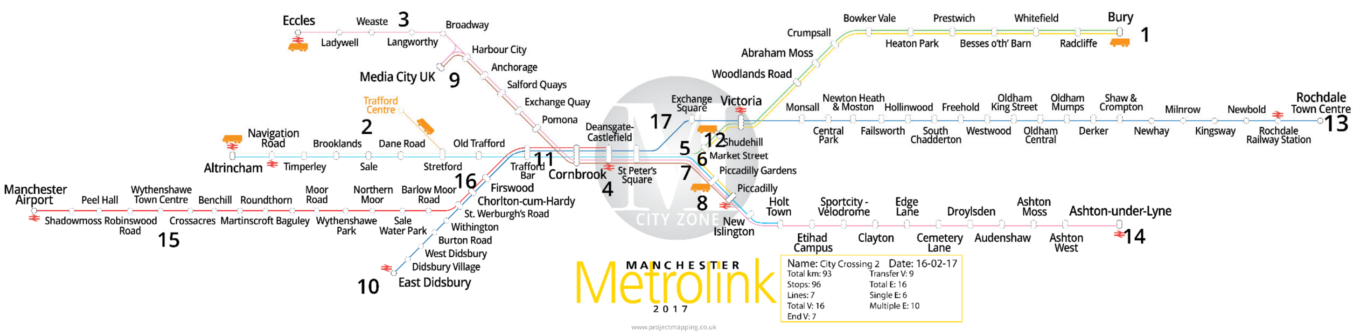
Network Position and Property Value:

Applying a network theory perspective to study the relationship between accessibility and house prices



By Thomas Schram

Colophon

Title	**Network Position and Property Value: Applying a network theory perspective to study the relationship between accessibility and house prices**
Version	Final version
Author	Thomas Schram
Student number	S2941589
E-mail	t.m.schram@student.rug.nl
Supervisor	Prof. dr. ir. A.J. van der Vlist
Second Reader	Dr. M. Van Duijn

Abstract

This thesis examines the influence of new metro stops on house prices in the Greater Manchester Metropolitan Area between 2010 and 2019. In this period, the metropolitan light rail network was expanded. First, the effect of new stops on house prices is examined. This examination is performed by a model that compares the transactions, which occurred before a new stop was built, to transactions which occurred after the opening of a new stop. Secondly, in contrast to other studies, a difference-in-difference model is used that includes a unique treatment and control group for every metro stop. The main finding is that the new stops in the Metrolink network of Greater Manchester have distinct effect sizes among each other. The effects of new stops on house prices range from positive to negative. In addition, a geographically weighted regression GWR is used. The reason for including a GWR is that the difference-in-difference model only includes coefficients for the nearest stop whereas the GWR provides more insight into spatial patterns of a variable effect. To measure the influence of network position on house prices, gravity-based network parameters are used. The GWR concludes that an improved network position increases house prices. However, not all variation in the effect of Metrolink stops on house prices over the region is explained by network position. Further research could focus on further developing complexity and network approaches in house prices models.

Keywords: Network Theory, Difference-in-Difference, Public Transit, Property Value, Manchester, Real Estate, Light Rail

Contents

1. INTRODUCTION 4
    1.1. Motivation 4
    1.2 Literature review 4
    1.3 Problem statement 5
    1.4 Outline 7
2. THEORY 8
    2.1 House prices 8
    2.2 Infrastructure networks and accessibility 9
    2.3 Hypotheses 11
3. METHODOLOGY 13
    3.1 Case Study 13
    3.2 Data 14
    3.3 Empirical model 18
4. RESULTS 25
    4.1 Diff-in-Diff results 25
    4.2 Robustness 28
    4.3 GWR Results 29
    4.4 Spatial non-stationarity 32
    4.5 Spatial Patterns 33
5. CONCLUSIONS 38
    5.1 Limitations & Recommendations 39
REFERENCES 42
Appendices 47

1. INTRODUCTION

1.1. Motivation

People use infrastructure for a wide variety of activities and services. New infrastructure is designed and built to improve accessibility but comes with high, typically public, expenses (Boardman et al., 1993). Due to these investment costs, policy makers commission studies to determine which projects are the most cost-effective (Boardman et al., 1993). These studies often require cost-benefit analyses (CBA), which list all direct and indirect costs and benefits, including negative and positive externalities. Externalities are usually harder to calculate than investment costs and have to be inferred indirectly because externalities are not always monetary.

Real estate value can be used to measure the indirect effects of public sector investments (Suzuki et al, 2015). Because infrastructure development shortens the relative distance to market, employment, or recreational opportunities, the economic potential of a neighborhood increases. In turn, the economic potential increases the demand for space in newly connected areas (Evans, 2004). Furthermore, the increase in demand drives property prices up through bidding. This increase in real estate value can be beneficial to an infrastructure CBA. A public entity benefits from real estate value through real estate taxes. Thereby, knowledge on the effect of infrastructure on real estate prices can help inform decision makers in designing and selecting infrastructure projects through better estimates of prospective real estate value and thereby real estate tax revenue increases.

1.2 Literature review

Existing research identifies "accessibility" as a key driver in house prices dynamics (Evans, 2004; Percoco, 2016; Stoilova and Stoev, 2015; Debrezion et al., 2007; Heddebaut and Palmer, 2014; Murray, 2017; Dengg, 2018). In this thesis, accessibility refers to people's ability to reach desired services and activities (Litman, 2008). Accessibility is improved as a result of overcoming barriers. In the case of physical barriers, accessibility is mainly driven by infrastructure developments and the available travel modes. This relationship will be expanded upon in Chapter 2. It has long been understood that new rail connections drive up land prices and by extension house prices. Even speculators have used this principle for financial gain regularly for more than a century (i.e., Ripley, 1911). The effect of most modern types of infrastructure networks on house prices has been evaluated in the past decades. Some researchers have focused on highways (Percoco, 2016), metro lines (Stoilova and Stoev, 2015), railway stations (Debrezion et al., 2007), or public infrastructure in general (Chandra and Thompson, 2000).

The variety of research into the effect of infrastructure on house prices indicates that many scholars expect a significant effect on house prices. However, due to the complex nature of infrastructure case studies, a consistent effect on house prices is yet to be estimated (Heddebaut and Palmer, 2014). Debrezion et al. (2007) present an overview of the effects of railway stations on the value of commercial and residential properties with a meta-review. In addition, Mohammed et al. (2013) conduct a meta-review for different types of rail transport. Both meta-reviews conclude that there is an inconsistency in findings and Mohammed et al. (2013) found varying sizes of effects per mode of transport. This results in a knowledge gap regarding the underlying relationship between transport opportunities and house prices. This knowledge gap inhibits the application of this theory in future projects in infrastructure and research.

The aforementioned literature uses “distance to infrastructure” as a proxy of the effect of accessibility on house prices (Murray, 2017; Dengg, 2018; Mohammed et al.,2013). A possible solution to the previously discussed knowledge gap could be the inclusion of a different approach to accessibility. A neglected part of accessibility is the position of a house in the whole network. For example, a central position in a transport network allows for shorter relative distances to more places. This could provide an accessibility premium to properties around more central positions in the networks as opposed to the peripheral positions in the network. This approach includes the changes in the whole network as opposed to “presence of nearby infrastructure” or “distance to nearest infrastructure”. So, more distant changes in the network are considered. The added benefit of this approach is that there are more data points over time as there are more potential changes to measure than the opening of a nearby metro stop.

A network approach has been used to evaluate urban transport networks in existing research (such as graph theory on transport networks: Derrible & Kennedy, 2009; Derrible, 2010; Derrible & Kennedy, 2010; Dimitrov & Ceder, 2016; Mishra et al., 2012; Stoilova & Stoev, 2015). In this thesis, the evaluation of networks and positions in the network is applied to the effect of new metro stops on house prices in the Manchester Metropolitan Area. During the research period, a metropolitan light rail network was expanded, meaning that more locations are accessible by metro over time. This allows for applying a network approach to house prices developments.

1.3 Problem statement

This thesis aims to examine the impact of network changes on house prices. Current research has a miss-match between the underlying theory of the value through accessibility and the implicit assumptions in their methods. Examining the network position has the potential to offer insight into these underlying dynamics. To achieve this aim, the main research question is: *To what extent do distant changes in the network impact the effect of new metro stops on house prices?* This question is decomposed into three sub questions:

Sub question 1:

What is the relationship between public infrastructure networks and house prices?

The first sub question can be answered with literature research. Firstly, literature research is done on the conceptual relationship between accessibility, networks, and house prices. Secondly, by exploring different approaches to compiling network parameters, the most relevant ones can be selected and applied in an empirical model for the second sub question

Sub question 2:

What is the effect size of new stops on house prices using a difference-in-difference modeling approach?

The second research question is addressed using an empirical model. This empirical model uses methods that are comparable to existing literature (Murray, 2017; Dengg, 2018; Mohammed et al.,2013), namely a difference-in-difference model to examine the treatment effect of new network connections. In addition, the treatment effects are examined per stop (heterogeneous model) as opposed to an estimation of a group of stops within a metropolitan area (homogeneous model). This provides the opportunity to compare the size of the effect at different locations, which will explain whether homogeneous models or heterogeneous models are better models for reality. The results of this question provide insight into the value of estimating different effect sizes for different stops. Following this question, the reasons for differences in effect sizes between stops are investigated using a network approach in the third research question.

Sub question 3:

To what extent can a network approach enrich insight into the effect of new metro stops on house prices for the case of Greater Manchester, 2010-2019?

The third research question examines the contribution of node and network parameters to the explanation of the spatial variation of the effect. This clarifies whether there is a relationship between network changes at a distance and house prices. For this question, a Geographically Weighted Regression (GWR) is used. This model allows for the testing of spatial variation in the estimated coefficients.

1.4 Outline

The remainder of this paper is organized as follows. Section 2 describes the theoretical background and concludes with a conceptual model. Section 3 describes the data and the methods used to answer the research questions. Section 4 presents the results of the empirical analysis, and section 5 concludes the study. Furthermore, the conclusions contain a discussion of limitations and recommendations for further research. In addition, an interview with an industry expert is conducted to explore the relevance of this thesis. The interview is included in Appendix 1. All the bookkeeping and workflow can be found in the appendices.

2. THEORY

In this chapter, the existing theories on house prices and network analysis are reviewed. Firstly, this section discusses how accessibility influences house prices. From this discussion, the theoretical reason for the indirect effects of network changes is identified. Subsequently, the theoretical basis for operationalizing a network approach is found, which answers the first research question: *What is the relationship between public infrastructure networks and house prices?*

2.1 House prices

House prices are often considered as the sum of its parts in hedonic house prices models. A house has several bedrooms, bathrooms, a garden, or a balcony with each of these parts contributing to the total house price (De Haan, 2013). In addition, contextual characteristics are included, such as the distance to schools, supermarkets, or the level of livability of the area (Ottensmann et al., 2008). Amenities can be valued using hedonic house price models. A house with a school in the vicinity is worth more than a house without schools nearby. The hedonic model provides coefficients for every variable. These coefficients can be interpreted as the premium or discount that an amenity or characteristic contributes to the house prices. In the context of this thesis, general accessibility is regarded as one of these amenities that contribute to total house prices.

House prices are related to the distance that the property has to the job market, according to Alonso (1964). This statement on the relationship between land price and accessibility is founded on research by Von Thünen (in Evans, 2004), who theorizes that agricultural land value should decrease as transport costs increase. This model was applied to urban functions by Alonso (1964) which led to the opening statement of this paragraph. Different urban functions sort themselves on how much each function can afford to bid for the land. Functions, such as retail and offices, are sorted towards the center of the city because they can make a higher bid for the land. Lower bidding functions, such as warehouses and residentials, are sorted towards the edges of the city. The difference in bids produces the bid-rent curve (see Figure 1). The bid rent curve is foundational to this thesis as it forms the underlying assumption that property value is related to distance from a central point.

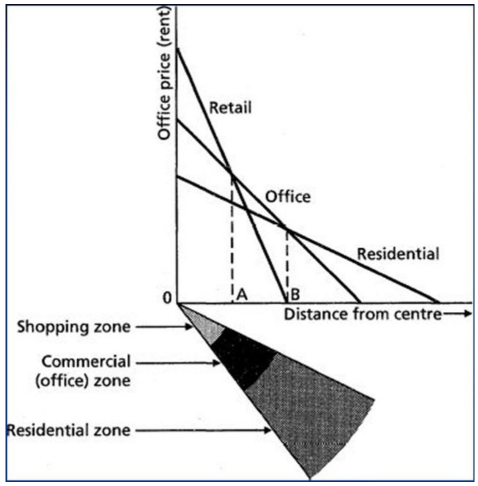


Figure 1 The bid rent curve (Alonso, 1964)

The aforementioned models assume an isotropic space. In monocentric and isotropic spaces, the most important aspect which determines price is the distance to the central market. However, on earth, space is non-isotropic in terms of transport costs. Transport costs also consist of travel time and the available modes of transport. In addition, a market is not the sole destination. This means that authors in transport literature use the term accessibility over distance to market. Accessibility refers to people's ability to reach desired services and activities (Litman, 2008).

2.2 Infrastructure networks and accessibility

Alonso's theory (1964) is based on the positive relationship between accessibility and land value. Land value is a part of the transaction price of houses. Therefore, improvements in people's ability to reach desired services and activities contribute to an increase in house prices. Accessibility can be improved. Using the aforementioned definition, improving people's ability to reach services and activities improves accessibility. From a network perspective, this "reaching" is performed by moving over links and traveling through nodes. Links can be road segments and metro lines. Nodes can be intersections or metro stops. Thus, improving the ability to reach services can be achieved by adding physical infrastructure. New infrastructure would create more and shorter travel paths. Furthermore, travel paths can also be made easier. A higher flux through the network, faster travel speeds, less waiting time, and lower monetary costs all improve the ability to reach desired services and activities fast and easily. In addition, cleanliness and perceived safety can also be resistance factors. In short, better transport options reduce travel frictions. This improves accessibility to desired services and activities. In turn, accessibility leads to higher transaction prices of houses (see Figure 2). However, externalities of transport networks are not exclusively positive. Levkovich et al. (2016) found that the infrastructure itself can cause a nuisance in the surrounding area through noise, pollution levels, or increased traffic intensity, thereby reducing house prices. In addition, Murray (2017) found that for an Australian light rail project the rise in property values in the first 100 meters from new stops is less than between 100-400 meters. They accredit this to nuisance. Despite these negative externalities, both studies conclude with net positive effects on house prices from infrastructure developments. Other research, which

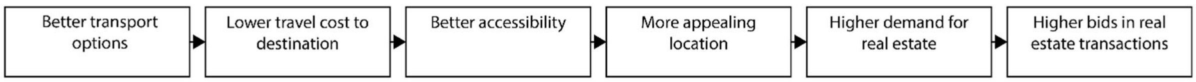


Figure 2 Schematic abstract on the theory of how transport results in higher real estate prices

attempts to estimate the relationship between house prices and the distance to the nearest metro stop, have similar methodological setups (e.g., Murray, 2017; Dengg, 2018). By comparing house prices before and after the addition of new metro or other light rail stops an effect size of the new stop is estimated. These authors start with the assumption that the classical homo economicus will choose to live in a place that provides accessibility to desired services and amenities (McFadden, 1978). By adding Alonso's (1964) theory of a bid rent curve this accessibility to services and amenities is coupled to land value. However, the existing literature, which estimates the impact of new metro stops on house prices, makes an implicit assumption. By measuring the distance to the metro stop but not the accessibility the metro stop provides, the literature assumes that the metro stop is an amenity by itself. It might be a useful simplification for large train stations with shops, but most simple stops are rarely a destination. In most cases, metro stops provide access to other amenities. Methodologically this assumption is implicitly made by grouping the data of house transactions around different stops (e.g., Murray, 2017; Dengg, 2018). This results in a homogeneous result where all stops have the same average effect. Dropping this implicit assumption allows for different levels of accessibility provided by different stops. By differentiating the effect of new stops based on the accessibility they provide, the underlying theory of the bid rent curve is reflected more accurately. Using the topology and usage of the network this accessibility can be characterized and included in empirical models.

Evaluating the topological development of a network statistically requires expressing the topological characteristics quantitatively. These results can answer the question: How well-positioned is the closest metro stop in the network? The next question is: How much did this position contribute to the transaction price of houses? In other words, what are the parameters to express the topological characteristics of nearby stops?

Networks are mathematically described as graphs. Some scholars have adapted graph theory metrics to transport networks (Derrible & Kennedy, 2009; Derrible, 2010; Derrible & Kennedy, 2010; Dimitrov & Ceder, 2016; Mishra et al., 2012; Stoilova & Stoev, 2015). One of the parameters derived from this train of thought is “centrality” (Sabidussi, 1966). Another approach is the construction of a connectivity index (e.g., Mishra et al., 2012). However, there are two major objections to the application of these parameters. Firstly, many connectivity indices and centrality parameters are relative. In the calculations of these parameters, characteristics are seen as a fraction of the network total, such as the number of neighboring nodes divided by the total number of nodes. However, in this research observations of absolute changes in the nodes are desired. Absolute changes make observations comparable as the network changes over time or when comparing between networks. Secondly, the connectivity indices or graph-based parameters do not always include information about the surrounding space. So, for example, does a stop provide access to services and activities or is it just a node? Thus, this research needs a different parameter to measure the ability to reach activities and services.

Interaction models, such as gravity models, can meet these criteria. These models are based on the existing or potential interaction between two entities in space (Rodrigue, 2020). These methods are used in economics to explain international trade, but also to explain the movement of people between or within cities (Rodrigue, 2020). The gravity model is a direct derivative from Newton’s work as it multiplies the power (Newton used mass) of two objects and divides it by the distance squared to represent the non-linear effect of distance on the interaction between two objects and is used to model commuter flows between two places (Ogura, 2010). Calculating the interaction potential between places avoids the previously listed objections. The interaction potential can be measured in absolute terms and contains information regarding the activities and services around the network. The equation for calculating this is included in chapter three.

2.3 Hypotheses

In other real estate papers, the effect sizes of different stops are assumed to be the same (Dengg, 2018). These papers use models that assume a homogenous effect of new stops on house prices. The theoretical basis in Alonso (1964) and the involvement of network positions allow for a heterogeneous effect. In models that allow for a heterogenous effect, different stops can have different effect sizes. Following this, three hypotheses are formulated:

1. On a global level, the newly opened stops have a positive impact on house prices.
2. When a model does not assume a homogenous effect on a global level it estimates widely different effect sizes.
3. There is a positive relationship between improvements in network position and house prices.

The first two hypotheses are tested with the second research question and the third hypothesis is tested with the third research question.

In addition to the aforementioned hypotheses on relationships, different patterns in spatial patterns can be argued to arise. In Figure 3 different patterns are visualized. Figure 3a shows the conventional models. Figures 3b and 3c show how effect size can differ and how the area of influence differs. In the visualization, the effect size or area of influence diminishes with distance from a CBD (surrogate for important services and activities). However, the inverse can also be theorized: suburban places gain a relatively high benefit from a new connection to important services and activities. Figure 3d visualizes a preexisting transport option that diminishes the added benefit of a new transport option. However, as this thesis is limited to the Metrolink network, the pattern in Figure 3d is not investigated in this thesis.

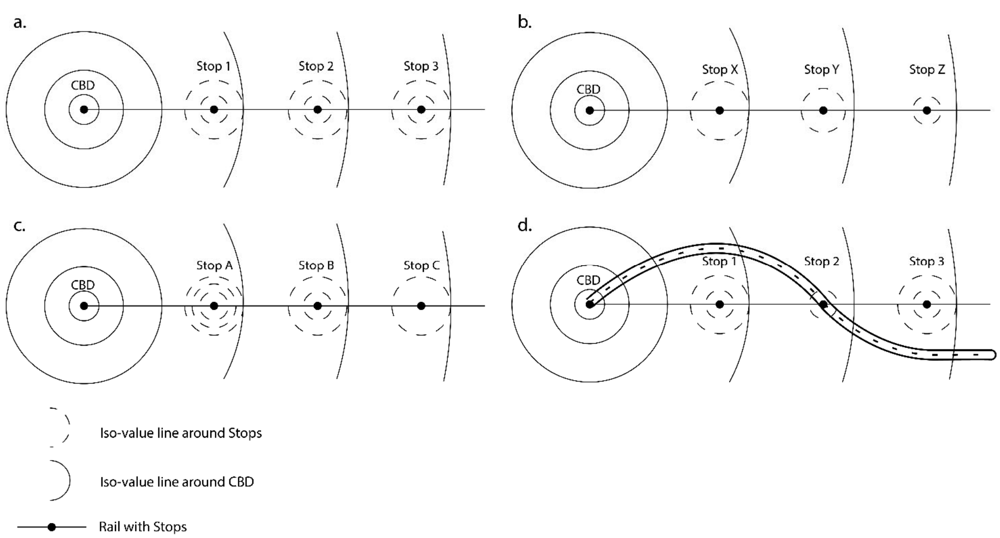


Figure 3 Hypothetical patterns of value premiums around public transport access points

3. METHODOLOGY

3.1 Case Study

This study uses a case study of Greater Manchester. This section provides the reasoning behind the case selection. The goal of case selection is to find a case or set of cases with changes in network and node parameters and accompanying transactions of houses. The case must contain many network developments in a period where consistent supporting data are available. This thesis uses the case of the Metrolink light rail network in Greater Manchester over the period 2010-2019. The UK provides transaction prices publicly, which is a necessary condition for case selection. Furthermore, the UK and Manchester provide diverse data sources over the last decades with documentation in English. In addition, this study requires a metropolitan area with an extensive and quickly evolving network. The number of large metropolitan areas with extensive networks is limited in the UK. Greater Manchester has been chosen because of its rapid development of the Metrolink metro network in a short period of time. The first line opened in 1992 and in 2020 there are more than 90 kilometers of rail within Greater Manchester. This allows for including many stops in the empirical model. However, between 1995 and 2010 the number of network adjustments has been minimal (Appendix 2). In addition, finding consistent data for covariates for the period 1995-2019 is challenging. Therefore, the period that is regarded in this study is 2010-2019.

For this thesis, the Metrolink network had to be reconstructed at every stage of its development. The reconstruction is derived from existing maps of the Metrolink network that have been made throughout the years and which are collectively stored on www.projectmapping.co.uk. This collection of maps largely consists of uploads of original maps made by Transport for Greater Manchester. However, some of the maps are personal interpretations by hobbyists that try to make a better map than the official publications by the transport authority. Some of these contained minor mistakes, which are resolved by referencing between the different sources continuously, resulting in the complete reconstruction in Appendix 3. The different stages in the development are referred to as "steps". The steps range from 1-19. The first step is the opening of the first line and step 19 is the opening of the most recent stops. However, only steps 7 through 19 happened in the 2010s and are part of the case study. So, for the diff-in-diff models only 58 out of 94 Metrolink stops are included. For the remaining 36 stops, it is not possible to have consistent "before" data. The GWR does not require "before" data. So, this includes all 94 stops.

Case Background

Manchester used to have an electric metro network from 1901 to 1949 (Museum of Transport Greater Manchester, 2004). At its peak, the network had a total length of 262 km (Museum of Transport Greater Manchester, 2004). The network was decommissioned during the advent of cheap diesel buses and trolleybuses starting in 1966. There have been multiple proposals for urban light rail and a connection between the north and south termini for heavy rail (Kessel, 2011). Furthermore, the lack of a good public transport connection to a newly developed office area motivated the demand for a light rail network (Museum of Transport Greater Manchester, 2004). The first proposed line was a converted heavy rail line north of the city (to Bury) and a heavy rail line south of the city (to Altrincham) connected with an on-street section through the city center. This is the Bury line, which was completed in 1992 (Kessel, 2011). Secondly, the Eccles line was constructed and completed in phases in 1999 and 2000 (Kessel, 2011). In the '00s plans were developed for three extra lines. These lines had to be built in phases (phase 3a and 3b) as funding requirement by the national government (Kessel, 2011). These lines are now the East Manchester Line, The South Manchester Line, and the Airport Line and were constructed during the 2010s. This expansion period is the main focus of this thesis. The evolution of the network can be found in Appendix 3.

3.2 Data

There are multiple data sources collected to test the hypotheses and answer the research questions. These data sources are presented in Table 1. Furthermore, this section describes the data on the Manchester Metrolink network and its evolution. After the description of the network, the transaction microdata are presented. Lastly, the most important variables from these datasets are presented with descriptive statistics.

Table 1 Data source

Dataset	Source
Metrolink Network Topology	www.projectmapping.co.uk (historical maps) https://tfgm.com/public-transport/tram/network-map (current map)
Metrolink patronage data	https://www.gmtu.gov.uk/reports/default.htm
Transaction prices	https://landregistry.data.gov.uk/app/ppd
CPI series	https://www.ons.gov.uk/economy/inflationandpriceindices
Postcode centroids	http://geoportal.statistics.gov.uk/datasets/75edec484c5d49bcadd4893c0ebca0ff_0
UK 2011 Census	https://data.mendeley.com/datasets/389scnndjy/1

Manchester Metrolink Data

To construct the network parameters from section 3.3 data on stop usage are required. Transport for Greater Manchester (TfGM) provides annual reports on Metrolink usage, including the number of boarders and alighters (people that disembark the vehicle) per stop. These data have been collected in a spreadsheet throughout the reports. The yearly data have been coupled to the evolution step of the network that was closest to it in time.

Micro Transaction Data

The observations in the model are individual transactions of houses. The data are collected by HM Land Registry as part of its legal obligation (GOV.UK, 2019). These data contain the observations of residential property transfers and include the price paid, type of house, type of ownership, address, and postcode, among other variables. HM Land Registry (GOV.UK, 2019) claims that it only includes residential properties transferred at full market value. The used data span from 2010 to 2019 and are limited to the Greater Manchester area. This query results in approximately one million transactions in that period and area. The Metrolink network does not cover the whole Greater Manchester region. Therefore, the distance to Metrolink stops is too large for many observations to be relevant. The distance criteria for not using observation in either the treatment or control group is set to 1200 meters. This distance is based on Mohammed et al. (2013) and Dengg (2018). In addition, considering that this thesis assumes a 500-meter radius of the effect and the sufficiency of the observations in the 500 to1200-meter range, there is no reason to extend the radius of the control group. Extending the range of the control group would only include properties that are in widely different spatial contexts (e.g., outside of the buildup areas of the towns in the Greater Manchester Area). These observations would be unsuitable for comparison. Subsequently, the remaining 372.900 observations are reviewed on outliers using a histogram (Appendix 4). Some multi-million transactions are included in the data. Upon further investigation, some commercial properties are included in the data of HM Land Registry contrary to what is reported. To prevent the inclusion of cases such as an £80.000.000 mall, outliers are identified. Using a boxplot, properties from £500.000 and above are labeled as outliers. However, after manual confirmation, these properties seem to fall within the criteria of the empirical model. So, only transaction value exceeding £2.000.000 are dropped. In addition, properties traded for extremely low values are dismissed as properties that are not traded at market value. Only transactions at market value are of interest for hedonic price house price models. Therefore, observations that are traded for less than £10.000 are dropped. This is an arbitrary cutoff point. These two decisions drop 2.472 observations. In addition to this exclusion of data points, the data also contain a category of transactions that are traded below market price according to HMS Registry. This thesis uses market value and therefore excludes this category (23.106 observations). These selection steps leave 206,940 observations.

A normal distribution of the dependent variable is desired in regressions. However, the distribution of price has a right-side tail. Using a natural logarithmic transformation this is transformed to a normal distribution (Figure 4). Other OLS assumptions are tested for in Appendix 6. In addition to the logarithmic transformation, prices have been adjusted to the 1995 price level using the monthly consumer price index published by the Office for National Statistics (2020), to make prices comparable over the case study period.

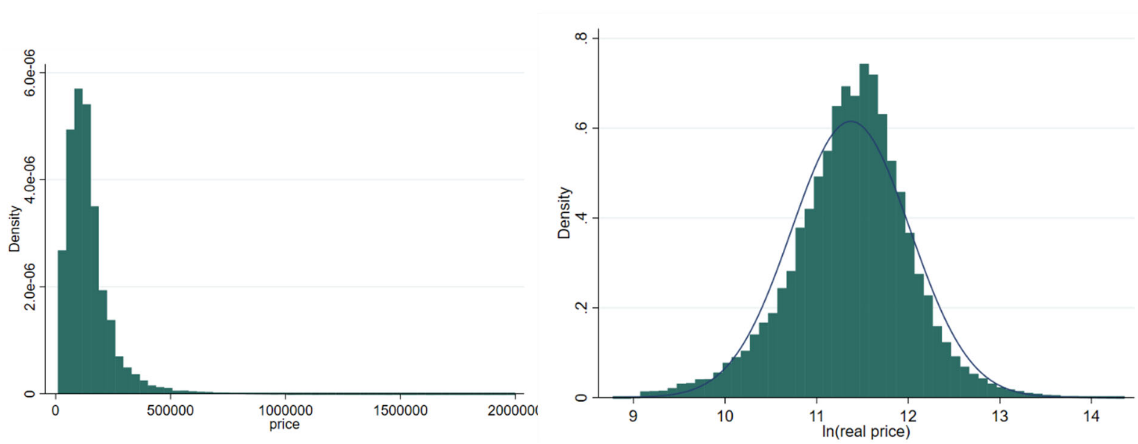


Figure 4 Left: Histogram of transaction prices. Right: Histogram of the natural logarithm of the deflated transaction prices

The model also contains area fixed effects. The area's dummies are by local authority. There are ten local authorities in the sample (see Table 2). This administrative level is chosen because these areas are governed by their separate council. Local governments can determine their policy on several topics that could influence real estate prices including property tax, business rates, planning permits, licensing, parking fees, and social housing provision.

The type of the employed model is a difference-in-difference (Diff-in-Diff). This type of model aims to compare the effect of a treatment (e.g., medical, policy, or event) between two groups. This comparison is achieved by grouping all the observations with two dummies. The dummies capture whether an observation happened before or after the treatment and whether the observation falls within the treatment group or the control group. The fundamental assumption of a diff-in-diff is that the two groups would have undergone the same trends if the treatment would not have been applied. A diff-in-diff requires sufficient observations before and after the implementation of a treatment. The treatment in this thesis is the opening of a nearby metro stop. In this case, the number of house transactions is quite evenly divided between before and after cases (Table 2).

Table 2 Observations split by being before or after the construction of the nearest Metrolink stop

After treatment	Freq.	Percent
Before	86,393	41.75
After	120.547	58.25
N	206,940	

This is further examined in Appendix 5. This appendix includes the observation per stop. The examination of the distribution of observations does not raise any concerns. However, in this appendix, the stops with less than 500 transaction observations are marked red. These stops are not excluded from the analysis. They remain included as insignificant results from a few stops do not critically influence conclusions.

Additional variables are included in the models to capture the effects of socioeconomic characteristics of the neighborhoods. Many authors presume or investigate the effect of socioeconomic characteristics in hedonic price models (e.g., Clauretie, Neill, 2000; Ozus et al., 2007; Koster & Rouwendal, 2012; Dijk et al., 2016). The primary data source for the socioeconomic covariates is the UK census. The UK census provides information on a small spatial scale, the Lower Statistical Output Area (LSOA). The potentially relevant variables are Unemployment, Car Ownership, and Home

Ownership. These are measured in percentage of the LSOA population. Car Ownership is of extra interest as this study examines the influence of a transport mode. Making more trips by car is assumed to influence the appreciation of the accessibility provided by a new metro stop. This relationship is not further investigated in this thesis. Table 4 provides descriptive statistics on the variables. Table 5 contains the descriptive statistics on the network parameters that are used in the GWR.

Table 4 Descriptive statistics of main diff-in-diff variables

	mean	sd	min	max
price	143,103.32	112,420.22	10,000.00	2,000,000.00
real price	11.347	0.678	8.775	14.363
Percentage unemployment	4.9	2.2	0.7	14.7
Percentage no car	33.7	15.6	5.0	77.0
Percentage homeownership	54.7	21.6	2.5	95.2
Percentage Freehold	0.477	0.499	0	1
Distance	643	296	27	1199
N	206,940			

Note: Percentages are measured as the share of the population in the LSOA
Note: Distance is measured in meters from the nearest Metrolink stop

Table 5 Descriptive statistics of network parameters in GWR

	Obs.	Mean	Std. Dev.	Min	Max
Distance to CBD	40.000	7196.213	4021.979	106.9334	17.89596
ln(networktotal)	40.000	5.699508	1.163027	1.925578	7.629455

Notes: The number of observations is decreased for computational reasons. It is a random subset from the observation in table 4. The calculation for reach and networktotal are described in equation 5. The units are abstract so there is no unit, but for reference appendix 1b provides an overview of the reach of all stops. Reach and network total are log-transformed to comply with OLS assumptions. Distance to CBD is measured in meters.

3.3 Empirical model

Difference-in-difference approach

To identify the effects-sizes of new metro stops on house prices, a difference-in-difference hedonic price model is estimated to capture the price change after the opening of a new stop. When the stop is within 500-meters of the house transaction, the transaction is categorized in the target group. The 500-meter radius has been selected based on Dengg (2018) and Mohammed et al. (2013). These studies find that within walkable distance the effect is the most significant and that 500 meters is a good approximation of a walkable distance. The house transaction is part of the control group when a stop is 500 to 1200 meters away. This is the radius between the already discussed upper boundary of 1200 meters and the inner radius of walkability.

Firstly, a homogenous model is formulated (equation 1) based on Van Duijn et al. (2016). This model estimates one treatment effect for all the new stops in Greater Manchester. Secondly, a heterogeneous model (equation 2) is formulated. This model estimates a treatment effect for every stop. This methodology is inspired by the methodology in equation 1. The main difference between the equations is that the treatment effect dummies have been replaced by treatment effect dummies for every stop separately. This allows for dropping the assumption, which is made implicitly by existing research, that all stops have a similar effect on house prices. How the following equations handle the data points differently is illustrated in Figures 5 and 6.

$$\ln(Price_i) = a + \beta_1 T_s + \beta_2 A_s + \beta_3 D_{is} + \beta_4 (T_s * A_s) + \beta_5 (T_s * A_s * D_{is}) + \sum(\gamma_k(C_i)) + \sum(\delta_j(L_i)) + \sum(\delta_j(Y_i)) + \varepsilon \tag{1}$$

Where:

$\ln(Price_i)$ = The log of the transaction price of house transaction i

a = The constant

T_s = Dummy for measuring if the transaction occurred within 500 meters of the nearest stop s

A_s = Dummy for measuring if the transaction occurred after the opening of the nearest stop s

D_{is} = The distance between transaction i and stop s

$\sum(\gamma_k(C_i))$ = The sum of all covariates and their respective coefficients

$\sum\left(\delta_j(L_i)\right)$ = The sum of all local authority dummies and their respective coefficients

$\sum(\delta_j(Y_i))$ = The sum of all year fixed effect dummies and their respective coefficients

ε = The error term

$$\ln(Price_i) = a + \sum(\theta_n(T_n)) + \beta_1 A_s + \beta_2 D_{is} + \sum(\vartheta_n(T_n * A_s)) + \sum(\rho_n(T_n * A_s * D_{is})) + \sum(\gamma_k(C_i)) + \sum(\delta_j(L_i)) + \sum(\delta_j(Y_i)) + \varepsilon \tag{2}$$

Where:

n = stop1, … , stop58

$\sum(\theta_n(T_n))$ = The sum of treatment dummies per stop_n and their respective coefficients

$\sum(\vartheta_n(T_n * A_s))$ = The sum of $T_n * A_s$ interaction terns per stop_n and their respective coefficients

$\sum(\rho_n(T_n * A_s * D_{is}))$
$= \text{The sum of } T_n * A_s * D_{is} \text{ interaction terns per stop}_n \text{ and their respective coefficients}$

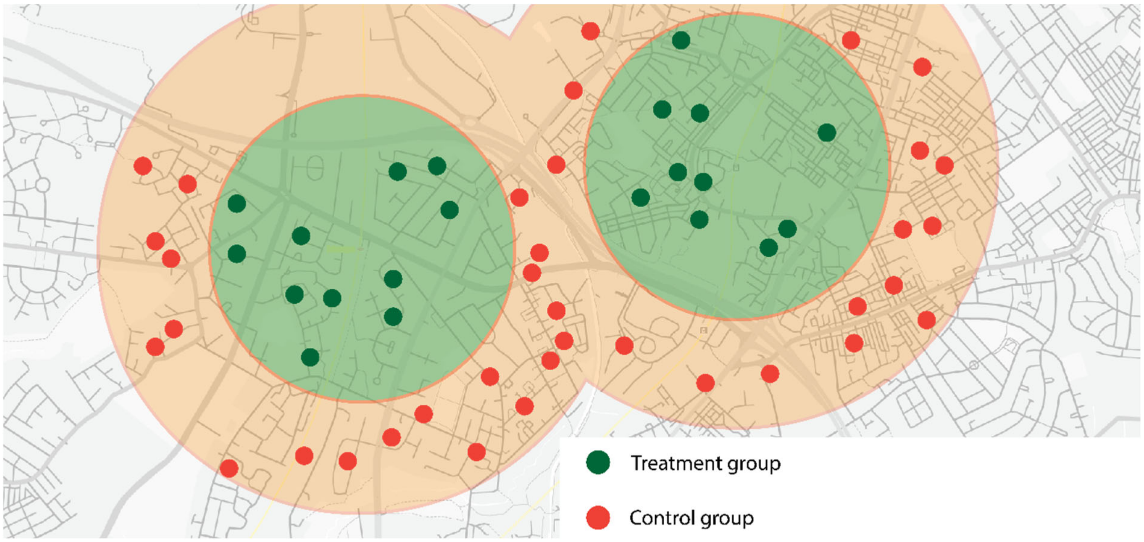


Figure 5: Common, homogeneous diff-in-diff method used in equation (1). Observations are grouped in a treatment group and control group. This results in an estimation of the average treatment effect of these stops. Distances are not to scale.

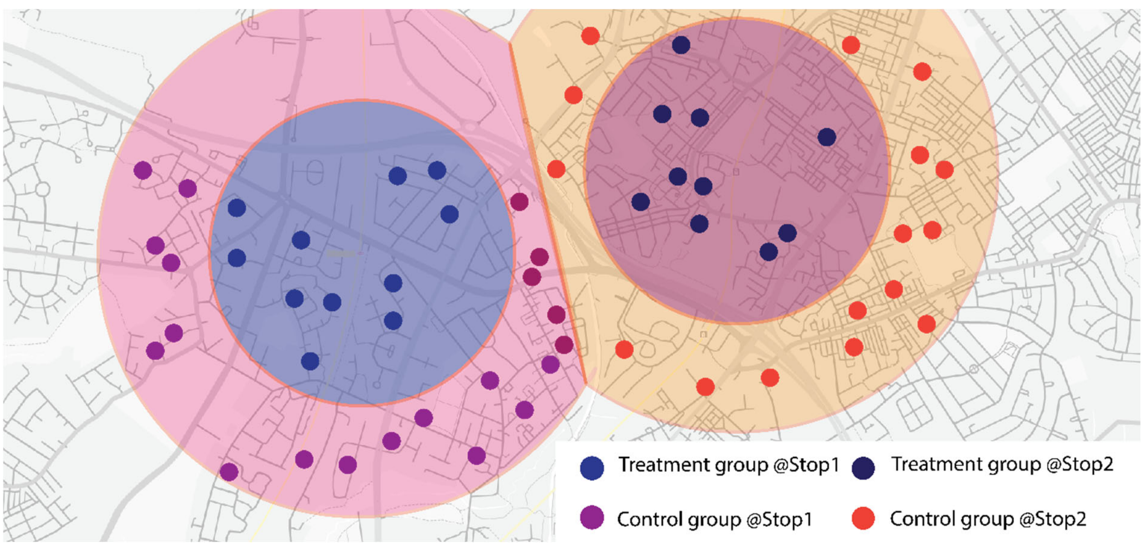


Figure 6: Heterogeneous diff-in-diff method used in equation (2). Observations are grouped in a treatment group and control group per stop. This allows for estimating a treatment effect per stop. Distances are not to scale

In equation (1), the β_4 coefficient is of special interest. This coefficient is estimated to describe the treatment effect of a new stop in the Greater Manchester area on house prices. The $\beta_5$coefficient is estimated to discover the interaction between the aforementioned treatment effect and the distance to the nearest stop. $\sum(\gamma_k(C_i))$ is an abbreviated notation for all the covariates, including car ownership, house type, unemployment, and home ownership. In equation (2), the ϑ_n coefficients are of special interest. These coefficients are estimated to describe the treatment effect per stop. The ρ_n coefficient is estimated to explore the relationship between the local treatment effects and the distance from the stop. Both equations 1 and 2 rely on the assumption of parallel trends. Equation 1 assumes that you can estimate one coefficient for all the stops, whereas equation 2 assumes you can estimate coefficients for every stop separately.

Other variables in the models are the covariates and the fixed effects. The covariates include characteristics of the transacted house (e.g., house type categories). The fixed effects include area and year fixed effects to control for variance due to omitted characteristics of the area the residential properties are in and the time in which they are transacted. Both the prices are deflated and year fixed effects are used in parallel. This is because year fixed effects can add more diverse information than the inflation rate, such as changes in local council policies.

A Wald-test is used to statistically establish differences in effect size of treatment effects (treatment*after). The Wald test requires a large number of random effects (Dickey, 2020). This requirement is satisfied in the data due to a large number of observations and metro stops. An advantage of the Wald test is that it only requires the estimation of the unrestricted model. This test is conducted for both the treatment effect and the interaction between the treatment effect and distance. The more coefficients that are added to the model the more likely some of them are significantly different. That is why the results are also reported with only the significant coefficients in the results section of this thesis. The null-hypotheses of these tests are:

$$\vartheta_1 = \vartheta_2 = \vartheta_{\dots} = \vartheta_{58} \tag{3}$$

$$\rho_1 = \rho_2 = \rho_{\dots} = \rho_{58} \tag{4}$$

The third research question requires an investigation into the spatial variation of the effect sizes. These models are presented in the next section.

Geographically weighted regression

This thesis applies geographically weighted regression (GWR) to examine the heterogeneity in network effects. This method is an addition to the insights from the diff-in-diff method. The diff-in-diff method shows the presence and the size of effects of new metro stops on house prices. The goal of a GWR is to find spatial variability. The GWR allows for adding variables that might explain the spatial variability and evaluating their effect on spatial variability. In addition, the results of the GWR can be presented spatially with a higher resolution and can estimate the coefficients in different directions as opposed to the diff-in-diff that can only measure one coefficient and has a single dimension (distance from the stop). This provides further insights into the spatial variability and into whether network parameters are related to this spatial variability. For the purpose of the GWR, the coefficients of many local models are estimated. Consequent tests indicate which coefficients have significant spatial variability. The number of regression points and the size of the kernels is decided by the bandwidth (Figure 7). This bandwidth can be manually specified or calculated by the command in STATA. In this study, the bandwidth is initially determined using the automatic method. After the first model, the sample is maintained, as well as the initially calculated bandwidth. This is done to maintain comparability between model specifications.

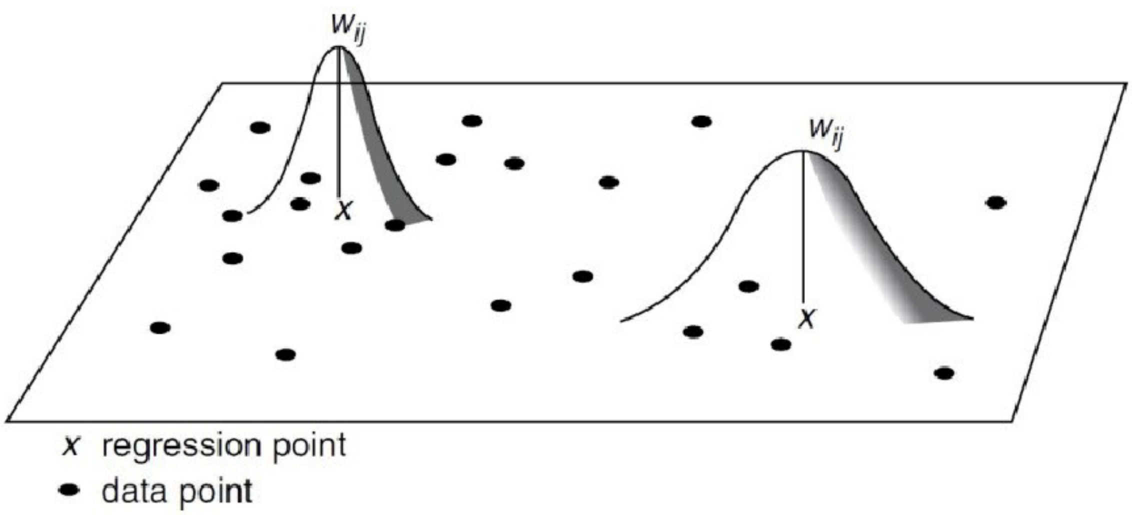


Figure 7 Two regression points with weight kernels (Fotheringham, 2009)

The GWR software has some limitations. One limitation is that only continuous variables can be included for the spatial variation testing. So, categorical variables, such as house type, cannot be included. Another limitation is that interaction terms cannot be included. These aspects limit the comparability with the earlier diff-in-diff models. In addition, when a GWR indicates spatial variation in a coefficient the underlying reasons are not yet given. In many instances, the variation might be due to an unknown variable. Theoretically, adding this unknown variable could remove spatial variability.

Besides the limitations, GWR also requires more computational power. Due to limitations in available computational power, the GWR is conducted with a random subsample from the data. This subset will be sufficiently large for statistical purposes and leaves enough computational capacity for the inclusion of model iterations and multiple model specifications. This selection explains the differences between N in Table 4 and Obs. in Table 5.

Network parameters

Based on the theoretical investigation, network parameters are determined using potential interaction models. This is a required step to include network parameters in the GWR. There are two types of network parameters used: (1) stop-specific, local parameters and (2) network-wide, global parameters. The network parameters are applied in the GWR. The addition of network parameters will or will not influence the spatial variability and can show a spatial pattern in the coefficients. Multiple network parameters are experimented with in different model formulations. The next paragraph explains how the network parameters are calculated.

The potential interaction model can provide a parameter that reflects the network position. A simple potential model is used for this purpose.

$$R_s = \sum \frac{Alighters_i}{Traveltime_i^2} \tag{5}$$

The reach at stop sis called R_s. Where I = 1, …, 94, but excluding s. So, for stop2 i = 1, 3,…,94. This interaction is defined as the alighters at all the other stops divided by the travel time to those other stops squared. This gives the potential reach as it contains information on how many appealing locations there are related to travel distance over the network's edges.

The denominator represents the decay function. Using travel time between stops is more relevant than using the distance between stops. Distance between stops does not account for different travel speeds. There are on-street segments and exclusive segments in the Metrolink network. These different types of rail allow for different travel speeds. Therefore, using travel time between stops provides a more uniformly applicable parameter (in more complex models one could include more friction factors for the decay, such as monetary costs). However, this version uses travel time as a decay function.

The nominator represents the attracted masses. In this study, the number of alighters during morning rush hour at every other stop is used. Other arguably valid data include the number of jobs, schools, shops, inhabitants, transfers, and hospitals. Stop usage data are used because they carry information regarding all activities surrounding the stop. The usage data carry this information because people leave the metro network to visit the aforementioned services and activities.

More network parameters can be calculated using the patronage data. As information is available on boarders and alighters of the Metrolink network at every stop, it is possible to calculate an inbound and outbound interaction. Equation 5 shows Outbound interaction, more intuitively referred to as "reach". Inbound interaction can be calculated by replacing "alighters" with "boarders" in Eq. 5. This is later referred to as (potential) "attraction". Reach and attraction are calculated at every stop or node in the network. For the global network parameters, a total of all the potential interactions in the network is calculated by summing all the interaction potentials of each stop. In addition, the distance from the house to the "CBD" is included. Despite the argumentation of the Los Angeles school of urbanists for polycentric urban fabric (Dear and Flusty, 2000), the metro network of Manchester is relatively monocentric. This is seen in Appendix 3, where most metro lines converge towards the city center. This will inevitably result in more attraction in the city center because of a higher density of stops and lines. The goal is to measure the effect through potential interaction in the network. Distance to CBD is included to exclude the proximity to the city center through other means from the interaction model coefficients. Technically, the location chosen as CBD is St. Peter's Square in the city of Manchester. This results in the following model for the GWR:

$$\ln(Price_i) = a + \beta_1(ln(R_s)) + \beta_2(\ln(P_s)) + \beta_3(\ln(Net)) + \beta_4(D_{iCBD}) + \beta_5(D_{is}) + \beta_6(A_s) + \beta_7(T_n) + \sum(\gamma_k(C_i)) + \varepsilon \quad (6)$$

Where:

$R_s = Reach\ of\ stop\ s\ as\ defined\ by\ equation\ 5$

$P_s = the\ sum\ of\ Reach\ and\ Attraction\ of\ stop\ s$

$Net = The\ sum\ of\ P\ for\ every\ stop\ in\ the\ network$

$D_{iCBD} = The\ distance\ from\ the\ house\ transaction\ to\ the\ CBD$

The method is derived from Du & Mulley (2007). The after and treatment dummies are not essential but are included to improve the comparability with the results of the diff-in-diff models. Several of the explanatory variables will undergo a logarithmic transformation to control for the skewness in their distributions. The data of the network parameters are presented in Table 5.

4. RESULTS

In this section, the results of the empirical models are presented. The results allow for an answer to research questions two and three. The results from the diff-in-diff models answer research question two. These results are presented in section 4.1. After the results of model specifications (3) and (4) are presented, the results of the Wald-test are discussed. Before continuing to the next main tests, a robustness test is conducted in section 4.2. This robustness test checks whether the moment of treatment is valid in the main models. If there is a significant difference among the coefficients of the treatment effects found in the Wald-test, there is a reason to investigate the spatial variation. This is done using a GWR in section 4.3. Lastly, the results of the GWR are further examined and displayed on maps in sections 4.4 and 4.5. These sections answer research question three.

4.1 Diff-in-Diff results

The results of the diff-in-diff models are summarized in Table 6. The table contains the estimated coefficients of model specifications (1), (2), (3), and (4). This paragraph discusses some of the noteworthy results from Table 6. Firstly, models (1) and (2) both estimate a negative coefficient for the treatment effect. For the interpretation of the coefficients, the log-transformation has to be taken into account. In model (2), the coefficient can be interpreted as follows: a traded residential property within 500 meters of an operational Metrolink stop likely has a $(e^{-0.0925} - 1) * 100\% =$ 8.8% lower price than a property in the control group (>500, <1200 meter from any stop). As most authors suggest a positive effect, this goes against the common hypotheses on the treatment effect. Mohammed et al. (2013) compared 23 studies with similar goals to this study, including three from the UK: two for the London underground and one for the Tyne and Wear Light rail. The two London Underground studies report positive effects (Gibbons & Machin, 2003; Chesterton, 2000), whereas the Tyne and Wear study reports effects between -42% and +50% (Du & Mulley, 2007). Thus, the negative results contradict the theory on accessibility, but these are not unique results.

Table 6 Coefficient estimates for diff-in-diff models

	(1) Homogeneous Treatment	(2) Homogeneous Treatment	(3) Heterogeneous Treatment	(4) Heterogeneous Treatment
Freehold	0.133***	0.0382***	0.127***	0.0510***
	(0.002)	(0.002)	(0.002)	(0.002)
Distance	-0.0000233*	0.0000259**	-0.0000336**	0.00000346
	(0.000)	(0.000)	(0.000)	(0.0)
Treatment group	0.151*** (0.012)	0.154*** (0.011)	-	-
57 heterogeneous treatment groups	-	-	X	X
After treatment	0.203*** (0.011)	0.0727*** (0.011)	0.242*** (0.000)	0.110*** (0.000)
Treatment group * After treatment	-0.122*** (0.014)	-0.0925*** (0.014)	-	-
57 treatment groups * After treatment	-	-	X	X
Treatment group * After treatment * distance	0.0000302 (0.412)	-0.0000221 (0.000)	-	-
57 treatment groups * After treatment * distance	-	-	X	X
Unemployment	-0.123*** (0.000)	-0.114*** (0.001)	-0.114*** (0.001)	-0.0980*** (0.001)
Car Ownership	-0.00247*** (0.000)	-0.00247*** (0.000)	-0.00276*** (0.000)	-0.00393*** (0.000)
Constant	12.40***	12.59***	12.37***	12.52***
	(0.011)	(0.011)	(0.011)	(0.011)
Area FE (#8)	-	X	-	X
Time FE (#10)	-	X	-	X
House type categories	X	X	X	X
Observations	206.940	206.940	206.940	206.940
Adjusted R-squared	0.421	0.598	0.477	0.636

Notes: The dependent variable is the log of the transaction price. The reference category for House type is Detached. The reference locality for Area FE is Bury. The data covers 2010-2019. Unemployment and Car Ownership are measured as a percentage of the LSOA population. Standard Errors are in parentheses.
* significant with p<0,05, ** significant with p<0,01, *** significant with p<0,001.

Models (1) and (2) assume that all stops have a similar effect size. Models (3) and (4) include a unique effect size coefficient for every stop. The estimated coefficients for models (3) and (4) are shown in Appendix 7. In the appendix, the positive and negative coefficients are marked with colors and the boldly printed cells represent the coefficients with a p-value<0.05 (for p-values see Appendix 8). In model (3), 20 of 58 treatment effect coefficients are significant, in model (4) 22 of 58. Among the significant coefficients, the majority is negative and three are positive. This is in line with the results from models (1) and (2). However, besides the treatment effects (treatment*after coefficient), the interactions between the treatment effects and distance are estimated. In model (3), 11 of 58 (3) stops have significant interaction coefficients and in model (4) 13 of 58 stops are significant. The estimated coefficients of the interaction terms are plotted in Figure 8. In this figure, only the significant coefficients are plotted with the stop name in the legend. These graphs show the wide band of the effects size that is observed in the Greater Manchester Area. For reference, the yellow line is the coefficient for the interaction effect between the treatment effect and distance of model specification (2), the city-wide model. The vertical line at 500 meters depicts the border of the target radius. So, the coefficient can only be correctly interpreted on the left-hand side. However, this raises concerns with regard to the assumption of effect range used in this thesis. This is further elaborated on in the conclusions as a shortcoming of this thesis.

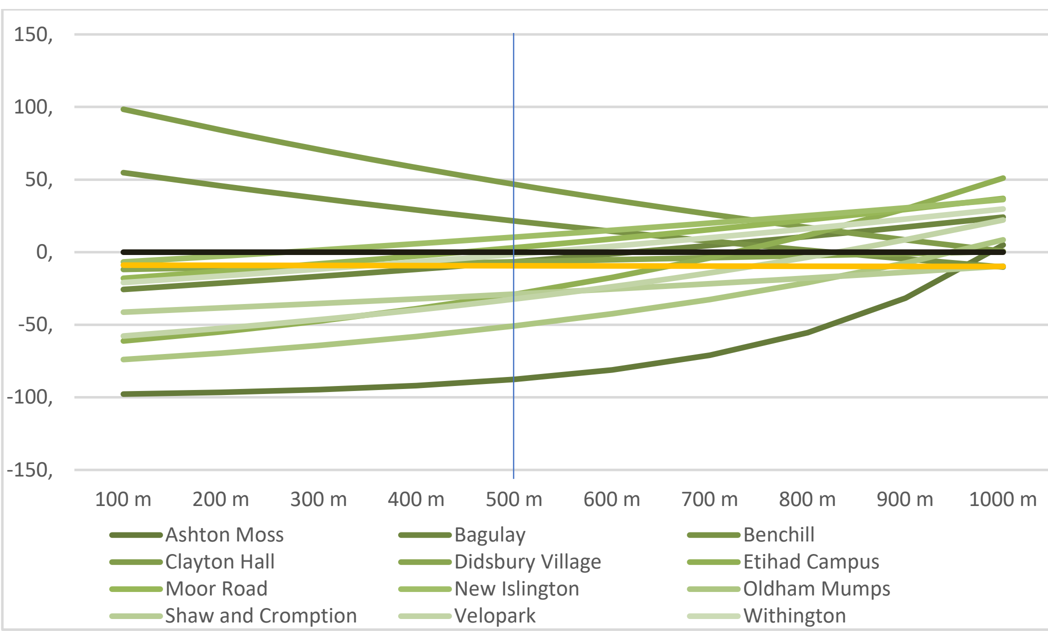


*Figure 8 Treatment effect over distance for individual stops that show a significant trend in model specification (4). On the X-axis the distance from the nearest Metrolink stop is displayed in steps of 100 meters. On the Y-axis the treatment effect is displayed. The effect over distance is measured by the interaction between Treat * After * Distance.*

To statistically confirm the inequality in effect sizes between stops the Wald-test is conducted. This test uses the treatment*after coefficient from the model (4). The results are summarized in Table 7. The table contains four versions of the test. Each version has a slightly different null hypothesis:

(1) H_0: treatment*after coefficients are equal

(2) H_0: interaction of treatment*after and distance coefficients are equal

(3) H_0: treatment*after coefficients with $p<0.05$ are equal

(4) H_0: interaction of treatment*after and distance coefficients with $p<0.05$ are equal

All the test results in table 7 allow for rejection of the null hypotheses with $p<0.001$. Rejection means that there are significant differences in treatment effect size per stop.

4.2 Robustness

This section presents the results from the robustness testing. As discussed in Chapter 3, previous research shows that the moment of treatment can be defined differently. The moment of treatment is the moment in time that is chosen to define which observations are labeled “before” and “after”. Dengg (2018) showed an anticipation effect in a Vienna metro expansion. This means that house prices are influenced by the metro line before operations commence. This effect is measured in Dengg (2018) by adding a treatment dummy based on the announcement day of the metro expansion. The models in this thesis use the opening date of the metro stop. The robustness test is executed to confirm that the opening date of the nearest stop is a substantiated moment to define a treatment effect. Two alternative moments of treatment are formulated. Table 8 shows the results. In the table, the stop opening treatment is the same as in model (2). The other results are from models that have omitted the stop opening treatment and exchanged it for an adjusted variant. The first alternative moment of

Table 8 Summary of estimation results

	Coefficient of treatment effect	R squared
Stop opening treatment	-0.0924765***	0.5975
Anticipation treatment	-0.0983277***	0.5975
Lagged treatment	-0.0896541***	0.5974

Notes: * significant with p<0,05 ** significant with p<0,01 *** significant with p<0,001. This table summarizes the results of three additional versions of the Diff-in-Diff model. The definition of the treatment group is changed. The goal is to see the effect of different definitions on coefficients and model performance. The rest of the model is the same as the model (4) from table 6. The following three definitions of the variable are used:
Stop opening treatment: the moment of treatment is equal to the opening day of the stop (as used in the earlier model specifications).
Anticipation treatment: the moment of treatment is placed one year before the opening of the Metrolink stop.
Lagged treatment: the moment of treatment is one year after the opening of the stop.

treatment is one year before opening and is called “anticipation treatment. The second is one year after the opening of the stop and is called “lagged treatment”. One year has been chosen as most construction activities took approximately a year to complete. As the results indicate, there is only a small difference between coefficients. To evaluate the different alternatives, the R-squared. The R-squared indicates explaining power. The R-squared is similar to both alternative definitions. Therefore, these different moments of treatment do not influence the interpretations of the results.

4.3 GWR Results

This section presents the results from the geographically weighted regression (GWR). The GWR is a test that comprises many local regressions and a test for spatial stationarity over the results of these local regressions. This provides insights into which coefficients are high or low in which area and if they are non-stationary over space. The difference-in-difference models were aimed at determining heterogeneity. The different model specifications of the GWR are geared towards testing several node parameters and one network parameter. These parameters are tested to explore their influence on the spatial non-stationarity of the treatment effect of stops. Table 9 shows the results of the global model and Table 10 shows the corresponding test results for spatial non-stationarity.

The different models are specified as follows:

1. Baseline GWR
2. GWR with the total potential of the nearest node (reach+attraction)
3. GWR with the reach of the nearest node
4. GWR with the sum of the potential of all nodes in the network at that stage of network development
5. GWR including the total network value and the reach of the nearest node
6. GWR that includes the distance to the central square in Manchester (nicknamed: CBD)
7. GWR that includes distance to CBD, the total network value, and the reach of the nearest node

Table 9 Results of the global GWR

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Freehold	0,1960032***	0,1733041***	0,1791608***	0,1999682***	0,1638150***	0,1641503***	0,1724905***
After treatment	0,0147831	-0,9598954***	-0,7211104***	0,0865460***	-0,1129427***	-0,0636190***	-0,1693652***
Treatment group	0,0143704	0,0082346	0,0089113	0,0126933	0,0115910	0,0086995	0,0089044
Unemployment	-0,1393957***	-0,1129541***	-0,1128689***	-0,1364514***	-0,1124410***	-0,1126440***	-0,1110200***
Car ownership	-0,0024346***	-0,0044983***	-0,0054630***	-0,0025819***	-0,0059999***	-0,0062149***	-0,0063090***
Distance	-0,00002470	-0,00000618	-0,00000469	-0,0000276*	0,00000880	0,00000568	0,00000578
Year FE	X	X	X	X	X	X	X
House Type categories	X	X	X	X	X	X	X
Ln(total potential)	-	0,1752701***	-	-	-	-	-
Ln(reach)	-	-	0,15639180***	-	0,20376350***	-	0,12288880***
Ln(total network value)	-	-	-	-0,00000616***	0,00001590***	0,00022160***	0,00000893***
Distance to CBD	-	-	-	-	-	-0,00001880***	-0,00000865***
Constant	12,314***	12,138***	12,167***	12,204***	12,185***	12,314***	11,858***
Adj R-squared	0,393	0,443	0,437	0,395	0,445	0,440	0,448
Prob F	0,000	0,000	0,000	0,000	0,000	0,000	0,000

Notes: Unemployment and Car ownership are measured as a share of the LSOA population. Ln(total potential) is the calculated reach + attraction, ln(total network value) is the aggregate of all the interactions in the network. The dependent variable is the log of the transaction price. The reference category for House type is Detached. The reference locality for Area FE is Bury. The data cover 2010-2019. Standard Errors are in parentheses.
* significant with p<0,05 ** significant with p<0,01 *** significant with p<0,001.

Positive coefficients

Negative coefficients

Table 10 P-values for test for non-stationarity over the local regressions

	1	2	3	4	5	6	7
	P-Value	P-Value	P-Value	P-Value	P-Value	P-Value	P-Value
Freehold	0,000	0,000	0,000	0,000	0,000	0,000	0,000
After treatment	0,002	0,000	0,000	0,000	0,000	0,000	0,000
Treatment Group	0,000	0,000	0,000	0,000	0,001	0,001	0,001
Unemployment	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Car ownership	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Year 1	0,522	0,002	0,263	0,161	0,511	0,000	0,000
Year 2	0,272	0,000	0,068	0,016	0,285	0,044	0,032
Year 3	0,342	0,000	0,068	0,028	0,380	0,002	0,002
Year 4	0,678	0,539	0,652	0,725	0,744	0,010	0,008
Year 5	0,814	0,679	0,752	0,787	0,825	0,041	0,029
Year 6	0,752	0,668	0,706	0,766	0,803	0,062	0,032
Year 7	0,753	0,695	0,770	0,778	0,804	0,117	0,051
Year 8	0,737	0,657	0,687	0,767	0,763	0,124	0,055
Year 9	0,680	0,610	0,674	0,702	0,740	0,127	0,063
Year 10	0,633	0,578	0,639	0,648	0,678	0,119	0,050
distance	0,000	0,000	0,000	0,000	0,000	0,141	0,055
Ln total potential	-	0,000	-	-	-	-	-
Ln reach	-	-	0,000	-	0,000	-	0,000
Ln total network value	-	-	-	0,012	0,000	0,000	0,000
Distance to CBD	-	-	-	-	-	0,000	0,000
House type	X	X	X	X	X	X	X
Constant	0,000	0,000	0,000	0,000	0,000	0,000	0,000

Notes: Results from the test for non-stationarity of the GWR. Obs.: 40.000. The data cover 2010-2019

Insignificant test results for spatial non-stationarity (p>0.05)

Several noteworthy characteristics of the results are highlighted in this section. Firstly, the variable "treatment group" does not have a significant coefficient as opposed to the global diff-in-diff models from Table 6. However, the treatment effect is not included in the GWR because the STATA package cannot include interaction terms. Secondly, the "distance" (to the nearest Metrolink stop) is only significant in the fourth (out of seven) GWR model and not by a wide margin. In the diff-in-diff models, this variable also has varying levels of significance over the different model formulations. This can be interpreted as the lack of a clear relationship between the distance to the Metrolink stop and the house prices. Instead, significant coefficients are consistently estimated for the network parameters. This, together with the wide range of results from the heterogeneous diff-in-diffs from Figure 8, hints at a more pronounced influence of the position of the stop in the network instead of the distance to the stop. This can be exemplified by the effect of the "reach" of the Metrolink stop. This variable is in a positive log-log relation with the dependent variable. This means that a 1% change in reach results in a 0.12% change in transaction price (in model specification 7). Adding one standard deviation to the mean equals a 24,5% change (see table 5). So, the effect of $+1\sigma = 24{,}5 * 0.12288 = 3\%$. So, the reach a stop provides has a noticeable effect on house prices. The total network value is also significant, but the effect size is smaller. The same calculation for $+1\sigma$ of the total network value results in a change of 0.002% in the transaction price. A smaller effect size is expected as this total network value measures the increased number of popular stops and users throughout the whole metropolitan area. Hence, the increased appeal of these properties is spread over a large supply of homes. This would, for a given demand in the region, spread the increase in bids over many properties. In conclusion, only a small premium can be derived from an increase in the value of the whole network.

By including the distance to CBD variable, the positive relationship with transaction price and the reach of the nearest Metrolink stop cannot be attributed to closeness to the city center. This ensures that reach is measured as an independent effect of the network position on the transaction price. In addition, the interpretation of the coefficient of "distance to CBD" supports the theory by Alonso (1964) as mentioned in section 2. The distance to the CBD shows a negative and significant relationship with transaction prices of houses with $(e^{-0.00000865} - 1) * 100\% =$ -0,9% per kilometer further away from the central square of Manchester.

4.4 Spatial non-stationarity

The test for spatial non-stationarity (Table 10) indicates that spatial variation in the coefficient is observed over the Manchester metropolitan area. This is in line with the conclusions of the diff-in-diff models, confirming earlier results. The variation is observed in the covariates as well as among the network characteristics. This means that including the network characteristics does not resolve the spatial variation among the other covariates. The spatial variation among the coefficients remains. The network characteristic variables are therefore not the only, primary, or conclusive reason for the variation in the treatment effects from the diff-in-diff models. However, the GWR does not include the treatment effect variable itself. Therefore, the relationship between the spatial variation in treatment effects and the network position cannot be ruled out.

An additional noteworthy observation is presented in Table 10. The table displays the p-values of the year FE dummy variables. As more network characteristics are added to the models, the p-values of these dummies decrease substantially. This means that without the network characteristics there is likely no spatial variation in the effect of heterogeneous year characteristics on the transaction price of houses. However, as more network characteristics are added to the model, a more distinguishable spatial variation of these effects is observed.

4.5 Spatial Patterns

This section looks at the maps and the presence of spatial patterns. The variability of the three coefficients from the seventh GWR model specification is mapped in Figures 9 to 11. The Metrolink network is displayed. There is an interpolated raster around the stops (inverse weighted distance) of the coefficient estimates of the local regressions. The legend has color coding to show how much an area deviates from the global model using standard deviations as units. Several noteworthy features of the maps are: (1) there seems to be a great similarity in the locations of positive and negative deviating areas between the three maps. So, when an area has a negative deviation from the global coefficient, there is also a negative deviation in the area for the other coefficients. This can be interpreted as areas with a higher or lower susceptibility to the network characteristics. (2) The topography of these maps is most extreme in the areas around the end of metro lines. Here, you can find the most positive and the most negative deviations in coefficients. For example, in the last map (Figure 11), the two northern, the eastern, the western, and one of the southern end stations show the most extreme deviations. The maps for the other coefficients show similar patterns.

As these two patterns appear in the maps for all three variables, bias in the data or model is suspected. Different sources are investigated. Firstly, extreme values in the raster can be caused by the Inverse Weighted Distance (IDW) interpolation method. Outside of the point cloud this method extrapolates and creates extremities. This problem could be suspected as the areas in question are at the edge of the dataset. To mitigate this, the areas outside of the point cloud have been clipped. This means that the presentation and interpretation are only conducted with data in the point cloud. This point cloud of local regressions can be seen in Appendix 9. In the case the pattern would be caused by the IDW, the pattern should not be present in the point data and only in the resulting raster. By focusing on the areas in question and labeling the points, the pattern can be confirmed in the point data. So, deformations due to the IDW interpolation are ruled out. Another reason for a bias in the model could be a low number of local regressions with outlier coefficient estimates. This can be further caused by a low number of observations in those areas on which the local regressions are based. However, further investigation in the form of a point density map (Appendix 10) shows that these areas are not necessarily the least populated in terms of observations or local regressions. Another reason could be that a sample of the full data was used to save processing time. A random sample of 40.000 observations was used. No reasons for a sampling bias have been identified. In conclusion, different reasons for a model or display bias are ruled out. Therefore, these extremities in topography are likely to be caused by patterns in the data. So, the results can be continued to be interpreted.

A possible explanation for the more extreme topography around the end stops could be that the addition of metro connections from a suburb to the city center has more impact than a connection to a neighborhood that is already well integrated into the urban network. How come that some transaction prices in neighborhoods react positively in some and negatively in other neighborhoods? This thesis cannot answer this question. Therefore, this question remains open to further research.

A visualization of the spread of regression points can be found in Appendix 9. This appendix shows that in the subsequent raster maps of the coefficients (Figure 9 – Figure 11), the raster image is based on the individual regressions.

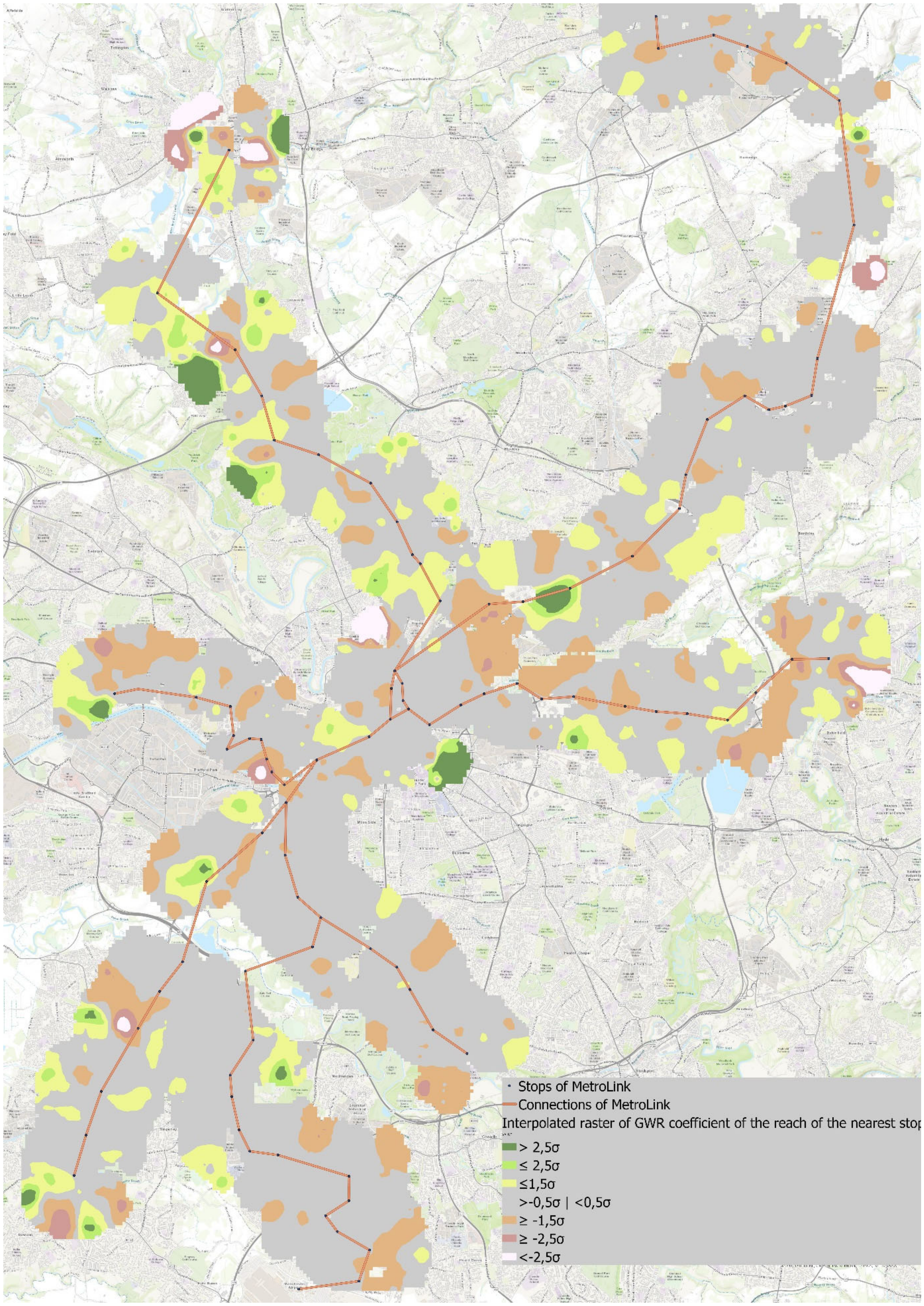


Figure 9 Map of deviations from the global coefficient of the "reach" variable

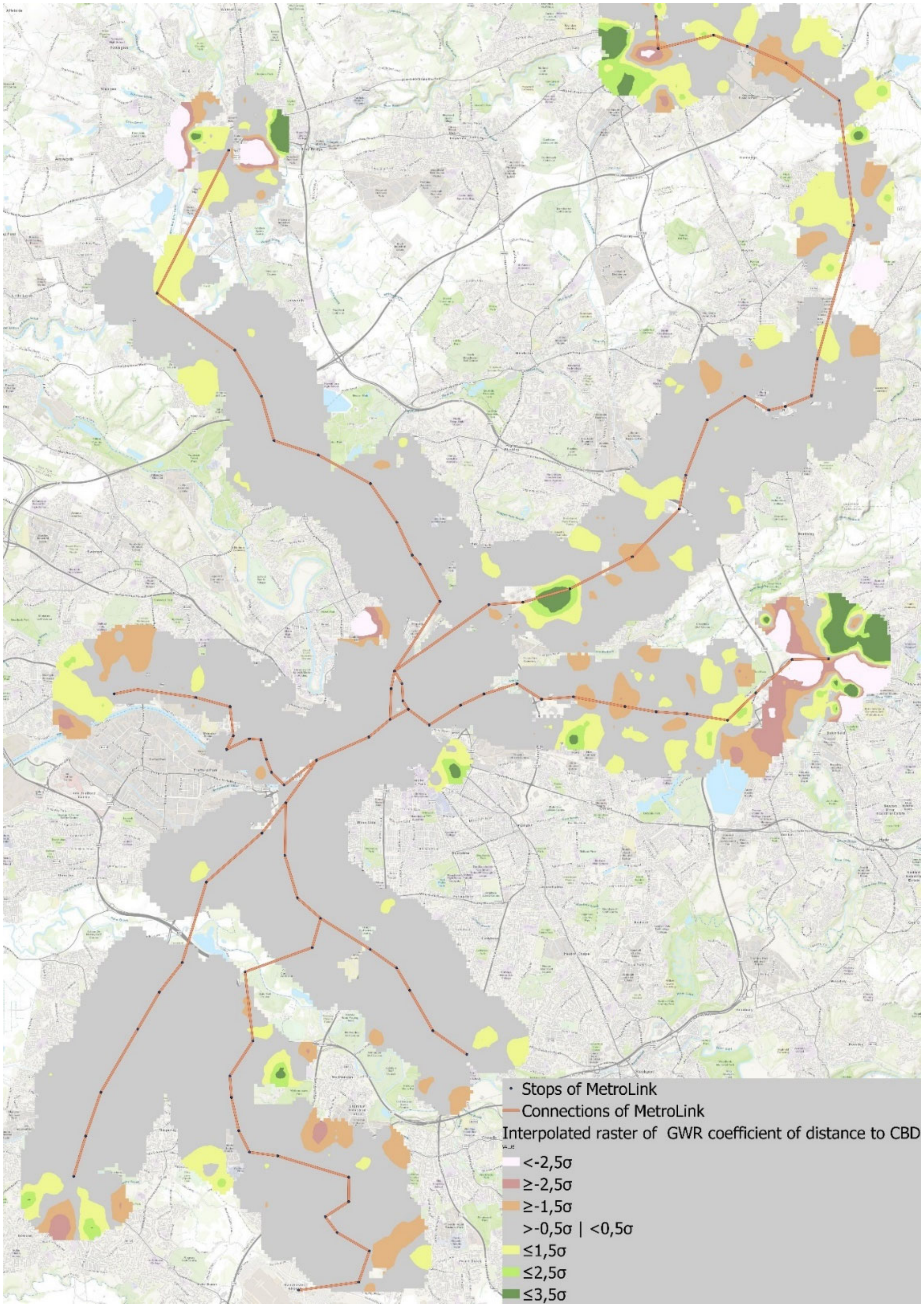


Figure 10 Map of deviations from the global coefficient of the "Distance to CBD" variable

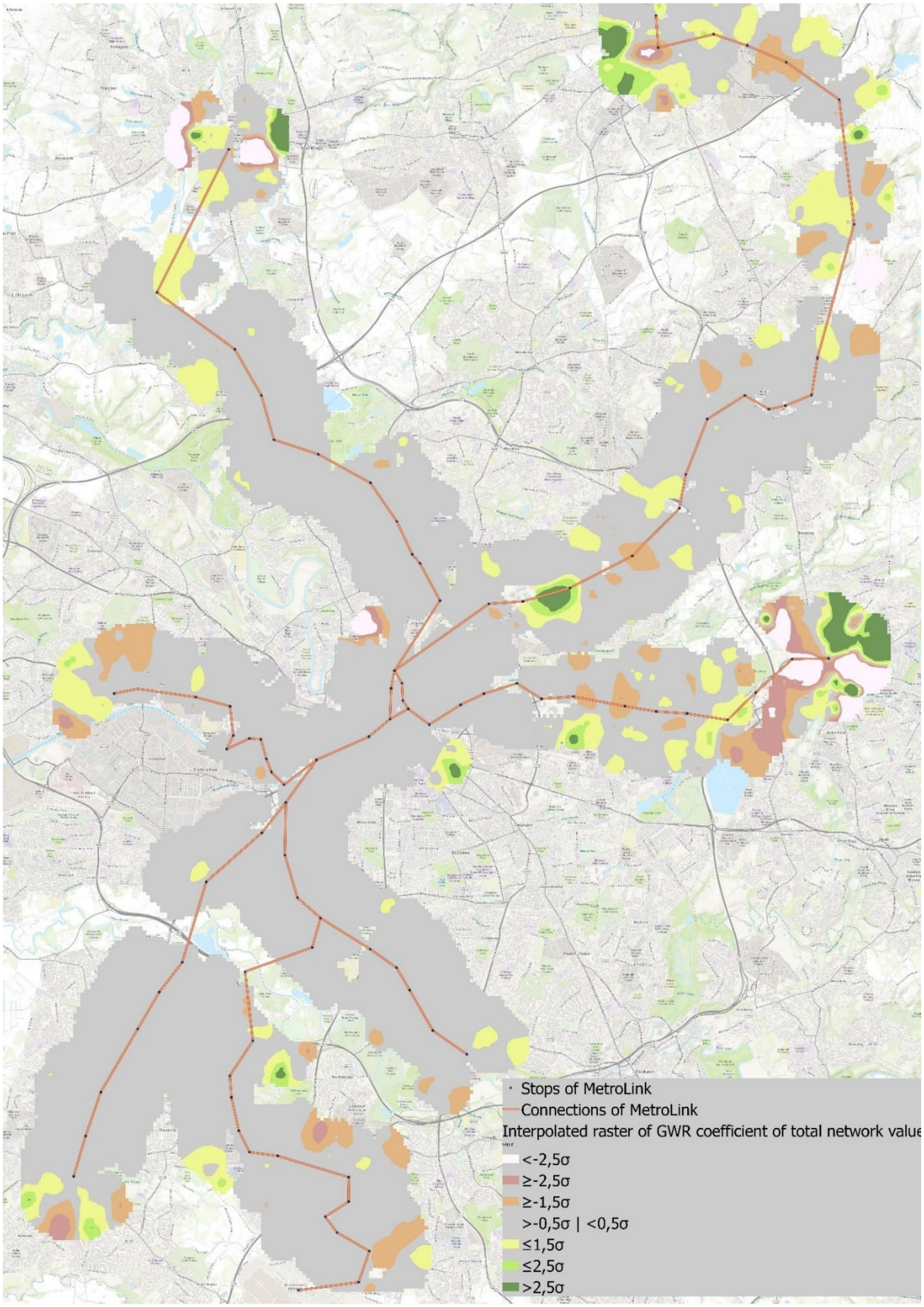


Figure 11 Map of deviations from the global coefficient of the "Total network value" variable

5. CONCLUSIONS

This thesis focuses on how new metro stops impact the house prices in the entire network. This is motivated by the theoretical relationship between an increase in accessibility and its effect on house prices. This theory hints at a different way of measuring than is commonly used in the literature. According to the theory, metro stops at different positions in the network should offer different accessibility premiums to house prices. Therefore, a network approach is introduced. This approach leads to the main research question: *To what extent do distant changes in the network impact the effect of new metro stops on house prices?*

The main findings of this thesis are twofold. First, it becomes clear that attempting to determine an effect size for a group of metro stops on house prices is a spurious generalization. Estimating one percentage of house prices premium for a group provides insight into the average effect of a completed project but does not provide insight into the local effect of house prices. This means that grouping the treatment effects of multiple stops does not provide an insight into the underlying dynamics. The second main finding is that by using gravity-based network parameters a relationship between house prices and the relative location of the Metrolink stop is found. This means that using these parameters in hedonic house prices models is a trustworthy addition to the hedonic house price models. In addition, it is apparent that network position is not the only variable that influences the differences in effect sizes of new metro stops. The complete dynamics of these varying effect sizes remain a subject for further research.

Three methods are used to achieve the main findings. Firstly, a literature study determined which network parameters are the most applicable to this research. Secondly, a difference-in-difference model is made, comparable to other literature. This model is compared to a difference-in-difference model that handles the treatment effect of every metro stop separately. This shows the varying effect sizes of the different stops. The third method is geographically weighted regression (GWR). This technique also estimates coefficients in a global model and consecutively estimates many local regressions. The coefficients of the local regressions are then tested on their spatial non-stationarity. The GWR included purposefully designed network parameters. These network parameters are gravity-based and use patronage data from the metro network. Hereby, the movements in the network are used to calculate accessibility from every stop's position in the network. This shows that the network position is a significant contribution to hedonic house prices models.

The main findings have implications for further research. Existing literature indicates a knowledge gap in the underlying principles of the effect of new rail infrastructure on house prices. The meta-reviews find that different case studies yield different results. The findings of this study show that there are differences in effect size between metro stops. These are differences within the same metropolitan area and not only between different case studies. This shows that the reasons for inconsistent results in the meta-reviews might not be based on the case cities but on more local factors. These factors are still poorly understood. This means that this thesis further emphasizes the characteristics of this knowledge gap.

Lastly, there are implications for planning practice, specifically for public infrastructure planners who aim to estimate the externalities of projects. Estimated benefits in CBAs from rail infrastructure projects are still coarse estimations. This thesis shows that the current methods from the literature do not provide a reliable method to estimate the effect of new metro stops on house prices. When this research field is further developed, it can help to understand the full benefits of infrastructure improvements. As the positive externalities of infrastructure are not always explored, it is recommended to fund research into the complete benefits of urban infrastructure projects beyond congestion relief. This will give policymakers and developers more insights into returns on infrastructure investments.

5.1 Limitations & Recommendations

Further research is also needed due to some of the limitations of this study. Solving these limitations can be a part of advancing this exploration. The following section is divided into two. The first paragraph discusses the limitations and further additions to the models in this thesis. The second paragraph discusses questions that require further conceptual thinking and modeling.

The models used in this thesis contain points of improvement. There were some limitations in the execution and there are some possible additions. There are four possible improvements. The first improvement can be made in the method of measuring the distance from the houses to the light rail stops. The used method took the absolute distance instead of the distance over roads, bridges, etc. This introduces inaccuracies in the observed values. This can be avoided in further research. The second improvement can be made in the geolocation of house transactions. Inaccuracies are introduced in the geolocation of houses transactions. In this research, postcode centroids are used. The centroids are used because they are publicly available data. In combination with the previous inaccuracies, the insignificant results in the variable “distance to the light rail stop” might be explained. However, with access to the coordinates of house addresses and a (walkable) road network this measure is easily improved. The third improvement could focus on the interaction between the effect of infrastructure on house prices and other socioeconomic and mobility variables, like the network parameters. Even on a neighborhood level, these variables could provide clues regarding dynamics which explain the pattern of spatial variation in the results of this study. Adding the interaction between these could give more insight into the relationship between the effect size of metro stops and other variables. The fourth possible improvement adds robustness. In further research, the patronage-based network parameters can be compared to urban function-based network parameters. Theoretically, these parameters measure a similar phenomenon The urban function-based network parameters are more complex to measure, whereas the network parameter from this study is more of a proxy by measuring the flow of people disembarking at light rail stops. The benefit of the method used in this study is that the model uses only one variable to describe the accessibility to different types of activities and services. Whereas including the distance to every single type or category of service would introduce many variables to a model. This would raise data requirements and complexity to models. Does combining these two approaches raise issues of multicollinearity or improve the explaining power of the model? Which of the approaches has the best explaining power?

Besides its limitations, this thesis raises several new research questions. There are five of these further recommended investigations. The first research recommendation is into the range of effect sizes. This study establishes that different stops have different effect sizes on house prices. However, this study did not explore whether different stops have varying ranges of their effects. Varying ranges have been researched in the literature, but the ranges were not allowed to differ between stops. In this thesis, the treatment group was within a 500-meter range. Could this range be linked to a variable? Do hub stations have a larger range than mid-line or end of line stops? So, is this 500-meter range valid for every (type of) stop? Perhaps the range can vary with a network parameter like reach or can be fitted using more variables. These questions remain for further research. The second research recommendation is to include more modes of transport. The underlying theory for this and similar studies is the hypothesis that better accessible homes go for a premium. Exploring this by only regarding the metro network or even a single extension of a metro line, is destined to be an incomplete exploration. This neglects preexisting travel options. Future research should move towards models with a comprehensive metropolitan network. This metropolitan network would include car, cycle, walking mobility, bus, light rail, heavy rail but also the possibility of multi-modal trips with Park and Ride facilities. A comprehensive mobility model might improve the estimation of the effect of new infrastructure on house prices. It is complicated to create a longitudinal model for this as transport systems are under constant change. For example, rerouting of lines, changes in frequency and, new business districts or neighborhoods. There is a large number of these changes throughout the span of a decade. This requires an intensive investigation. The third research recommendation is into the causal relationship between new metro stops and house prices premiums. Is the effect of new infrastructure on house prices direct or indirect? Does a new light rail stop affect house prices because it offers better mobility? Alternatively, does it influence house prices because it increases footfall/changes the audience that supports different services around the stop that in turn influence the bid rent curve? A comparison between new services (shopping streets, offices, etc.) around the stops versus stops in solely residential areas could help to resolve this question. The fourth research recommendation is about the spatial pattern that is observed in this thesis. A surprising trend is the large variation of this effect around the terminal stops of the light rail lines. Observing both weaker and stronger relationships between network parameters and house prices in neighborhoods adjacent to the terminal stops. This pattern was not among the hypothesized patterns for spatial variation. The reason for this pattern cannot be proven with this study and remains a subject for further research. The fifth recommendation for further research is a comparative study. High on the research agenda should be a similar study that includes multiple metropolitan areas. Not in a meta-review but in a panel data model. This objective has difficult data requirements. However, such a study could be the start of creating comparable results and understanding underlying dynamics.

REFERENCES

Alonso, W. (1964). A theory of the urban land market. *Papers and proceedings of the regional science association*, 6, 149-157.

Boardman, A., Vining, A., Waters, W. G. (1993). Costs and Benefits through Bureaucratic Lenses: Example of a Highway Project. *Journal of Policy Analysis and Management*. 12(3):532-555

Chandra, A.,Thompson, E. (2000). Does Public Infrastructure Affect Economic Activity? *Regional Science and Urban Economics*, 30(4):457–490

Chesterton (2000). Property Market Study – Working Paper No. 32. Report Jubilee Line Extension Impact Study Unit University of Westminster.

Clauretie, T. M., Neill, H. R. (2000). Year-round school schedules and residential property values. *The Journal of Real Estate Finance and Economics*, 20(3): 311–322. https://doi.org/10.1023/A:1007841326833

Dear, M. J., Flusty, S. (2000). Postmodern Urbanism. *Annals of the Association of American Geographers*. 90(1): 50-72

De Haan, J., Diewert, E. (2013). "Hedonic Regression Methods", in OECD, et al., Handbook on Residential Property Price Indices, Eurostat, Luxembourg

Dengg, S. (2018). Master's thesis: *How a Metro Development Project Affects Land Values – Evidence from Vienna*

Derrible, S., Kennedy, C. (2009). Characterizing metro networks: state, form, and structure. *Transportation*, 37:275–297 DOI 10.1007/s11116-009-9227-7

Derrible, S., Kennedy, C. (2010). Applications of Graph Theory and Network Science to Transit Network Design. *Transport Reviews*, 31(4): 495–519, July 2011

Derrible, S. (2010). The Properties and Effects of Metro Network Designs. Graduate Department of Civil Engineering, University of Toronto

Dimitrov, S., Ceder, A (2016). A method of examining the structure and topological properties of public-transport networks. *Physica* A, 451: 373–387

Du, H., Mulley, C. (2007). Transport Accessibility and Land Values: A Case Study of Tyne and Wear. Report RICS Research Paper Series.

Evans (2004). *Economics, Real Estate and the Supply of Land.* Blackwell Publishing: Oxford

Fejarang, R. A. (1994). *Impact on property values: a study of the Los Angeles Metro Rail*. London: PTRC Education and Research Services Ltd.

Galati, G., Teppa, F. (2017). Heterogeneity in House prices Dynamics. *De Nederlandsche Bank*, Working Paper 564: 1-23.

Garretsen, H., Marlet, G. (2017). Amenities and the attraction of Dutch cities. *Regional Studies*, *51*(5): 724–736. https://doi.org/10.1080/00343404.2015.1135239

Gibbons, S., Machin, S. (2003). *Rail Access and House prices: An Evaluation of the Wider Benefits of Transport Improvements*. Report.

GOV.UK (2019). How to access HM Land Registry Price Paid Data. Retrieved from https://www.gov.uk/guidance/about-the-price-paid-data#explanations-of-column-headers-in-the-ppd on 8-05-2019

Gross, J., Jellen, J. (2005). *Graph Theory and its applications.* Taylor & Francis Group: New York

Heddebaut, O., Palmer, D. (2014). Multimodal city-hubs and their impact on local economy and land use. L'archive ouverte pluridisciplinaire HAL.

Hodge, T., Sands, G., Skidmore, M. (2017). The Land Value Gradient in a (Nearly) Collapsed Urban Real Estate Market. *Land Economics*, 93(4):549-566.

Levkovich, O., Rouwendal, J., van Marwijk, R. (2015). The effects of highway development on housing prices. *Transportation*, 43 (2): 379-405

Litman, T. (2008). Evaluating Accessibility in transport planning. *87th Transportation Research Board Annual Meeting*

Kessel, C. (2011). Manchester Metrolink 20 Years of Evolution. Retrieved from https://www.railengineer.co.uk/2011/11/30/manchester-metrolink-20-years-of-evolution/ on 20-03-2019

Koster, H. R. A., Rouwendal, J. (2012). The impact of mixed land use on residential property values*. *Journal of Regional Science*, *52*(5): 733–761. https://doi.org/10.1111/j.1467-9787.2012.00776.x

McFadden, D. (1977). Modelling the Choice of Residential Location. *The Economics of Housing*, 1, 531-552.

Mishra, S., Welch, T.F., Jha, M.K. (2012). Performance indicators for public transit connectivity in multi-modal transportation networks*. Transportation Research,* Part A 46:1066–1085

Mohammad, S. I., Graham, D. J., Melo, P. C., Anderson, R. J. (2013). A Meta-Analysis of the Impact of Rail Projects on Land and Property Values. *Transportation Research Part A*, 50: 158–170.

Moridpour, S., Toran Pour, A., Saghapour, T. (2019). Big Data Analytics in Traffic and Transportation Engineering: Emerging Research and Opportunities: Emerging Research and Opportunities. *IGI Global*, 11 Jan 2019 - Technology & Engineering - 188 pages

Murray, CK. (2017). Light Rail, Land Values and Taxes. *Economic Record,* 93 (302): 448-464.

Museum of Transport Greater Manchester (2004). A Short History of Public Transport in Greater Manchester. Retrieved from https://web.archive.org/web/20101110051835/http://www.gmts.co.uk/explore/history/history.html on 23-03-2019

Office for National Statistics (2018). ONS Postcode Directory (Latest) Centroids. Released on 31-5-2018

Office for National Statistics (2011). Census geography. Via: https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography

Office for National Statistics (2020). Consumer Price Index. Retrieved in February 2020 via: https://www.ons.gov.uk/economy/inflationandpriceindices

Ottensmann, J., Payton, S., Man, J. (2008). Urban Location and Housing Prices within a Hedonic Model. *The Journal of Regional Analysis and Policy.* 38(1):19-35

Ozus, E., Dokmeci, V., Kiroglu, G., Egdemir, G. (2007). Spatial analysis of residential prices in Istanbul. *European Planning Studies*, *15*(5): 707–721.

Percoco, M. (2016). Highways, Local Economic Structure and Urban Development. *Journal of Economic Geography*, 16(5):1035–1054.

Projectmapping.co.uk *Greater Manchester MetroLink Maps.* retrieved in July 2019 via: http://www.projectmapping.co.uk/Reviews/manchester_train_maps.html

Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika,* 31, 581–603 (1966). https://doi.org/10.1007/BF02289527

Stoilova, S., Stoev, V. (2015). Application of the graph theory which examines the metro networks. *Transport problems,* 10 (2).

Suzuki, Hiroaki, Murakami, Yu-Hung Hong, Tamayose, B. (2015). *Financing Transit-Oriented Development with Land Values: Adapting Land Value Capture in Developing Countries.* Urban Development Series. Washington, DC: World Bank.

Ripley, W. (1911). Railway speculation. *The Quarterly Journal of Economics, 25*(2): 185-185. doi:10.2307/1884948

Rodrigue, J. P. (2020). *The Geography of Transport Systems*. New York: Routledge ISBN: 978-0-367-36463-2

TFGM (2019). Greater Manchester tram network map. Retrieved from https://tfgm.com/public-transport/tram/network-map on 20-03-2019

Van Dijk, D., Siber, R., Brouwer, R., Logar, I., Sanadgol, D. (2016). Valuing water resources in Switzerland using a hedonic price model. *Water Resources Research*, *52*(5): 3510–3526. https://doi.org/10.1002/2015WR017534

Van Duijn, M., Rouwendal, J., Boersema, R., (2016). Redevelopment of Industrial Heritage: Insights into External Effects on House prices, Regional Science and Urban Economics, 57:91-107.

Appendices

Appendix 0

Expert interview about relevance of research

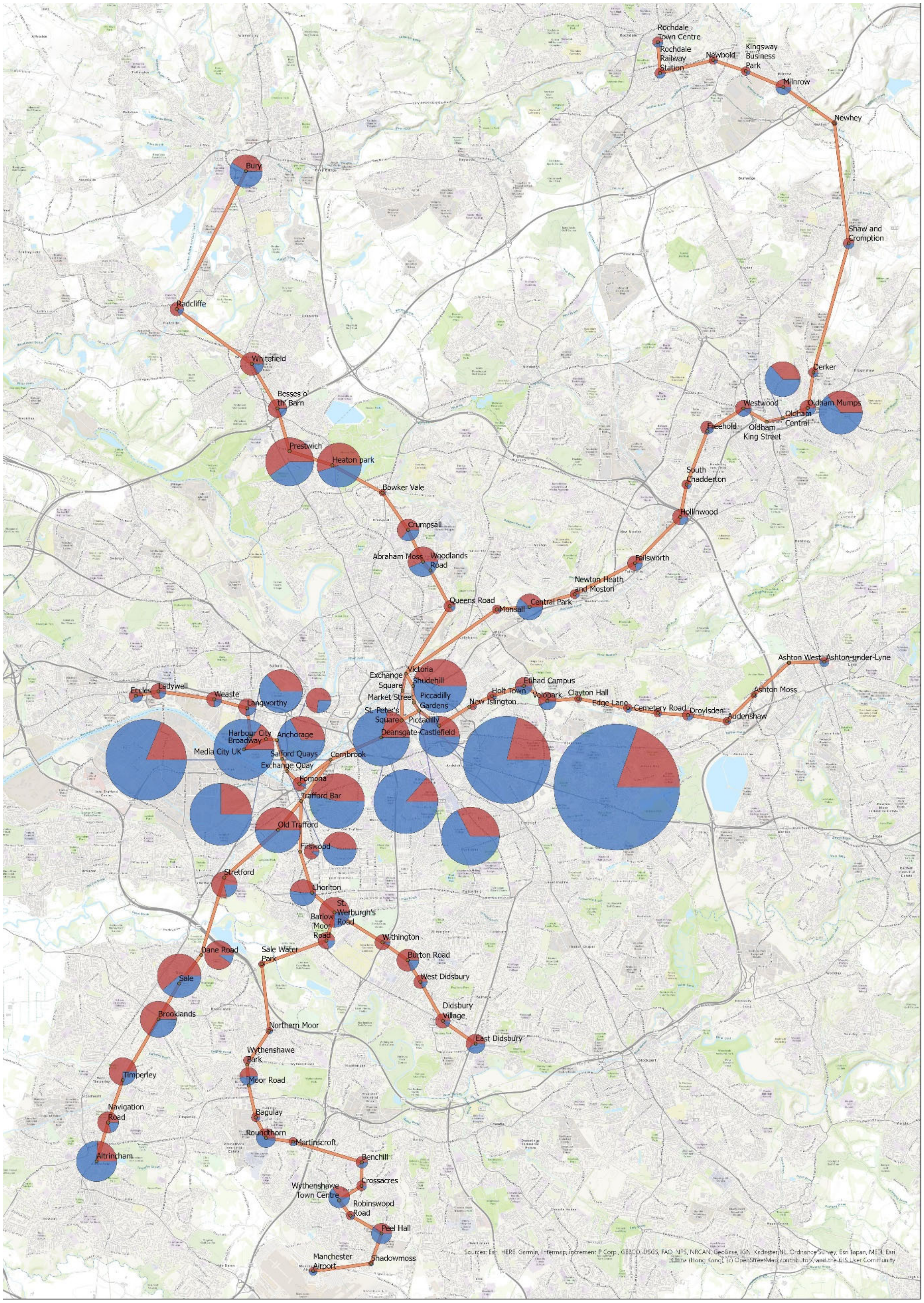
An expert from practice has been interviewed to discuss implications of this thesis. The expert that is interviewed is Wolbert van der Haar. He works for Arcadis and has a background in real estate and spatial planning. In his portfolio at Arcadis there are also infrastructure related projects.

Firstly, we discussed societal relevance. When asking Wolbert about the implications for Cost-Benefit Analysis. At first, he responded that it seems logical a local government wants more insights in its budgeting. However, on second thought he was doubtful how much real estate values are regarded in this type of decision making. Currently, these types of decisions are mainly based on congestion relief. For financing of projects, local governments usually look to higher levels of government and attempt to argue the regional, national, or transnational (EU) benefits. Wolbert estimates that the mindset and the different goals of public stakeholders make the direct application of effect size from infrastructure development limited. Real estate is (in NL) considered a private interest. However, social geographers and planners sometimes ask what a new metro line did or will do to an area. For example: “Did it kick start gentrification?”, and similar questions.

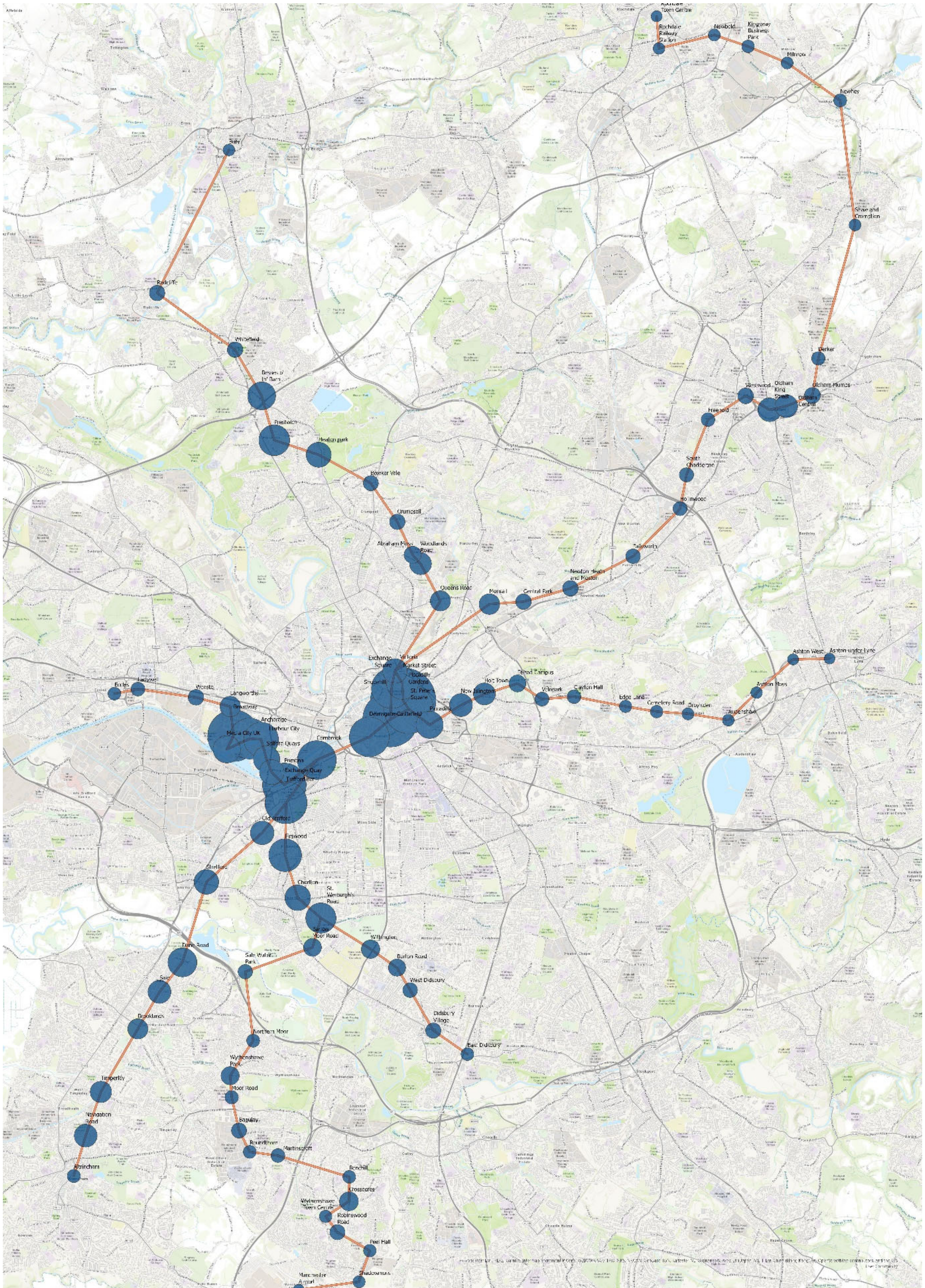
Secondly, we discussed whether similar calculations are done in business and real estate industry. Despite the fact that the main clientele of Arcadis are governments, he mentions that many organizations look thematically at issues. So, there is a team at Arcadis that calculates the economic costs of flood events. The methods in this study are at the edge of infrastructure planning and real estate. These are still often different sectors. However, in the public sector the cry for a comprehensive paradigm for area development is getting louder.

Thirdly, Wolbert and I discussed the results of this thesis. He expected a strong effect in green field locations. In those situations, the empty plots around infrastructure go for a premium. However, in redevelopment cases or purely transactions of homes the situation is much more complex. Perhaps this difference is what makes it difficult to pinpoint a single dynamic for the effect size of new metro stops. This raises the question about the possibility of doing a study with greenfield sites. In addition, Wolbert thinks that the variation in effect sizes is difficult to link to new infrastructure because the time horizons in business and infrastructure processes are different. Professional investors and developers cannot always afford to wait or gamble on the actual realization of the infrastructure. So, for them it is difficult to capitalize on infrastructure development. They can start bidding more when the whole area is about to rejuvenate, but possible infrastructure developments play only a part in that. For example, the developments around train station Holland Spoor. An old train station with a hinterland that exploded in land value in the last decade without physical infrastructure changes locally.

In conclusion, this interview shows that this type of research is too much of a niche to have mainstream application in industry and public planning. Public and private institutional structures are not shaped to do this type of research. Real estate businesses have many other things to consider, especially in redevelopment areas. In those cases, accessibility plays a part in the rejuvenation potential of a neighborhood but does not dominate. This means that accessibility potential can take decades to be capitalized in real estate prices. The effects for individual home buyers and how they value accessibility is not something Wolbert is an expert in. Manchester realtors did not respond to comment on these questions.



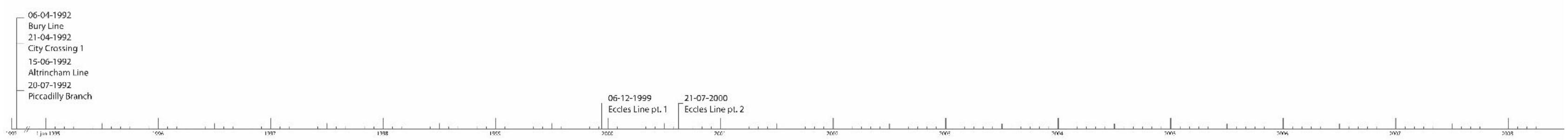
Appendix 1a Inbound (blue) & Outbound (red) users per stop on a weekday morning in 2017

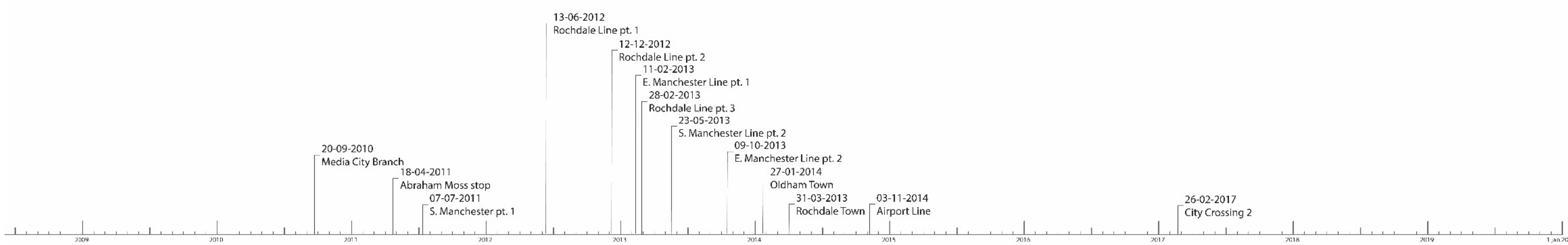


Appendix 1b Size of the "ln(reach)" variable at every Metrolink stop

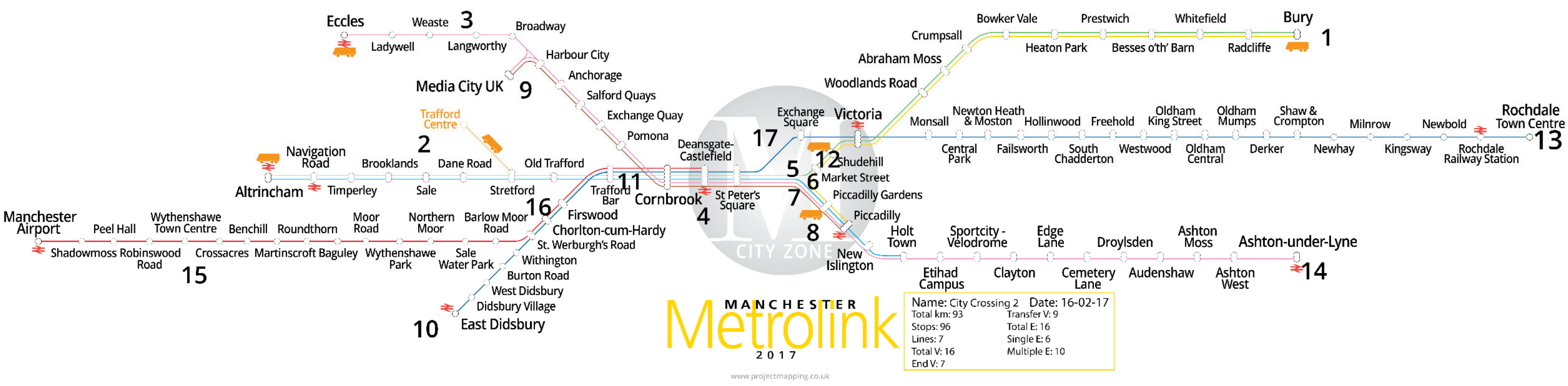
Appendix 2 Timeline of Metrolink Adjustments

The development of the MetroLink network
during the period of transaction data availability
Starting in 1995

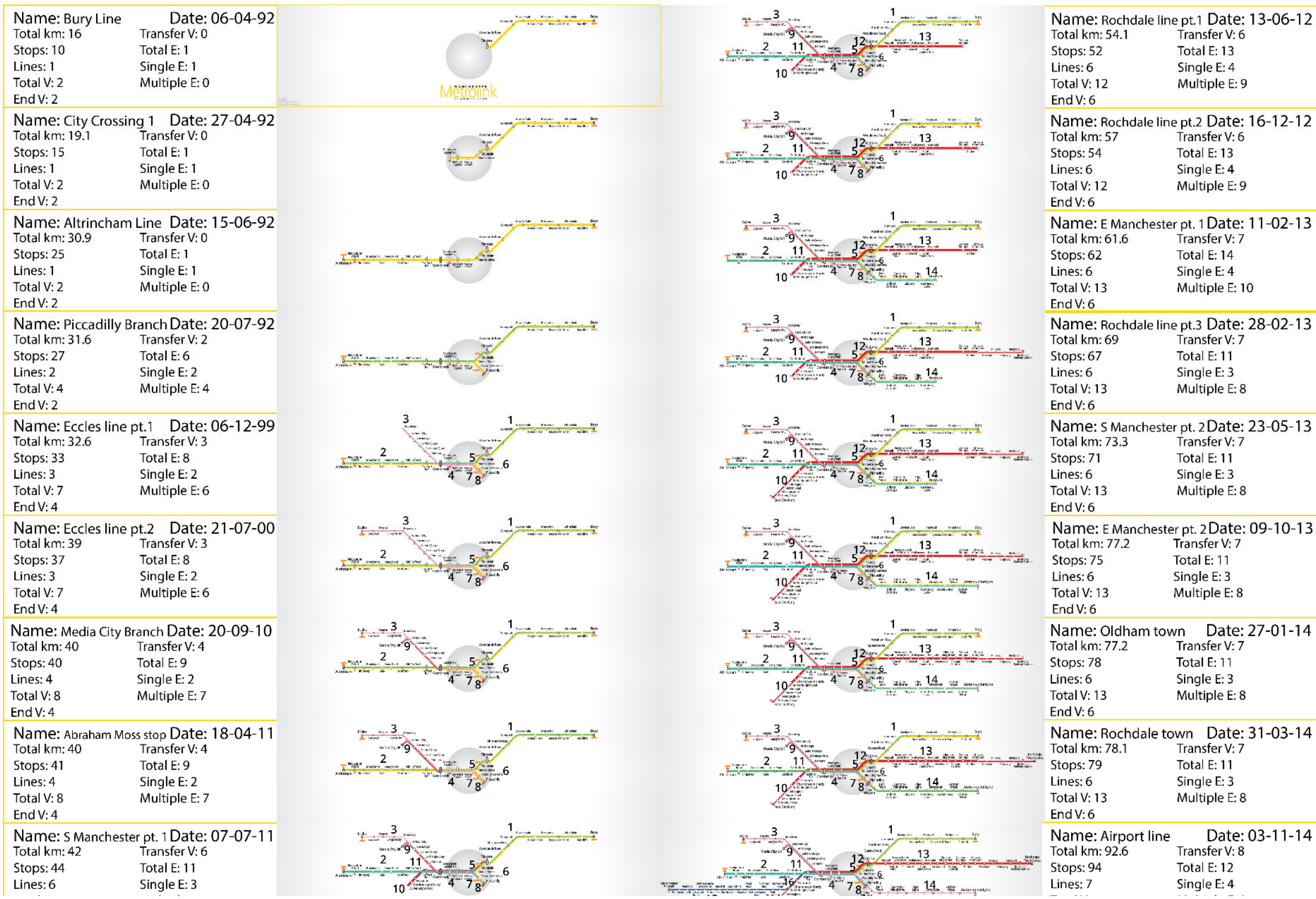




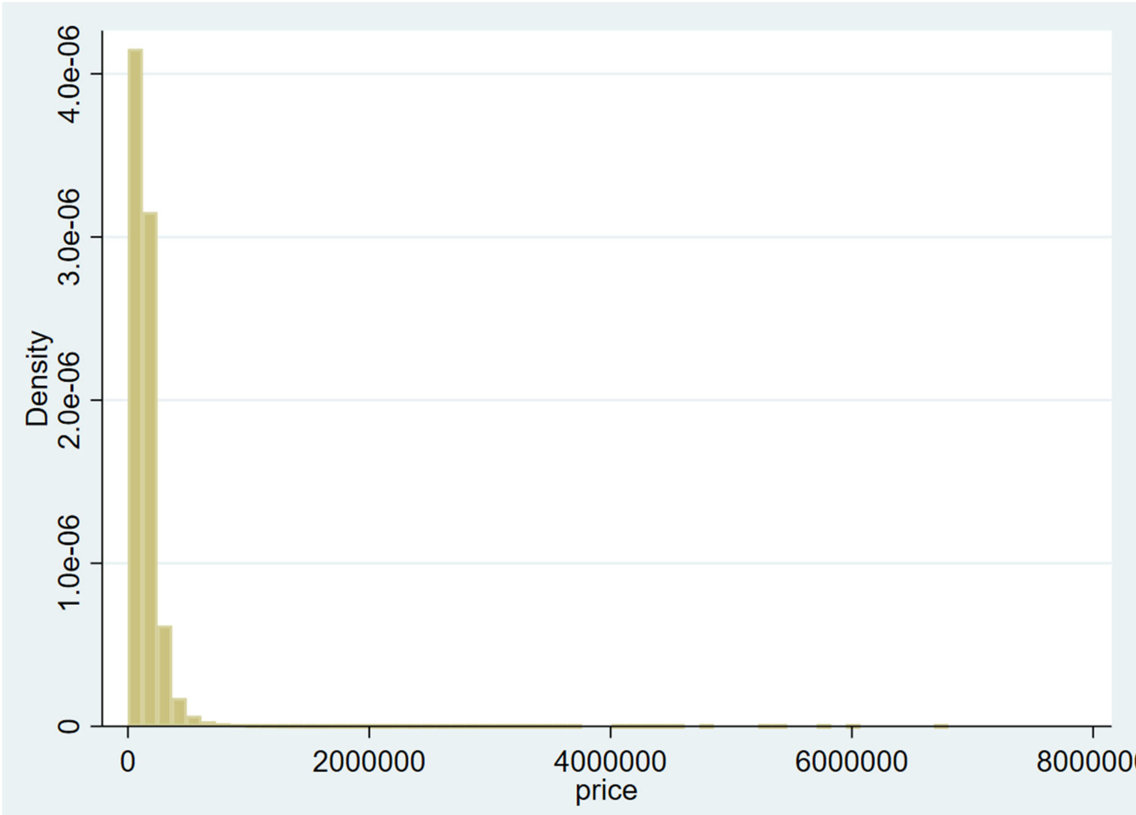
Appendix 3a Map of latest version of the Metrolink network



Appendix 3b Maps of all Metrolink Network adjustments



Appendix 4 Histogram of data before excluding outliers, log transformation and deflation



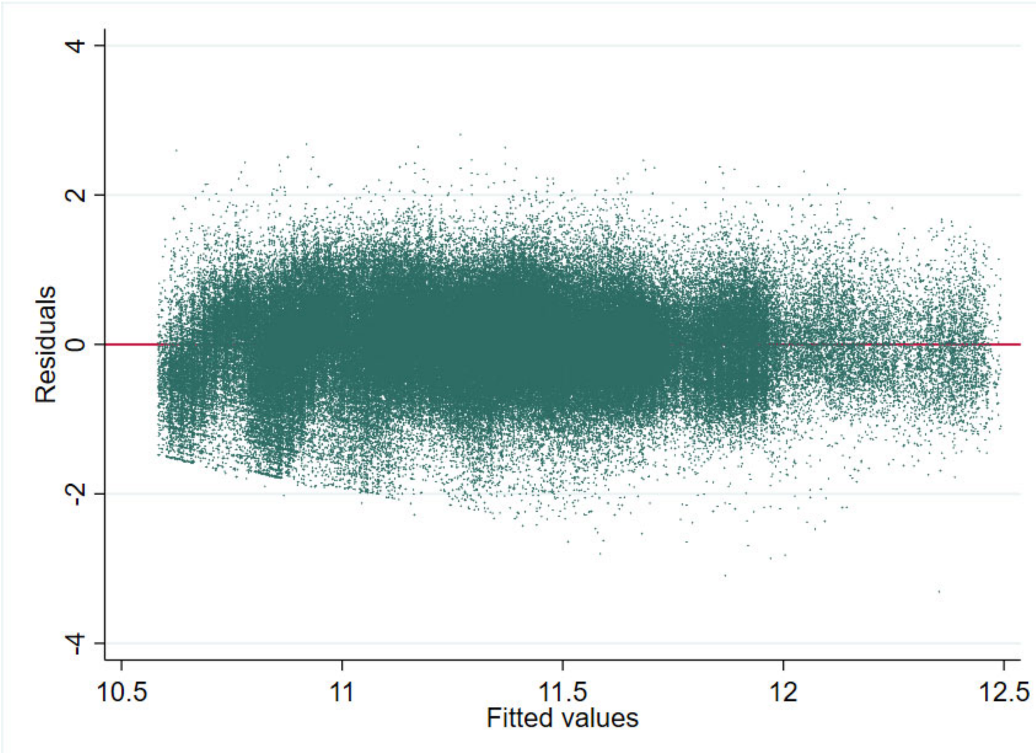
Appendix 5 Observations per Stop (n<500 is red)

Stop name	Freq.	Percent	Stop name cont.	Freq.	Percent
Abraham Moss	2,255	1.09	Media City UK	512	0.25
Altrincham	5,688	2.75	Milnrow	1,991	0.96
Anchorage	401	0.19	Monsall	816	0.39
Ashton Moss	2,063	1	Moor Road	1,250	0.6
Ashton West	1,407	0.68	Navigation Road	4,156	2.01
Ashton-under-Lyne	3,674	1.78	New Islington	1,583	0.76
Audenshaw	2,234	1.08	Newbold	1,705	0.82
Bagulay	1,417	0.68	Newhey	986	0.48
Barlow Moor Road	1,999	0.97	Newton Heath and Moston	3,552	1.72
Benchill	2,527	1.22	Northern Moor	2,743	1.33
Besses o' th' Barn	3,287	1.59	Old Trafford	1,568	0.76
Bowker Vale	2,301	1.11	Oldham Central	696	0.34
Broadway	109	0.05	Oldham King Street	1,408	0.68
Brooklands	4,236	2.05	Oldham Mumps	3,118	1.51
Burton Road	3,739	1.81	Peel Hall	1,365	0.66
Bury	4,263	2.06	Piccadilly	1,753	0.85
Cemetery Road	2,857	1.38	Piccadilly Gardens	1,184	0.57
Central Park	2,180	1.05	Pomona	22	0.01
Chorlton	4,615	2.23	Prestwich	2,875	1.39
Clayton Hall	1,790	0.86	Queens Road	1,054	0.51
Cornbrook	3,964	1.92	Radcliffe	4,607	2.23
Crossacres	707	0.34	Robinswood Road	646	0.31
Crumpsall	1,774	0.86	Rochdale Railway Station	2,632	1.27
Dane Road	1,816	0.88	Rochdale Town Centre	2,229	1.08
Deansgate-Castlefield	4,816	2.33	Roundthorn	622	0.3
Derker	2,501	1.21	Sale	4,350	2.1
Didsbury Village	3,873	1.87	Sale Water Park	964	0.47

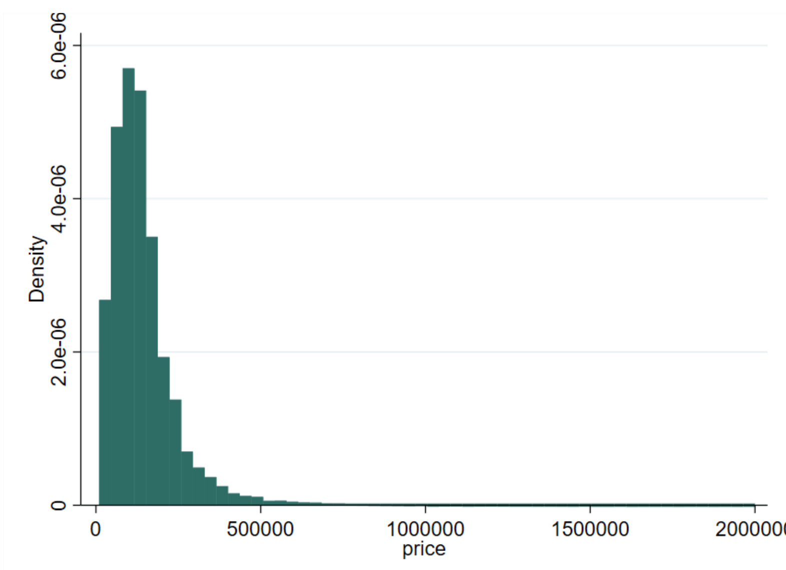
Droylsden	2,089	1.01	Salford Quays	376	0.18
East Didsbury	3,669	1.77	Shadowmoss	496	0.24
Eccles	2,826	1.37	Shaw and Cromption	4,449	2.15
Edge Lane	2,933	1.42	Shudehill	1,240	0.6
Etihad Campus	291	0.14	South Chadderton	3,055	1.48
Exchange Quay	440	0.21	St. Peter's Square	1,262	0.61
Exchange Square	2,439	1.18	St. Werburgh's Road	2,248	1.09
Failsworth	4,112	1.99	Stretford	4,658	2.25
Firswood	2,830	1.37	Timperley	4,253	2.06
Freehold	2,881	1.39	Trafford Bar	1,675	0.81
Harbour City	704	0.34	Velopark	630	0.3
Heaton Park	2,700	1.3	Victoria	2,026	0.98
Hollinwood	3,035	1.47	Weaste	2,130	1.03
Holt Town	582	0.28	West Didsbury	4,545	2.2
Kingsway Business Park	932	0.45	Westwood	1,729	0.84
Ladywell	1,588	0.77	Whitefield	3,842	1.86
Langworthy	2,732	1.32	Withington	1,322	0.64
Manchester Airport	157	0.08	Woodlands Road	1,179	0.57
Market Street	636	0.31	Wythenshawe Park	2,176	1.05
Martinscroft	1,923	0.93	Wythenshawe Town Centre	1,281	0.62
			Total	**206,951**	**100**

Appendix 6 OLS assumption testing and descriptive

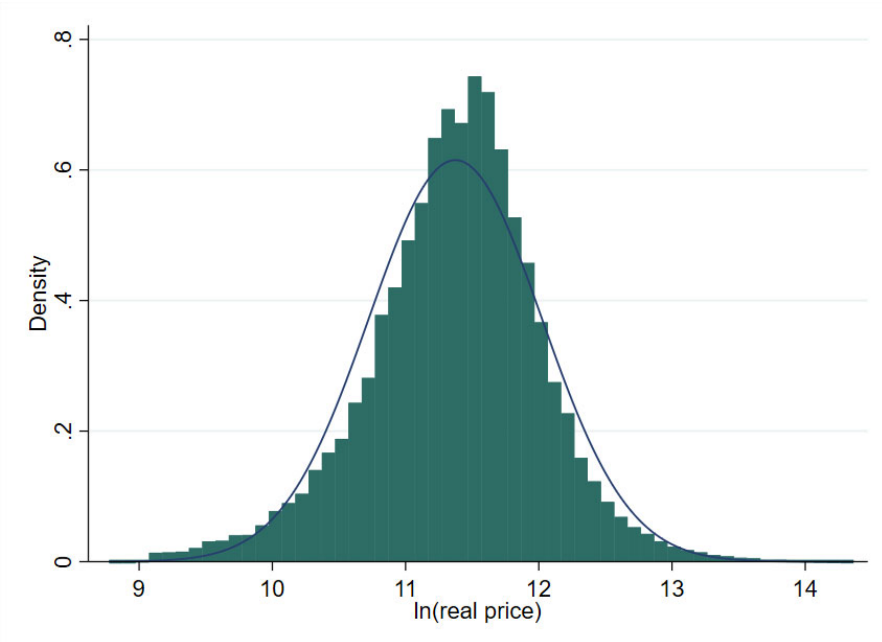
6a Observing homoskedasticity



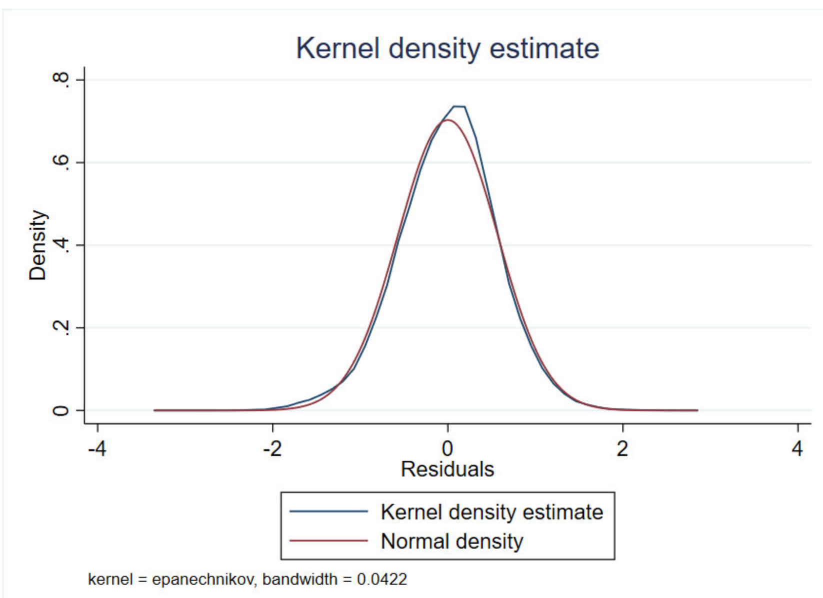
6b Observing skewness



6c Result of log transformation



6d remaining skewness & kurtosis



6e List of data collection conditions in transaction dataset

The dataset excludes (HM Land Registry, 2020):

- sales that have not been lodged with HM Land Registry
- sales that were not for value
- transfers, conveyances, assignments, or leases at a premium with nominal rent, which are:

    - ‘Right to buy’ sales at a discount
    - subject to an existing mortgage
    - to effect the sale of a share in a property, for example, a transfer between parties on divorce
    - by way of a gift
    - under a compulsory purchase order
    - under a court order
    - to Trustees appointed under Deed of appointment
- Vesting Deeds Transmissions or Assents of more than one property

Multiple variables from this source are included in the model:

- House type: the type of house (detached, semi-detached terraced, flats and other)
- Freehold: Whether the house has a freehold or leasehold ownership.
- Price paid: the market value transaction price
- Date: the date of transaction

6f Spatial accuracy of house transactions

To join the transaction data to a location and subsequently to the closest metro stop, the locations have to be georeferenced. Georeferencing on the address level is possible with a larger budget (especially for 400.000+ observations). Therefore, postcode centroids have been used. This diminishes the accuracy of the distance to the closest stop. However, the accuracy loss is relatively small as postcodes in the UK are available for every street segment. Figure 12 illustrates this postcode density (central City of Manchester).

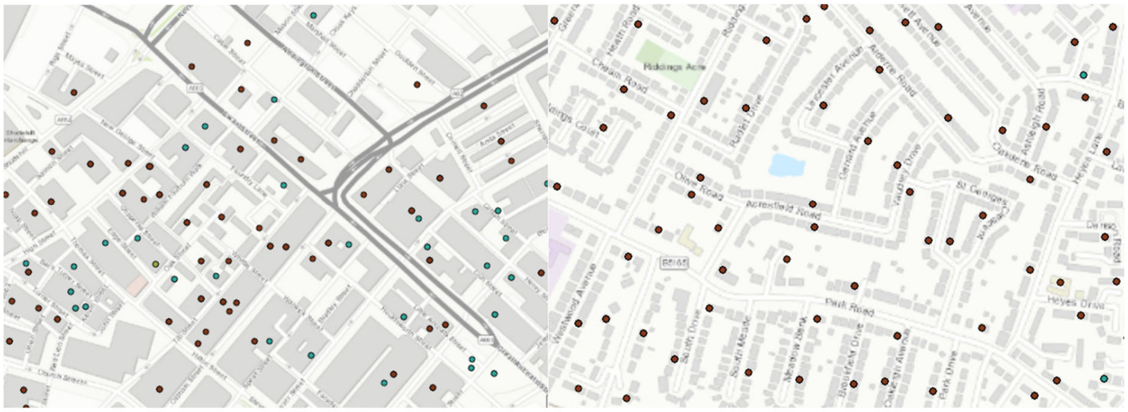


Figure 12 Postcode centroids in Greater Manchester (left: Manchester city center, right: Altrincham residential neighborhood)

6g VIF test

To obtain coefficients with small standard errors multicollinearity should be avoided. One indicator for this is a high correlation between explanatory variables. The correlation matrix (table 5) indicates that most variables have a correlation that does not lead to any suspicion of multicollinearity. However, the correlation between the percentage of people that have no car and that own their home is high (-0.85). This will be further investigated using a variance inflation factor test (table 6). This VIF-test indicates that there is an inflated variance when the percentage of home ownership is included. Running the VIF-test again, now excluding home ownership, shows an improvement.

	VIF	VIF*
PERCENTAGE NO CAR	4.98	2.48
PERCENTAGE HOMEOWNERSHIP	5.63	-

Table 6 VIF-test results of multicollinearity suspected variables

6h Distribution of observation over local areas

The coefficients of the dummies for Bolton and Wigan cannot be accurately estimated due to low numbers of observations and these observations will be omitted.

Table 2 Observations by local authority

TOWN	FREQ.	PERCENT
BURY	22,796	11.02
MANCHESTER	76,092	36.77
OLDHAM	25,728	12.43
ROCHDALE	10,453	5.05
SALFORD	16,815	8.13
STOCKPORT	2,831	1.37
TAMESIDE	14,779	7.14
TRAFFORD	37,446	18.09
WIGAN	0 (+2)	0
BOLTON	0 (+9)	0
TOTAL	206,942 (+11)	100

6i Correlation Matrix

	LN(REAL PRICE)	PERC UNEMPLOYMENT	PERC NO CAR	PERCENTAGE HOMEOWNERSHIP	FREEHOLD	DISTANCE
LN(REAL PRICE)	1					
PERCENTAGE UNEMPLOYMENT	-0.55	1				
PERCENTAGE NO CAR	-0.40	0.57	1			
PERCENTAGE HOMEOWNERSHIP	0.20	-0.41	-0.85	1		
FREEHOLD	0.12	-0.00	-0.17	0.19	1	
DISTANCE	-0.06	0.12	-0.13	0.24	0.04	1

Appendix 7 Coefficients of heterogeneous treatment effects

	Treatment Group * After Treatment		Treatment Group * After Treatment * Distance			Treatment Group * After Treatment		Treatment Group * After Treatment * Distance	
	3	**4**	**3**	**4**		**3**	**4**	**3**	**4**
Abraham Moss	0.168	0.0990	-0.000517	-0.000337	Media City UK	**-0.586*****	-0.203	0.000555	0.0000234
Altrincham	0	0	0	0	Milnrow	-0.0924	-0.200	-0.000245	-0.0000542
Anchorage	0	0	0	0	Monsall	0.0906	-0.0872	-0.000896	0.0000244
Ashton Moss	**-5.196****	**-4.237***	**0.0107***	**0.00857***	Moor Road	**-0.482*****	**-0.255***	**0.00180*****	**0.00114*****
Ashton West	-0.562	-0.586	0.00143	0.00136	Navigation Road	0	0	0	0
Ashton-under-Lyne	0.178	0.127	-0.00115	**-0.00114***	New Islington	**-0.422*****	-0.112	**0.00185*****	**0.000842****
Audenshaw	0.0723	-0.0290	-0.000410	-0.000286	Newbold	**-0.275***	**-0.289****	-0.0000999	0.0000646
Bagulay	-0.354	**-0.354***	**0.00130***	**0.00114***	Newhey	**-0.301***	**-0.321****	0.000489	0.000464
Barlow Moor Road	0.0596	0.0749	-0.00000849	-0.0000949	Newton Heath and Moston	-0.103	-0.0402	-0.0000658	-0.000110
Benchill	**0.668*****	**0.497****	**-0.00148****	**-0.00121****	Northern Moor	0.0537	-0.0509	-0.000262	-0.0000858
Besses o' th' Barn	0	0	0	0	Old Trafford	0	0	0	0
Bowker Vale	0	0	0	0	Oldham Central	-2.199	**-2.076***	0.00432	0.00398
Broadway	0	0	0	0	Oldham King Street	**-1.184***	**-0.998***	0.00213	0.00169

Brooklands	0	0	0	0	Oldham Mumps	**-1.697*****	**-1.504*****	**0.00360*****	**0.00317*****
Burton Road	**-0.131***	**-0.164****	0.0000922	0.000158	Peel Hall	0.000285	0.0232	-0.000191	-0.000476
Bury	0	0	0	0	Piccadilly	0	0	0	0
Cemetery Road	-0.134	**-0.186***	0.0000469	0.0000836	Piccadilly Gardens	0	0	0	0
Central Park	0.144	0.574	0.0000943	-0.000917	Pomona	0	0	0	0
Chorlton	-0.0208	0.0230	-0.0000328	-0.000128	Prestwich	0	0	0	0
Clayton Hall	**0.772*****	**0.761*****	**-0.00166****	**-0.00151*****	Queens Road	0.128	0.181	-0.000785	-0.000914
Cornbrook	0	0	0	0	Radcliffe	0	0	0	0
Crossacres	-0.204	-0.127	**0.000835***	0.000590	Robinswood Road	0.195	0.228	-0.000293	-0.000526
Crumpsall	0	0	0	0	Rochdale Railway Station	-0.173	-0.188	-0.000104	-0.000103
Dane Road	0	0	0	0	Rochdale Town Centre	-0.539	-0.626	0.000434	0.000771
Deansgate-Castlefield	0	0	0	0	Roundthorn	0.0931	0.0889	-0.0000368	-0.0000697
Derker	0.0416	-0.0244	-0.0000744	0.000142	Sale	0	0	0	0
Didsbury Village	**-0.133***	**-0.142****	**0.000371***	**0.000293***	Sale Water Park	1.034	1.971	-0.00217	-0.00430
Droylsden	0.183	0.115	-0.000532	-0.000439	Salford Quays	0	0	0	0
East Didsbury	0.0733	-0.0167	-0.000412	-0.000168	Shadowmoss	0.260	0.203	-0.000665	-0.000582
Eccles	0	0	0	0	Shaw and Crompton	**-0.487****	**-0.581*****	0.000745	**0.000959****

Edge Lane	-0.195*	-0.208**	0.000293	0.000289	Shudehill	0	0	0	0
Etihad Campus	-1.416**	-1.098**	0.00353*	0.00302*	South Chadderton	-0.0598	-0.0437	-0.000203	-0.000290
Exchange Quay	0	0	0	0	St. Peter's Square	0	0	0	0
Exchange Square	0.250	0.134	-0.00102	-0.000788	St. Werburgh's Road	0.0764	0.0575	-0.000254	-0.000195
Failsworth	-0.133	-0.145*	0.000136	0.0000635	Stretford	0	0	0	0
Firswood	-0.00140	0.0184	0.000208	0.000143	Timperley	0	0	0	0
Freehold	-0.0882	-0.130	-0.000123	0.0000754	Trafford Bar	0	0	0	0
Harbour City	0	0	0	0	Velopark	-1.424**	-0.980*	0.00317*	0.00236*
Heaton Park	0	0	0	0	Victoria	0	0	0	0
Hollinwood	0.280	0.471	-0.000900	-0.00134	Weaste	0	0	0	0
Holt Town	0.191	-0.0217	0.0000548	0.000728	West Didsbury	-0.204**	-0.256***	0.0000367	0.000108
Kingsway Business Park	-0.327*	-0.395**	0.000522	0.000547	Westwood	-0.243	-0.283*	0.0000778	0.000134
Ladywell	1.741***	1.320**	0	0	Whitefield	0	0	0	0
Langworthy	0	0	0	0	Withington	-0.359*	-0.290	0.00104	0.00110*
Manchester Airport	0	0	0	0	Woodlands Road	0	0	0	0
Market Street	0	0	0	0	Wythenshawe Park	-0.0186	-0.0330	0.000105	-0.00000879
Martinscroft	-0.188	-0.145	0.000121	-0.0000351	Wythenshawe Town Centre	0.484	0.379	-0.000915	-0.000709

					Total number of stops	94	94	94	94
Omitted	36	36	37	37	Not omitted	58	58	57	57
Insignificant	38	36	46	44	Significant	20	22	11	13
					Positive	3	3	9	10
					Negative	17	19	2	3

Bold printed coefficients are significant: *p<0.05, **p<0.01, ***p<0.001
Green cells contain positive coefficients
Red cells contain negative coefficients
White cells with a value of 0 are omitted automatically due to multicollinearity
*Treatment Group * After treatment* are the coefficients for the treatment effect
*Treatment Group * After treatment * Distance* are the coefficients for the interaction between treatment effect and distance
The bold printed numbers 3 & 4 indicate the model specification

Appendix 8 Difference-in-difference model: All coefficients, local and global specifications

	(1)	(2)	(3)	(4)
	Homogeneous Treatment	Homogeneous Treatment	Heterogeneous Treatment	Heterogeneous Treatment
Freehold	0.201***	0.0782***	0.168***	0.0753***
	(0.000)	(0.000)	(0.000)	(0.000)
D	0	0	0	0
	(.)	(.)	(.)	(.)
F	-0.556***	-0.620***	-0.686***	-0.734***
	(0.000)	(0.000)	(0.000)	(0.000)
S	-0.501***	-0.488***	-0.486***	-0.467***
	(0.000)	(0.000)	(0.000)	(0.000)
T	-0.975***	-0.891***	-0.924***	-0.852***
	(0.000)	(0.000)	(0.000)	(0.000)
year	0.0253***	0.0308***	0.0267***	0.0312***
	(0.000)	(0.000)	(0.000)	(0.000)
distance	-0.0000456***	-0.0000148	-0.0000600***	-0.0000420***
	(0.000)	(0.189)	(0.000)	(0.000)
treatgroup	0.173***	0.183***		
	(0.000)	(0.000)		
aftertreat	0.199***	0.0395**	0.234***	0.0626***
	(0.000)	(0.002)	(0.000)	(0.000)

treatgroup # aftertreat	-0.228***	-0.179***		
	(0.000)	(0.000)		
treatgroup # aftertreat # distance_int	0.000358***	0.000225***		
	(0.000)	(0.000)		
Unemployment	-0.123***	-0.114***	-0.114***	-0.0980***
	(0.000)	(0.000)	(0.000)	(0.000)
Car Ownership	-0.00247***	-0.00247***	-0.00276***	-0.00393***
	(0.000)	(0.000)	(0.000)	(0.000)
MANCHESTER		0.107***		0.0816***
		(0.000)		(0.000)
OLDHAM		-0.285***		-0.260***
		(0.000)		(0.000)
ROCHDALE		-0.231***		-0.225***
		(0.000)		(0.000)
SALFORD		-0.0462***		0.0237***
		(0.000)		(0.000)
STOCKPORT		0.434***		0.420***
		(0.000)		(0.000)
TAMESIDE		-0.199***		-0.144***
		(0.000)		(0.000)
TRAFFORD		0.463***		0.500***
		(0.000)		(0.000)

Abraham Moss	-0.497***	-0.537***
	(0.000)	(0.000)
Altrincham	0.504***	0.145*
	(0.000)	(0.032)
Anchorage	0.292***	0.456***
	(0.000)	(0.000)
Ashton Moss	0.658	1.038
	(0.299)	(0.081)
Ashton West	-0.631**	-0.478*
	(0.004)	(0.020)
Ashton-under-Lyne	-0.606***	-0.489***
	(0.000)	(0.000)
Audenshaw	-0.0638	0.124*
	(0.328)	(0.044)
Bagulay	-0.269*	-0.331**
	(0.015)	(0.002)

Barlow Moor Road		0.220**	0.157*
		(0.004)	(0.028)
Benchill		-1.019***	-1.025***
		(0.000)	(0.000)
Besses o' th' Barn		-0.237***	-0.101*
		(0.000)	(0.021)
Bowker Vale		-0.315***	-0.202***
		(0.000)	(0.000)
Broadway		1.593***	1.744***
		(0.000)	(0.000)
Brooklands		0.217***	-0.0773
		(0.000)	(0.164)
Burton Road		0.935***	0.867***
		(0.000)	(0.000)
Bury		0.190	0.372***
		(0.111)	(0.001)

Cemetery Road		-0.335***	-0.161***
		(0.000)	(0.000)
Central Park		0.201	0.132
		(0.604)	(0.716)
Chorlton		0.841***	0.779***
		(0.000)	(0.000)
Clayton Hall		-0.783***	-0.873***
		(0.000)	(0.000)
Cornbrook		0.0729	0.174**
		(0.206)	(0.001)
Crossacres		0.0317	-0.0321
		(0.658)	(0.633)
Crumpsall		-0.508***	-0.418***
		(0.000)	(0.000)
Dane Road		0.225***	-0.0996*
		(0.000)	(0.031)

Deansgate-Castlefield		0.371***	0.456***
		(0.000)	(0.000)
Derker		-0.507***	-0.315***
		(0.000)	(0.000)
Didsbury Village		0.878***	0.758***
		(0.000)	(0.000)
Droylsden		-0.258**	-0.158*
		(0.001)	(0.038)
East Didsbury		0.122	0.116
		(0.099)	(0.093)
Eccles		-0.596***	-0.500***
		(0.000)	(0.000)
Edge Lane		-0.109*	-0.00724
		(0.013)	(0.862)
Etihad Campus		0.496	0.508*
		(0.063)	(0.042)

Exchange Quay		-0.0110	0.144**
		(0.832)	(0.003)
Exchange Square		0.948***	0.913***
		(0.000)	(0.000)
Failsworth		-0.183***	0.0451
		(0.000)	(0.225)
Firswood		0.302***	-0.232***
		(0.000)	(0.000)
Freehold		-0.203***	-0.0254
		(0.001)	(0.644)
Harbour City		-0.186*	-0.0770
		(0.023)	(0.316)
Heaton Park		0.0422	0.110**
		(0.337)	(0.008)
Hollinwood		-0.403	-0.238
		(0.077)	(0.266)

Holt Town	-0.164	-0.235
	(0.314)	(0.124)
Kingsway Business Park	-0.642***	-0.356***
	(0.000)	(0.000)
Ladywell	-2.034***	-2.020***
	(0.000)	(0.000)
Langworthy	-0.453***	-0.308***
	(0.000)	(0.000)
Manchester Airport=0	0	0
	(.)	(.)
Market Street	0.180**	0.263***
	(0.009)	(0.000)
Martinscroft	-0.161*	-0.240***
	(0.029)	(0.001)
Media City UK	0.688***	0.680***
	(0.000)	(0.000)
Milnrow	0.151*	0.325***

	(0.045)	(0.000)
Monsall	-0.0784	-0.236
	(0.562)	(0.063)
Moor Road	0.471***	0.402***
	(0.000)	(0.000)
Navigation Road	0.154**	-0.168***
	(0.003)	(0.001)
New Islington	0.434***	0.363***
	(0.000)	(0.000)
Newbold	-0.240***	-0.0588
	(0.000)	(0.313)
Newhey	0.0582	0.189**
	(0.403)	(0.004)
Newton Heath and Moston	0.0140	-0.142*
	(0.849)	(0.040)
Northern Moor	-0.0278	-0.202***
	(0.619)	(0.000)

Old Trafford		0.0639	-0.319***
		(0.525)	(0.001)
Oldham Central		0.517***	0.710***
		(0.000)	(0.000)
Oldham King Street		0.673***	0.911***
		(0.000)	(0.000)
Oldham Mumps		0.601***	0.794***
		(0.000)	(0.000)
Peel Hall		0.0396	-0.0143
		(0.664)	(0.867)
Piccadilly		0.227***	0.304***
		(0.000)	(0.000)
Piccadilly Gardens		0.293***	0.393***
		(0.000)	(0.000)
Pomona		-0.490	-0.788
		(0.363)	(0.119)
Prestwich		-0.210***	-0.0418
		(0.000)	(0.384)

Queens Road		-0.382***	-0.433***
		(0.000)	(0.000)
Radcliffe		-0.382***	-0.324***
		(0.000)	(0.000)
Robinswood Road		-0.112	-0.216*
		(0.253)	(0.018)
Rochdale Railway Station		-0.0837	0.102
		(0.303)	(0.183)
Rochdale Town Centre		0.393**	0.588***
		(0.002)	(0.000)
Roundthorn		-0.492***	-0.570***
		(0.000)	(0.000)
Sale		0.112**	-0.218***
		(0.008)	(0.000)
Sale Water Park		-0.134	-0.737
		(0.910)	(0.506)
Salford Quays		0.200*	0.352***
		(0.017)	(0.000)

Shadowmoss		-0.229*	-0.311**
		(0.031)	(0.002)
Shaw and Cromption		-0.0785	0.169*
		(0.326)	(0.025)
Shudehill		0.100*	0.187***
		(0.035)	(0.000)
South Chadderton		-0.335***	-0.167*
		(0.000)	(0.015)
St. Peter's Square		0.376***	0.478***
		(0.000)	(0.000)
St. Werburgh's Road		0.553***	0.503***
		(0.000)	(0.000)
Stretford		0.0292	-0.341***
		(0.573)	(0.000)
Timperley		0.237***	-0.0685
		(0.000)	(0.147)

Trafford Bar		-0.0442	-0.385***
		(0.534)	(0.000)
Velopark		1.631***	1.640***
		(0.000)	(0.000)
Victoria		0.167*	0.242**
		(0.035)	(0.001)
Weaste		-0.605***	-0.468***
		(0.000)	(0.000)
West Didsbury		0.858***	0.779***
		(0.000)	(0.000)
Westwood		-0.224*	-0.0302
		(0.013)	(0.721)
Whitefield		0.0726	0.234***
		(0.218)	(0.000)
Withington		0.873***	0.798***
		(0.000)	(0.000)
Woodlands Road		-0.172*	-0.102
		(0.011)	(0.105)
Wythenshawe Park		0.0904	0.0594

	(0.106)	(0.258)
Wythenshawe Town Centre	-0.339	-0.410*
	(0.057)	(0.014)
Abraham Moss # aftertreat	-0.145	-0.0105
	(0.244)	(0.929)
Ashton Moss # aftertreat	-5.399*	-4.902*
	(0.011)	(0.014)
Ashton West # aftertreat	-1.167*	-0.985
	(0.044)	(0.070)
Ashton-under-Lyne # aftertreat	-0.141	-0.0151
	(0.639)	(0.957)
Audenshaw # aftertreat	-0.205	-0.0911
	(0.096)	(0.431)
Bagulay # aftertreat	-0.616**	-0.498**
	(0.002)	(0.008)
Barlow Moor Road # aftertreat	-0.138	-0.0222
	(0.406)	(0.887)

Benchill # aftertreat		0.501*	0.582**
		(0.020)	(0.004)
Burton Road # aftertreat		-0.470***	-0.342***
		(0.000)	(0.000)
Cemetery Road # aftertreat		-0.436***	-0.310***
		(0.000)	(0.001)
Central Park # aftertreat		-0.186	0.0983
		(0.791)	(0.881)
Chorlton # aftertreat		-0.261**	-0.129
		(0.007)	(0.154)
Clayton Hall # aftertreat		0.479*	0.590**
		(0.028)	(0.004)
Crossacres # aftertreat		-0.352**	-0.240*
		(0.005)	(0.043)
Derker # aftertreat		-0.287	-0.168
		(0.074)	(0.264)

Didsbury Village # aftertreat	-0.392***	-0.265***
	(0.000)	(0.000)
Droylsden # aftertreat	-0.150	-0.0183
	(0.287)	(0.890)
East Didsbury # aftertreat	-0.204	-0.182
	(0.088)	(0.104)
Edge Lane # aftertreat	-0.532***	-0.397***
	(0.000)	(0.000)
Etihad Campus # aftertreat	-1.704**	-1.497**
	(0.001)	(0.003)
Exchange Square # aftertreat	-0.0753	0.0417
	(0.773)	(0.865)
Failsworth # aftertreat	-0.426***	-0.294***
	(0.000)	(0.000)
Firswood # aftertreat	-0.306**	-0.152
	(0.004)	(0.128)

Freehold # aftertreat		-0.301**	-0.186
		(0.005)	(0.064)
Hollinwood # aftertreat		0.117	0.295
		(0.844)	(0.597)
Holt Town # aftertreat		-0.305	-0.182
		(0.342)	(0.546)
Kingsway Business Park # aftertreat		-0.605***	-0.501**
		(0.000)	(0.001)
Ladywell # aftertreat		1.573**	1.697***
		(0.003)	(0.001)
Martinscroft # aftertreat		-0.416**	-0.289*
		(0.008)	(0.050)
Media City UK # aftertreat		-0.791***	-0.655***
		(0.000)	(0.000)
Milnrow # aftertreat		-0.385**	-0.260*
		(0.005)	(0.044)

Monsall # aftertreat		-0.247	-0.155
		(0.429)	(0.598)
Moor Road # aftertreat		-0.605***	-0.479***
		(0.000)	(0.000)
New Islington # aftertreat		-0.559***	-0.412***
		(0.000)	(0.000)
Newbold # aftertreat		-0.545***	-0.422***
		(0.000)	(0.000)
Newhey # aftertreat		-0.586***	-0.435***
		(0.000)	(0.001)
Newton Heath and Moston # aftertreat		-0.363**	-0.196
		(0.006)	(0.112)
Northern Moor # aftertreat		-0.183	-0.106
		(0.095)	(0.304)
Oldham Central # aftertreat		-2.670*	-2.566*
		(0.027)	(0.024)

Oldham King Street # aftertreat	-1.174	-1.067
	(0.058)	(0.067)
Oldham Mumps # aftertreat	-2.151***	-1.909***
	(0.000)	(0.000)
Peel Hall # aftertreat	-0.210	-0.0871
	(0.248)	(0.610)
Queens Road # aftertreat	-0.207	-0.0984
	(0.287)	(0.589)
Robinswood Road # aftertreat	-0.237	-0.0836
	(0.198)	(0.629)
Rochdale Railway Station # aftertreat	-0.426*	-0.314
	(0.031)	(0.091)
Rochdale Town Centre # aftertreat	-0.785	-0.687
	(0.079)	(0.101)
Roundthorn # aftertreat	-0.163	-0.0201
	(0.530)	(0.934)

Sale Water Park # aftertreat		1.031	1.262
		(0.735)	(0.658)
Shadowmoss # aftertreat		0.0232	0.111
		(0.917)	(0.593)
Shaw and Cromption # aftertreat		-0.735***	-0.638***
		(0.000)	(0.000)
South Chadderton # aftertreat		-0.299	-0.181
		(0.051)	(0.207)
St. Werburgh's Road # aftertreat		-0.243*	-0.113
		(0.013)	(0.221)
Velopark # aftertreat		-1.656**	-1.579**
		(0.004)	(0.003)
West Didsbury # aftertreat		-0.457***	-0.340***
		(0.000)	(0.000)
Westwood # aftertreat		-0.637***	-0.518**
		(0.000)	(0.002)
Withington # aftertreat		-0.522**	-0.387*
		(0.006)	(0.031)

Wythenshawe Park # aftertreat		-0.295**	-0.154
		(0.006)	(0.128)
Wythenshawe Town Centre # aftertreat		0.169	0.286
		(0.626)	(0.378)
Abraham Moss # aftertreat # distance_int		-0.000416	-0.000421
		(0.233)	(0.199)
Ashton Moss # aftertreat # distance_int		0.0106*	0.00985*
		(0.021)	(0.022)
Ashton West # aftertreat # distance_int		0.00211	0.00201
		(0.129)	(0.124)
Ashton-under-Lyne # aftertreat # distance_int		-0.000868	-0.000848
		(0.234)	(0.215)
Audenshaw # aftertreat # distance_int		-0.000402	-0.000343
		(0.247)	(0.292)

Bagulay # aftertreat # distance_int		0.00135*	0.00139*
		(0.042)	(0.026)
Barlow Moor Road # aftertreat # distance_int		-0.000139	-0.0000912
		(0.755)	(0.827)
Benchill # aftertreat # distance_int		-0.00173**	-0.00162**
		(0.002)	(0.002)
Burton Road # aftertreat # distance_int		0.000543*	0.000550**
		(0.011)	(0.006)
Cemetery Road # aftertreat # distance_int		0.000116	0.000132
		(0.697)	(0.636)
Central Park # aftertreat # distance_int		0.000207	-0.000130
		(0.901)	(0.934)
Chorlton # aftertreat # distance_int		-0.0000357	-0.0000403
		(0.898)	(0.878)
Clayton Hall # aftertreat # distance_int		-0.00163**	-0.00155**
		(0.005)	(0.004)

Crossacres # aftertreat # distance_int		0.000361	0.000456
		(0.420)	(0.278)
Derker # aftertreat # distance_int		-0.000282	-0.000146
		(0.551)	(0.742)
Didsbury Village # aftertreat # distance_int		0.000361	0.000391*
		(0.054)	(0.026)
Droylsden # aftertreat # distance_int		-0.000401	-0.000374
		(0.298)	(0.302)
East Didsbury # aftertreat # distance_int		-0.000308	-0.0000284
		(0.336)	(0.925)
Edge Lane # aftertreat # distance_int		0.000502	0.000589*
		(0.097)	(0.038)
Etihad Campus # aftertreat # distance_int		0.00368*	0.00348*
		(0.021)	(0.020)
Exchange Square # aftertreat # distance_int		-0.000787	-0.000725

		(0.215)	(0.224)
Failsworth # aftertreat # distance_int		0.000233	0.000205
		(0.358)	(0.388)
Firswood # aftertreat # distance_int		0.000298	0.000277
		(0.311)	(0.314)
Freehold # aftertreat # distance_int		-0.000213	-0.000131
		(0.489)	(0.650)
Hollinwood # aftertreat # distance_int		-0.00115	-0.00126
		(0.398)	(0.327)
Holt Town # aftertreat # distance_int		0.000687	0.000749
		(0.499)	(0.432)
Kingsway Business Park # aftertreat # distance_int		0.000443	0.000573
		(0.344)	(0.192)
Martinscroft # aftertreat # distance_int		0.0000128	0.0000300
		(0.977)	(0.944)

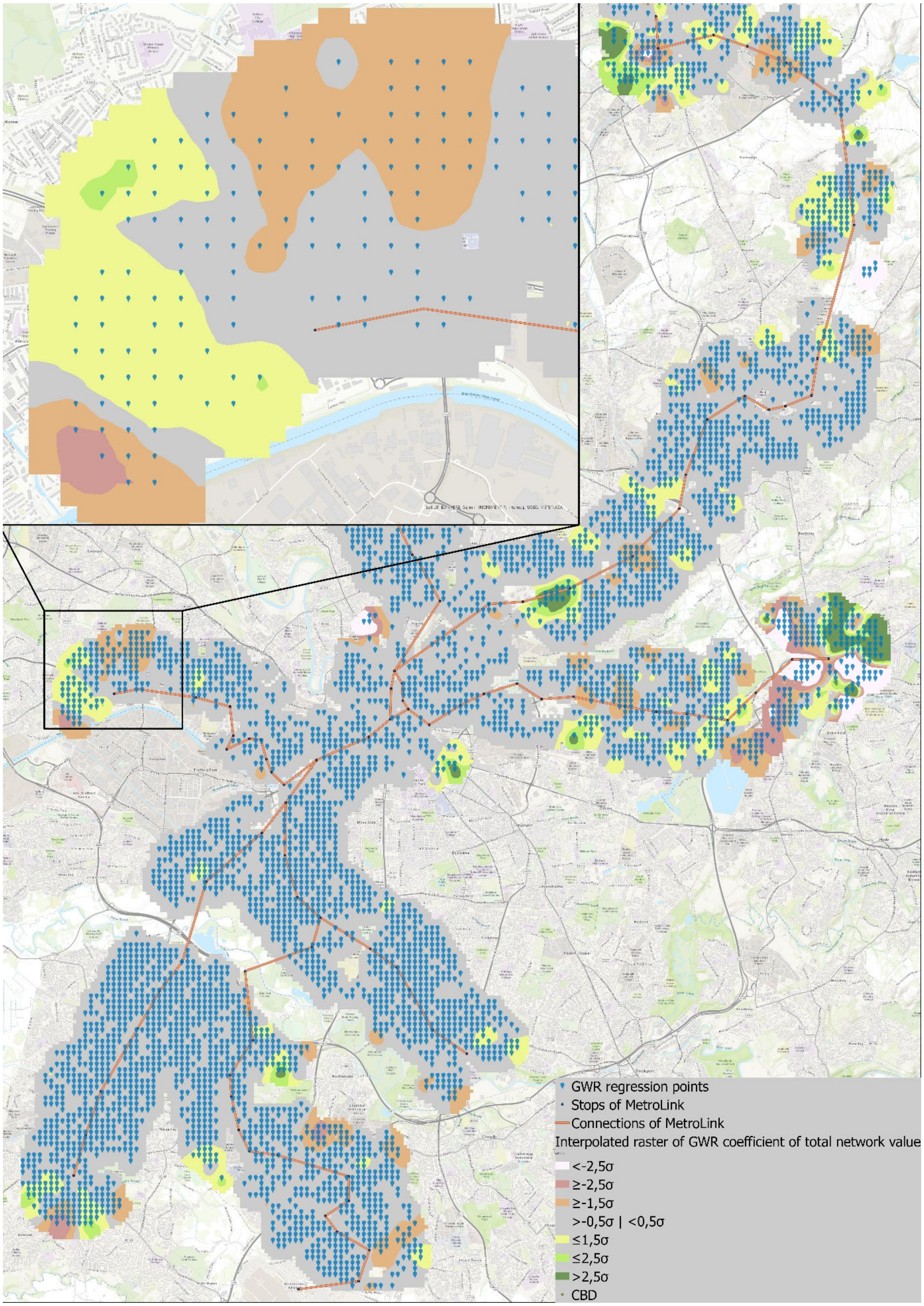
Media City UK # aftertreat # distance_int		0.000536	0.000561
		(0.295)	(0.243)
Milnrow # aftertreat # distance_int		-0.000230	-0.000179
		(0.534)	(0.606)
Monsall # aftertreat # distance_int		-0.000627	-0.000465
		(0.562)	(0.647)
Moor Road # aftertreat # distance_int		0.00132**	0.00133***
		(0.002)	(0.001)
New Islington # aftertreat # distance_int		0.00130**	0.00126***
		(0.001)	(0.001)
Newbold # aftertreat # distance_int		-0.0000250	0.0000457
		(0.943)	(0.888)
Newhey # aftertreat # distance_int		0.000465	0.000444
		(0.266)	(0.257)
Newton Heath and Moston # aftertreat # distance_int		0.0000895	0.0000337
		(0.828)	(0.931)
Northern Moor # aftertreat # distance_int		-0.000284	-0.000148
		(0.351)	(0.606)

Oldham Central # aftertreat # distance_int	0.00480	0.00479
	(0.078)	(0.060)
Oldham King Street # aftertreat # distance_int	0.00157	0.00167
	(0.294)	(0.236)
Oldham Mumps # aftertreat # distance_int	0.00424***	0.00395***
	(0.000)	(0.000)
Peel Hall # aftertreat # distance_int	-0.000411	-0.000370
	(0.408)	(0.427)
Queens Road # aftertreat # distance_int	-0.000530	-0.000379
	(0.406)	(0.527)
Robinswood Road # aftertreat # distance_int	0.000224	0.000159
	(0.675)	(0.752)
Rochdale Railway Station # aftertreat # distance_int	-0.000309	-0.000247
	(0.567)	(0.625)

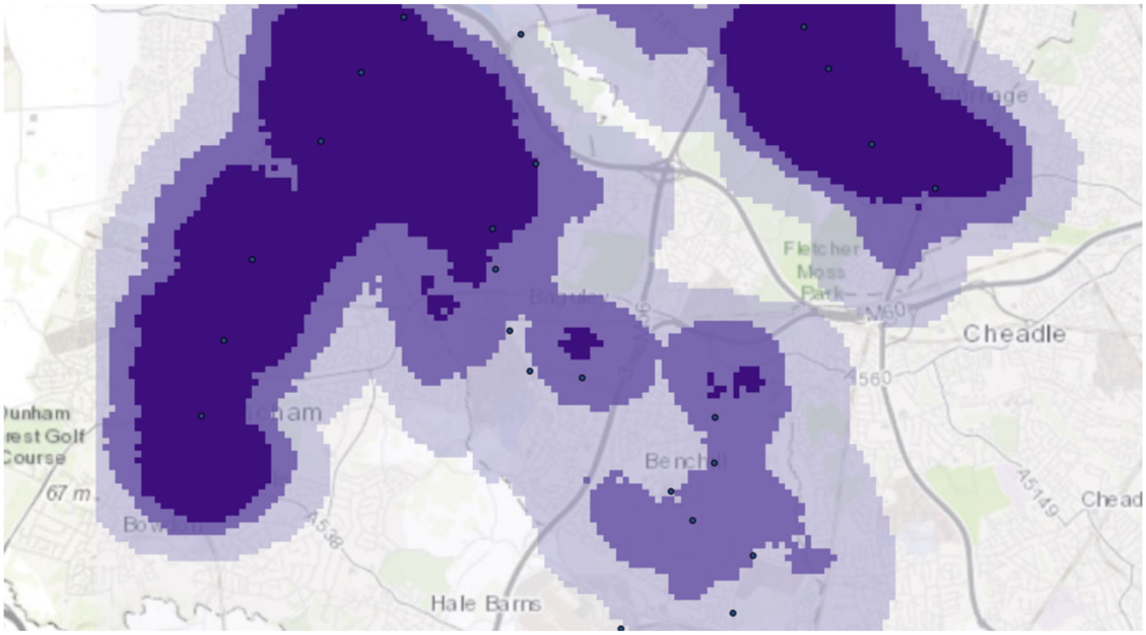
Rochdale Town Centre # aftertreat # distance_int		0.000136	0.000217
		(0.907)	(0.843)
Roundthorn # aftertreat # distance_int		-0.0000638	-0.0000816
		(0.943)	(0.922)
Sale Water Park # aftertreat # distance_int		-0.00264	-0.00282
		(0.702)	(0.663)
Shadowmoss # aftertreat # distance_int		-0.000630	-0.000581
		(0.351)	(0.360)
Shaw and Crompton # aftertreat # distance_int		0.000661	0.000762
		(0.132)	(0.064)
South Chadderton # aftertreat # distance_int		-0.000363	-0.000299
		(0.395)	(0.455)
St. Werburgh's Road # aftertreat # distance_int		-0.000108	-0.0000984
		(0.700)	(0.707)
Velopark # aftertreat # distance_int		0.00312*	0.00329*
		(0.034)	(0.017)

West Didsbury # aftertreat # distance_int			-8.94e-08	0.0000553
			(1.000)	(0.801)
Westwood # aftertreat # distance_int			0.000276	0.000341
			(0.598)	(0.488)
Withington # aftertreat # distance_int			0.00121	0.00121
			(0.067)	(0.051)
Wythenshawe Park # aftertreat # distance_int			0.0000346	0.0000123
			(0.911)	(0.966)
Wythenshawe Town Centre # aftertreat # distance_int			-0.000700	-0.000652
			(0.486)	(0.489)
Constant	-38.97***	-49.93***	-41.79***	-50.70***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	206940	206940	206940	206940
Adjusted R-squared	0.299	0.398	0.403	0.475
p-values in parentheses				
=" p<0.05*	** p<0.01	*** p<0.001"		

Appendix 9 Locations of regression points (interpolated raster of total network value as background)



Appendix 10 Density of observation in GWR sampled data set. Observing no correlating pattern of low observation in discussed areas



Appendix 11 Regression STATA Do-file

More DO-files and the workflow are available on

https://drive.google.com/drive/folders/1Jua_r6GNtdmxQ3-Lm9HCiCYrEu7giURl?usp=sharing

```
clear all
//some working room
set maxvar 10000
set matsize 10000
use "F:\Thesis\Data mergers\9. PP+stopname+patronage"


//generate dummies for heterogeneous treatment effects
tabulate stopname, generate(ddist)

replace ddist1  =      0 if distance>500     & stopname =="Abraham Moss"
replace ddist2  =      0 if distance>500     & stopname =="Altrincham"
replace ddist3  =      0 if distance>500     & stopname =="Anchorage"
replace ddist4  =      0 if distance>500     & stopname =="Ashton Moss"
replace ddist5  =      0 if distance>500     & stopname =="Ashton West"
replace ddist6  =      0 if distance>500     & stopname =="Ashton-under-Lyne"
replace ddist7  =      0 if distance>500     & stopname =="Audenshaw"
replace ddist8  =      0 if distance>500     & stopname =="Bagulay"
replace ddist9  =      0 if distance>500     & stopname =="Barlow Moor Road"
replace ddist10 =      0 if distance>500     & stopname =="Benchill"
replace ddist11 =      0 if distance>500     & stopname =="Besses o' th' Barn"
replace ddist12 =      0 if distance>500     & stopname =="Bowker Vale"
replace ddist13 =      0 if distance>500     & stopname =="Broadway"
replace ddist14 =      0 if distance>500     & stopname =="Brooklands"
replace ddist15 =      0 if distance>500     & stopname =="Burton Road"
replace ddist16 =      0 if distance>500     & stopname =="Bury"
replace ddist17 =      0 if distance>500     & stopname =="Cemetery Road"
replace ddist18 =      0 if distance>500     & stopname =="Central Park"
replace ddist19 =      0 if distance>500     & stopname =="Chorlton"
replace ddist20 =      0 if distance>500     & stopname =="Clayton Hall"
replace ddist21 =      0 if distance>500     & stopname =="Cornbrook"
replace ddist22 =      0 if distance>500     & stopname =="Crossacres"
replace ddist23 =      0 if distance>500     & stopname =="Crumpsall"
replace ddist24 =      0 if distance>500     & stopname =="Dane Road"
replace ddist25 =      0 if distance>500     & stopname =="Deansgate-Castlefield"
replace ddist26 =      0 if distance>500     & stopname =="Derker"
replace ddist27 =      0 if distance>500     & stopname =="Didsbury Village"
replace ddist28 =      0 if distance>500     & stopname =="Droylsden"
replace ddist29 =      0 if distance>500     & stopname =="East Didsbury"
replace ddist30 =      0 if distance>500     & stopname =="Eccles"
replace ddist31 =      0 if distance>500     & stopname =="Edge Lane"
replace ddist32 =      0 if distance>500     & stopname =="Etihad Campus"
replace ddist33 =      0 if distance>500     & stopname =="Exchange Quay"
replace ddist34 =      0 if distance>500     & stopname =="Exchange Square"
replace ddist35 =      0 if distance>500     & stopname =="Failsworth"
```

```
replace ddist36 =      0 if distance>500     & stopname =="Firswood"
replace ddist37 =      0 if distance>500     & stopname =="Freehold"
replace ddist38 =      0 if distance>500     & stopname =="Harbour City"
replace ddist39 =      0 if distance>500     & stopname =="Heaton Park"
replace ddist40 =      0 if distance>500     & stopname =="Hollinwood"
replace ddist41 =      0 if distance>500     & stopname =="Holt Town"
replace ddist42 =      0 if distance>500     & stopname =="Kingsway Business Park"
replace ddist43 =      0 if distance>500     & stopname =="Ladywell"
replace ddist44 =      0 if distance>500     & stopname =="Langworthy"
replace ddist45 =      0 if distance>500     & stopname =="Manchester Airport"
replace ddist46 =      0 if distance>500     & stopname =="Market Street"
replace ddist47 =      0 if distance>500     & stopname =="Martinscroft"
replace ddist48 =      0 if distance>500     & stopname =="Media City UK"
replace ddist49 =      0 if distance>500     & stopname =="Milnrow"
replace ddist50 =      0 if distance>500     & stopname =="Monsall"
replace ddist51 =      0 if distance>500     & stopname =="Moor Road"
replace ddist52 =      0 if distance>500     & stopname =="Navigation Road"
replace ddist53 =      0 if distance>500     & stopname =="New Islington"
replace ddist54 =      0 if distance>500     & stopname =="Newbold"
replace ddist55 =      0 if distance>500     & stopname =="Newhey"
replace ddist56 =      0 if distance>500     & stopname =="Newton Heath and Moston"
replace ddist57 =      0 if distance>500     & stopname =="Northern Moor"
replace ddist58 =      0 if distance>500     & stopname =="Old Trafford"
replace ddist59 =      0 if distance>500     & stopname =="Oldham Central"
replace ddist60 =      0 if distance>500     & stopname =="Oldham King Street"
replace ddist61 =      0 if distance>500     & stopname =="Oldham Mumps"
replace ddist62 =      0 if distance>500     & stopname =="Peel Hall"
replace ddist63 =      0 if distance>500     & stopname =="Piccadilly"
replace ddist64 =      0 if distance>500     & stopname =="Piccadilly Gardens"
replace ddist65 =      0 if distance>500     & stopname =="Pomona"
replace ddist66 =      0 if distance>500     & stopname =="Prestwich"
replace ddist67 =      0 if distance>500     & stopname =="Queens Road"
replace ddist68 =      0 if distance>500     & stopname =="Radcliffe"
replace ddist69 =      0 if distance>500     & stopname =="Robinswood Road"
replace ddist70 =      0 if distance>500     & stopname =="Rochdale Railway Station"
replace ddist71 =      0 if distance>500     & stopname =="Rochdale Town Centre"
replace ddist72 =      0 if distance>500     & stopname =="Roundthorn"
replace ddist73 =      0 if distance>500     & stopname =="Sale"
replace ddist74 =      0 if distance>500     & stopname =="Sale Water Park"
replace ddist75 =      0 if distance>500     & stopname =="Salford Quays"
replace ddist76 =      0 if distance>500     & stopname =="Shadowmoss"
replace ddist77 =      0 if distance>500     & stopname =="Shaw and Crompton"
replace ddist78 =      0 if distance>500     & stopname =="Shudehill"
replace ddist79 =      0 if distance>500     & stopname =="South Chadderton"
replace ddist80 =      0 if distance>500     & stopname =="St. Peter's Square"
replace ddist81 =      0 if distance>500     & stopname =="St. Werburgh's Road"
replace ddist82 =      0 if distance>500     & stopname =="Stretford"
replace ddist83 =      0 if distance>500     & stopname =="Timperley"
replace ddist84 =      0 if distance>500     & stopname =="Trafford Bar"
replace ddist85 =      0 if distance>500     & stopname =="Velopark"
replace ddist86 =      0 if distance>500     & stopname =="Victoria"
```

```
	replace ddist87 =	0 if distance>500	& stopname =="Weaste"
	replace ddist88 =	0 if distance>500	& stopname =="West Didsbury"
	replace ddist89 =	0 if distance>500	& stopname =="Westwood"
	replace ddist90 =	0 if distance>500	& stopname =="Whitefield"
	replace ddist91 =	0 if distance>500	& stopname =="Withington"
	replace ddist92 =	0 if distance>500	& stopname =="Woodlands Road"
	replace ddist93 =	0 if distance>500	& stopname =="Wythenshawe Park"
	replace ddist94 =	0 if distance>500	& stopname =="Wythenshawe Town Centre"

//outliers
	drop if price<10000
	drop if price>2000000

	//gen year dummies
	tabulate year, gen(dyear)

	//encoding variables
	encode housetype, gen(housetype_code)
	encode town, gen(locality_code)

	replace new="1" if new=="Y"
	replace new="0" if new=="N"
	destring new, replace

	rename ownership freehold
	replace freehold="1" if freehold=="F"
	replace freehold="0" if freehold=="L"
	drop if freehold=="no_value"
	destring freehold, replace

	//this is a category of transaction that
	drop if a_b=="B"

	//3828 observations had missing date field for the stop opening
	replace stop_opening_date=19750 if opening_date=="27/1/2014 12:00:00 AM"

	gen aftertreat=0
	replace aftertreat=1 if datedmy>stop_opening_date

	gen treatgroup=0
	replace treatgroup=1 if distance<500

	gen timesincetreat=datedmy-stop_opening_date

	gen distance_int= round(distance)
	recast int distance_int

	//new houses are usually not considered as traded at market value
	drop if new==1
	drop if distance>1200
	drop if town=="BOLTON" | town=="WIGAN"
```

```
**some histograms and tables
	//hist price
	//hist ln_defl_pr, normal
	//estpost sum price
	//esttab using "F:\Thesis\Estimation results\Other tables\sumprice.csv", replace
cells("mean(fmt(2)) sd(fmt(2)) min(fmt(2)) max(fmt(2))") nomtitle
	//tabulate stopname, matcell(freq)
	//esttab using "F:\Thesis\Estimation results\Other tables\obsperstop.csv", replace
cells("freq(fmt(2)) stopname(fmt(2))") nomtitler
	//tabulate town, matcell(freq)

** does the estimation effect change with a kind of time lag since construction? NO
	//keep if aftertreat==0 | timesincetreat>1000
//regress ln_defl_pr freehold i.housetype_code year treatgroup##aftertreat
treatgroup#aftertreat#c.datediff //Area FE:  i.locality_code	|	Time FE in years:
	dyear*

	pause
	//prepare for ~20 minutes calculating time
	//regress ln_defl_pr c.(after_s7) c.(ddist1-ddist94) c.(dyear1-dyear20) c.(ddist1-
ddist94)# c.(dyear1-dyear20)
	//eststo
	//test c.(after_s7) c.(ddist1-ddist94) c.(dyear1-dyear20) c.(ddist1-ddist94)#
c.(dyear1-dyear20)

//keep if new_stop==1
//drop r
regress ln_defl_pr freehold i.housetype_code year c.distance_int treatgroup##aftertreat
treatgroup##aftertreat#c.distance_int punemployment pnocar//Area FE:	i.locality_code
	|	Time FE in years:	dyear*
eststo
regress ln_defl_pr freehold i.housetype_code year i.locality_code c.distance_int
treatgroup##aftertreat treatgroup##aftertreat#c.distance_int punemployment pnocar//Area FE:
	i.locality_code	|	Time FE in years:	dyear*
eststo
regress ln_defl_pr freehold i.housetype_code year c.distance_int ddist*##aftertreat
ddist*##aftertreat#c.distance_int punemployment pnocar//Area FE:   i.locality_code	|
	Time FE in years:	dyear*
eststo
regress ln_defl_pr freehold i.housetype_code year i.locality_code c.distance_int
ddist*##aftertreat ddist*##aftertreat#c.distance_int punemployment pnocar//Area FE:
	i.locality_code	|	Time FE in years:	dyear*
eststo
esttab using "F:\Thesis\Estimation results\bookniceDID.csv", replace p ar2 label

** No difference in coefficients by rounding of distances
//gen distancein100= round(distance_int, 100)/100
//regress ln_defl_pr freehold i.housetype_code year i.locality_code c.distance_int
ddist*##aftertreat ddist*##aftertreat#c.distance_int //Area FE:    i.locality_code	|
	Time FE in years:	dyear*
```

```
//eststo
//esttab using "F:\Thesis\Estimation results\bookniceDIDroundeddistance.csv", replace p ar2
label

testparm ddist*#aftertreat, equal //wald test are coefficients equal? NO!
eststo
testparm ddist*#aftertreat#c.distance_int, equal
eststo
testparm ddist4#aftertreat#c.distance_int    ddist8#aftertreat#c.distance_int
        ddist10#aftertreat#c.distance_int    ddist15#aftertreat#c.distance_int
        ddist20#aftertreat#c.distance_int    ddist31#aftertreat#c.distance_int
        ddist51#aftertreat#c.distance_int    ddist53#aftertreat#c.distance_int
        ddist61#aftertreat#c.distance_int    ddist85#aftertreat#c.distance_int
        ddist27#aftertreat#c.distance_int ddist32#aftertreat#c.distance_int
ddist53#aftertreat#c.distance_int, equal
eststo
testparm ddist1#aftertreat    ddist4#aftertreat     ddist8#aftertreat     ddist9#aftertreat
        ddist10#aftertreat    ddist15#aftertreat    ddist17#aftertreat    ddist20#aftertreat
        ddist22#aftertreat    ddist27#aftertreat    ddist31#aftertreat    ddist32#aftertreat
        ddist36#aftertreat    ddist37#aftertreat    ddist42#aftertreat    ddist47#aftertreat
        ddist48#aftertreat    ddist49#aftertreat    ddist51#aftertreat    ddist53#aftertreat
        ddist54#aftertreat    ddist55#aftertreat    ddist56#aftertreat    ddist59#aftertreat
        ddist61#aftertreat    ddist70#aftertreat    ddist77#aftertreat    ddist81#aftertreat
        ddist85#aftertreat    ddist88#aftertreat    ddist89#aftertreat    ddist91#aftertreat
        ddist93#aftertreat    , equal
esttab using "F:\Thesis\Estimation results\waldtestresults.csv", replace


//diagnostics
predict r, resid
sktest r //skewness and kurtosis are present
kdensity r, normal
pnorm r
qnorm r
rvfplot, yline (0) //residuals have heteroscadasity, decreasing coefficient precision and
lowering the p-value. standard errors ,robust optie in de regressie
eststo
esttab using "F:\Thesis\Estimation results\bookNewStop.csv", replace p ar2 label
//estout using "F:\Thesis\Estimation results\book1.csv", replace

//regress wihtout Area FE and Time FE
regress ln_defl_pr new freehold i.housetype_code after_s* ddist*
eststo esttab using "F:\Thesis\Estimation results\bookzonderFEs.csv", replace p ar2 label
regress ln_defl_pr new freehold i.housetype_code year ddist* aftertreat distance_int
ddist*#aftertreat ddist*#aftertreat#distance_int //Area FE:        i.locality_code        |
        Time FE in years:     dyear*
```