Mind the Gap: Windfall Gains in Housing Values along London's Elizabeth Line

Lawrence Lee

University of Cambridge

Abstract

This dissertation measures the impact of the Elizabeth line on residential property values. Using a difference-in-differences model on 18 years of transaction data within 3 kilometres of all 41 stations, average price uplifts of 3.15%, 5.31% and 3.01% are identified during the line's announcement, construction and post-opening phases respectively. Properties with higher relative prices, closer proximity to stations and near major interchanges experience substantially larger effects. With a lower-bound estimate of £3.08 billion in aggregate windfall gains — equivalent to 16.3% of total construction costs — the findings support greater use of land value capture mechanisms to help fund future infrastructure investments.

Acknowledgements

I would like to thank my supervisor, Professor Helen Bao, for all her invaluable guidance in writing this dissertation.

I am also very grateful to my Director of Studies, Professor Andreas Kontoleon, for his mentorship these past three years.

Lastly, I want to thank my friends and family for their constant encouragement and support.

"Everything is related to everything else, but near things are more related than distant things."
— Tobler's First Law of Geography

Contents

1 Introduction

Transport infrastructure remains a pillar of cities. Beyond reducing travel times for its own and competing users, public transport helps boost total factor productivity (Anas & Chang, 2023), relieve congestion (Anderson, 2013), reduce car-related accidents and injuries (Bauernschuster et al., 2017), enhance health outcomes (Sener et al., 2016) and improve air quality at certain sites (Ma et al., 2021). One widely-studied outcome of transport infrastructure expansion is an increase in the values of properties near affected stations — a phenomenon referred to as land value uplift (LVU). The well-established socioeconomic and environmental benefits present clear grounds for public investment into improving inter- and intra-city connectivity; however, limited public funds naturally constrain the number of transport projects ultimately executed.

The existence of positive capitalisation effects from rail transit systems is very well documented in research covering many cities worldwide in the past five decades. Since 1970, there have been over 200 studies exploring transit-induced LVU (Rennert, 2022), with over 60 studies conducted in North America alone (Higgins & Kanaroglou, 2016). This has important implications for capturing these gains to finance ongoing or future projects via land value capture (LVC). Much of the empirical literature on transit-induced LVU remains dominated by US-focused studies, with a growing number in East Asia, particularly China. In contrast, UK-centric research on this subject is relatively limited in scope and scale.

One of the most notable examples of transport innovations in recent years is the Elizabeth line — a major project aimed at improving connectivity, reducing travel times and addressing regional inequality. The opening of the Elizabeth line presents a unique opportunity to study LVU in a British context. Using a high-resolution dataset larger than most studies globally and advanced econometric methods, this dissertation aims to quantify the extent and distribution of windfall gains to residential homeowners. In addition, this study applies a hedonic pricing model to identify the treatment zone and uses a formal test for parallel trends — both techniques employed by relatively few authors in the literature — and introduces a new distance variable, the Manhattan distance, as a potential substitute for network distance.

Applying a difference-in-differences framework reveals that the Elizabeth line has positively affected housing values. Across the entire study boundary, prices rose by 3.15% on average at the announcement phase, 5.31% during construction and 3.01% after opening. Further models highlight that higher-value properties and properties within 0–900m of stations experience even stronger uplift, particularly if the station has zero or at least two connecting lines. This may explain why capitalisation effects are more prominent in Central London. The results satisfy the parallel trends assumption, and are robust to reweighting on observed characteristics and clustered standard errors.

These findings imply total LVU of £3.08 billion for the entire dataset of transactions, or 16.3% of construction costs, with the true figure for all transacted and untransacted properties likely surpassing this estimate, suggesting that LVC mechanisms on residential properties, similar to the Business Rate Supplement applied to commercial properties, would aid in funding essential transport infrastructure. Such a move would shift London closer to the beneficiary pays principle used by Hong Kong, Singapore and other jurisdictions.

The paper is organised as follows. Section 2 reviews the related literature surrounding rail infrastructure expansion on residential property values. Section 3 provides a brief history and overview of the Elizabeth line. Section 4 describes the data and section 5 details the empirical methodology. Section 6 presents the results. Section 7 performs several robustness checks. Section 8 discusses the results and the policy implications. Section 9 concludes.

2 Literature Review

2.1 Transport Network Expansion and Land Value Uplift

Bid-rent theory connects the expansion of transportation infrastructure to LVU. Building upon Von Thünen's (1826/1966) early work on farmland values and accessibility, Alonso (1964), Muth (1969) and Mills (1972) extended this idea to urban land markets in the monocentric city model. Under this framework, the spatial distribution of firms and households in a city is explained by the trade-off between land use and commuting costs, illustrated by the negative relationship between land prices and distance from the central business district known as bid-rent curves. This creates a utility maximisation problem whereby firms and households locate at a distance that balances travel costs and land rents based on their preferences. Since the physical characteristics of land are assumed given, this theory posits that location is the primary factor determining land values. As transport infrastructure reduces commuting costs, in terms of both time and money, improved accessibility should lead to increased bidding between users for properties close to stations, resulting in LVU and a localised, negatively sloped bid-rent curve away from stations.

The majority of studies have found positive LVU (Mohammad et al., 2013; Rennert, 2022). Contemporaneously, negative effects (Landis et al., 1995; Nelson, 1992; Wagner et al., 2017) and insignificant results (Du & Mulley, 2007; Gatzlaff & Smith, 1993) have also been identified in a few cases. Furthermore, a plethora of authors have found that LVU manifests long before the official opening of a new line (Bao et al., 2021; Billings, 2011; Damm et al., 1980; McDonald & Osuji, 1995; McMillen & McDonald, 2004; Yen et al., 2018).

The first major meta-analysis, by Debrezion et al. (2007), attempted to explain some of this heterogeneity in the magnitude and sign of results, and concluded that LVU is stronger for commercial property for short distances from stations, but accessibility premia persist across longer distances for residential property. In addition, commuter rail transit stations produce stronger uplift than stations pertaining to other transport modes. Mohammad et al. (2013) build upon this foundation by incorporating more contextual and methodological factors, and extend the geographical scope beyond North America by including studies in Europe and Asia. Their results confirm the key findings by Debrezion et al., and indicate stronger LVU effects in Europe and East Asia compared to North America, possibly due to greater reliance on public transit systems in these regions.

Most recently, Rennert (2022) separates these two factors by controlling for public transport expenditure share, concluding that more expensive transit services, relative to income, reduce accessibility benefits. Moreover, larger transit networks generate marginally higher effects, and results from Europe, Asia and Oceania are statistically similar to North America, whereas East Asia produces lower results.

Furthermore, Bowes and Ihlanfeldt (2001) theorise that LVU could be the effect of positive benefits such as accessibility and retail development net of negative externalities and crime, particularly for properties situated closest to stations. Capitalisation effects may also be dependent on the project's success; for example, the study by Wagner et al. (2017) was distinct in that the light rail transit system examined was among the worst performing transit lines in the US at the time in terms of ridership, profitability and relative accessibility benefits, which contextualises the land value depreciation of 7.8%. Akin to the Miami Metrorail (Gatzlaff & Smith, 1993), limited rail usage in Greater Manchester may have dampened any significant LVU associated with the Metrolink (Forrest et al., 1996). In Sunderland, Du and Mulley (2007) suggest that weak economic conditions, coupled with a limited data timeframe, likely contributed to the absence of significant uplift.

Beyond station-specific variations in effects, a few studies have discovered network effects of various forms — whether through increased connectivity at a former terminus station gaining integration (Higgins, 2019), enhanced accessibility benefits extending beyond immediate catchment areas (He, 2020) or uplift along pre-existing lines at interchange stations (Zhu & Diao, 2024). On balance, the magnitude of LVU appears to reflect a complex interplay of dynamic and spatially-varying factors. Two main categories of methods seek to identify these effects: hedonic pricing models and difference-in-differences models.

2.2 Hedonic Analysis

Hedonic pricing theory, originating from the seminal work of Rosen (1974), has long been a popular method in the applied economics literature due to its wide applicability. By incorporating housing and neighbourhood controls along with an accessibility variable, researchers can estimate LVU using cross-sectional or before-and-after ordinary least squares (OLS) regressions. One of the earliest reported instances of transport-induced LVU in the UK comes from the South Yorkshire Supertram in Sheffield, though the effect size was small (Henneberry, 1998).

However, because unobserved factors may simultaneously influence both property values and accessibility, OLS suffers from endogeneity. An example of omitted variable bias in early studies involves the omission of competing accessibility factors, such as distance to highways, leading to an overestimation of proximity effects (Debrezion et al., 2007; Mohammad et al., 2013). Omitting local crime statistics among other neighbourhood amenities could also bias estimates (Bowes & Ihlanfeldt, 2001).

Repeat-sales models control for endogeneity by removing biases caused by time-invariant omitted variables. Although this requires lower-resolution data, it introduces further problems. Not only is the sample size reduced in the process, but there may also be sample selection biases if properties that are sold at least twice are materially different from those that sell just once (Gatzlaff & Haurin, 1997). Moreover, many housing and neighbourhood characteristics do change over time, even if slowly; therefore, differencing housing prices relies on a flawed assumption of time-invariant attributes that may not adequately address endogeneity.

2.3 Difference in Differences

Difference-in-differences (DID) studies have become increasingly adopted in the transit-induced LVU literature. The essence of DID lies in combining the strengths of before-and-after and with-and-without analyses by testing the pre- and post-intervention changes between treated and control groups following an exogenous shock, whilst controlling for observed and unobserved confounders. Using DID in lieu of a standard OLS regression mitigates concerns of endogeneity, but in turn introduces the crucial new assumption of parallel trends.[1] The ability to establish causality makes DID a powerful tool for impact evaluation, and like OLS, DID's versatility makes it a popular methodology for many fields beyond economics.

Gibbons and Machin (2005) followed the DID method to estimate the effect of new Docklands Light Railway (DLR) and Jubilee line stations on property values. Using data from Nationwide Building Society, the authors restrict their sample to properties within 30 kilometres from Holborn (a proxy for London's central business district), and find that a 1km reduction in distance to the nearest Jubilee line or DLR station increases property values by around 2.1% for properties within 2km of a station. Their paper is one of the closest to this study given its focus on the London residential property market. However, its methodology can be extended by incorporating a treatment–phase interaction term to test for anticipation effects, and by using property-level, rather than postcode-aggregated, data to improve spatial precision.

[1] As DID is estimated using OLS, "OLS" is used to refer specifically to standard cross-sectional and before-and-after models that do not incorporate a treatment–control design.

In a similar vein, Ahlfeldt (2013) examines the 1999 Jubilee Line and DLR extensions using a gravity-based labour market accessibility model. By linking changes in travel times to employment centres with property prices, the model estimates that doubling accessibility yields a utility gain equivalent to £383 per month in household income (2001 prices). This translates to residential property value uplift of approximately £716 million (1999 prices). Therefore, although academic research on transport-induced LVU in the UK and London exists, it largely pertains to older transport interventions.

2.4 Spatial Econometrics

Researchers often assume away spatial interactions between observations by including spatial fixed effects. However, there is a growing body of literature arguing for the need to control for possible spatial dependencies (LeSage & Pace, 2009). The existence of spatial lag is supported by the intuition that property valuation is not only based on intrinsic and local characteristics, but also on the transaction prices of nearby comparable properties. This is the premise behind the increasing adoption of spatial difference-in-differences (SDID) models, which is perhaps the most novel innovation in this field.

Much of the early work in spatial econometrics was pioneered by Anselin (1988). Although not formally part of the spatial econometrics canon, geographically weighted regression (GWR) represents an early attempt to model spatially varying relationships. Focusing on the light rail system in Tyne and Wear, Du and Mulley (2012) apply both OLS and GWR to examine how public transport accessibility is capitalised into property prices. The GWR reveals that the effect varies greatly across space, with the largest premia observed in poorer neighbourhoods, and provides a better model fit than OLS.

Dubé et al. (2014) applied spatial econometric techniques to the study of transportation on housing values using a spatial lag (SAR) model, whereas Diao et al. (2017) included both a spatial lag and error parameter in their spatial autocorrelation combined (SAC/SARAR) model. Spatial models incorporating a temporal dimension are sometimes distinguished as spatio-temporal autoregressive (STAR) models. Subsequent studies employing spatial regressions of varying forms generally find significant spatial parameters (Arku et al., 2024; Higgins, 2019; Higgins et al., 2024; Huang et al., 2024; Hyun & Milcheva, 2019; Ke & Gkritza, 2019; Qiu & Tong, 2021; Zhu & Diao, 2024).

The primary issue with spatial regressions involves the lack of a mathematically-rigorous, first-best method that is universally accepted, and therefore largely remains an empirical question. The memory-intensive nature of calculating the matrices, owing to its quadratic time complexity, further stifles its diffusion into the literature. Though the lack of standardisation across spatial models may undermine their application in guiding policymaking, it is nonetheless worthwhile to include when modelling LVU to observe whether spatial dependencies exist, and to benchmark against conventional models.

2.5 Research Gap

There have been several commercial reports on the capitalisation effects of the Elizabeth line before its inauguration. For example, GVA and Crossrail Ltd (2018) estimate realised LVU of 19% and 8% for residential and commercial properties within 1km of an Elizabeth line station respectively, representing an additional £10.6 billion in residential property value. Research by CBRE (2024) finds a growth premium of 6% in property values and 2% in rental values between 2008 and 2023 along London stations specifically (excluding those in Zone 1) compared to wider boroughs. Transport for London (TfL, 2017) found no evidence of anticipated LVU on residential property during the line's construction, but uplifts on commercial values of around 1–2.5% per year. While the results are generally positive but varied, the exact methodologies are not always specified. Moreover, these reports tend to focus on the short-term impacts of the post-announcement years, with none providing a full view of the Elizabeth line's entire life cycle, which requires data predating its official announcement in 2008.

To the best of my knowledge, there are no existing academic studies that analyse the capitalisation effects of the Elizabeth line on residential property values across its entire lifespan (from pre-announcement to operation), nor are there any that use the Manhattan distance measure. While existing UK-based research provides useful insight into local impacts of transport infrastructure, these studies often suffer from narrow geographic scope, relatively short temporal windows or aggregated data structures. Leveraging a comprehensive dataset spanning 18 years, this study contributes to the existing literature by applying a DID approach to add a datapoint on the magnitude of LVU resulting from the Elizabeth line, and to introduce the Manhattan distance measure as a feasible substitute for network distance. The results can help inform policymaking in terms of cost-benefit analysis, and support the usage of funding methods for future rail transport infrastructure.

3 The Elizabeth Line

Named after HM Queen Elizabeth II, the Elizabeth line is a 118km-long railway that runs in an east–west direction across London, with its eastern periphery extending to Shenfield and its western terminus in Reading. At an estimated cost of £18.9 billion, the line and its 41 stations have been expected to expand central London's rail capacity by 10% and contribute £42 billion to the UK economy (TfL, 2022). Jointly sponsored by TfL and the Department for Transport, the project aligns with the government's broader transport decarbonisation goal and Levelling Up mission to ameliorate regional inequality.

Under the alias of Crossrail, plans first began in 2001 when the Crossrail Ltd joint venture formed between the Department for Transport and TfL. The business case, setting out the line's objectives and funding model, was published in 2005.[2] The system was formally approved by the Crossrail Act 2008, receiving royal assent on 22$^{\text{nd}}$ July 2008.[3] Construction works began soon after on 15$^{\text{th}}$ May 2009, starting in Canary Wharf. Originally planned to open in 2018, the project faced repeated delays, including as a result of the COVID-19 pandemic. Service officially commenced on 24$^{\text{th}}$ May 2022. Bond Street station, however, opened five months later on 24$^{\text{th}}$ October 2022.

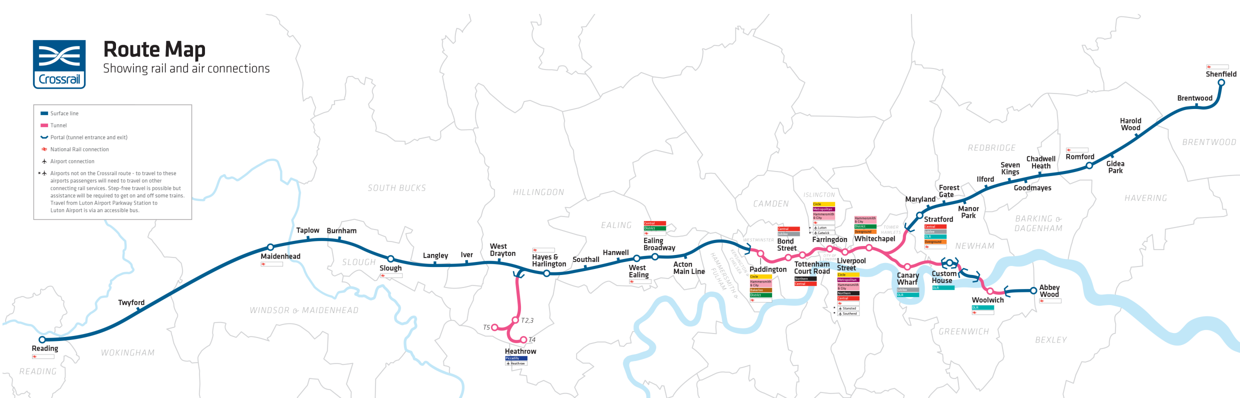

Figure 1: The Elizabeth Line Route

Since opening, the line has made over 350 million journeys. The cumulative benefits of improving train and highway journey times, emissions, noise and air quality, congestion relief, accident savings and tax revenues are estimated to outweigh capital and operating costs by a ratio of 1.9 (TfL, 2024).

[2]See Bennett (2017) and Tucker (2017).

[3]https://www.legislation.gov.uk/ukpga/2008/18/contents

4 Data

4.1 Study Area

The dataset consists of residential property transactions within a 3km straight-line buffer zone from all 41 Elizabeth line stations between January 2006 and December 2024. The pooled cross-sectional data is sourced from Zoopla, one of the UK's largest real estate property portals, and consists of 309,876 housing transactions in total.

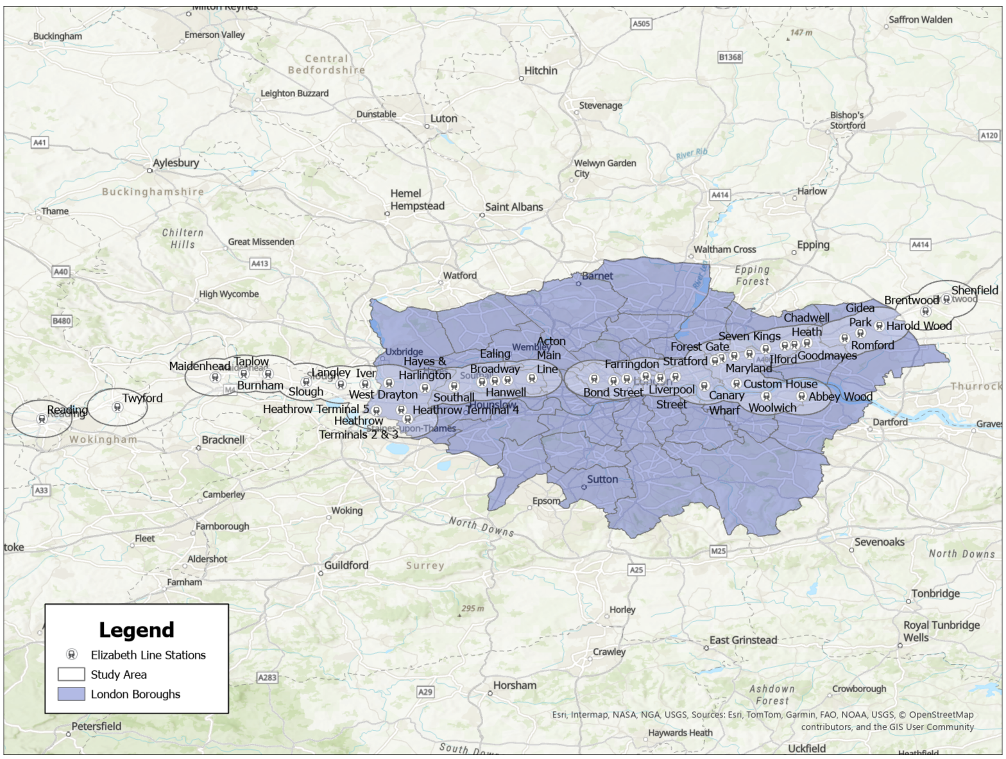

Figure 2: Study Boundary

4.2 Housing Characteristics

Each observation includes the floor area (in square metres), number of bedrooms, bathrooms and living rooms, property type, tenure, Energy Performance Certificate (EPC) rating and longitude and latitude coordinates.

The property type category includes flats, terraced, semi-detached and detached properties; tenure is either leasehold or freehold. EPC ranges from A to G, where A represents the best possible energy efficiency rating. The natural logarithm of the price and floor area variables is taken in order to normalise the distributions, and to capture diminishing marginal utility in attributes.

In addition, the distance to the nearest Elizabeth line station is calculated for each property using the network, Euclidean and Manhattan distance measures. Elizabeth line station coordinates are obtained from the Transport for London website. Road network shapefile data is obtained from Ordnance Survey.

4.3 Neighbourhood Characteristics

Population density and population age composition (0–29, 30–64 and 65+) data from the Office for National Statistics are matched to each property using their Lower Layer Super Output Area (LSOA), a small geographical unit created for statistical purposes. LSOA boundaries are available on the Open Geography Portal, allowing each property to be matched to its 2001, 2011 and 2021 LSOA. Demographic time series data is available from 2006 to 2022; however, pre-2011 values are matched using 2011 LSOAs, and the rest using 2021 LSOAs. Using 2011–2022 data, the ARIMA (Autoregressive Integrated Moving Average) model is applied to forecast 2023 and 2024 values for each 2021 LSOA.

The Index of Multiple Deprivation (IMD) is an index that captures various dimensions of socioeconomic disadvantage, including income, employment, education, health, crime, environment and barriers to housing, providing a comprehensive measure of deprivation, whereby a higher score reflects a more deprived area. Given the IMD's multifaceted nature, median income, crime statistics and educational performance variables are not required, as including them results in high multicollinearity. Data on IMD scores are provided for 2004, 2007 and 2010 under 2001 LSOAs, and in 2015 and 2019 for 2011 LSOAs by the Department for Levelling Up, Housing and Communities. Consumer Data Research Centre additionally provides 2010 data adjusted for 2011 LSOAs. This means that IMD scores for all years between 2004 and 2019 can be linearly interpolated. As there are comparatively fewer years of published data for IMD, 2020–2024 values are assumed to match those of 2019 to avoid introducing noise through extrapolation.

4.4 Fixed Effects and Multicollinearity

To control for time-invariant local attributes unaccounted for by existing neighbourhood data, 168 spatial fixed effects are introduced in total using outward postcodes. Time fixed effects are included on a quarterly basis, with 76 quarterly time dummies spanning from Q1 2006 to Q4 2024 to account for inflation in housing prices. Due to high correlation between floor area and bedrooms, bedrooms is omitted from the models given its discrete nature. The tenure category is also dropped as flats, freehold and leasehold are highly correlated.

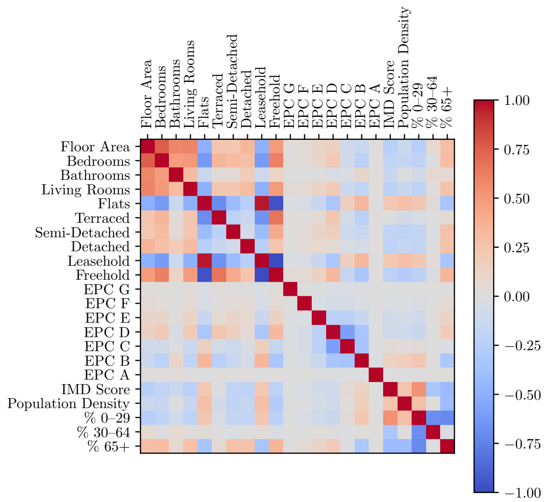

Figure 3: Covariate Correlation Matrix

For each categorical variable, one dummy is dropped to serve as a reference category, preventing perfect multicollinearity. For property type, the flats category is dropped, and for EPC, the G rating. For the population age bands, the 30–64 group serves as the baseline.

4.5 Distance Measures

Suppose there are two observations i and j with respective longitude and latitude coordinates (x_i, y_i) and (x_j, y_j). The Euclidean and Manhattan distance measures are calculated in two-dimensional space using the formulas:

$$d_{ij}^{\text{Euclidean}} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

$$d_{ij}^{\text{Manhattan}} = |x_i - x_j| + |y_i - y_j|$$

Except for when i and j are identical or perfectly aligned, $d_{ij}^{\text{Manhattan}}$ must necessarily be greater than $d_{ij}^{\text{Euclidean}}$ due to the triangle inequality. Diao et al. (2017) argues that because the Euclidean distance underestimates the true distance faced by commuters, using it can mislead results. However, the Euclidean metric may be conceptualised as the perceived distance to stations, and thus still considered relevant (Hess & Almeida, 2007).

In the below example, the network distance between 1 Park Drive and Canary Wharf Elizabeth line station is calculated as 768m in ArcGIS Pro. The Euclidean distance is 526m, and the Manhattan distance, 719m. The Manhattan distance is a closer approximation of the network distance despite not requiring network layout data and Geographic Information System, and unlike Euclidean distance, does not assume commuters traverse water or buildings.

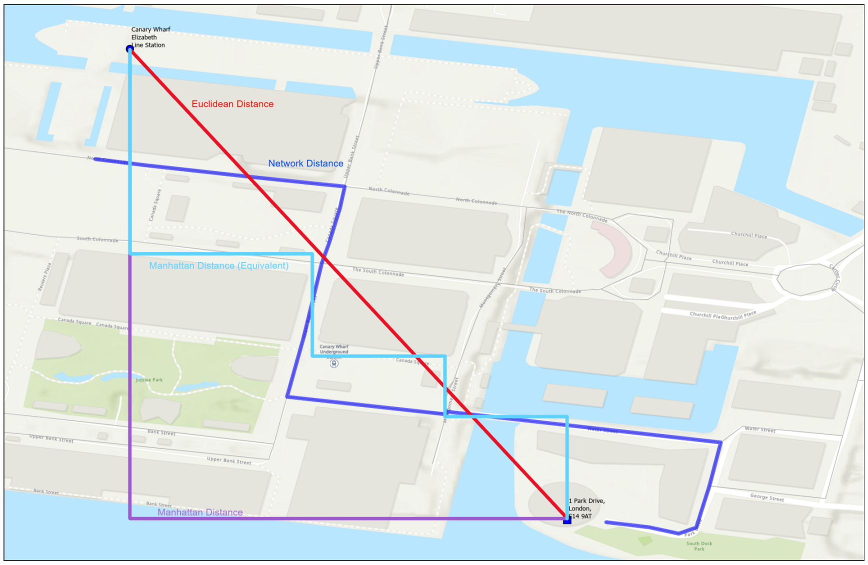

Figure 4: Distance Measure Comparison

In general, the Manhattan distance becomes a better proxy for the network distance the more grid-like the study area's network is. However, both Euclidean and Manhattan distance will fall short if there is a large insurmountable obstacle, such as crossing the River Thames in areas without a bridge or tunnel nearby.

4.6 Treatment and Control Zones

Many studies use 1/2 mile or 800m as the cutoff distance to sort between groups; the line of reasoning is that this figure represents a 10-minute walkable distance (Guerra et al., 2012). Although this argument has merit, the true treatment zone may not be homogenous across study areas. Rather than arbitrarily assigning a threshold, a more robust selection method may involve dynamically defining a boundary using a hedonic pricing model. Xu and Zhang's (2016) methodology is adapted by sorting observations into 150m distance bands and setting properties beyond 2850m as the reference category for the Euclidean and Manhattan distances. As the maximum value under the network distance is 7093m, 300m distance bands are used, and transactions beyond 5700m are set as the reference category. Including housing and neighbourhood controls alongside spatial and time fixed effects enables the price premium associated with each band to be identified. It is expected that the coefficients should be systematically positive and significant up until a certain group, after which it remains insignificant or negative.

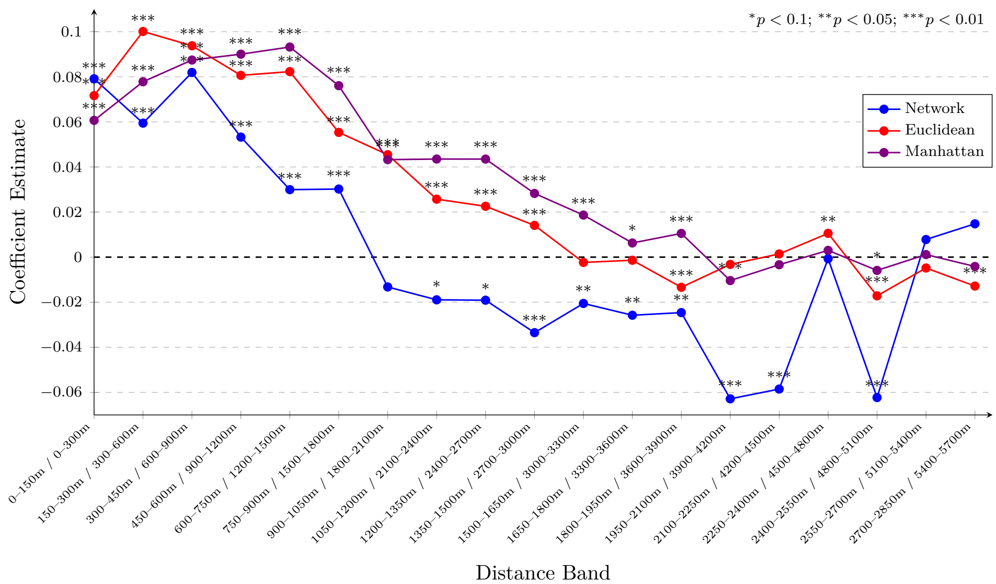

Figure 5: Location Premium by Distance Band

Figure 5 shows that the hedonic price for proximity to stations initially increases with distance, before gradually descending. The coefficients are positive and significant up until 1800m, 1500m and 1950m for the network, Euclidean and Manhattan distance measures respectively. This result is similar to a report by Nationwide (2021), finding that the location premium associated with London Underground stations extends up to a straight-line distance of 1500m.

4.7 Descriptive Statistics

Table 1 presents the descriptive statistics for the entire dataset, and the treatment and control groups defined by network distance. Prima facie, the characteristics of all groups appear similar.

	Full Dataset			Treatment Group			Control Group		
Statistic	N	Mean	St. Dev.	N	Mean	St. Dev.	N	Mean	St. Dev.
Price	309,876	508,688.00	739,005.70	120,331	481,896.00	629,137.20	189,545	525,696.60	800,508.40
Floor Area	309,876	90.92	45.51	120,331	88.95	44.08	189,545	92.18	46.35
Bedrooms	309,876	2.36	1.05	120,331	2.29	1.05	189,545	2.40	1.05
Bathrooms	309,876	1.41	0.66	120,331	1.40	0.64	189,545	1.41	0.67
Living Rooms	309,876	1.26	0.54	120,331	1.25	0.53	189,545	1.27	0.55
IMD Score	309,876	23.65	11.67	120,331	23.26	11.47	189,545	23.90	11.79
Population Density	309,876	10,669.78	8,224.42	120,331	10,581.48	8,398.32	189,545	10,725.83	8,111.62
% 0–29	309,876	42.37	7.66	120,331	43.05	8.09	189,545	41.93	7.33
% 30–64	309,876	47.61	5.42	120,331	47.30	5.56	189,545	47.81	5.33
% 65+	309,876	10.03	5.56	120,331	9.65	5.44	189,545	10.28	5.63
Network Distance	309,876	2,232.12	1,143.95	120,331	1,124.96	427.56	189,545	2,934.99	866.66
Euclidean Distance	309,876	1,588.09	772.03	120,331	824.68	350.61	189,545	2,072.73	539.91
Manhattan Distance	309,876	1,977.45	963.87	120,331	1,044.33	450.83	189,545	2,569.84	697.22

Table 1: Summary Statistics

5 Estimation Strategy

5.1 Difference-in-Differences Model Specification

The baseline results begin with a standard DID model:

$$\begin{aligned}\ln(P) = \alpha + \beta_1(\text{Treat} \times \text{Announcement}) + \beta_2(\text{Treat} \times \text{Construction}) \\ + \beta_3(\text{Treat} \times \text{Post}) + H'\gamma + N'\theta + \varphi + \xi + \varepsilon\end{aligned} \tag{1}$$

where $\ln(P)$ is log transaction price, $H'\gamma$ represents housing characteristics, $N'\theta$ represents neighbourhood characteristics, φ are spatial fixed effects, ξ are time fixed effects and ε is the i.i.d. error term. Treat is a binary variable which equals 1 if the property is within the threshold distance identified in Figure 5, and 0 otherwise. Announcement equals 1 if the transaction date falls on or after 22$^{\text{nd}}$ July 2008 and before 15$^{\text{th}}$ May 2009. Construction equals 1 if the date is on or after 15$^{\text{th}}$ May 2009 and before 24$^{\text{th}}$ May 2022. Post equals 1 if the transaction date occurs on or after 24$^{\text{th}}$ May 2022 (or on or after 24$^{\text{th}}$ October 2022 if the nearest station is Bond Street); all indicators are 0 otherwise. The coefficients of interest — β_1, β_2 and β_3 — are the ceteris paribus average treatment effect on the treated.

5.1.1 Quantile Regression

Although equation (1) reveals the sample-average LVU over time, there may be significant variation in effects across subgroups. To test for heterogeneous effects in LVU across the transaction-price distribution, quantile regressions are estimated at the 10$^{\text{th}}$, 25$^{\text{th}}$, 50$^{\text{th}}$, 75$^{\text{th}}$ and 90$^{\text{th}}$ percentiles:

$$Q_{\ln(P)}(\tau \mid X) = X\beta_\tau \tag{2}$$

where $Q_{\ln(P)}(\tau|X)$ is the τ-th conditional quantile of $\ln(P)$ given explanatory covariates, X, from equation (1) using the Frisch–Newton interior point method (Portnoy & Koenker, 1997).

5.1.2 Treatment Effects over Space

Since a multitude of studies have observed that LVU is uneven over space, spatial variations in treatment effects can be identified by dividing the treatment group into 5 zones:

$$\begin{aligned}\ln(P) = \alpha + \sum_{i=1}^{5}\beta_i(\text{Zone } i \times \text{Announcement}) + \sum_{j=1}^{5}\delta_j(\text{Zone } j \times \text{Construction}) \\ + \sum_{k=1}^{5}\sigma_k(\text{Zone } k \times \text{Post}) + H'\gamma + N'\theta + \varphi + \xi + \varepsilon\end{aligned} \tag{3}$$

where each Zone extends 300m, except for Zone 5, which covers 1200–1800m/1500m/1950m based on treatment threshold identified for each distance measure in Figure 5.

5.1.3 Treatment Effects across Regions

To identify whether capitalisation effects are stronger in some areas compared to others, the dataset is grouped into 5 regions based on each transaction's closest station in terms of network distance, and equation (1) is re-estimated on each subdataset.

The regions are divided into West of London (Reading to Iver), West London (West Drayton and Heathrow to Acton Main Line), Central London (Paddington to Liverpool Street), East London (Whitechapel to Harold Wood and Abbey Wood) and East of London (Brentwood and Shenfield).

5.2 Spatial Difference-in-Differences Model Specification

Integrating the base DID regression into the STAR model yields the SDID specification, estimated via maximum likelihood, to capture spatial lag and error in housing prices:

$$\begin{aligned} \ln(P) &= \rho W \ln(P) + X'\eta + u \\ u &= \lambda W u + \varepsilon \end{aligned} \tag{4}$$

where ρ is the spatial autoregressive parameter, $W \ln(P)$ is the spatially lagged dependent variable, X represents all explanatory variables in equation (1) with coefficient vector η, u is the spatially correlated error term, λ is the spatial error parameter and Wu is the spatially lagged error term.

Due to computer memory constraints from the large sample size, the SDID model is applied to the three stations with the most properties nearest to them by network distance: Paddington, Canary Wharf and Reading. The aspatial DID regression from equation (1) is estimated on the same subdatasets to benchmark against the SDID results.

5.2.1 The Spatial Weight Matrix and Spatial Weights

The spatial weight matrix, S, is an $n \times n$ positive, symmetric and non-stochastic matrix, where n is the number of observations and the element s_{ij} acts as the spatial weight between i and j. The magnitude of the weights represents the pairwise degree of proximity. By convention, the diagonal cells equal 0, as self-neighbourhood is excluded. This is the key term that formalises spatial dependence into a model:

$$S = \begin{bmatrix} 0 & s_{12} & \cdots & s_{1n} \\ s_{21} & 0 & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & 0 \end{bmatrix}$$

As there is a temporal dimension, the dataset is oftentimes sorted chronologically before calculations are performed for practicality. The negative exponential model is then used to compute distance-based spatial weights:

$$s_{ij} = \begin{cases} \exp(-d_{ij}) & \text{if } i \neq j \text{ and } d_{ij} \leq \overline{d}, \\ 0 & \text{otherwise,} \end{cases}$$

where d_{ij} is the Euclidean distance in kilometres between properties i and j and $\overline{d}$ is a threshold distance determined by fitting an empirical semivariogram. Figure 6 illustrates that the range — the distance at which spatial correlation plateaus — is approximately 1000m.

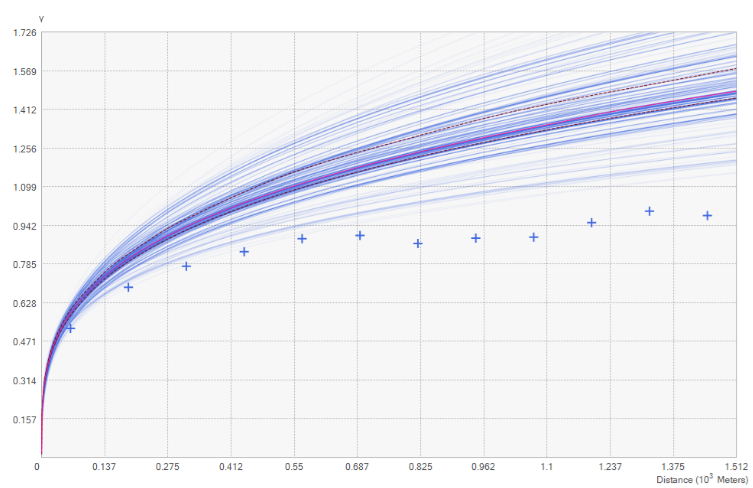

Figure 6: Empirical Semivariogram

5.2.2 The Temporal Weight Matrix

The temporal weight matrix is structurally analogous to the spatial weight matrix, with each element representing the temporal distance between two transactions:

$$T = \begin{bmatrix} 0 & t_{12} & \cdots & t_{1n} \\ t_{21} & 0 & \cdots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{n1} & t_{n2} & \cdots & 0 \end{bmatrix}$$

The temporal weights are calculated as follows:

$$t_{ij} = \begin{cases} (v_i - v_j)^{-1} & \text{if } i > j \text{ and } 0 < v_i - v_j \leq \overline{v}, \\ 1 & \text{if } v_i = v_j \text{ and } i \neq j, \\ (v_j - v_i)^{-2} & \text{if } i < j \text{ and } 0 < v_j - v_i \leq \underline{v}, \\ 0 & \text{otherwise,} \end{cases}$$

where v_i is the number of months between i and the earliest month of the dataset and $\overline{v}$ and $\underline{v}$ denote the thresholds for past and future influence. This formulation assumes that transactions may be influenced by sales occurring up to 6 months prior and 3 months after, enforcing a harsher penalty on future influences. Whilst this does not strictly adhere to the unidirectionality of time emphasised by Thanos et al. (2016), implementing their approach results in isolates within the final matrix. Higgings' (2019) method is therefore followed, which permits limited future influence to capture expectation and peer effects.

The spatio-temporal weight matrix is computed as the Hadamard product of the spatial and temporal weight matrices (Smith & Wu, 2009):

$$W = S \odot T$$

which is then row standardised.

5.3 Triple Difference Model Specification

Lastly, station-level heterogeneity may be partially explained by network effects accentuating LVU for stations with more connecting lines. Adding a three-way interaction term creates a difference-in-difference-in-differences (DDD), or triple difference, setting to test whether connection-driven network effects exist for interchange stations:

$$\begin{aligned} \ln(P) = \alpha &+ \beta_1(\text{Treat} \times \text{Announcement}) + \beta_2(\text{Treat} \times \text{Construction}) \\ &+ \beta_3(\text{Treat} \times \text{Post}) + \sum_{i=1}^{6} \delta_i(\text{Treat} \times \text{Announcement} \times \text{Connections}) \\ &+ \sum_{j=1}^{6} \sigma_j(\text{Treat} \times \text{Construction} \times \text{Connections}) \\ &+ \sum_{k=1}^{6} \psi_k(\text{Treat} \times \text{Post} \times \text{Connections}) + H'\gamma + N'\theta + \varphi + \xi + \varepsilon \end{aligned} \tag{5}$$

where Connections is a categorical variable representing the number of lines the transaction's closest station connects to. Each DDD coefficient signals the additional premium for interchange stations above and beyond the base premium, β_i, for non-interchange stations.

6 Results

6.1 DID Model Results

Table 2 presents the results from equation (1). All coefficients are positive and significant at the 1% level. The model achieves an adjusted R^2 of 0.80, suggesting a strong overall fit. In line with the majority of the empirical literature, the capitalisation of accessibility benefits occurs well before inauguration. The results indicate that LVU in treatment areas were highest during the construction phase (2.86–5.31%), followed by the announcement phase (2.21–3.15%) and lowest post-opening (0.81–3.01%). This implies that expectations were priced in early on, peaking during construction as the line's completion became more certain. However, post-opening, the benefits have been mostly exhausted, and potential negative externalities from operations may temper further gains.

	Network	Euclidean	Manhattan
Treat × Announcement	0.0315***	0.0285***	0.0221***
	(0.0078)	(0.0075)	(0.0074)
Treat × Construction	0.0531***	0.0347***	0.0286***
	(0.0017)	(0.0017)	(0.0016)
Treat × Post	0.0301***	0.0103***	0.0081***
	(0.0032)	(0.0031)	(0.0031)
Property Characteristics	Yes	Yes	Yes
Neighbourhood Attributes	Yes	Yes	Yes
Spatial Fixed Effects	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes
Observations	309,876	309,876	309,876
R^2	0.7995	0.7991	0.7991
Adjusted R^2	0.7994	0.7990	0.7989
Residual Std. Error (df = 309614)	0.2976	0.2978	0.2979
F Statistic (df = 261; 309614)	4,730.8300***	4,719.7950***	4,717.2050***

Note: All tables report standard errors in parentheses. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 2: Capitalisation Effects over Time

Furthermore, the network distance measure produces higher coefficient estimates than the Euclidean distance. While this corroborates with the findings of Diao et al. (2017), it contrasts with Hess and Almeida (2007). Although Diao et al. applied a consistent 600m treatment threshold across both measures, this analysis uses different thresholds for each distance measure. Importantly, these thresholds represent different proportions of their respective distance ranges: 1500m is half of the maximum Euclidean distance (3000m), while 1800m represents about a quarter of the maximum network distance (7093m). As such, the treatment group defined using network distance likely includes properties that are, on average, closer to stations than those included using the Euclidean metric.

Contrary to expectations, the Euclidean distance yields estimates closer to the network distance than Manhattan distance for all three coefficients. This suggests that while Manhattan distance generally better approximates actual travel patterns than Euclidean, using the former may not generate closer results to network distance.

6.1.1 Hedonic Prices

Although the housing and neighbourhood characteristics are not the primary focus of this study, they are not without interest. The coefficients for floor area, bathrooms, living rooms and property type are positive and highly significant, as expected. The EPC coefficients warrant more interest. Whilst an EPC rating of B commands a 14.21% premium compared to properties with the lowest, G, rating, all the other grades are statistically insignificant. This could be because official EPC data includes expired certificates that no longer reflect a property's current condition, and buyers may instead value a property's potential EPC rating — neither of which is accounted for in the model. Additionally, the lack of a significant effect for EPC A may reflect the limited number of observations (230).

The results suggest that local factors too shape housing prices, albeit to a lesser extent. Areas with a higher proportion of residents aged 0–29 or 65+ (relative to the 30–64 age group) are associated with lower property values. Population density is likewise linked with lower prices, though the effect size is minimal. Deprivation is also negatively correlated with prices, with each additional point on the IMD score resulting in a 0.59% reduction in housing values.

Moreover, the Variance Inflation Factors for each covariate across all models are below 4, suggesting that multicollinearity is not a significant issue.[4]

[4]While Table 3 is based on the network-distance model in Table 2, the values under the other distance measures are virtually identical, and the same conclusions hold.

	Dependent Variable: $\ln(P)$
ln(Floor Area)	0.6120***
	(0.0021)
Bathrooms	0.1002***
	(0.0011)
Living Rooms	0.0373***
	(0.0013)
Terraced	0.1375***
	(0.0017)
Semi-Detached	0.1789***
	(0.0023)
Detached	0.2525***
	(0.0032)
EPC F	0.0239
	(0.0154)
EPC E	0.0157
	(0.0140)
EPC D	−0.0005
	(0.0139)
EPC C	−0.0019
	(0.0139)
EPC B	0.1421***
	(0.0140)
EPC A	−0.0005
	(0.0241)
IMD Score	−0.0059***
	(0.0001)
Population Density	−0.000003***
	(0.000000)
% 0–29	−0.0018***
	(0.0001)
% 65+	−0.0023***
	(0.0002)
Constant	9.3262***
	(0.0188)

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 3: Hedonic Prices

6.1.2 Treatment Effects across Percentiles

Price effects are markedly concentrated at the right tail of the distribution, indicating that LVU is skewed towards higher-value properties, particularly in the 75$^{\text{th}}$ percentile and above. This may reflect socioeconomic sorting if wealthier buyers and investors place greater value on improved connectivity. Whether this behaviour is generalisable is contested. Hess and Almeida (2007) found the highest positive effects in high-income neighbourhoods, and the largest negative in low-income neighbourhoods: a result consistent with Bowes and Ihlanfeldt (2001) but contrary to Nelson (1992).

	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}
Treat × Announcement	-0.0152	0.0122	0.0189^{***}	0.0444^{***}	0.0672^{***}
	(0.0129)	(0.0088)	(0.0071)	(0.0084)	(0.0117)
Treat × Construction	0.0205^{***}	0.0322^{***}	0.0454^{***}	0.608^{***}	0.0698^{***}
	(0.0021)	(0.0015)	(0.0013)	(0.0016)	(0.0019)
Treat × Post	0.0113^{**}	0.0209^{***}	0.0251^{***}	0.0318^{***}	0.0310^{***}
	(0.0049)	(0.0031)	(0.0028)	(0.0028)	(0.0036)
Property Characteristics	Yes	Yes	Yes	Yes	Yes
Neighbourhood Attributes	Yes	Yes	Yes	Yes	Yes
Spatial Fixed Effects	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes
Distance Measure	Network	Network	Network	Network	Network
Observations	309,876	309,876	309,876	309,876	309,876
Pseudo R^2	0.4724	0.5628	0.6146	0.6505	0.6845

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 4: Distributional Effects

Due to long computational runtime, primarily from the inclusion of fixed effects, standard errors were bootstrapped using 100 replications. While this is on the lower end of common practice, it provides a reasonable approximation of the standard error (Fox, 2015). Additional replications on a repeat-sales subdataset yielded similar standard errors, suggesting that results are not sensitive to the number of bootstrap draws.

6.1.3 Treatment Effects over Space

Across all distance measures, the largest and most significant price increases are observed in Zones 1–3 (within 900m), especially during construction. This aligns with the literature, which finds that LVU peaks at around 500m (Mohammad et al., 2013; Rennert, 2022). When significant, these coefficients usually exceed the average estimate for the entire treatment group. Zones 4 and 5 (beyond 900m), in contrast, generally exhibit weaker and less consistent effects, particularly after completion. This diminishing pattern of effects over space reflects the localised nature of accessibility gains. The lower LVU estimates during construction compared to at announcement for Zone 1 under Euclidean and Manhattan distances hint at the existence of negative externalities associated with construction efforts.

	Network	Euclidean	Manhattan
Zone 1 × Announcement	0.0397	0.0769***	0.0630**
	(0.0362)	(0.0252)	(0.0305)
Zone 1 × Construction	0.0768***	0.0714***	0.0596***
	(0.0060)	(0.0040)	(0.0049)
Zone 1 × Post	0.0596***	0.0349***	0.0231*
	(0.0149)	(0.0095)	(0.0119)
Zone 2 × Announcement	0.0388*	0.0491***	0.0365**
	(0.0210)	(0.0143)	(0.0186)
Zone 2 × Construction	0.0544***	0.0666***	0.0685***
	(0.0036)	(0.0028)	(0.0033)
Zone 2 × Post	0.0202**	0.0367***	0.0443***
	(0.0079)	(0.0059)	(0.0072)
Zone 3 × Announcement	0.0435***	0.0470***	0.0771***
	(0.0152)	(0.0124)	(0.0139)
Zone 3 × Construction	0.0812***	0.0561***	0.0663***
	(0.0029)	(0.0024)	(0.0027)
Zone 3 × Post	0.0491***	0.0193***	0.0321***
	(0.0062)	(0.0050)	(0.0058)
Zone 4 × Announcement	0.0162	−0.0013	0.0016
	(0.0147)	(0.0128)	(0.0137)
Zone 4 × Construction	0.0612***	0.0262***	0.0306***
	(0.0026)	(0.0023)	(0.0025)
Zone 4 × Post	0.0347***	0.0071	0.0013
	(0.0057)	(0.0048)	(0.0054)
Zone 5 × Announcement	0.0360***	0.0342***	0.0148
	(0.0101)	(0.0123)	(0.0094)
Zone 5 × Construction	0.0418***	0.0104***	0.0136***
	(0.0019)	(0.0022)	(0.0018)
Zone 5 × Post	0.0263***	0.0016	0.0041
	(0.0040)	(0.0048)	(0.0037)
Property Characteristics	Yes	Yes	Yes
Neighbourhood Attributes	Yes	Yes	Yes
Spatial Fixed Effects	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes
Observations	309,876	309,876	309,876
R^2	0.7997	0.7995	0.7995
Adjusted R^2	0.7995	0.7993	0.7993
Residual Std. Error (df = 309602)	0.2975	0.2976	0.2976
F Statistic (df = 273; 309602)	4,526.5060***	4,522.3580***	4,521.0150***

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 5: Spatial Variation in Effects

Some patterns may reflect differences in sample composition. The Euclidean metric yields the most evenly distributed sample across zones, producing the smallest standard errors. Given its linearity, it is unsurprising that LVU consistently declines with distance (with the exception of Zone 5 × Announcement). On the other hand, few properties fall into Zone 1 under the network and Manhattan definitions, while Zone 5 accounts for the largest share of treated transactions. This skew arises because non-Euclidean measures assign higher values than straight-line distance. Many properties near stations on a planar basis — typically Zones 1 to 3 — are reclassified into Zones 2 to 5 under alternative metrics. As a result, these zones capture a greater share of well-connected properties, helping explain why they often display similar or stronger capitalisation effects than closer zones across all distance measures.

	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Total
Network	3,864	12,435	21,336	25,989	56,707	120,331
	(3.2%)	(10.3%)	(17.7%)	(21.6%)	(47.1%)	
Euclidean	9,653	25,082	36,579	39,198	36,798	147,310
	(6.6%)	(17.0%)	(24.8%)	(26.6%)	(25.0%)	
Manhattan	5,991	16,474	26,123	30,987	76,300	155,875
	(3.8%)	(10.6%)	(16.8%)	(19.9%)	(48.9%)	

Similarly to Table 2, network distance is most strongly associated with uplift. Though Manhattan distance coefficients now better resemble network distance than Euclidean distance in many cases, there is no consistent pattern of outperformance.

6.1.4 Treatment Effects across Regions

Table 6 reveals substantial regional variation in LVU. Central London and East see the highest gains during construction, at 8.25% and 10.99% respectively. These regions also exhibit strong post-opening effects (6.18% and 6.64%), whereas the West region faces negative effects of –1.68%. The results suggest that accessibility improvements amplified pre-existing advantages in Central London, and brought lasting benefits to newly connected areas in East and East London. Conversely, muted or negative effects in West may reflect overpricing prior to opening, or weaker demand after May 2022.

	West	West London	Central London	East London	East
Treat × Announcement	0.0311^{**}	-0.0274	0.0143	0.0391^{***}	0.1433^{***}
	(0.0132)	(0.0181)	(0.0205)	(0.0113)	(0.0378)
Treat × Construction	0.0252^{***}	0.0503^{***}	0.0825^{***}	0.0636^{***}	0.1099^{***}
	(0.0027)	(0.0037)	(0.0062)	(0.0022)	(0.0075)
Treat × Post	-0.0168^{***}	0.0415^{***}	0.0618^{***}	0.0261^{***}	0.0664^{***}
	(0.0054)	(0.0068)	(0.0106)	(0.0044)	(0.0143)
Property Characteristics	Yes	Yes	Yes	Yes	Yes
Neighbourhood Attributes	Yes	Yes	Yes	Yes	Yes
Spatial Fixed Effects	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes
Distance Measure	Network	Network	Network	Network	Network
Observations	46,918	54,768	65,123	133,834	9,233
R^2	0.8283	0.8008	0.7757	0.7230	0.8239
Adjusted R^2	0.8280	0.8004	0.7751	0.7227	0.8221
Residual Std. Error	0.2104 (df = 46811)	0.2694 (df = 54645)	0.3774 (df = 64947)	0.2687 (df = 133689)	0.2411 (df = 9136)
F Statistic	$2{,}131.0240^{***}$ (df = 106; 46811)	$1{,}801.1690^{***}$ (df = 122; 54645)	$1{,}283.7930^{***}$ (df = 175; 64947)	$2{,}422.8330^{***}$ (df = 144; 133689)	445.3435^{***} (df = 96; 9136)

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 6: Regional Variation in Effects

6.2 SDID Model Results

Table 7 depicts that SDID estimates are consistently lower and less significant than those from the standard DID model, aligning with Diao et al. (2017) but diverging from Dubé et al. (2014). This likely reflects model differences; Diao et al. use SAC/SARAR, whereas Dubé et al. use SAR, with higher coefficients in Diao et al.'s SAR model supporting this. The significant spatial lag coefficients confirm that property values are influenced by neighbouring areas, while the strongly significant spatial error terms evidence spatially correlated unobserved factors. The generally lower Bayesian information criterion (BIC), substantially lower Akaike information criterion (AIC) and highly significant likelihood ratio (LR) test statistic provide strong evidence that SDID offers a superior fit for the subdatasets.

	Paddington		Canary Wharf		Reading	
	DID	SDID	DID	SDID	DID	SDID
Treat × Announcement	0.0098	-0.0100	0.1601***	0.0848*	0.0605**	0.0256
	(0.0363)	(0.0417)	(0.0314)	(0.0401)	(0.0204)	(0.0268)
Treat × Construction	0.0458***	0.0453***	0.1654***	0.0900***	0.0352***	0.0148*
	(0.0106)	(0.0120)	(0.0057)	(0.0080)	(0.0047)	(0.0061)
Treat × Post	0.0271	0.0406	0.0874***	0.0241	-0.0033	0.0154
	(0.0185)	(0.0221)	(0.0137)	(0.0197)	(0.0090)	(0.0127)
ρ		0.1888***		0.4550***		-0.1222***
		(0.0243)		(0.0457)		(0.0342)
λ		0.3028***		0.8896***		0.9127***
		(0.0423)		(0.0214)		(0.0139)
Property Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Neighbourhood Attributes	Yes	Yes	Yes	Yes	Yes	Yes
Spatial Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Distance Measure	Network	Network	Network	Network	Network	Network
Observations	21,834	21,834	21,392	21,392	18,224	18,224
Adjusted / Pseudo R^2	0.7722	0.7760	0.5874	0.6630	0.7918	0.8103
AIC	22,109.3517	21,890.2757	9,060.7064	5,168.1944	$-7,490.8283$	$-8,680.6260$
BIC	22,988.3863	22,785.2928	5,973.2424	9,849.8128	$-6,717.5893$	$-7,891.7661$
LR Test Statistic	223.0760***		3,896.5120***		1,193.7978***	

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 7: Station-Level SDID

Paddington and Canary Wharf illustrate that regional-level analysis can still mask local variation. Most strikingly, Canary Wharf exhibits LVU approximately double the regional average and triple the overall mean. This likely reflects a combination of accessibility benefits that amplify its existing Jubilee line and DLR connections, strong residential and commercial demand and investments tied to local redevelopment plans.

6.3 DDD Model Results

The DDD model reveals positive network effects, but nonlinearities in marginal gains from connectivity. Stations with no interchanges experienced no significant announcement effects, and little to no uplift during construction (up to 2.2%), yet experienced strong post-opening uplift (4.8–7.2%), likely because the Elizabeth line delivered substantial transformations in accessibility for these relatively isolated stations. Excluding the six-connection station, this is reinforced by the negative post-opening coefficients for stations with interchanges, and the negative or insignificant coefficients for stations with one connection — reflecting the fact that the greatest beneficiaries of the line post-opening were those with the fewest pre-existing connections. As usual, the network distance generally produces the highest estimates, with the Mahattan and Euclidean metrics providing close approximations at times. The lack of a predictive pattern ultimately reflects the fact that they are three distinct measures, providing three different interpretations of the same underlying phenomenon.

As connectivity increases, anticipatory effects heighten. At a high level, the existence of positive network effects — in the form of additional LVU for interchange stations — is supported by the positive announcement and construction coefficients for stations with two to six connections. However, marginal gains are not always positive when comparing adjacent groups. This stems from the key limitation that the DDD model implicitly assumes all connections are equally valuable, which is rarely true in practice. Lines with higher ridership, greater frequency and better central access are likely to drive stronger effects; a more fine-tuned model would weight connectivity using these variables.

Canary Wharf exemplifies this caveat. Despite only two connections, its extreme uplift skews not only skews the DDD estimates, but also brings attention to the broader reality that a myriad of other factors besides interchange count play a significant role in determining capitalisation effects. Liverpool Street, the sole station with six connections, recorded the highest uplift effects, peaking at an estimated 27.97% during construction.[5] This epitomises its role as London's central hub, with exceptional multimodal connectivity, employment density and sustained real estate demand — reiterating the notion that both local context and network topology matter. Nevertheless, even with its simplifying assumptions, the DDD framework offers valuable insight by helping uncover patterns of network effects as one pivotal influence of the spatial heterogeneity in LVU observed in Tables 6 and 7.

[5]Though using DID for Liverpool Street based on network distance yields more modest estimates of 16.01%, 16.54% and 8.74% at announcement, construction and post respectively.

	Network	Euclidean	Manhattan
Treat × Announcement	-0.0127	-0.0175	0.0114
	(0.0152)	(0.0150)	(0.0162)
Treat × Construction	0.0220^{***}	0.0001	0.0173^{***}
	(0.0033)	(0.0032)	(0.0033)
Treat × Post	0.0716^{***}	0.0475^{***}	0.0573^{***}
	(0.0059)	(0.0058)	(0.0063)
Treat × Announcement × 1	0.0308^{*}	0.0355^{**}	0.0081
	(0.0176)	(0.0169)	(0.0187)
Treat × Construction × 1	-0.0099^{***}	0.0044	-0.0131^{***}
	(0.0038)	(0.0036)	(0.0038)
Treat × Post × 1	-0.0522^{***}	-0.0378^{***}	-0.0472^{***}
	(0.0068)	(0.0065)	(0.0072)
Treat × Announcement × 2	0.1226^{***}	0.0784^{***}	0.0890^{***}
	(0.0210)	(0.0197)	(0.0228)
Treat × Construction × 2	0.0917^{***}	0.0567^{***}	0.0671^{***}
	(0.0048)	(0.0045)	(0.0048)
Treat × Post × 2	-0.0384^{***}	-0.0637^{***}	-0.0506^{***}
	(0.0092)	(0.0082)	(0.0094)
Treat × Announcement × 3	0.0150	0.0165	0.0057
	(0.0271)	(0.0258)	(0.0299)
Treat × Construction × 3	0.0908^{***}	0.0873^{***}	0.0822^{***}
	(0.0058)	(0.0057)	(0.0060)
Treat × Post × 3	-0.0756^{***}	-0.0596^{***}	-0.0761^{***}
	(0.0105)	(0.0100)	(0.0114)
Treat × Announcement × 4	0.0865^{**}	0.1234^{***}	-0.0135
	(0.0339)	(0.0319)	(0.0374)
Treat × Construction × 4	0.1707^{***}	0.1916^{***}	0.1187^{***}
	(0.0094)	(0.0083)	(0.0099)
Treat × Post × 4	-0.0372^{**}	-0.0155	-0.0578^{***}
	(0.0163)	(0.0152)	(0.0183)
Treat × Announcement × 5	0.0844^{***}	0.1008^{***}	-0.0068
	(0.0253)	(0.0235)	(0.0280)
Treat × Construction × 5	0.1032^{***}	0.0874^{***}	0.0711^{***}
	(0.0059)	(0.0054)	(0.0060)
Treat × Post × 5	-0.0399^{***}	-0.0575^{***}	-0.0451^{***}
	(0.0113)	(0.0103)	(0.0116)
Treat × Announcement × 6	0.2374^{***}	0.1870^{***}	0.1712^{***}
	(0.0442)	(0.0453)	(0.0492)
Treat × Construction × 6	0.2577^{***}	0.2581^{***}	0.2464^{***}
	(0.0088)	(0.0078)	(0.0090)
Treat × Post × 6	0.0408^{**}	0.0328^{*}	0.0189
	(0.0182)	(0.0168)	(0.0195)
Property Characteristics	Yes	Yes	Yes
Neighbourhood Attributes	Yes	Yes	Yes
Spatial Fixed Effects	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes
Observations	309,876	309,876	309,876
R^2	0.8009	0.8005	0.8003
Adjusted R^2	0.8007	0.8003	0.8001
Residual Std. Error (df = 309596)	0.2965	0.2969	0.2970
F Statistic (df = 279; 309596)	$4{,}464.5050^{***}$	$4{,}451.9180^{***}$	$4{,}446.5530^{***}$

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 8: Marginal Effect of Connections

7 Robustness Checks

7.1 Parallel Trends

Given the choice of station location is almost always influenced by factors such as population density, deprivation levels and potential for development, the non-random nature of station placement complicates proving parallel trends and establishing causal relationships compared to randomised controlled trials. A formal test for common trends prior to an intervention involves adding interaction terms between the treatment variable and time fixed effects to estimate the temporal trend in the treated group relative to the control group:

$$\ln(P) = \alpha + \beta \text{Treat} + \sum_{i=1} \delta_i (\text{Treat} \times \xi_i) + H'\gamma + N'\theta + \varphi + \xi + \varepsilon \tag{6}$$

If the coefficients of the interaction terms are statistically insignificant, this indicates no diverging patterns between the treatment and control groups in the pre-treatment period (Dubé et al., 2024; Huang et al., 2024; Zhu & Diao, 2024).

	Network	Euclidean	Manhattan
Treat × 2006 Q2	0.0013	0.0018	−0.0070
	(0.0103)	(0.0100)	(0.0100)
Treat × 2006 Q3	0.0040	0.0027	0.0067
	(0.0099)	(0.0097)	(0.0096)
Treat × 2006 Q4	0.0202**	0.0165*	0.0140
	(0.0100)	(0.0097)	(0.0097)
Treat × 2007 Q1	0.0028	−0.0004	−0.0041
	(0.0103)	(0.0100)	(0.0100)
Treat × 2007 Q2	0.0056	0.0065	0.0063
	(0.0101)	(0.0098)	(0.0098)
Treat × 2007 Q3	0.0012	0.0010	−0.0038
	(0.0100)	(0.0097)	(0.0097)
Treat × 2007 Q4	−0.0099	−0.0114	−0.0111
	(0.0104)	(0.0102)	(0.0101)
Treat × 2008 Q1	0.0005	−0.0001	0.0004
	(0.0114)	(0.0112)	(0.0112)
Treat × 2008 Q2	−0.0119	−0.0073	−0.0112
	(0.0114)	(0.0112)	(0.0112)
Treat × 2008 Q3	0.0073	0.0077	−0.0046
	(0.0234)	(0.0230)	(0.0230)
Property Characteristics	Yes	Yes	Yes
Neighbourhood Attributes	Yes	Yes	Yes
Spatial Fixed Effects	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes
Observations	50,995	50,995	50,995
R^2	0.7855	0.7844	0.7841
Adjusted R^2	0.7847	0.7836	0.7833
Residual Std. Error (df = 50799)	0.2456	0.2462	0.2464
F Statistic (df = 195; 50799)	953.9007***	947.7870***	946.2954***

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 9: Parallel Trends Test

Table 9 illustrates that all coefficients under Manhattan distance are insignificant, and all but 2006 Q4 are insignificant for network and Euclidean distance. However, as this is the only significant pre-intervention quarter, it is safe to say that the parallel trends assumption is broadly satisfied. This assumption is essential for causal identification, as it implies that any divergence in trends after treatment can be attributed to the intervention itself.

The quarterly fixed effects enable the construction of a housing price index for both groups. In addition to visualising pre-treatment common trends, Figure 7 also traces the evolution of treatment effects within individual phases. During the construction phase, the size of LVU starts to slowly decline between late 2013 and early 2016, before sharply increasing and stabilising from 2017 onwards. The vertical gap between trends reveals that capitalisation effects are stronger towards the end, rather than start, of construction, and appear to peak in the years directly before the line opened.

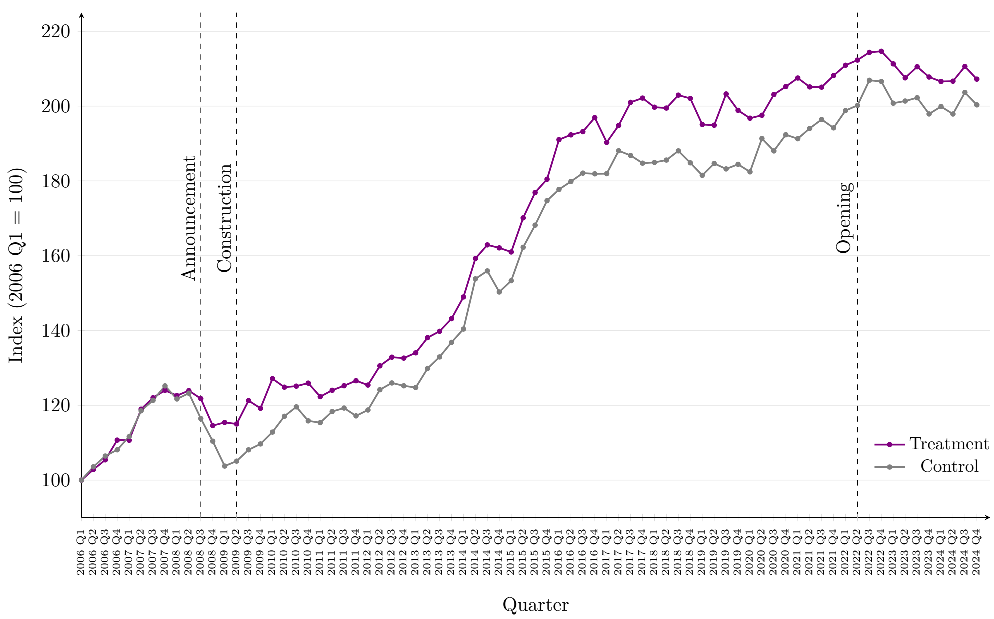

Figure 7: Quarterly Housing Price Index

7.2 Propensity Score Weighting

As an additional robustness check, the baseline DID model is re-estimated using inverse probability weighting to address potential selection bias, which can undermine identification of the treatment effect. Weights are derived from a logistic regression predicting treatment assignment based on pre-treatment housing and neighbourhood characteristics to adjust for observable differences between treated and control properties:

$$\hat{p} = \Pr(\text{Treat} = 1 \mid H, N) = \Lambda\left(\alpha + H'\gamma + N'\theta\right)$$

Inverse probability weights are then constructed:

$$\omega_i = \begin{cases} 1 & \text{if Treat} = 1, \\ \frac{\hat{p}}{1-\hat{p}} & \text{if Treat} = 0. \end{cases}$$

The resulting weighted estimates (Table 10) closely resemble those from the original unweighted model (Table 2) in both magnitude and statistical significance, suggesting that the main findings are not sensitive to selection on observables.

	Network	Euclidean	Manhattan
Treat × Announcement	0.0196***	0.0222***	0.0174**
	(0.0074)	(0.0075)	(0.0075)
Treat × Construction	0.0559***	0.0410***	0.0354***
	(0.0017)	(0.0017)	(0.0016)
Treat × Post	0.0441***	0.0305***	0.0284***
	(0.0032)	(0.0032)	(0.0031)
Property Characteristics	Yes	Yes	Yes
Neighbourhood Attributes	Yes	Yes	Yes
Spatial Fixed Effects	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes
Observations	309,876	309,876	309,876
R^2	0.7956	0.7895	0.7906
Adjusted R^2	0.7954	0.7893	0.7905
Residual Std. Error (df = 309614)	0.2637	0.2911	0.2996
F Statistic (df = 261; 309614)	4,617.6720***	4,449.7340***	4,479.8210***

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

Table 10: Inverse Probability Weighted DID

7.3 Clustered Standard Errors

Clustered standard errors are omitted from the main specifications due to the empirically defined treatment zones based on distance-based hedonic modelling (Figure 5), which create concentrated and overlapping spatial units. This violates the assumption of independent, numerous clusters. Significant spatial lag in the SDID model further confirms interdependence in housing prices, suggesting that clustering may distort the true error structure. Given these issues, and the inclusion of detailed spatial and temporal fixed effects, robust standard errors provide a more appropriate basis for inference. Table 11 presents clustered standard errors as a sensitivity check. While announcement and post-opening effects weaken, the construction phase effect remains strongly significant, an outcome similar to the SDID results, reinforcing the core finding that LVU is strongest during construction.

	Network	Euclidean	Manhattan
Treat × Announcement	0.0315	0.0285	0.0221
	(0.0215)	(0.0210)	(0.0185)
Treat × Construction	0.0531***	0.0347***	0.0286***
	(0.0149)	(0.0105)	(0.0095)
Treat × Post	0.0301**	0.0103	0.0081
	(0.0130)	(0.0101)	(0.0101)
Property Characteristics	Yes	Yes	Yes
Neighbourhood Attributes	Yes	Yes	Yes
Spatial Fixed Effects	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes
Clustered Standard Errors	Yes	Yes	Yes
Observations	309,876	309,876	309,876
R^2	0.7995	0.7991	0.7991
Within R^2	0.5823	0.5815	0.5813

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

Table 11: Clustered Standard Errors

7.4 Machine Learning Prediction Models and Buyer Perceptions

To further examine the choice of distance metric, the mean absolute percentage error is calculated for each proxy. Manhattan distance deviates from actual network distance by 16.2% on average, while Euclidean deviates by 27.3%, confirming that Manhattan provides a closer numerical approximation.

However, when using a gradient boosted trees model (XGBoost) to predict network distance, Euclidean distance accounts for 88.3% of the model's total gain, compared to just 11.7% for Manhattan, suggesting that Euclidean captures a greater share of variation. The model achieves a root mean squared error of 565.7, indicating strong overall fit. The same conclusion holds when using a Lasso regression to predict network distance, where Euclidean has a noticeably higher coefficient than Manhattan (0.949 versus 0.255). The Euclidean distance's greater predictive power likely reflects a stronger overall correlation with network distance, despite being less accurate on average.

In evaluating which distance measure best captures how buyers truly value distance accessibility, XGBoost and Lasso models are used to predict log price using all three distance types, alongside controls and fixed effects. The relative importance of each distance variable indicates which measure is most strongly reflected in observed price formation. Both models find that Euclidean distance contributes the most towards explaining price variation, achieving the highest feature gain in XGBoost (0.0028 versus 0.0012 for both Manhattan and network) and the largest retained coefficient in Lasso (–2.24e–05, 5.57e–06 and –2.17e–05 respectively). While all effects are small, this consistent outperformance suggests that buyers are influenced more by the perceived straight-line proximity than the actual or city block distance.

These observations combined help explain why Manhattan distance, despite its definitional advantage, does not consistently yield closer results to network than Euclidean.

8 Discussion

8.1 Expectations

The timing of uplift across phases suggests that housing markets price in anticipated benefits well before infrastructure becomes operational. Across all models, construction effects tend to be the strongest and most significant. This implies that expectations play a central role in shaping LVU. For policymakers, this emphasises the need to consider early-stage intervention — such as implementing LVC before benefits are fully capitalised — and to manage public expectations to avoid overvaluation. Anticipatory uplift may not always align with realised outcomes, especially where benefits are diffuse or slower to materialise, as implied in areas West of London.

While Manhattan distance was expected to approximate network distance more closely than Euclidean, this was not consistently the case. Manhattan's lower predictive power in the machine learning models helps reconcile these inconsistencies, suggesting that Euclidean may better capture variation and serve as a perceptual heuristic for buyers. Another contributing factor may be that the Manhattan threshold covered a larger share of its total range (1950m versus 4238m) than the network threshold (1800m against 7093m). This likely resulted in a broader treatment group, with properties further out on average, that diluted treatment effects. The threshold itself was derived from a hedonic pricing model, where the significance of the penultimate distance band was marginal; had it been insignificant, the Manhattan threshold would have been 1650m, which may have produced closer approximations. Moreover, Manhattan distance assumes orthogonal movement along a rectilinear path — an approach not always satisfied by London's irregular and historical streets. Taken together, these factors suggest that each distance measure captures a distinct spatial logic, and should be interpreted as complementary rather than hierarchical.

8.2 Cost-Benefit Analysis

Table 2 states that residential properties near an Elizabeth line station experienced, on average, uplift of 5.31% during the 15-year construction phase using network distance. Given that the average transaction value within the treatment group is £481,896, this implies an average LVU of £25,589. Since there are 120,331 transactions in the treatment group, the total LVU resulting from the Elizabeth line amounts to £3.08 billion. With an estimated final cost of £18.9 billion, this represents approximately 16.3% of total costs.

Naturally, this result will vary depending on which distance measure and model one applies, and a more nuanced analysis would involve accounting for each property's phase, relative price, location and nearest station. Having said that, this figure is likely to fall towards the lower bound of estimates. This is because the dataset only contains transactions for which there is data for all housing characteristics on Zoopla; thus, the true number of transactions in the treated group will be higher. Furthermore, many untransacted properties will have experienced a windfall gain yet to be realised upon sale. For a given average treatment effect on the treated, this will necessarily lead to an even higher aggregate estimate. Nevertheless, even with this conservative estimate, the bottom line is that the aggregate windfall gains arising from accessibility improvements reflect a non-trivial proportion of construction costs.

8.3 Relative Beneficiaries

Despite the aggregate benefits derived from the Elizabeth line, these gains do not necessarily benefit everyone, nor are they universally distributed. Although the taxpayer bears a cost, the net gain or loss one experiences will depend on to what extent the line directly and indirectly benefits them. Non-users who do not experience any indirect benefits will bear the greatest burden — most likely those living particularly far from the line.

As we have seen, even for homeowners within the treatment zone, windfall gains are far from homogenous. Higher-value properties, properties within 900m of stations and those near well-connected stations experience above-average LVU. Given that UK median annual earnings in 2023–24 were £37,430, the average windfall gain represents around 68% of the median person's yearly income. For more advantaged transactions, LVU will easily exceed 100%. As wealthier individuals are more likely to be homeowners (Gregg & Kanabar, 2023), such a result would indicate that, all else equal, the line exacerbates wealth inequality and housing affordability via the LVU channel. If rents were to rise in response to the line, as suggested by CBRE (2024), this would represent an additional redistribution of resources from tenants to landlords, heightening income inequality. The wealth disparities will be unevenly distributed across regions, with Central London, East London and East generating more wealth than other Elizabeth line areas. Though one could argue that East London should accumulate more wealth due to its greater deprivation, the reality, as shown by Canary Wharf and the quantile regression, is that such wealth creation is likely concentrated within a select few stations, and at the higher end of property values.

It is important to note, however, that these results must be contextualised with the broader socioeconomic and environmental benefits beyond LVU provided by the Elizabeth line. The line helps connect previously underserved communities in East and South East London to the city centre, and in the process, stimulates local economies and offers easier access to opportunities and resources elsewhere. While property-driven wealth inequalities may arise, they must not overshadow the overarching benefits and goal of local and regional development.

8.4 Land Value Capture

The prior discussion presents a clear case for capturing some of these windfall gains. Besides cost recovery and fairness, a strong rationale for implementing LVC is that it may enhance project efficiency (Ingram & Hong, 2012). While designing a perfect LVC mechanism goes beyond the scope of this paper, it is worth discussing the current state of affairs. The UK currently has four property taxes: Council Tax, Business Rates, Stamp Duty Land Tax and Capital Gains Tax. However, all four are ineffective at capturing LVU because they are relatively unresponsive to changes in property values (TfL, 2017). London, however, has already used a levy on commercial property, the Business Rates Supplement, which helped raise £4.1 billion specifically towards Crossrail; the Community Infrastructure Levy, a charge imposed on new developments, and developer contributions each raised an additional £300 million (Buck, 2017).

Hong Kong's "Rail + Property" model offers a celebrated example of successful LVC, whereby the Mass Transit Railway Corporation obtains development rights for land around new stations at pre-rail values, and uses post-rail values to recoup its costs, generating over half of the operator's total income (Cervero & Murakami, 2009). However, transposing this model to the UK would be challenging, as Hong Kong has a public leasehold system, empowering them to easily grant long-term leases. TfL (2017) have proposed several potential LVC methods. This includes Community Infrastructure Levy and Section 106 requirements for new developments near major transport projects to help capture uplift without directly taxing homeowners. Additionally, allowing a portion of Stamp Duty Land Tax growth in transport-enhanced areas to be reinvested locally would provide a steady revenue stream without imposing new levies. Political opposition and disputes over charges based on property valuations will inevitably rise against these measures if executed, paralleling some of the demerits of a land value tax, a measure proposed since the late 19^{th} century, but never implemented in the UK. Nonetheless, a hybrid strategy that leverages these mechanisms could help London systematically capture the value it creates, ensuring future transport projects are more self-sustaining.

9 Conclusion

This dissertation uses DID methods to evaluate the Elizabeth line's impact on residential property prices across its development. The findings from the 18-year dataset indicate a positive effect on housing values within the treatment zone, with results robust to propensity score weighting and clustered standard errors, and consistent with the parallel trends assumption. Although average LVU is valued at 3.15%, 5.31% and 3.01% for the announcement, construction and operational phases respectively, it is ultimately the location and timing of transactions that matter most. In general, the largest uplift is experienced towards the end of the construction phase, particularly for higher-priced properties, properties within 900m of stations and those near interchange stations with zero or at least two connecting lines.

The network distance measure produces the highest estimates, and despite its conceptual advantage, there is little evidence supporting the hypothesis that the Manhattan distance measure is systematically better than the Euclidean distance metric in approximating network distance results. This highlights the unique and distinct properties of all three spatial measures. Additionally, machine learning algorithms suggest that Euclidean, or perceived, distance influences buyer decision-making the most.

Moreover, SDID models yield lower LVU estimates than aspatial DID models, provide a better fit and suggest that spillover effects exist in fact, an unsurprising result given London's active housing market. With a minimum estimated aggregate LVU of £3.08 billion based on the dataset, or 16.3% of the Elizabeth line's total construction costs, there is strong reason to implement LVC mechanisms on residential properties experiencing uplift, both in the rental and sales market.

LVU is merely one of the many benefits offered by the Elizabeth line, and the full impact of such a transformative and long-anticipated system will almost certainly cover its costs multiple times over. Therefore, future research exploring LVU effects of other major upcoming transport projects — most notably High Speed 2 — and using alternative variants of accessibility measures (including gravity-, travel time- and labour market-based approaches) will deepen our understanding of LVU dynamics in the UK, providing crucial evidence for decision-makers to invest in infrastructure that will go on to improve and shape the cities we live in.

10 Bibliography

Ahlfeldt, G. M. (2013). If we build it, will they pay? Predicting property price effects of transport innovations. *Environment and Planning A*, *45*(8), 1977–1994. https://doi.org/10.1068/a45429

Alonso, W. (1964). *Location and land use: Toward a general theory of land rent*. Harvard University Press.

Anas, A., & Chang, H. (2023). Productivity benefits of urban transportation megaprojects: A general equilibrium analysis of «Grand Paris Express». *Transportation Research Part B: Methodological*, *174*, 102746. https://doi.org/10.1016/j.trb.2023.03.006

Anderson, M. L. (2013). *Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion* (Working Paper No. 18757). National Bureau of Economic Research. https://doi.org/10.3386/w18757

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Kluwer Academic Publishers.

Arku, R. N., Higgins, C. D., Fischer, J., & Farber, S. (2024). Do affluent neighbourhoods pay more for transit access? Exploring the capitalization of employment accessibility within different housing submarkets in Vancouver. *Journal of Transport Geography*, *121*, 104038. https://doi.org/10.1016/j.jtrangeo.2024.104038

Bao, H. X., Larsson, J. P., & Wong, V. (2021). Light at the end of the tunnel: The impacts of expected major transport improvements on residential property prices. *Urban Studies*, *58*(14), 2971–2990. https://doi.org/10.1177/0042098020967308

Bauernschuster, S., Hener, T., & Rainer, H. (2017). When labor disputes bring cities to a standstill: The impact of public transit strikes on traffic, accidents, air pollution, and health. *American Economic Journal: Economic Policy*, *9*(1), 1–37. https://doi.org/10.1257/pol.20150414

Bennett, S. (2017). Crossrail project to deliver London's Elizabeth line: The parliamentary bill process. *Proceedings of the Institution of Civil Engineers - Civil Engineering*, *170*(6), 10–14. https://doi.org/10.1680/jcien.17.00014

Billings, S. B. (2011). Estimating the value of a new transit option. *Regional Science and Urban Economics*, *41*(6), 525–536. https://doi.org/10.1016/j.regsciurbeco.2011.03.013

Bowes, D. R., & Ihlanfeldt, K. R. (2001). Identifying the impacts of rail transit stations on residential property values. *Journal of Urban Economics*, *50*(1), 1–25. https://doi.org/10.1006/juec.2001.2214

Buck, M. (2017). Crossrail project: Finance, funding and value capture for London's new Elizabeth line. *Proceedings of the Institution of Civil Engineers - Civil Engineering*, *170*(6), 15–22. https://doi.org/10.1680/jcien.17.00005

CBRE. (2024). *The Elizabeth line: The impact on London's housing market*. https://www.cbre.co.uk/insights/reports/the-elizabeth-line-the-impact-on-london-s-housing-market

Cervero, R., & Murakami, J. (2009). Rail and property development in Hong Kong: Experiences and extensions. *Urban Studies*, *46*(10), 2019–2043. https://doi.org/10.1177/0042098009339431

Damm, D., Lerman, S. R., Lerner-Lam, E., & Young, J. (1980). Response of urban real estate values in anticipation of the Washington Metro. *Journal of Transport Economics and Policy*, *14*(3), 315–336.

Debrezion, G., Pels, E., & Rietveld, P. (2007). The impact of railway stations on residential and commercial property value: A meta-analysis. *The Journal of Real Estate Finance and Economics*, *35*(2), 161–180. https://doi.org/10.1007/s11146-007-9032-z

Diao, M., Leonard, D., & Sing, T. F. (2017). Spatial-difference-in-differences models for impact of new mass rapid transit line on private housing values. *Regional Science and Urban Economics*, *67*, 64–77. https://doi.org/10.1016/j.regsciurbeco.2017.08.006

Du, H., & Mulley, C. (2007). The short-term land value impacts of urban rail transit: Quantitative evidence from Sunderland, UK. *Land Use Policy*, *24*(1), 223–233. https://doi.org/10.1016/j.landusepol.2005.12.003

Du, H., & Mulley, C. (2012). Understanding spatial variations in the impact of accessibility on land value using geographically weighted regression. *Journal of Transport and Land Use*, *5*(2). https://doi.org/10.5198/jtlu.v5i2.225

Dubé, J., Le Gallo, J., Des Rosiers, F., Legros, D., & Champagne, M.-P. (2024). An integrated causal framework to evaluate uplift value with an example on change in public transport supply. *Transportation Research Part E: Logistics and Transportation Review*, *185*, 103500. https://doi.org/10.1016/j.tre.2024.103500

Dubé, J., Legros, D., Thériault, M., & Des Rosiers, F. (2014). A spatial difference-in-differences estimator to evaluate the effect of change in public mass transit systems on house prices. *Transportation Research Part B: Methodological*, *64*, 24–40. https://doi.org/10.1016/j.trb.2014.02.007

Forrest, D., Glen, J., & Ward, R. (1996). The impact of a light rail system on the structure of house prices: A hedonic longitudinal study. *Journal of Transport Economics and Policy*, *30*(1), 15–29. https://jtep.org/wp-content/uploads/2021/02/Volume_XXX_No_1_15-29-1.pdf

Fox, J. (2015). *Applied regression analysis and generalized linear models* (3rd). Sage Publications.

Gatzlaff, D. H., & Haurin, D. R. (1997). Sample selection bias and repeat-sales index estimates. *The Journal of Real Estate Finance and Economics*, *14*(1), 33–50. https://doi.org/10.1023/A:1007763816289

Gatzlaff, D. H., & Smith, M. T. (1993). The impact of the Miami Metrorail on the value of residences near station locations. *Land Economics*, *69*(1), 54–66. https://doi.org/10.2307/3146278

Gibbons, S., & Machin, S. (2005). Valuing rail access using transport innovations. *Journal of Urban Economics*, *57*(1), 148–169. https://doi.org/10.1016/j.jue.2004.10.002

Gregg, P., & Kanabar, R. (2023). *Parental homeownership and education: The implications for offspring wealth inequality in GB* (CEPEO Working Paper Series No. 22-01). UCL Centre for Education Policy and Equalising Opportunities. https://repec-cepeo.ucl.ac.uk/cepeow/cepeowp22-01r2.pdf

Guerra, E., Cervero, R., & Tischler, D. (2012). Half-mile circle: Does it best represent transit station catchments? *Transportation Research Record*, *2276*(1), 101–109. https://doi.org/10.3141/2276-12

GVA & Crossrail Ltd. (2018). *Crossrail property impact & regeneration study 2012-2026*. https://learninglegacy.crossrail.co.uk/wp-content/uploads/2018/07/4D-003-crossrail_property_impact_regeneration_study.pdf

He, S. Y. (2020). Regional impact of rail network accessibility on residential property price: Modelling spatial heterogeneous capitalisation effects in Hong Kong. *Transportation Research Part A: Policy and Practice*, *135*, 244–263. https://doi.org/10.1016/j.tra.2020.01.025

Henneberry, J. (1998). Transport investment and house prices. *Journal of Property Valuation and Investment*, *16*, 144–158. https://doi.org/10.1108/14635789810212913

Hess, D. B., & Almeida, T. M. (2007). Impact of proximity to light rail rapid transit on station-area property values in Buffalo, New York. *Urban Studies*, *44*(5-6), 1041–1068. https://doi.org/10.1080/00420980701256005

Higgins, C. D. (2019). A 4D spatio-temporal approach to modelling land value uplift from rapid transit in high density and topographically-rich cities. *Landscape and Urban Planning*, *185*, 68–82. https://doi.org/10.1016/j.landurbplan.2018.12.011

Higgins, C. D., Arku, R. N., Farber, S., & Miller, E. J. (2024). Modelling changes in accessibility and property values associated with the King Street Transit Priority Corridor project in Toronto. *Transportation Research Part A: Policy and Practice*, *190*, 104256. https://doi.org/10.1016/j.tra.2024.104256

Higgins, C. D., & Kanaroglou, P. S. (2016). Forty years of modelling rapid transit's land value uplift in North America: Moving beyond the tip of the iceberg. *Transport Reviews*, *36*(5), 610–634. https://doi.org/10.1080/01441647.2016.1174748

Huang, Y., Parker, D. C., Babin, R., & Kong, F. (2024). Causal identification of transit-induced property value uplift in Canada's Waterloo Region: A spatio-temporal difference-in-differences method application. *Cities*, *145*, 104676. https://doi.org/10.1016/j.cities.2023.104676

Hyun, D., & Milcheva, S. (2019). Spatio-temporal effects of an urban development announcement and its cancellation on house prices: A quasi-natural experiment. *Journal of Housing Economics*, *43*, 23–36. https://doi.org/10.1016/j.jhe.2018.09.008

Ingram, G. K., & Hong, Y.-H. (Eds.). (2012). *Value capture and land policies*. Lincoln Institute of Land Policy.

Ke, Y., & Gkritza, K. (2019). Light rail transit and housing markets in Charlotte-Mecklenburg County, North Carolina: Announcement and operations effects using quasi-experimental methods. *Journal of Transport Geography*, *76*, 212–220. https://doi.org/10.1016/j.jtrangeo.2019.03.011

Landis, J., Guhathakurta, S., Huang, W., & Zhang, M. (1995). *Rail transit investments, real estate values, and land use change: A comparative analysis of five California rail transit systems* (University of California Transportation Center, Working Papers). University of California Transportation Center. https://escholarship.org/uc/item/2hf9s9sr

LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC. https://doi.org/10.1201/9781420064254

Ma, L., Graham, D. J., & Stettler, M. E. (2021). Air quality impacts of new public transport provision: A causal analysis of the Jubilee Line Extension in London. *Atmospheric Environment*, *245*, 118025. https://doi.org/10.1016/j.atmosenv.2020.118025

McDonald, J. F., & Osuji, C. I. (1995). The effect of anticipated transportation improvement on residential land values. *Regional Science and Urban Economics*, *25*(3), 261–278. https://doi.org/10.1016/0166-0462(94)02085-U

McMillen, D. P., & McDonald, J. (2004). Reaction of house prices to a new rapid transit line: Chicago's midway line, 1983–1999. *Real Estate Economics*, *32*(3), 463–486. https://doi.org/10.1111/j.1080-8620.2004.00099.x

Mills, E. S. (1972). *Studies in the structure of the urban economy*. Johns Hopkins University Press.

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: Policy and Practice*, *50*, 158–170. https://doi.org/10.1016/j.tra.2013.01.013

Muth, R. F. (1969). *Cities and housing: The spatial pattern of urban residential land use*. University of Chicago Press.

Nationwide. (2021). *Transport special feature: June 2021*. https://www.nationwidehousepriceindex.co.uk/reports/london-sees-biggest-house-price-premium-as-households-still-attach-significant-value-to-transport-links-despite-the-pandemic

Nelson, A. C. (1992). Effects of elevated heavy-rail transit stations on house prices with respect to neighborhood income. *Transportation Research Record*, (1359). http://onlinepubs.trb.org/Onlinepubs/trr/1992/1359/1359-016.pdf

Portnoy, S., & Koenker, R. (1997). The Gaussian hare and the Laplacian tortoise: Computability of squared-error versus absolute-error estimators. *Statistical Science*, *12*(4), 279–300. https://doi.org/10.1214/ss/1030037960

Qiu, F., & Tong, Q. (2021). A spatial difference-in-differences approach to evaluate the impact of light rail transit on property values. *Economic Modelling*, *99*. https://doi.org/10.1016/j.econmod.2021.03.015

Rennert, L. (2022). A meta-analysis of the impact of rail stations on property values: Applying a transit planning lens. *Transportation Research Part A: Policy and Practice*, *163*, 165–180. https://doi.org/10.1016/j.tra.2022.06.013

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 34–55. https://doi.org/10.1086/260169

Sener, I. N., Lee, R. J., & Elgart, Z. (2016). Potential health implications and health cost reductions of transit-induced physical activity. *Journal of Transport & Health*, *3*(2), 133–140. https://doi.org/10.1016/j.jth.2016.02.002

Smith, T. E., & Wu, P. (2009). A spatio-temporal model of housing prices based on individual sales transactions over time. *Journal of Geographical Systems*, *11*, 333–355. https://doi.org/10.1007/s10109-009-0085-9

Thanos, S., Dubé, J., & Legros, D. (2016). Putting time into space: The temporal coherence of spatial applications in the housing market. *Regional Science and Urban Economics*, *58*, 78–88. https://doi.org/10.1016/j.regsciurbeco.2016.03.001

Transport for London. (2017). *Land value capture: Final report*. https://www.london.gov.uk/sites/default/files/land_value_capture_report_transport_for_london.pdf

Transport for London. (2022). *Elizabeth line: Evidencing the value*. https://tfl.gov.uk/corporate/publications-and-reports/elizabeth-line-benefits-framework

Transport for London. (2024). *Evidencing the value of the Elizabeth line: An update on the outcomes of London's transformational railway*. https://tfl.gov.uk/corporate/publications-and-reports/elizabeth-line-benefits-framework

Tucker, W. (2017). Crossrail project: The execution strategy for delivering London's Elizabeth line. *Proceedings of the Institution of Civil Engineers - Civil Engineering*, *170*(5), 3–14. https://doi.org/10.1680/jcien.16.00021

Von Thünen, J. H. (1966). *The isolated state: An English translation of Der isolierte Staat* (P. Hall, Ed.; C. M. Wartenberg, Trans.). Pergamon Press. (Original work published 1826)

Wagner, G. A., Komarek, T., & Martin, J. (2017). Is the light rail "Tide" lifting property values? Evidence from Hampton Roads, VA. *Regional Science and Urban Economics*, *65*, 25–37. https://doi.org/10.1016/j.regsciurbeco.2017.03.008

Xu, T., & Zhang, M. (2016). Tailoring empirical research on transit access premiums for planning applications. *Transport Policy*, *51*, 49–60. https://doi.org/10.1016/j.tranpol.2016.03.003

Yen, B. T., Mulley, C., Shearer, H., & Burke, M. (2018). Announcement, construction or delivery: When does value uplift occur for residential properties? Evidence from the Gold Coast Light Rail system in Australia. *Land Use Policy*, *73*, 412–422. https://doi.org/10.1016/j.landusepol.2018.02.007

Zhu, Y., & Diao, M. (2024). The local and network effects of rail transit network expansion on retail property values. *Journal of Planning Education and Research*, *44*(3), 1820–1834. https://doi.org/10.1177/0739456X221121247