, multicollinearity and influential outliers and there are no additional assumptions for the fixed effects modelling. The instrumental variables are applied in a 2-stage least squares algorithm, with each stage being a linear model. As such the fixed effects baseline model will be thoroughly tested for fit and robustness, and where applicable additional tests will be conducted on the 2SLS model. Both models are presented at the end for comparison.

### 3.3.6 Spatial Autocorrelation

Spatial autocorrelation implies that local points are more likely to be similar than points which are further away. This non-independence can affect standard errors which in turn will affect the confidence of the results (Mouw, 2000). There are regression models which can account for the spatial dependence of the error term, but they rely upon a spatial weighting matrix which cannot be estimated empirically. The areal fixed effects modelling will reduce the spatial autocorrelation of the residuals, but will not remove it entirely, and as such the residuals will be tested for Global Moran's I statistics to compare with that of the independent and dependent variables (Jin and Paulsen, 2018). Geospatial studies of unemployment and transport accessibility typically use a binary spatial contiguity matrix for geographical weighting (Mouw, 2000; Jin and Paulsen, 2018). This applies a weighting of 1 to all adjacent neighbours and 0 to all others, therefore removing any distance relationship between the LSOAs. This is a significant simplification, but it is transparent and objective, has been used elsewhere and is only used in this analysis to understand the level of spatial autocorrelation remaining.

## 3.4 Limitations

As with any modelling, there are areas where simplifications and assumptions must be made in order to construct a mathematical model of reality.

PTJA can only model a snapshot in time, and therefore does not incorporate the full diversity of accessibility across the city region, including nighttime and weekend commuting behaviours. This could be addressed using more complex indexing models (Anderson, Owen and Levinson, 2012) but a cursory sensitivity test of departure time was conducted and considered acceptable for the level of accuracy of this study.

The model does not account for competition for jobs from the local population. This can be modelled using a similar accessibility algorithm to reduce the 'attractiveness' of job-numbers in an LSOA depending on the population which can reach that LSOA, decayed according to the same travel time function as used for the job accessibility. It was first developed by Shen (1998) and appears in some of the more contemporary models.

The employment opportunity modelling is an area of significant simplification. The shortcomings of using the employed positions as a proxy for job opportunities are described in Section 3.2.3, but it also takes no account for working from home. There is also the job-skill mismatch (i.e. not all jobs are suitable for all people) and thus this is a significant area of weakness for this study.

The IVs exogeneity is not possible to establish empirically, and therefore relies upon theory only. The complex nature of unemployment leads to theoretical relationships with many concepts, and accordingly finding a truly exogenous instrument is extremely difficult. This is also found in all other studies deploying IVs.

# 4 Results and Discussion

This section reports the results and discusses the implications of the analysis. There are three sub-sections: descriptive statistics of the dataset, geographical representations of distributions across GMCA, and statistical analysis results.

## 4.1 Descriptive Statistics

The descriptive statistics presented below are at the LSOA level across GMCA. The unemployment rate across GMCA is above the national average and is significantly higher within the urban centres in the city region.

| Variable | Mean (Inter-quartile Range) |
|---|---|
| Unemployment Rate (%) | 5.73 (3.14) |
| PTJA Index ($x10^3$) | 132 (119) |
| Population Density (persons / sq.km) | 4,732 (3,302) |
| Population | 1,685 (341) |
| No Car Rate (%) | 26 (14) |
| White Ethnicity (%) | 78 (21) |
| Single Parent Households (%) | 12.8 (6.0) |
| Low Qualified (%) | 49 (12) |
| Socio-economic Class - Management (%) | 30 (13) |

*Table 4-1. Descriptive statistics of the Greater Manchester Combined Authority (GMCA) at the Lower Super Output Areas (LSOAs) level.*

## 4.2 Geographic Distributions

### 4.2.1 Job Locations

As discussed in the data and methods section, employment opportunities are modelled by existing job locations. There are approximately 1,507,715 jobs across GMCA, distributed as shown in Figure 4-1. It highlights the advantage of using an index to reflect accessibility, compared with travel time to a single employment centre, for several reasons. Job locations may be more densely packed in employment centres but are also distributed throughout urban and suburban areas, thus using the travel time to employment centre is a significant simplification. A location equidistant between Manchester and Oldham has a similar travel time as a location to the East of Oldham, though the former has 2 significant towns (with

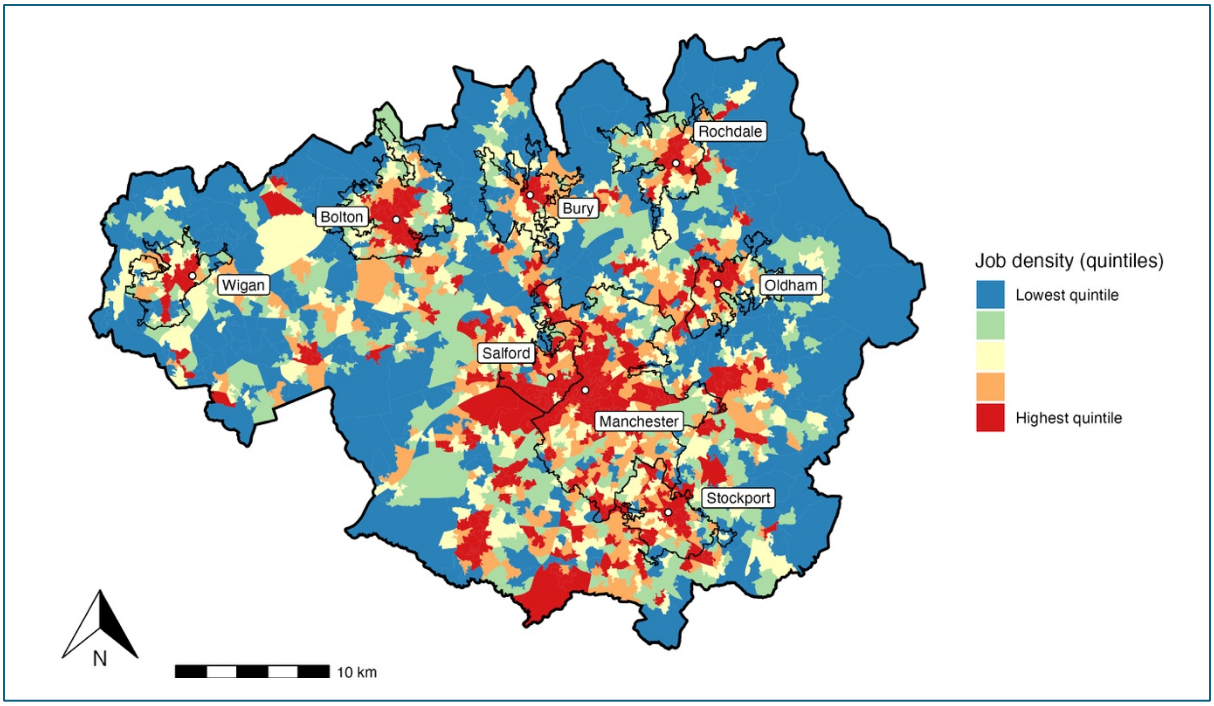



*Figure 4-1. Choropleth maps of job density distribution across Greater Manchester Combined Authority (GMCA) shown in quintiles, including the major town and city boundaries.*

commensurate employment opportunities) close by, and the latter is effectively a rural location.

Figure 4-1 also demonstrates the complexity around larger conurbations. Jobs within the Manchester-Salford centre are dispersed geographically from the city centres. The more jobs there are distant from the centre, the less accurate using travel time to the centre as a proxy for accessibility becomes.

Table 4-2 outlines the numbers of jobs and populations of the major cities and towns in GMCA, with Salford and Manchester having the highest, and both being co-centred in the city region. This presents a simplification, as all towns and cities have surrounding populations, and the delineation of the boundary is a subjective task; populations and jobs move around over time, but it does show that the larger cities are all hosting more jobs per person than the smaller conurbations.

| | Jobs | Economically Active Population | Jobs per person |
|---|---|---|---|
| Salford | 135,550 | 104,009 | 1.30 |
| Manchester | 449,450 | 559,951 | 0.80 |
| Stockport | 72,125 | 104,038 | 0.69 |
| Wigan | 38,790 | 74,354 | 0.52 |
| Bury | 26,180 | 59,523 | 0.44 |
| Oldham | 43,865 | 116,125 | 0.38 |
| Bolton | 64,660 | 174,224 | 0.37 |
| Rochdale | 33,675 | 97,065 | 0.35 |

*Table 4-2. Employment and Population in Greater Manchester Combined Authority (GMCA) Cities and Towns.*

### 4.2.2 Unemployment

The unemployment rate is derived from the 2021 census (ONS, 2024c). The geographic distribution in figure 4-2 is presented in quintiles as there are significant outliers which distort the distribution. All towns and cities have areas of high unemployment. The overall UK unemployment rate from this dataset is 4.9%, compared with 5.6% for GMCA. Within GMCA, the urban (within a town or city boundary line) unemployment rate is 7.4% compared with a suburban (outside of a town or city boundary) unemployment rate of 5.1%, which is to be expected as urban centres typically suffer multiple deprivations to a higher extent than more suburban areas (Bailey and Minton, 2017). It is also visually evident that unemployment levels within urban centres are more similar than if compared with suburban areas, which is confirmed analytically later. The inset map of Manchester City Centre has been coloured according to its own quintiles to show the variation within the city centre more clearly.

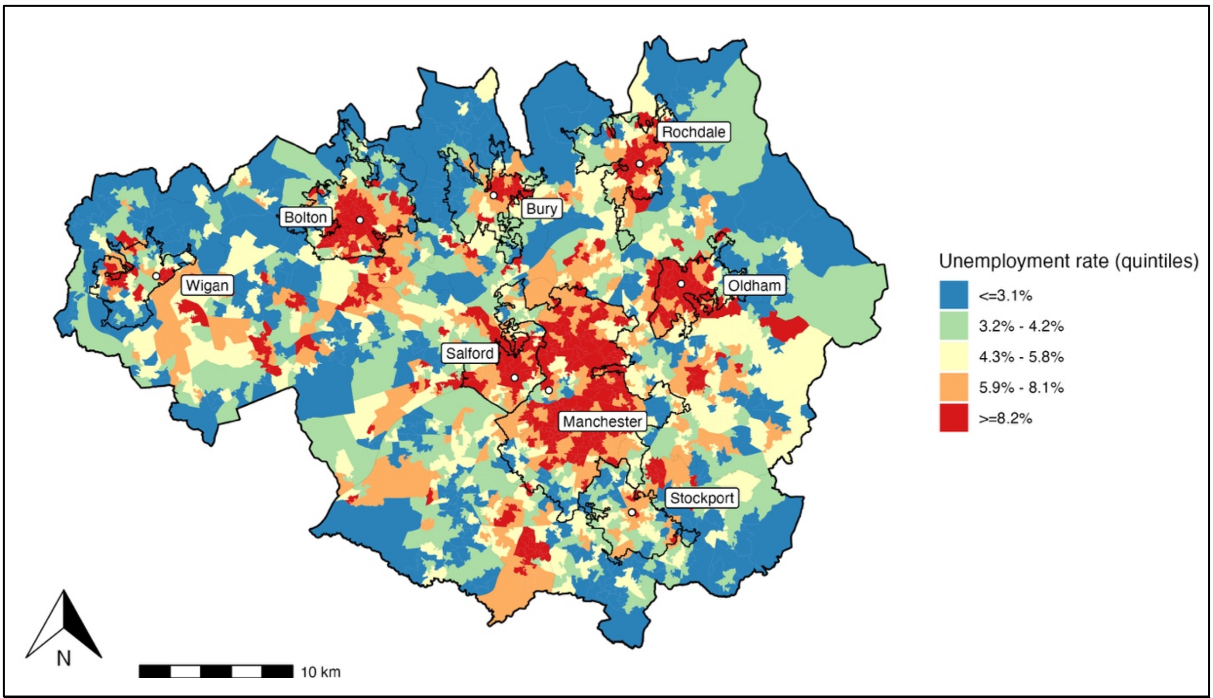
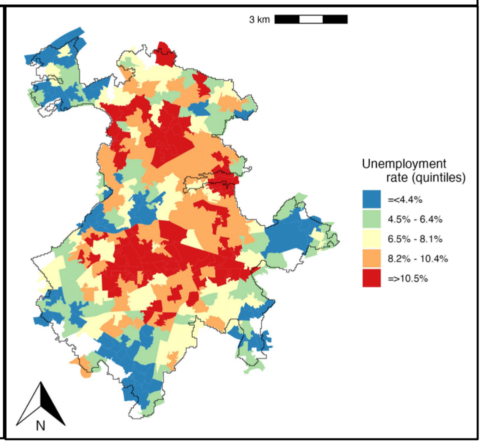



*Figure 4-2. Choropleth maps showing (above) unemployment rates at Lower Super Output Area (LSOA) level across Greater Manchester Combined Authority (GMCA) including the boundaries of the major towns and cities. Manchester city centre (right) with quintiles adjusted to highlight distribution.*




### 4.2.3 Travel Time to Employment Centre

The employment centres displayed on Figure 4-3 and used in the calculation of travel time are the centres of the major towns and cities as published by the ONS. There are 10 LADs in GMCA, 8 of which have a major town or city as their administrative centre. The other 2 members of GMCA (Tameside and Trafford) do have towns as their administrative centre, but they are not large enough to feature on the ONS list. This highlights one of the assumptions implicit in measuring travel time to a local employment centre; determining what constitutes an employment centre. There are 6 more employment centres within a 20km radius outside the GMCA boundary (Blackburn and Burnley to the North, Halifax and Huddersfield to the East and St Helens and Warrington to the West) which were included in the calculation to mitigate boundary effects. To the West of Wigan on the map you can see some evidence of travel times towards Warrington which shows the importance of including the additional areas in the buffer region. This approach using the same 20km buffer region is implemented into the PTJA calculation as well.

The travel times to the employment centres show a consistent story whether close to a Metrolink line or otherwise; the public transport connections to the employment centres are not fast; just 37% of the population can reach their local employment centre within 30minutes, though rising to 77% within a 45-minute threshold. This demonstrates one of the limitations of tram networks; lower transfer speed. Whilst being an efficient and low cost mass-transit mode, being limited to road speeds and safe acceleration levels for streetscapes means transfer times can be longer and will not compete with an underground rail network. There is also only a limited indication on the map of the benefit inferred by the Metrolink; there are some slightly lighter areas along the Metrolink corridors in some areas but it would be difficult to know the route if not explicitly marked on the map. The Metrolink has advantages which will not be recognised in this model. It is more able to adhere to its timetable than the bus service, as it takes priority at intersections and parts of its journey are segregated from road traffic. The advantageous reliability over the bus service could modelled as future works in using the actualised timetables of the public transport modes.

The Urban Big Data Centre at the University of Glasgow (Verduzco Torres and McArthur, 2022) found that approximately 40% of the population of Great Britain are within 45 minutes by public transport of their nearest main urban centre. There is a subjective set of

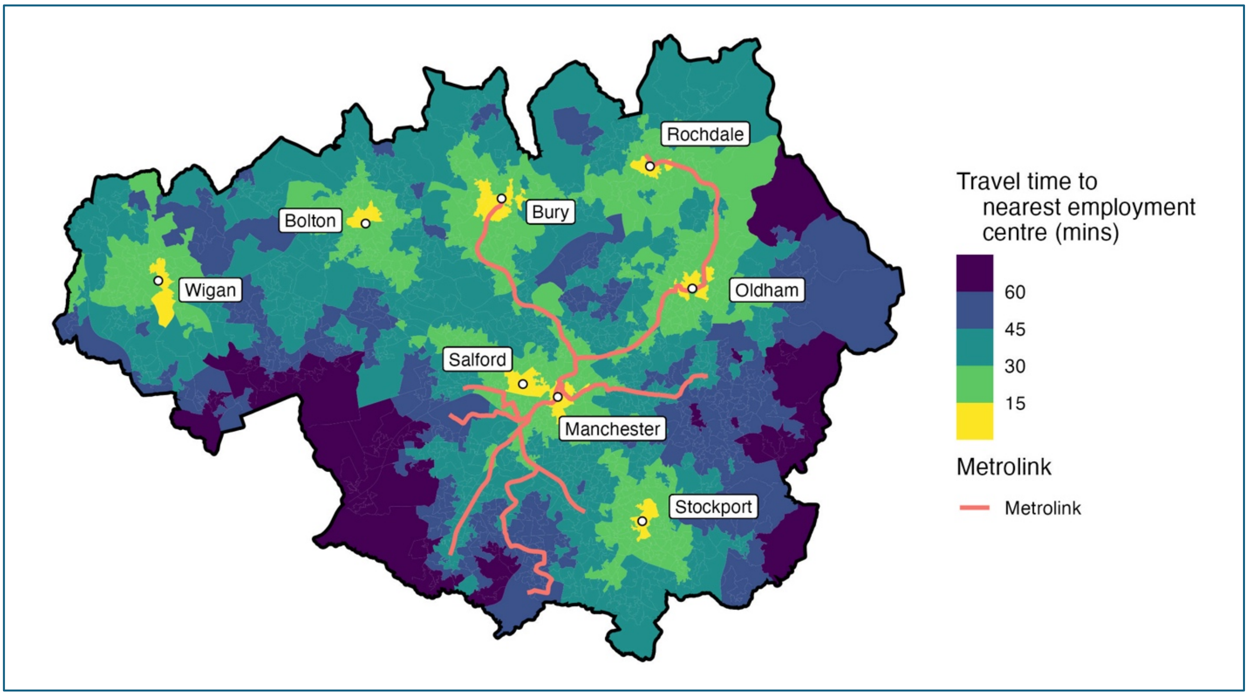



*Figure 4-3. Choropleth map of GMCA showing travel time to nearest employment centre and Metrolink tram lines.*

towns which should qualify as a major conurbation, and consequently the UBDC have used a dataset with 42% more main centres than are published in the ONS dataset. This indicates that the public transport access to the local town centre in GMCA significantly exceeds the average across the country, which is as expected due to the extensive tram network and the higher density than average across the country eliciting a more effective bus network. There's also a significant proportion, approximately 17% according to the 2011 census (GOV.UK, 2021b), of the UK population who live in rural areas, which have less effective PT and would expect a significantly longer transit time to their local employment centre.

The travel time to nearest employment centre has provided useful information and a firm basis on which to test aspects of the transport model, but similar PTJA it is a measure of accessibility, and we cannot use both in the regression model.

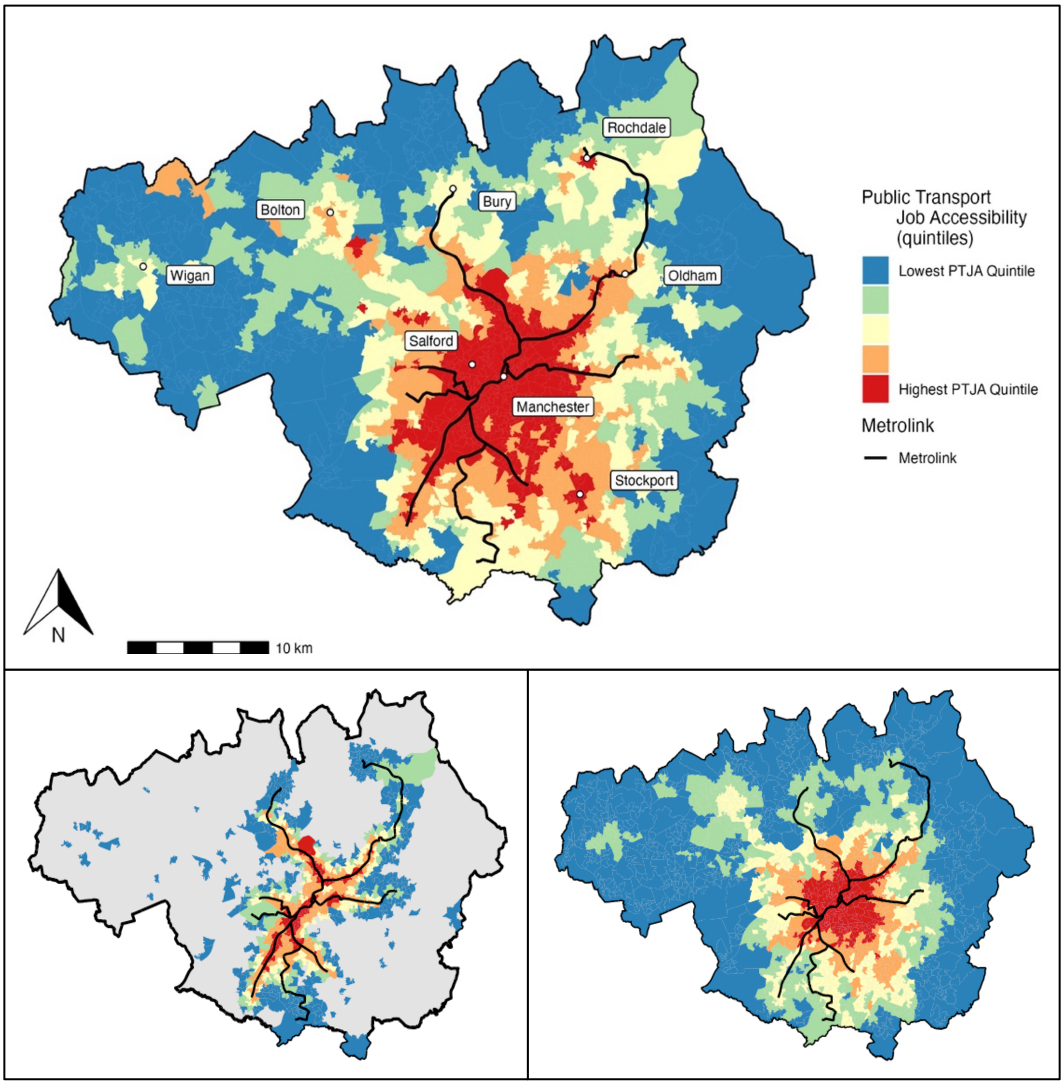



*Figure 4-4. Choropleth Map of Public Transport Job Accessibility (PTJA) across Greater Manchester Combined Authority (GMCA) showing Metrolink tram network lines. PTJA contribution from tram network (bottom left) and bus network (bottom right) shown for comparison. Colour categories have been kept identical.*

### 4.2.4 Public Transport Job Accessibility

The index is calculated using the distribution of jobs and the public transport travel time matrix as described earlier in the data section and is shown in Figure 4-4. It is displayed in quintiles as the absolute values vary significantly, with the more accessible areas being highly concentrated in the city centres, where excellent public transport links coincide with high numbers of jobs. It is noted that the PTJA outside of Manchester and Salford City Centre is low, even in the peripheral town centres where public transport access is good, as per figure 4-3. This would suggest that having expeditious public transport accessibility to your nearest employment centre is not comparable to having good links with several employment centres or having a fast link with a large employment centre. The mono-centricity of Figure 4-4 implies that Manchester city centre is the employment hub of the region, and the smaller peripheral employment centres are less important in providing employment opportunities. This is due to the very high density of jobs and public transport in Manchester city centre.

The impact from the Metrolink is more clearly visible in Figure 4-4 (PTJA) compared with the Figure 4-3 (travel time to employment centre). This might suggest that the Metrolink connects residences to job locations more specifically than to the town centres, or that employment opportunities have migrated to areas close to the Metrolink itself, as some of the lines have been opened for over 20 years. Figure 4-4 also shows the contribution to PTJA from the tram and bus (bottom maps). The quintile cutoffs have been kept the same for all three plots, and this demonstrates the ubiquitous contribution of the urban bus network in comparison with the highly unequal benefits inferred by the Metrolink. But also, that the Metrolink alone provides high (red) job accessibility to the areas it serves, whereas the bus network can only provide such high access to areas which are in close proximity of very high job densities.

### 4.2.5 - RQ-1: How does access to jobs by public transport vary across the Greater Manchester Combined Authority (GMCA) area?

It is important to consider Figures 4-1, 4-3 and 4-4 in discussing RQ-1 as there are a number of factors to consider. The travel time to employment centre towards Manchester is not visibly different to the towns at the periphery (per Figure 4-3); the public transport access seems visually similar by the relative sizes of the isochrones. But due to the relative lack of employment in the outer regional towns (Figure 4-1), that public transport access isn't providing the same job access benefit to the residents as the Metrolink or radial bus routes going into Manchester/Salford city centres. This results in the PTJA plot showing a highly monocentric city region with all of the high access areas being in Manchester or along the Metrolink corridor. The Metrolink has contributed to distributing access to those inner-city jobs to the suburbs, and extending the tram network will continue to distribute, but the local government could also consider policies to encourage business activity in the regional towns as well to expand employment opportunities outside of Manchester/Salford centres.

## 4.3 Regression Model Testing

### 4.3.1 Fixed Effects

Unemployment studies often face the challenge of autocorrelation; it is not evenly distributed throughout any geography. There are areas of high unemployment, and they have geographic relationships which are highly nuanced and contextual. This violates the observation independence assumption for OLS. Accordingly, and as used in several other studies of this nature (Johnson, Ercolani and Mackie, 2017; Jin and Paulsen, 2018; Hernandez, Hansz and Massobrio, 2020), a fixed effects model is used to control for the spatial relationships. The areal unit is intrinsic to the analysis and accordingly some testing was carried out at three areal groupings; the LAD, the Towns and City boundaries and a combination of the two (see Method section for more detail).

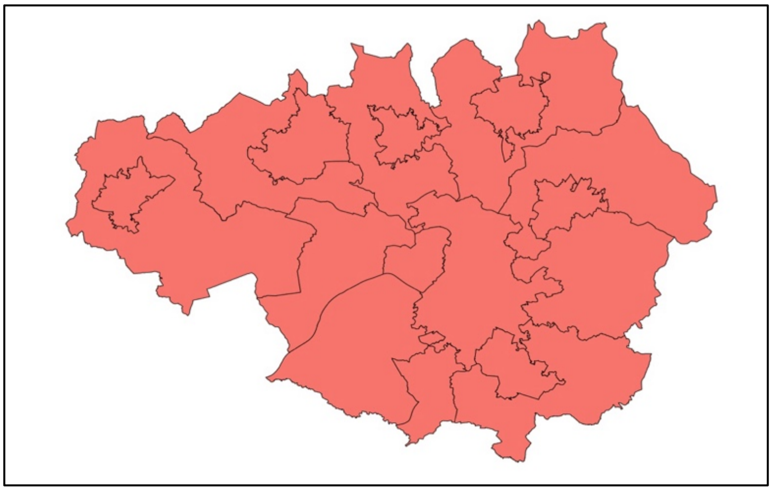


*Figure 4-5. Greater Manchester Combined Authority (GMCA) divided by major town and city boundaries then Local Authority Districts (LADs) for fixed-effect modelling.*

The test results for fixed effect modelling (Field, Miles and Field, 2021) are shown in table 4-3. The Akaike Information Criterion (AIC) describes the accuracy of the fit of the model to the data and decreases as the fit improves. In these models there are no covariates, but it is demonstrated that the fixed effects are offering increasing improvements as the geographies become smaller in area (i.e. moving down the list). The p-values are from model comparisons with the Intercept Only model and infer that the improvement of each model is significant. The combined area will be used for fixed effects throughout all modelling.

| Model | AIC | p-value |
|---|---|---|
| Baseline Model | 8,733 | NA |
| LAD Fixed Effects | 8,487 | <0.001 |
| Town and City Boundary Fixed Effects | 8,394 | <0.001 |
| Combination Area Fixed Effects | 8,358 | <0.001 |

*Table 4-3. Fixed-effect geographical level comparison.*

### 4.3.2 Linear Regression Assumptions

Fixed-effect regression is an adaptation of linear modelling, and as such the Best Linear Unbiased Estimator (BLUE) assumptions will be used to test the model. These are described in the methods section, and the results presented here.

### 4.3.3 Exogeneity

Assumption of exogeneity notes that covariates should be uncorrelated with the error term of the model. Endogeneity can be caused by several paths, including omitted variables and reverse causality, as discussed previously in the literature review section.

Travel times to nearest employment centre by public transport is providing an alternative measure of accessibility by public transport. It is not adding anything novel to the model but is causing endogeneity similar reasons to PTJA. In order to reduce this endogeneity, the variable will be removed from the model. The jobcentres are distributed quite evenly through the city region, and accordingly there is not so much variability in the travel time to them, so this variable will be removed from the final model as it is not adding anything and will also reduce the endogeneity through public transport accessibility.

Instrumental Variables (IV) are used, as discussed in the methods section, to reduce the endogeneity between PTJA and unemployment, and no-car rate and unemployment.

Population density has been used as an IV for PTJA and SEC: Management has been used for the No-car rate variable. There are two important criteria for a good instrumental variable: strength of relationship with the endogenous variable and exogeneity from the dependent variable. Exogeneity cannot be tested statistically (Ullah, Zaefarian and Ullah, 2021; Bastardoz et al., 2023), only theoretically described as in the methods section. But the strength of the instrument is assessable statistically by reporting the F-statistic of a regression with just the two variables in the model, along with the degrees of freedom for transparency (Bastardoz et al., 2023). The F-statistic should exceed 10 for a strong instrument (Stock, 2001). The Pearson correlation coefficient is also quoted to further assert the strength of the relationships.

| Endog.Var. ~ IV | F-Statistic | Degrees of Freedom | Pearson Correlation Coefficient |
|---|---|---|---|
| PTJA ~ Population Density | 485 | 1,700 | 0.47 |
| No-car rate ~ SEC Upper Management | 913 | 1,700 | -0.59 |

*Table 4-4. Instrument strength test results*

Endogeneity is the primary cause of inconsistency in studies of employment and transport accessibility and could be present in many different guises in the final model. Mitigation is a highly complex problem and could consume an entire study. The instrumental variable approach to resolving endogeneity is explored here but removing the issue entirely is conceptually implausible.

### 4.3.4 Linearity

Figure 4-6 shows the scatterplot of residuals from the FE-models. The mean line, shown in blue, should run along the zero line in order for the model to meet the linearity assumption, but in the FE-model it is entirely positive and curved upward quite significantly at higher fitted values of unemployment. The log-transformed model (middle) shows an improvement of the and is approximately linear, and that improvement is also visually reduced by the different axis scales between the plots (see appendix for plots shown on identical axes). The IV model is shown (bottom) as the output is also assumed to be linear. There is a deviation at low and high fitted values of unemployment in the IV model, but the residuals do appear linear through the middle values. This indicates model non-linearity and violates an assumption of linear modelling. The other continuous independent variables were transformed to check for improvement in the fit of the model, but all transformations made the component + residual plots (also shown in the appendix) less linear. Log-transforming the unemployment rate has improved the linearity and will be used for the rest of the modelling. The output from the IV model is significantly non-linear, and it consistently underestimates lower values unemployment and overestimates higher values.

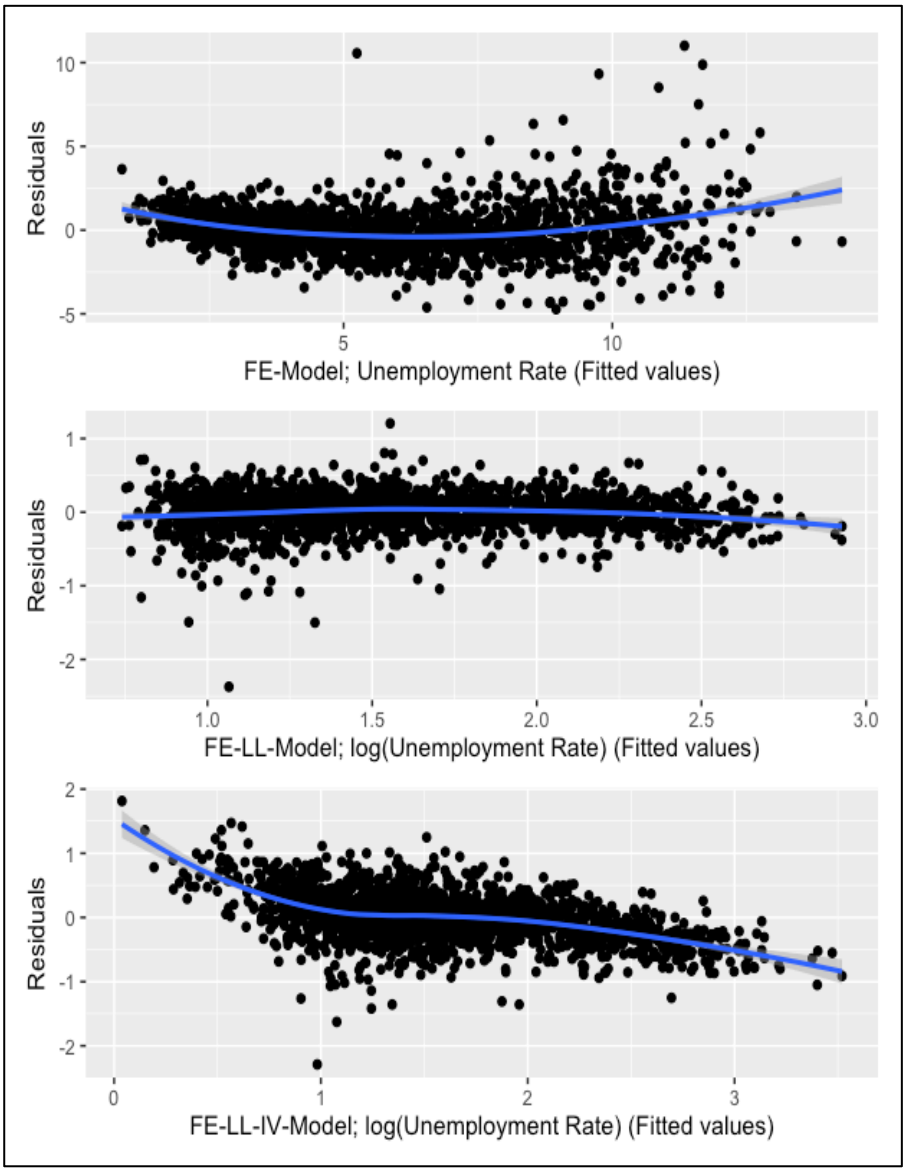



*Figure 4-6. Residuals vs Fitted plots for the Fixed Effects (FE) regression models with a linear dependent variable (FE-model - top), log-transformed dependent variable (FE-LL-Model - middle) and Instrumental Variable model (FE-LL-IV) (bottom)*

The Ramsey Regression Equation Specification Error Test (RESET) tests for linearity by comparing the base model with models including variables with higher orders of themselves to see if the fit is improved (Ramsey, 1969) indicating non-linearity. The test calculates an F-statistic and a p-value for the null hypothesis that higher orders of at least one variable provide no improvement to the model fit. Thus, rejecting the null hypothesis when $p<0.05$ means the model is improved by higher orders of one of the variables therefore is probably non-linear. The Ramsey RESET test statistics and p-values for the three models are presented in table 4-5 and demonstrates the log transform has a positive effect, though doesn't quite achieve linearity in the test, and also that the instrumental variable model is highly non-linear. There are component-residual plots for the FE and FE-IV models in the appendix, which also show same non-linearity in the IV model. Several transformations were trialled on the FE-IV model but none made any improvement over the log transform of the dependent variable.

| Model | RESET Statistic | p-value |
|---|---|---|
| OLS Model | 78 | 4.4e-33 |
| FE Model | 76 | 2.7e-32 |
| FE-LL Model | 17 | 7.2e-08 |
| FE-LL-IV Model | 50 | 7.5e-22 |

*Table 4-5. Ramsey RESET test results for OLS model (for comparison), FE Linear Model, FE-LL model and FE-LL-IV model.*

### 4.3.5 Homoskedasticity

The errors should have approximately constant variance across the dataset, or be homoskedastic, which can be assessed by the Breusch-Pagan test. The null hypothesis is homoskedasticity, so a p-value less than 0.05 (a rejection of the null hypothesis) means a probability of heteroskedasticity. Table 4-6 below shows the models' Breusch-Pagan statistics and highlights that the log-transformed model is a significant improvement, though there is still heteroskedasticity in the model. Robust standard errors (Eicker, 1963) are presented in the model to correct for heteroskedasticity, though it should be recognised that standard errors may still be underestimated. The results from the IV model are also presented and it’s observed that the errors show more heteroskedasticity than the FE-loglinear model.

| Model | Breusch-Pagan Statistic | p-value |
|---|---|---|
| FE Model | 147.8 | 3.9e-30 |
| FE-LL Model | 60.6 | 1.8e-05 |
| FE-LL-IV Model | 74.3 | 5.1e-16 |

*Table 4-6. Breusch-Pagan test results for heteroskedasticity in Fixed Effect Linear and Log-Linear models and Instrumental Variable FE Loglinear model. A p-value less than 0.05 indicates probable heteroskedasticity.*

### 4.3.6 No Autocorrelation

We have controlled for some spatial autocorrelation using the fixed effects at the local authority/city boundary level, though this has not completely eliminated the spatial interdependency of the observations. Table 4-7 shows the Global Moran’s I statistics for the key independent variable (PTJA) and dependent variable (unemployment rate) both showing a high Moran’s I statistic. Perfectly clustered data i.e. all similar values grouped together spatially, gives a statistic of +1 and perfectly distributed gives -1. A Moran’s I statistic equal to the expected value represents randomly distributed, and the null hypothesis of the test is that no spatial autocorrelation is present in the data. We reject the null hypothesis and infer that there is spatial autocorrelation in all 4 cases, though the benefits of the fixed effect modelling are evident in reducing the Morans’ I statistic to less than 0.1 for the residuals of the model. The loglinear model has slightly higher Moran’s I (and hence spatial autocorrelation) than the linear model, thus has removed slightly less autocorrelation, but the fixed effects have reduced a significant amount of the autocorrelation as is highlighted by comparison with the linear OLS model residuals.

| Variable / Model | Moran's I Statistic | Expected Moran's I | p-value |
|---|---|---|---|
| Unemployment Rate | 0.547 | -0.001 | 0.0e+00 |
| Public Transport Job Accessibility | 0.915 | -0.001 | 0.0e+00 |
| OLS Model Residuals | 0.136 | -0.001 | 1.0e-22 |
| FE-Linear Model Residuals | 0.107 | -0.001 | 8.6e-15 |
| FE-LL (LogLinear) Model Residuals | 0.094 | -0.001 | 1.0e-11 |
| FE-LL-IV Model Residuals | 0.292 | -0.001 | 0.0e+00 |

*Table 4-7. Results of Global Moran's I tests on the dependent variable, the key independent variable and the residuals from the linear and loglinear regression models*

### 4.3.1 - RQ2: Is fixed effects regression a suitable approach and what geographical level should it be set at?

Utilising fixed effects for geographical areas has offered significant advantages over OLS regression models. The approach reduces the geographic interdependence of observations, improving model fit and reducing heterogeneity. Table 4-2 shows the improvements of the fixed effects over the baseline model fit, and the improvement increases as the fixed effect geographical areas reduce. Testing the model at the MSOA (Mid-Super Output Area - collections of 4-8 LSOAs) would be an expeditious experiment but may result in reduced advantage as the areal effects are likely to span wider regions than individual MSOAs. The fixed effects model also has a positive effect on homoskedasticity as per table 4-6 and the residuals of the model are less spatially autocorrelated, as shown in table 4-7. Testing suggests that the fixed effects approach is a positive improvement to OLS regression but that it doesn't fully meet the assumptions of a best linear unbiased estimator due to endogeneity (though that will be present in any model), and the linearity is improved by the log-transformed dependent variable but does not meet the RESET test threshold and the heteroskedasticity can be mitigated using robust standard errors.

### 4.3.1 Testing Summary

Some other testing (multicollinearity / variance inflation factors, influential outliers, normality of residuals) has been carried out on the FE-LL and FE-LL-IV models, and the results are presented in the Appendix.

The assumptions testing has revealed that a log transformation on the dependent variable (unemployment rate) improves the FE model linearity of both the dependent variable and several covariates (see Appendix for covariate-plots) and reduces the heteroskedasticity, though doesn't eliminate it. Applying instrumental variables in a 2-stage regression makes the output less linear when compared with the FE-loglinear model, and also increases the heteroskedasticity, but it will reduce the endogeneity as described in the regression table below. There are several reasons to doubt the robustness of the IV model, and the results will be presented only as an indication of the probable direction of the bias due to endogeneity. The fixed effect loglinear (FE-LL) model is the most accurate model available following this investigation and will be presented and discussed in the results.

## 4.4 Regression results

Table 4-8 shows the coefficients and goodness-of-fit values for the FE-LL model and the FE-IV model alongside. Both models are displayed to demonstrate the effect that the instrumental variable has on the results. The Wu-Hausman statistic for the IV model is reported in the table and indicates a highly statistically significant result ($p<0.00005$). Thus, the null hypothesis of no improvement is rejected, and we can note that instrumenting the two IVs for is likely to have inferred a reduction in the endogeneity in the model.

As the instrumental variable has reduced the endogeneity-bias in the model, the direction of the effect is said to be indicative of the opposite of that bias. As the implementing the IV increases the magnitude of the coefficient on PTJA it can be said that the bias associated with endogeneity has a reducing effect, and it follows that the FE-model is underestimating the magnitude of the relationship between PTJA and unemployment rate.

The PTJA variable is shown as ($x10^3$) which represents a 1000-employee company in close vicinity to the LSOA origin point. This is considered straightforward to conceptualize, though as travel time increases the gravity decay function reduces the attractiveness; a 10-minute bus journey reduces the PTJA of that destination by approximately 10% and a 60-minute journey reduces by approximately 50%.

An additional 1,000 jobs in the immediate vicinity would be related to a reduction of 0.06% in the unemployment rate in the area, when controlling for rate of car ownership, ethnicity, rate of single parent households, and rate of low qualification levels. This would mean the unemployment rate reducing from 8% to 7.9952%, for example.

As the proportion of households without a car increases in the LSOA, the unemployment level also increases. This is expected as unemployed people do often not have resources to purchase or lease a vehicle, and also aligns with results of other studies (Ozbay, Ozmen and Berechman, 2006; Johnson, Ercolani and Mackie, 2017). When instrumented by proportion of senior managers to reduce the endogeneity, the effect size trebles indicating that the endogeneity is likely to be reducing the magnitude of the coefficient in the model.

All other coefficients are statistically significant and show the as-expected polarity in the FE-LL model. Single-parent household and low-qualified rate become statistically insignificant when the IVs are introduced, suggesting that the endogeneity is increasing the effect-size of those two variables. For each additional 1% of white people in the LSOA, the unemployment rate falls by 0.8% For 1% more single-parent households the unemployment rate increases by 1.2%. For 1% increase in the proportion of low qualified people, the unemployment rate would increase 0.65%. These results are all as expected and in accordance with other studies on the topic.

The $R^2$ value for the FE-LL model is 0.78, indicating that a large amount of the variation in the data is explained by the model. The $R^2$ calculation becomes meaningless for IV models, hence why it is shown as ‘not applicable’ (Waddell, 2019).

| | *Dependent variable:* Log (Unemployment rate) | |
|---|---|---|
| | FE-LL | FE-LL-IV |
| Public Transport Job Accessibility (PTJA) ($x10^3$) | -0.0006 (0.0001) | |
| | **p = 0.0000*** | |
| Estimated PTJA - IV ($x10^3$) | | -0.0059 (0.0016) |
| | | **p = 0.0002*** |
| No-car rate (%) | 0.0219 (0.0008) | |
| | **p = 0.0000*** | |
| Estimated no-car rate - IV (%) | | 0.0467 (0.0056) |
| | | **p = 0.0000*** |
| White people (%) | -0.0078 (0.0004) | -0.0124 (0.0019) |
| | **p = 0.0000*** | **p = 0.0000*** |
| Single-parent households (%) | 0.0194 (0.0017) | 0.0059 (0.0037) |
| | **p = 0.0000*** | p = 0.1132 |
| Low-qualified rate (%) | 0.0065 (0.0011) | -0.0099 (0.0051) |
| | **p = 0.0000*** | p = 0.0505 |
| AIC | 6160 | 1746 |
| Observations | 1,702 | 1,702 |
| $R^2$ | 0.7771 | NA |
| Residual Std. Error (df = 1679) | 0.2614 | 0.4012 |
| *Note:* | Standard errors in parentheses. * = p < 0.05 | |

*Table 4-8. Regression results for fixed-effects-loglinear (FE-LL) and instrumental variable (FE-LL-IV) models. Analysis unit is Lower Super Output Area in Greater Manchester Combined Authority. Dependent variable is log(unemployment rate). Standard errors and p-values are transformed for heteroskedasticity robustness.*

The testing results in table 4-9 confirm as described earlier in table 4-4 that the instruments are strong, according to the ex-ante testing. The Wu-Hausman statistic is presented, confirming the null hypothesis rejected, and that the consistency is improved when using the IVs in 2-stage regression model.

| Test | Statistic | p-value |
|---|---|---|
| Weak instruments (PT_Job_Access_Index) | 70 | 6.6e-30 |
| Weak instruments (No_car_rate) | 338 | 4.9e-124 |
| Wu-Hausman test for exogeneity | 54 | 2.7e-23 |

*Table 4-9. Instrument strength tests ex-ante*

### 4.4.1 - RQ-3 How does public transport job accessibility relate to unemployment rate across GMCA?

The regression model output suggests a negative relationship between public transport job accessibility and unemployment rate, when controlling for ethnicity, qualification level, proportion of single-parent households and rate of unavailability of private vehicles. This is in accordance with the findings from similar case studies in different contexts (see literature review, table 2-1). The magnitude of the relationship is small, which also agrees with other studies (Johnson, Ercolani and Mackie, 2017; Bastiaanssen, Johnson and Lucas, 2022) There is (and always will be) endogeneity bias present in the model, but the instrumental variables to address it have suggested that it infers a negative bias, resulting in underestimation of the size of the relationship.

The plots and model results combined together describe a complex environment in GMCA and make a significant contribution to the empirical literature, and present answers to the research questions. It is useful in describing the relationships between socio-economic factors and unemployment rate, a key economic indicator. There are supplementary plots and tables available in the Appendix for further information.

# 5 Conclusions

Labour market outcomes, and specifically unemployment levels in this instance, are a measure of the effectiveness of the urban economy and infrastructure at supporting its citizens. Public transport networks move people more effectively than private vehicles in urban conurbations, and many Global North cities are strategizing to transition their urban transport networks towards public transport and active travel as a result. Therefore, the interaction between our public transport networks and urban economic systems will become increasingly important for feeding the planning process and supporting investment decisions.

The context of Greater Manchester Combined Authority (GMCA) city region is polycentric and provides a diverse range of rural, urban and peri-urban environs in close proximity across a diverse landscape. It has a significant fixed transport infrastructure, namely the Metrolink Tramway and national railway lines, and local and regional bus services. The local transport authority has recently taken back control of the local bus network, after 40 years of private sector operation, and will also franchise the train network from 2028. Therefore, the capability to integrate and adjust the public transport network and policies to support economic development will be back in full public control. The results of this study can be developed to support transport infrastructure cost-benefit analyses for investments or adjustments to maximise economic gains.

The notable findings from the geographic distributions were that the accessibility to the regional towns is not significantly less than to the major conurbations at the centre of the city region, but that the lack of jobs in the peripheral conurbations makes the public transport less effective at providing accessibility to employment opportunities.

The statistical modelling of public transport job accessibility and unemployment in GMCA has provided a quantified relationship, and it suggests that endogenous effects are causing an underestimation of the relationship. The size of the relationship is small but statistically significant and suggests that positive labour market outcomes are achievable from improving public transport travel times.

This study only captures availability aspects of public transport accessibility, which are important and also quantifiable from publicly available datasets. But public transport accessibility is much more than timetables, and incorporates safety and security, journey reliability, affordability, physical accessibility, inclusivity, comfort, and whilst the study is agnostic to these concepts, they should not be underestimated and must be considered in any public transport improvements.

## 5.1 Study Limitations

The size of the relationship between PTJA and unemployment is likely to be related with the proportion of people using public transport for commuting, and the influence would be expected to increase as more people commute by public transport. The transport influence on unemployment would be split between the different travel modes, though allowance has been made for the proportion of households which have no access to a private vehicle.

This model is only illustrating relationship between unemployment and public transport and is not intended to make any inference as to the actual benefit of the public transport network, as the complexity around development of individual travel patterns is built on decades of infrastructure development, land use patterns, cultural norms and all the other accessibility issues not modelled herein.

Any conclusions derived from transport studies should not be applied to different contexts directly, as

## 5.2 Future Works

There are some improved modelling approaches suggested at the end of the Data and Methods section which are not repeated here but could represent future works to develop the regression model.

The bus and tram timetables could be replaced with actualised timetables, derived from the position data available from the Bus Open Data Service (BODS). This would give a better view of how the public transport system is operating than the timetabled services. However, the difference between timetabled and actual schedules is more likely to influence the decision whether to use public transport over a private vehicle for commuting, whereas unemployment may be influenced more by job application and acceptance, which in turn is more likely centred on timetabled routing.

It would be interesting to test the efficacy of the PTJA index developed here against other contemporary accessibility measures, such as travel time to employment centre, or jobs within an isochrone. And to assess the relationship with population density to understand how the accessibility is related with residential location (Saghapour, Moridpour and Thompson, 2016).

Finally, there is a possibility to extend this work to other combined authorities across the UK to conduct comparisons between cities and assess equity of provision.

The strategy for de-prioritising private vehicles in urban areas will result in increased patronage on public transport, and understanding how interactions with our urban economy will become increasingly important. This study reviewed the literature and techniques for correlating employment outcomes with public transport job accessibility and draws inference from the data from Greater Manchester. However, there are many more avenues to explore, and the analysis techniques can be refined to give more robust results.

## References

Abrantes, P., Fuller, R. and Bray, J. (2013) *The Case for the Urban Bus*. Available at: https://www.urbantransportgroup.org/system/files/general-docs/pteg%20Case%20for%20bus%20report%20FINAL.pdf (Accessed: 18 March 2024).

Anderson, P., Owen, A. and Levinson, D. (2012) *The Time Between: Continuously-defined accessibility functions for schedule-based transportation systems*. Available at: https://transportist.org/2012/08/13/the_time_between_continuously/ (Accessed: 18 June 2024).

Bastardoz, N. *et al.* (2023) 'Instrumental variables estimation: Assumptions, pitfalls, and guidelines', *Leadership Quarterly*, 34(1). Available at: https://doi.org/10.1016/j.leaqua.2022.101673.

Bastiaanssen, J., Johnson, D. and Lucas, K. (2020) 'Does transport help people to gain employment? A systematic review and meta-analysis of the empirical evidence', *Transport Reviews*, 40(5), pp. 607–628. Available at: https://doi.org/10.1080/01441647.2020.1747569.

Bastiaanssen, J., Johnson, D. and Lucas, K. (2022) 'Does better job accessibility help people gain employment? The role of public transport in Great Britain', *Urban Studies*, 59(2), pp. 301–322. Available at: https://doi.org/10.1177/00420980211012635.

Bauer, J. and Groneberg, D.A. (2016) 'Measuring spatial accessibility of health care providers-introduction of a variable distance decay function within the floating catchment area (FCA) method', *PLoS ONE*, 11(7). Available at: https://doi.org/10.1371/journal.pone.0159148.

Breach, A. and Swinney, P. (2024) *Climbing the Summit Big cities in the UK and the G7*. Available at: https://www.centreforcities.org/wp-content/uploads/2024/06/Climbing-the-Summit-June-2024.pdf (Accessed: 18 May 2024).

References

Breuer, M. and Dehaan, E. (2024) ‘Using and Interpreting Fixed Effects Models’, *Journal of Accounting Research*, 62(4), pp. 1183–1226. Available at: https://doi.org/10.1111/1475-679X.12559.

Carey, D. (2024) ‘Metrolink Expansion Plans Announced’, *Manchester Evening News*. Available at: https://www.manchestereveningnews.co.uk/news/greater-manchester-news/new-metrolink-map-tram-manchester-29510064 (Accessed: 13 August 2024).

Carr, C.M.H. and Whitehand, J.W.R. (2001) ‘The Scale and Causes of Suburban Growth’, in *Twentieth-Century Suburbs*. Taylor & Francis Group, pp. 20–40. Available at: http://ebookcentral.proquest.com/lib/gla/detail.action?docID=1666801.

Centre for Cities (2022) *Billions lost in productivity as Northern cities’ transport networks lag behind European equivalents*. Available at: https://www.centreforcities.org/press/northern-cities-transport-networks-lag-behind-european-equivalents/ (Accessed: 18 July 2024).

Cervero, R., Rood, T. and Appleyard, B. (1995) *Job Accessibility as a Performance Indicator: An Analysis of Trends and Their Social Policy Implications in the San Francisco Bay Area*. Available at: https://escholarship.org/uc/item/6mp941d9.

Clark, T., Foster, L. and Sloan, L. (2021) *Bryman’s Social Research Methods*. Fifth Edition. Oxford University Press. Available at: https://read.kortext.com/reader/epub/974675?page= (Accessed: 18 August 2024).

Cooke, S. and Behrens, R. (2017) ‘Correlation or cause? the limitations of population density as an indicator for public transport viability in the context of a rapidly growing developing city’, in *Transportation Research Procedia*. Elsevier B.V., pp. 3003–3016. Available at: https://doi.org/10.1016/j.trpro.2017.05.229.

Costa, A.B., Ramos, C. and Zheng, S. (2022) ‘Subway expansion, job accessibility improvements, and home value appreciation in four global cities: Considering both local and network effects’, *Journal of Transport and Land Use*, 15(1), pp. 613–634. Available at: https://doi.org/10.5198/jtlu.2022.2146.

References

Curl, A., Nelson, J.D. and Anable, J. (2011) 'Does accessibility planning address what matters? A review of current practice and practitioner perspectives', *Research in Transportation Business and Management*, 2, pp. 3–11. Available at: https://doi.org/10.1016/j.rtbm.2011.07.001.

Darby, J., McIntyre, S. and Roy, G. (2022) 'What can analysis of 47 million job advertisements tell us about how opportunities for homeworking are evolving in the United Kingdom?', *Industrial Relations Journal*, 53(4), pp. 281–302. Available at: https://doi.org/10.1111/irj.12375.

Department for Transport (2024) 'National Travel Survey, 2002-2022', *UK Data Service*, 17th Edition. Available at: https://doi.org/10.5255/UKDA-SN-5340-13.

DiMento, J. and Ellis, C. (2012) 'Urban Freeways and National Policy 1939 - 1945', in *Changing Lanes: Visions and Histories of Urban Freeways*. MIT Press. Available at: http://ebookcentral.proquest.com/lib/gla/detail.action?docID=3339557 (Accessed: 18 August 2024).

Dixon, S., Johnson, D. and Batley, R. (2019) 'A job accessibility index to evaluate employment impacts in isolated regions now restored to the rail network', *Transportation Planning and Technology*, 42(5), pp. 515–537. Available at: https://doi.org/10.1080/03081060.2019.1609223.

Econometrics with R (2024) *Fixed Effects Regression*. Available at: https://www.econometrics-with-r.org/10.3-fixed-effects-regression.html (Accessed: 18 August 2024).

Eicker, F. (1963) 'Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions', *The Annals of Mathematical Statistics.*, 34(2), pp. 447–456. Available at: https://www.jstor.org/stable/2238390 (Accessed: 18 August 2024).

Field, A., Miles, J. and Field, Z., (2012) *Discovering Statistics Using R,* London: Sage Publications

References

Fox, J., Kleiber, C. and Zeileis, A. (2023) *Diagnostics for 2SLS Regression*, *ivreg*. Available at: https://zeileis.github.io/ivreg/articles/Diagnostics-for-2SLS-Regression.html#introduction (Accessed: 10 August 2024).

Galí, J. (2011) 'The Return of the Wage Phillips Curve', *Journal of the European Economic Association*, 9(3), pp. 436–461. Available at: https://doi.org/10.1111/j.1542-4774.2011.01023.x.

Glasgow City Council (2022) *Glasgow Transport Strategy - Policy Framework*. Available at: www.glasgow.gov.uk/transportstrategy (Accessed: 18 July 2024).

GMCA (2021) *Greater Manchester Transport Strategy 2040*. Available at: https://assets.ctfassets.net/nv7y93idf4jq/01xbKQQNW0ZYLzYvcj1z7c/4b6804acd572f00d8d728194ef62bb89/Greater_Manchester_Transport_Strategy_2040_final.pdf (Accessed: 18 June 2024).

GMCA (2024) *This is Greater Manchester*. Available at: https://www.greatermanchester-ca.gov.uk/ (Accessed: 13 August 2024).

GOV.UK (2021) *National Travel Survey 2021: Household car availability and trends in car trips*, *National Travel Survey*. Available at: https://www.gov.uk/government/statistics/national-travel-survey-2021/national-travel-survey-2021-household-car-availability-and-trends-in-car-trips (Accessed: 14 August 2024).

GOV.UK (2024a) *Annual Bus Statistics*. Available at: https://www.gov.uk/government/statistics/annual-bus-statistics-year-ending-march-2023/annual-bus-statistics-year-ending-march-2023 (Accessed: 12 August 2024).

GOV.UK (2024b) 'Bus Open Data Service', *BODS* [Preprint]. Available at: https://www.bus-data.dft.gov.uk/ (Accessed: 18 August 2024).

GTFS (2024) *GTFS: Making Public Transit Data Universally Accessible*. Available at: https://gtfs.org/ (Accessed: 18 August 2024).

References

Hansen, W.G. (1959) ‘How Accessibility Shapes Land Use’, *Journal of the American Planning Association*, 25(2), pp. 73–76. Available at: https://doi.org/10.1080/01944365908978307.

Hernandez, D., Hansz, M. and Massobrio, R. (2020) ‘Job accessibility through public transport and unemployment in Latin America: The case of Montevideo (Uruguay)’, *Journal of Transport Geography*, 85. Available at: https://doi.org/10.1016/j.jtrangeo.2020.102742.

Hill, A.D. *et al.* (2021) ‘Endogeneity: A Review and Agenda for the Methodology-Practice Divide Affecting Micro and Macro Research’, *Journal of Management*, 47(1), pp. 105–143. Available at: https://doi.org/10.1177/0149206320960533.

Holzer, H.J. (1991) ‘The Spatial Mismatch Hypothesis : What Has the Evidence Shown?’, *Urban Studies*, 28(1), pp. 105–122.

Hu, L. (2017) ‘Job accessibility and employment outcomes: which income groups benefit the most?’, *Transportation*, 44(6), pp. 1421–1443. Available at: https://doi.org/10.1007/s11116-016-9708-4.

Hu, L. and Giuliano, G. (2017) ‘Poverty concentration, job access, and employment outcomes’, *Journal of Urban Affairs*, 39(1), pp. 1–16. Available at: https://doi.org/10.1111/juaf.12152.

Ihlanfeldt, K.R. and Sjoquist, D.L. (1991) ‘The Effect of Job Access on Black and White Youth Employment : A Cross-sectional Analysis’, *Urban Studies*, 28(2), pp. 255–265.

Internet Geography (2024) *How has urban change created challenges in Manchester?* Available at: https://www.internetgeography.net/topics/how-has-urban-change-created-challenges-in-manchester/ (Accessed: 12 June 2024).

Jin, J. and Paulsen, K. (2018) ‘Does accessibility matter? Understanding the effect of job accessibility on labour market outcomes’, *Urban Studies*, 55(1), pp. 91–115. Available at: https://doi.org/10.1177/0042098016684099.

References

Johnson, D., Ercolani, M. and Mackie, P. (2017) ‘Econometric analysis of the link between public transport accessibility and employment’, *Transport Policy*, 60, pp. 1–9. Available at: https://doi.org/10.1016/j.tranpol.2017.08.001.

Johnson, R.C. (2006) ‘Landing a job in urban space: The extent and effects of spatial mismatch’, *Regional Science and Urban Economics*, 36(3), pp. 331–372. Available at: https://doi.org/10.1016/j.regsciurbeco.2005.11.002.

Kain, John.F. (1968) ‘Housing Segregation, Negro Employment, and Metropolitan Decentralization’, *Quarterly Journal of Economics*, LXXXII(2), pp. 175–197. Available at: https://academic.oup.com/qje/article/82/2/175/1894117.

Korsu, E. and Wenglenski, S. (2010) ‘Job Accessibility, Residential Segregation and Risk of Long-Term Unemployment in The Paris Region’, *Urban Studies*, 47(11), pp. 2279–2324. Available at: https://doi.org/10.1177/0042098009357962.

Liverpool City Region Combined Authority (2022) *Developing a vision for local transport to 2040*. Available at: https://liverpoolcityregion-ca.moderngov.co.uk/documents/s60735/Enc.%201%20for%20Developing%20the%20new%20Local%20Transport%20Plan%20for%20the%20LCR.pdf (Accessed: 12 August 2024).

Livingstone, N., Fiorentino, S. and Short, M. (2021) ‘Planning for residential “value”? London’s densification policies and impacts’, *Buildings and Cities*, 2(1), pp. 203–219. Available at: https://doi.org/10.5334/bc.88.

Matas, A., Raymond, J.L. and Roig, J.L. (2010) ‘Job accessibility and female employment probability: The cases of Barcelona and Madrid’, *Urban Studies*, 47(4), pp. 769–787. Available at: https://doi.org/10.1177/0042098009352364.

Miner, P. *et al.* (2024) ‘Car harm: A global review of automobility’s harm to people and the environment’, *Journal of Transport Geography*, 115. Available at: https://doi.org/10.1016/j.jtrangeo.2024.103817.

References

Morgan, M. (2024) *UK2GTFS: Converts UK transport timetable datasets to GTFS format.*, *https://itsleeds.git.io/uk2gtfs*. Available at: https://github.com/itsleeds/uk2gtfs (Accessed: 18 August 2024).

Mouw, T. (2000) 'Job Relocation and the Racial Gap in Unemployment in Detroit and Chicago, 1980 to 1990', *American Sociological Association*, 5, pp. 730–753. Available at: https://www.jstor.org/stable/2657544 (Accessed: 13 August 2024).

Naro, G., Biraghi, C.A. and Lenzi, E. (2024) 'City Transport Analyzer: A Powerful Qgis Plugin For Public Transport Accessibility And Intermodality Analysis', in. International Society for Photogrammetry and Remote Sensing, pp. 113–119. Available at: https://doi.org/10.5194/isprs-archives-XLVIII-4-W12-2024-113-2024.

Naseer, Nyla. (2023) *At Home with Work: Understanding and Managing Remote and Hybrid Work*. Business Expert Press. Available at: https://portal-igpublish-com.ezproxy2.lib.gla.ac.uk/iglibrary/obj/BEPB0001273 (Accessed: 18 August 2024).

NBER (2022) *Cross-Country Evidence on Labor Market Fluidity and Wage Growth*, *National Bureau of Economic Research*. Available at: https://www.nber.org/digest/202203/cross-country-evidence-labor-market-fluidity-and-wage-growth (Accessed: 18 March 2024).

ONS (2018) 'Major Town and City Boundaries 2015', *Data.gov* [Preprint]. Available at: https://www.data.gov.uk/dataset/5e1ec4f2-2bd5-4a20-a442-ac19cfd0fceb/major-towns-and-cities-december-2015-boundaries (Accessed: 18 August 2024).

ONS (2023) *Quality and methodology information (QMI) for Census 2021*. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/qualityandmethodologyinformationqmiforcensus2021#quality-summary (Accessed: 18 August 2024).

ONS (2024a) 'Business Register and Employment Survey : open access', *nomisweb.co.uk* [Preprint]. Available at: https://www.nomisweb.co.uk/datasets/newbres6pub (Accessed: 18 August 2024).

ONS (2024b) *Business Register and Employment Survey (Rounding)*, *nomisweb.co.uk*. Available at: https://www.nomisweb.co.uk/articles/1103.aspx (Accessed: 18 August 2024).

ONS (2024c) 'Census 2021', *nomisweb.co.uk* [Preprint]. Available at: https://www.nomisweb.co.uk/sources/census_2021 (Accessed: 18 May 2024).

ONS (2024d) 'Combined Authority Boundaries', *Geoportal Statistics* [Preprint]. Available at: https://geoportal.statistics.gov.uk/datasets/3f826d170e224924bcf602d2d0d53b3c_0/explore (Accessed: 18 May 2024).

ONS (2024e) 'LSOA Boundaries 2021', *Geoportal Statistics* [Preprint]. Available at: https://geoportal.statistics.gov.uk/datasets/bb427d36197443959de8a1462c8f1c55_0/explore (Accessed: 18 May 2024).

OpenStreetMap (2018) 'OpenStreetMap Data Extracts', *Geofabrik* [Preprint]. Available at: https://download.geofabrik.de/europe/united-kingdom.html (Accessed: 18 August 2024).

Ozbay, K., Ozmen, ; D and Berechman, J. (2006) 'Modeling and Analysis of the Link between Accessibility and Employment Growth', *Journal of Transport Engineering*, 132(5), pp. 385–393. Available at: https://doi.org/10.1061/ASCE0733-947X2006132:5385.

Page, T. (2011) *WikiMedia Commons - Greater Manchester Metrolink*. Available at: https://commons.wikimedia.org/wiki/File:Greater_Manchester_Metrolink_-_tram_3009A.jpg (Accessed: 13 August 2024).

Di Paolo, A., Matas, A. and Raymond, J.L. (2017) 'Job accessibility and job-education mismatch in the metropolitan area of Barcelona', *Papers in Regional Science*, 96, pp. S91–S112. Available at: https://doi.org/10.1111/pirs.12179.

Patacchini, E. and Zenou, Y. (2005) 'Spatial mismatch, transport mode and search decisions in England', *Journal of Urban Economics*, 58(1), pp. 62–90. Available at: https://doi.org/10.1016/j.jue.2005.01.005.

References

Pereira, R.H.M. *et al.* (2021) 'r5r: Rapid Realistic Routing on Multimodal Transport Networks with R5 in R', *Findings* [Preprint]. Available at: https://doi.org/10.32866/001c.21262.

Pinna, F. and Murrau, R. (2018) 'Age factor and pedestrian speed on sidewalks', *Sustainability (Switzerland)*, 10(11). Available at: https://doi.org/10.3390/su10114084.

Pont, M.B. *et al.* (2021) 'Systematic review and comparison of densification effects and planning motivations', *Buildings and Cities*, 2(1), pp. 378–401. Available at: https://doi.org/10.5334/bc.125.

Qiang, Y., Xu, J. and Zhang, G. (2020) 'The shapes of US cities: Revisiting the classic population density functions using crowdsourced geospatial data', *Urban Studies*, 57(10), pp. 2147–2162. Available at: https://doi.org/10.1177/0042098019871191.

Ramsey, J.B. (1969) 'Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis', *Source: Journal of the Royal Statistical Society. Series B (Methodological)*, 31(2), pp. 350–371. Available at: https://www.jstor.org/stable/2984219.

RDG (2024) *Rail Delivery Group Data Download*, *Rail Delivery Group* . Available at: https://www.raildeliverygroup.com/ (Accessed: 18 August 2024).

Rogers, C.L. (1997) 'Job Search and Unemployment Duration: Implications for the Spatial Mismatch Hypothesis', *Journal of Urban Economics*, 42, pp. 109–132. Available at: https://www-sciencedirect-com.ezproxy2.lib.gla.ac.uk/science/article/pii/S0094119096920173 (Accessed: 18 August 2024).

Saghapour, T., Moridpour, S. and Thompson, R.G. (2016) 'Public transport accessibility in metropolitan areas: A new approach incorporating population density', *Journal of Transport Geography*, 54, pp. 273–285. Available at: https://doi.org/10.1016/j.jtrangeo.2016.06.019.

Schork, J. (2024) *Fixed Effects in Linear Regression (Example in R) | Cross Sectional, Time & Two-Way*, *Statistics Globe*. Available at: https://statisticsglobe.com/fixed-effects-linear-regression (Accessed: 18 August 2024).

References

Scientific American (2015) *Exxon Knew about Climate Change almost 40 years ago*, *Scientific American*. Available at: https://www.scientificamerican.com/article/exxon-knew-about-climate-change-almost-40-years-ago/ (Accessed: 13 August 2024).

Shen, Q. (1998) ‘Location characteristics of inner-city neighborhoods and employment accessibility of low-wage workers’, *Environment and Planning B: Planning and Design*, 25, pp. 345–365. Available at: https://journals-sagepub-com.ezproxy2.lib.gla.ac.uk/doi/pdf/10.1068/b250345 (Accessed: 18 June 2024).

Stock, J. and Watson, M. (2019) ‘Regression with a Single Regressor Hypothesis Tests and Confidence Intervals’, in *Introduction to Econometrics*, pp. 178–210. Available at: https://ebookcentral.proquest.com/lib/gla/detail.action?docID=5640381. (Accessed: 10 March 2024).

Stock, J.H. (2001) *Testing for Weak Instruments in Linear IV Regression*. Department of Economics, Harvard University.

Taylor, B.D., Morris, E.A. and Brown, J.R. (2023) ‘The Rise of the Freeway Era’, in *The Drive for Dollars*. New York: Oxford University Press, pp. 197–213. Available at: https://doi.org/10.1093/oso/9780197601518.003.0008.

UK Government (2024) *Key Economic Indicators*, *House of Commons Library*. Available at: https://commonslibrary.parliament.uk/research-briefings/cbp-9040/ (Accessed: 12 August 2024).

Ullah, S., Zaefarian, G. and Ullah, F. (2021) ‘How to use instrumental variables in addressing endogeneity? A step-by-step procedure for non-specialists’, *Industrial Marketing Management*. Elsevier Inc., pp. A1–A6. Available at: https://doi.org/10.1016/j.indmarman.2020.03.006.

United States Census Bureau (2021) *US Census*. Available at: https://data.census.gov/table/ACSDP1Y2021.DP04 (Accessed: 14 August 2024).

References

Verduzco Torres, J.R. and McArthur, D.P. (2024) ‘Public transport accessibility indicators to urban and regional services in Great Britain’, *Scientific Data*, 11(1), pp. 1–14. Available at: https://doi.org/10.1038/s41597-023-02890-w.

Verduzco Torres, R. and McArthur, D. (2022) ‘Accessibility Indicators for Great Britain 2021 Accessibility Indicators for Great Britain-Technical appendix’. Available at: https://doi.org/10.5281/zenodo.6759240.

de Vries, J.J., Nijkamp, P. and Rietveld, P. (2009) ‘Exponential or power distance-decay for commuting? An alternative specification’, *Environment and Planning A*, 41(2), pp. 461–480. Available at: https://doi.org/10.1068/a39369.

Waddell, G. (2019) *Instrumental variables*, *Micro-Econometrics*. Available at: https://pages.uoregon.edu/waddell/metrics/index.html (Accessed: 18 August 2024).

## Appendix – Supplemental Plots and Tables

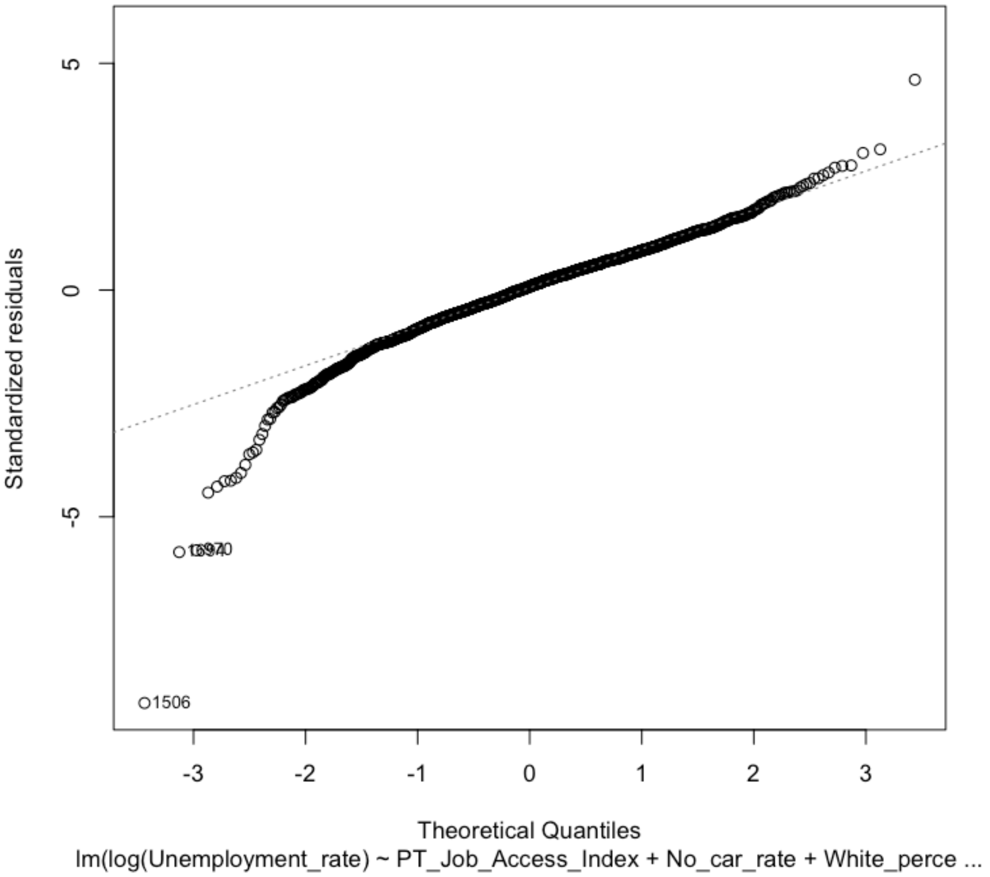



*Figure A- 1. Quantile-Quantile Plot of Normality of Residuals for FE-LL model*

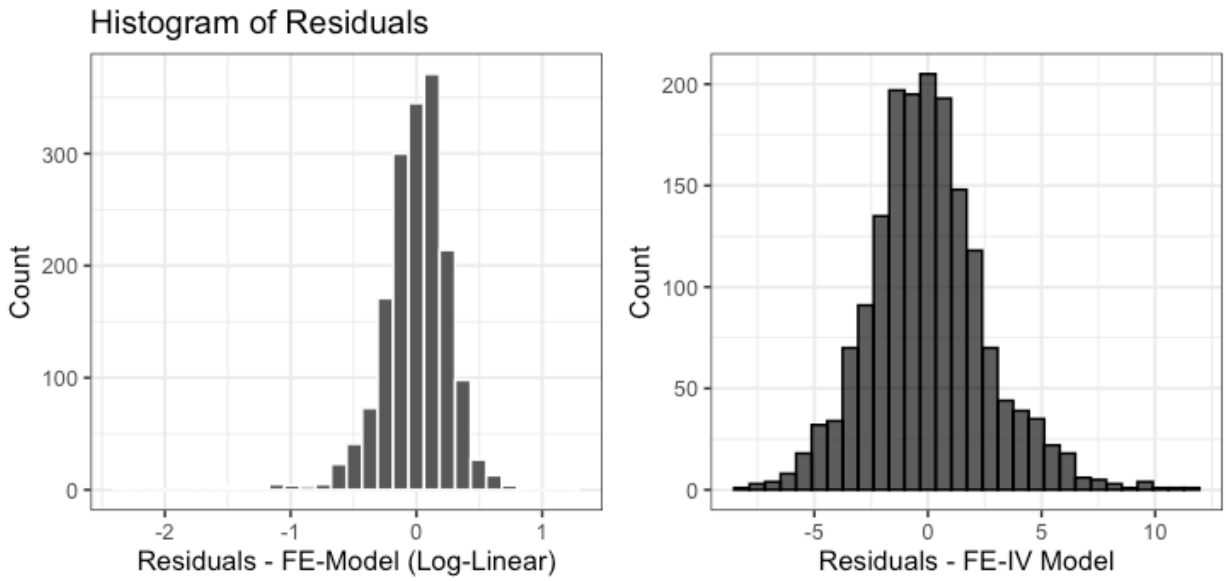



*Figure A- 2. Normality of Residuals plots for FE-LL and FE-LL-IV models.*

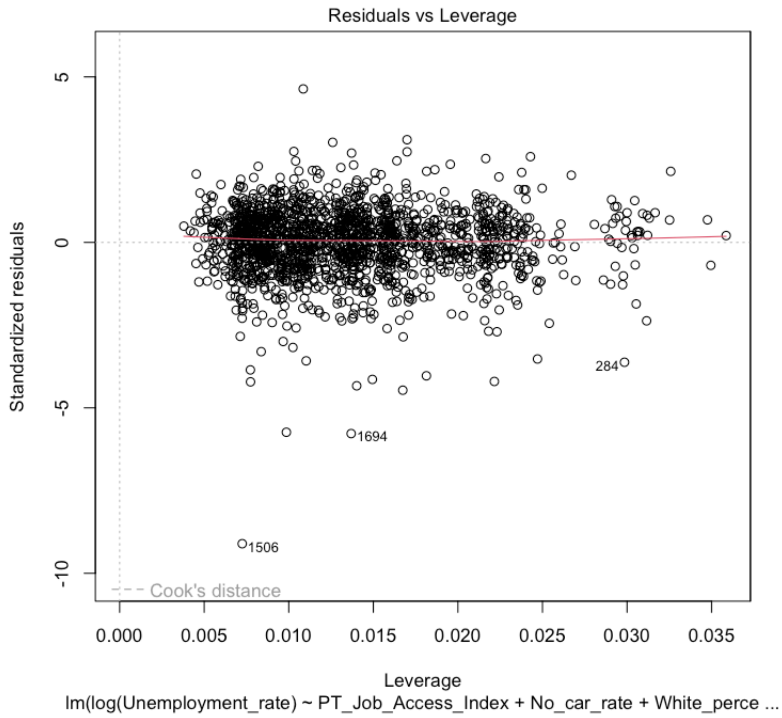



*Figure A- 3. Leverage vs. Residuals Plot for FE-LL model, demonstrating no influential outliers are present.*

| Variable | Generalised Variance-Inflation Factor |
|---|---|
| Public Transport Job Accessibility (PTJA) | 3.01 |

| Variable | Generalised Variance-Inflation Factor |
|---|---|
| No-car rate (%) | 2.46 |
| White people (%) | 1.43 |
| Single-parent Households (%) | 2.25 |
| Low-qualified rate (%) | 2.98 |

*Figure A- 4. Variance Inflation Factors for independent variables in the model. A value exceeding 5 indicates multicollinearity, which is not observed here.*

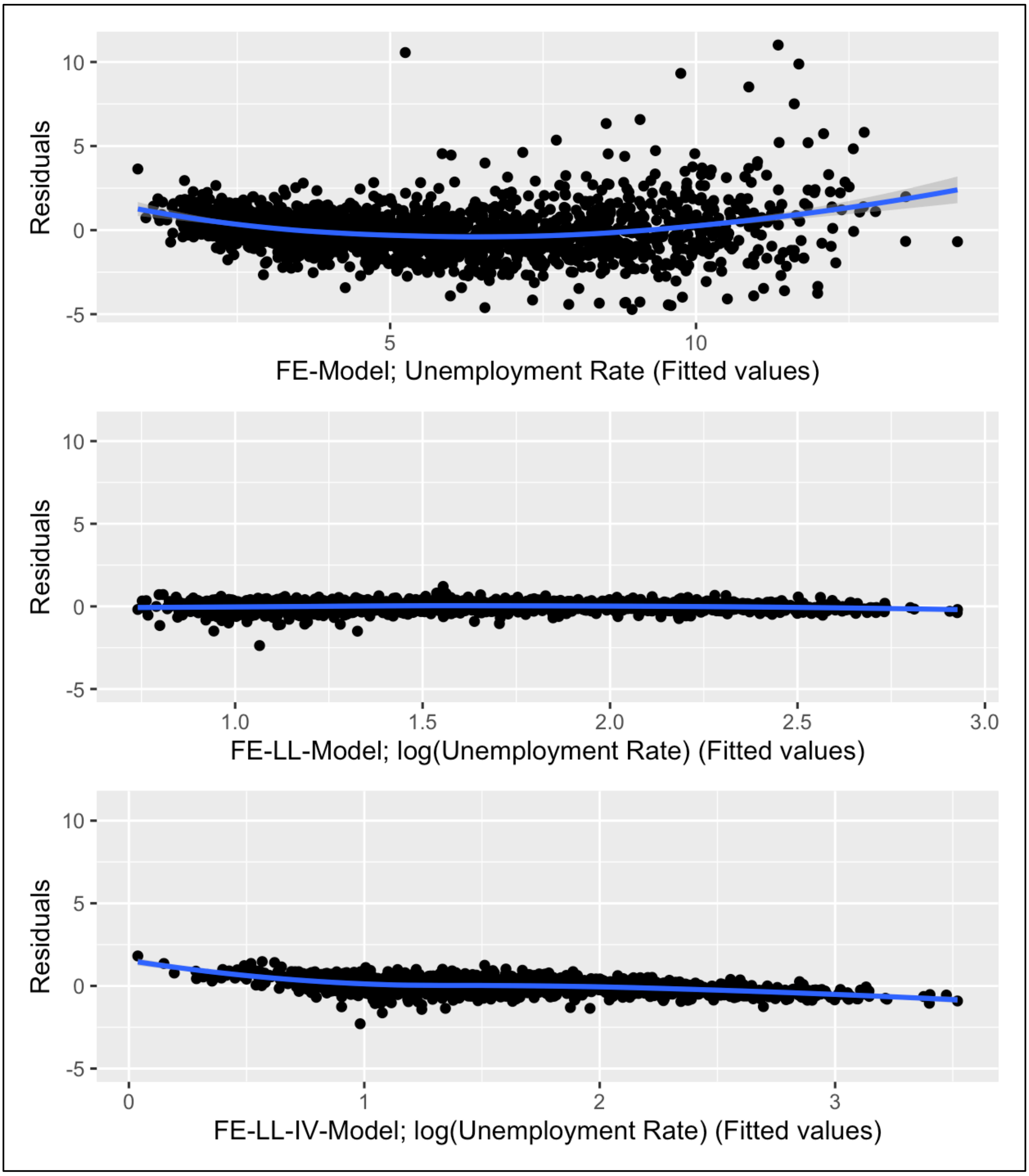



*Figure A- 5. Linearity of Residuals plots for FE-Model, FE-LL-Model, and FE-LL-IV-Model, shown on identical-scale y-axes.*

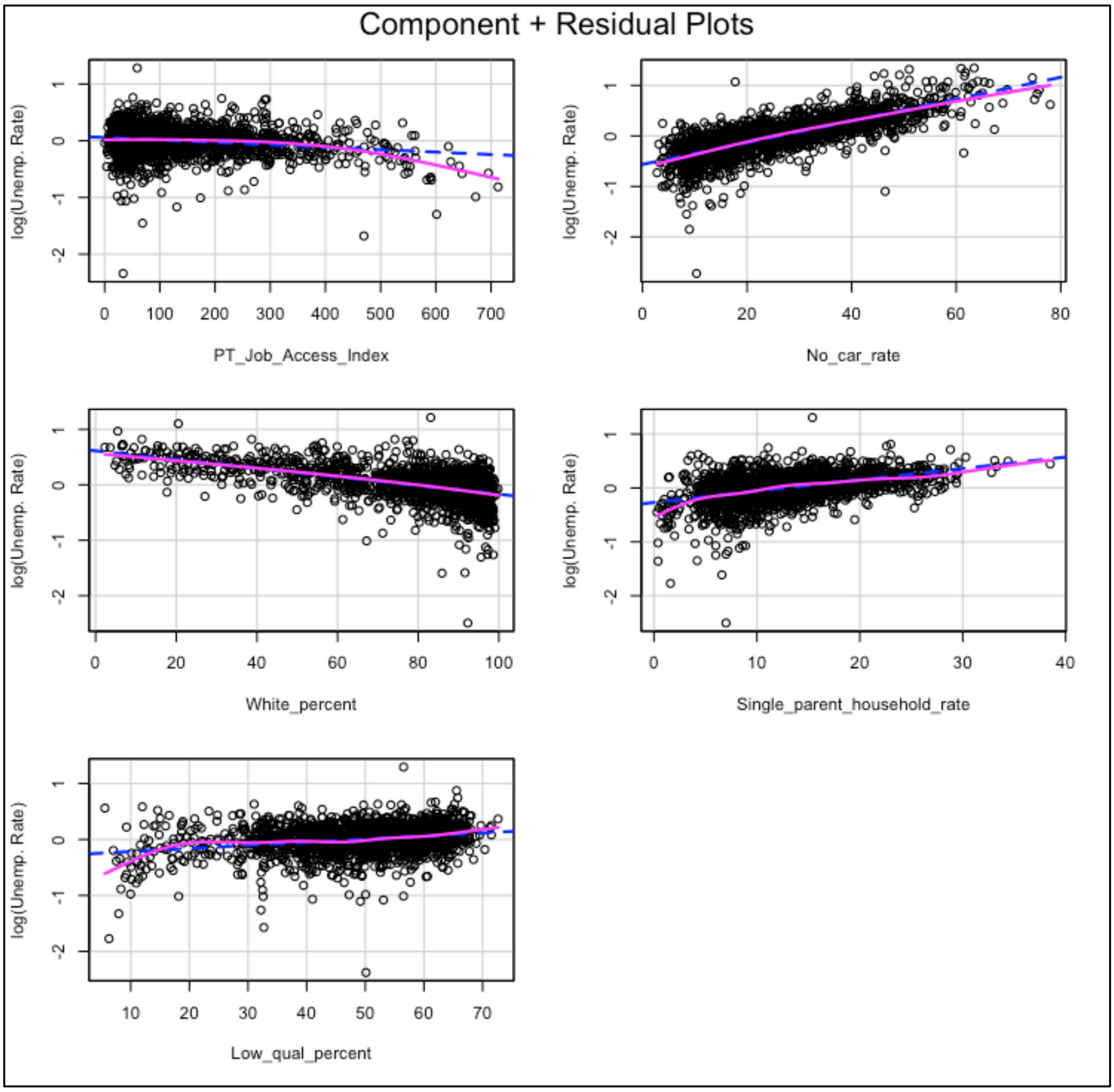



*Figure A- 6. Linearity of Component + Residuals Plot for independent variables in FE-LL Model.*

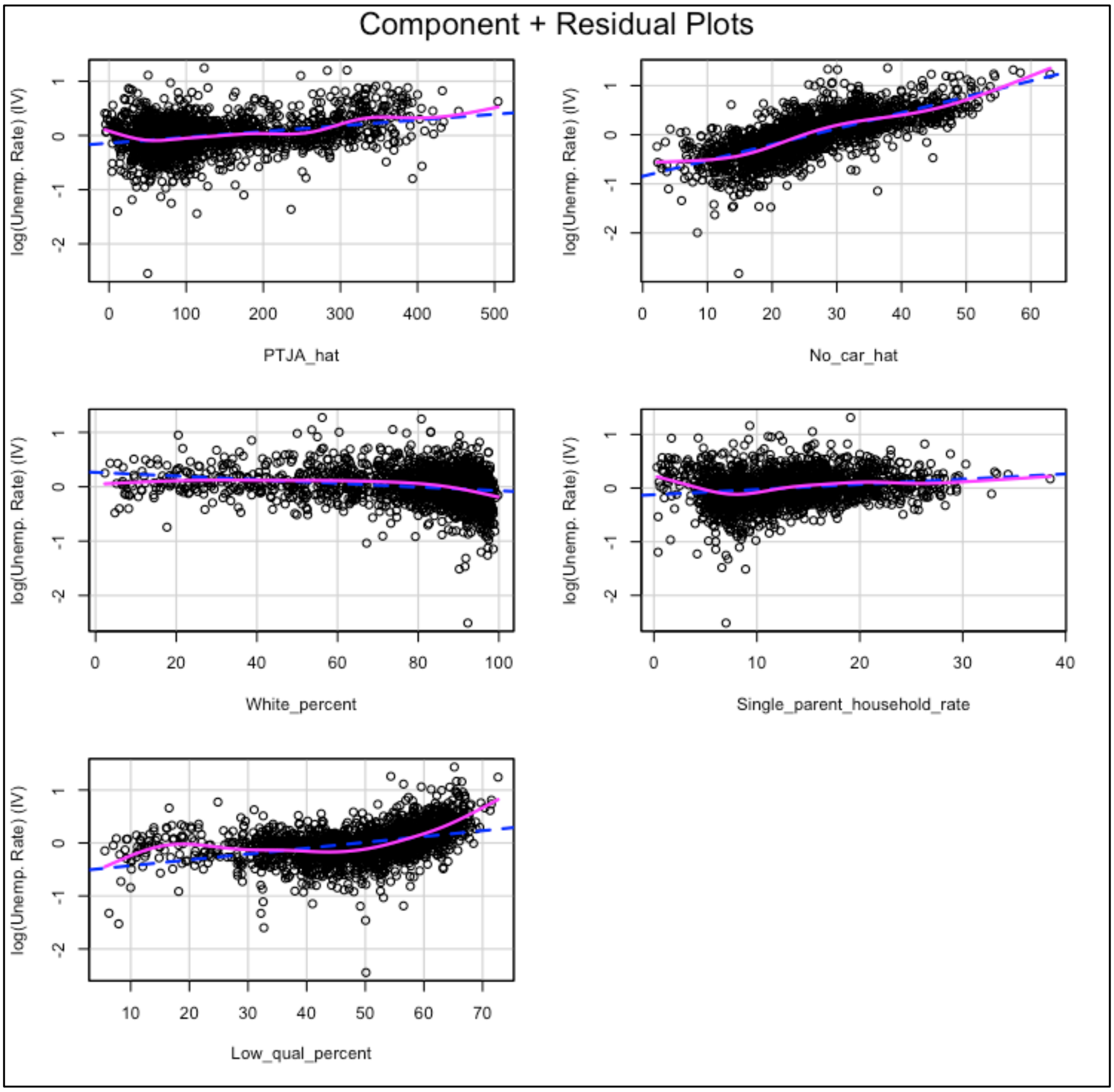



*Figure A- 7. Linearity of Component + Residuals plots for independent variables for FE-LL-IV model*