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

Lawrence Lee

University of Cambridge

Note

This is still a very early work in progress. I will be continuously working on this throughout the next few months (starting from late October until March/April), and so will hopefully become much more populated very soon.

Contents

1 Literature Review

This topic lies at the crossroads between urban economics, real estate economics and spatial econometrics. The statistical techniques employed are relatively novel and continually refined in a growing body of research, but are nonetheless built upon well-established concepts and models. By combining economic and geocoded data using Geographic Information System (GIS), we are able to consider complex spatial effects that would have otherwise been overlooked by traditional methods.

1.1 Theoretical Foundations of the City

The monocentric city model of urban form is one of the most fundamental models in urban economics. Alonso (1964), Muth (1969) and Mills (1972) explain the spatial distribution of firms and households in a city by examining the trade-off between land use and commuting costs. Transport costs are considered not only in monetary terms, but also include all forms of disutility and inconvenience associated with travelling. This maximisation problem is illustrated by the negative relationship between housing prices and distance from the city centre, known as bid-rent curves. The slope of the bid-rent function, the Alonso-Muth condition, represents the marginal increase in commuting costs incurred as households move away from the employment centre.

Many empirical studies have shown that new urban travel infrastructure investments could significantly flatten the bid rent gradient, and reduce the housing price gap between urban and rural areas. The increased accessibility brought by new travel investment is translated into housing wealth accrued to local residents via capitalisation effects.

1.2 Hedonic Pricing Models

Modern hedonic pricing theory, originating from the seminal work of Rosen (1974), has become a widely used methodology in the applied economics and real estate literature. Groundbreaking at the time, this technique enables researchers to quantify the implicit price of ordinarily unobservable attributes simply by running an Ordinary Least Squares (OLS) regression on observable variables. As a result, difficult questions such as determining the value of a statistical life, environmental quality and crime could now be answered with the appropriate dataset.

Expectedly, this approach was adopted into studies on the effects of transport infrastructure on housing prices. By including a rich set of housing and spatial attributes, researchers are able to determine the market price of location.

However, OLS suffers from serious endogeneity issues when applied to the study of RTS effects. This is because the models are unable to separate unobserved factors that could influence the covariance between housing price and RTS accessibility. This stems from the fact that the consumer simultaneously chooses both the quantity and marginal price of the characteristic. For OLS, the explanatory variables must truly be exogenous and uncorrelated with the error term.

Researchers use repeated sales data to control for endogeneity in modelling housing price changes. While the repeated-sales approach removes biases caused by time-invariant omitted variables by taking the first differencing in housing prices, the sample size is drastically reduced in the process. Furthermore, there may be selection biases if house that are sold at least two times possess different attributes from those that sell just once. Changes in structural characteristics and amenities in local areas can occur over a long period of time, and sometimes randomly, thus differencing the repeated housing prices alone may not adequately remove intertemporal effects of both observed and unobserved factors.

1.3 Difference in Differences

The quasi-experimental approach has become increasingly popular in the regional and urban economics literature. In a randomised experiment, sample houses are sorted into a treatment group and control group. A random event — such as the opening of a new line — is used to simulate an exogenous shock to housing prices. The difference-in-difference (DID) model then tests the before-and-after price changes between the two groups following the shock whilst controlling for observed and unobserved variations in housing and spatial factors. If the pre-existing within-group price variations change after the treatment, causality of the new stations on housing prices can be established.

1.4 Spatial Difference in Differences

In the context of areas undergoing changes following the opening of a new station, there is a need to control for possible autoregressive lag and error in the spatial interactions of houses. This is why spatial difference-in-differences (SDID) models have been increasingly used. Much of the early work in spatial econometrics was pioneered by Anselin (1988). In order to incorporate spatial and temporal aspects into the estimation, to move from DID to SDID, one must incorporate 1) a spatial lag term to account for spatially dependent responses and 2) a spatial error term.

Dubé et al. (2014) were among the first to use SDID to study the impact of public mass transit system expansion on real estate values in Montreal, Canada. Diao et al. (2017) further calibrate the SDID specification to estimate the capitalisation effect of a new MRT line in Singapore that accounts for the spatial dependencies in the dependent variable and the error term.

1.5 Defining the Treatment Area and Distance Measures

Most of the early studies use a Euclidean distance measure from the property to the closest station as a variable in hedonic pricing models to capture cross-sectional variations in housing prices.

Furthermore, earlier studies define the treatment area using proximity measures such as linear distance or buffer zone, which suffers from omitted variable problems, most particularly when the selection of the station location is not exogenous, but correlated with some other factor including population density. The effect in a densely populated city centre may co-vary with other amenities in the area.

Diao et al. (2017) were among the first to propose using network distance as a more appropriate distance measure, as it accounts for topographic features including natural obstacles and spatial constraints. They also use a local polynomial regression to identify the treatment zone.

1.6 Related Studies

A large number of empirical studies find evidence of positive capitalisation effects of proximity to stations in housing values. Evidence of positive capitalisation effects of urban rail transit systems is shown in many studies across many countries over time. Debrezion et al. (2007) provide a meta-analysis of 57 cities and suggest that property values increase by 2.3% for every 250m closer to a railway station.

There exists a few studies that find insignificant or even negative effects of proximity to selected stations in cities. The negative externalities are usually caused by noise and high crime rates found in these areas near stations.

2 The Elizabeth Line

The Elizabeth line is a heavy rail transit service that runs across London, with its peripheries extending to Reading and Shenfield. Named after HM Queen Elizabeth II, the line opened on 17th May 2022 during her Platinum Jubilee year. Operations commenced soon after on 24th May 2022.

Plans for the Elizabeth line first began in 2001. Under a different alias of Crossrail, the system was approved in 2007. The Crossrail Act 2008 authorised the construction project, receiving royal assent on 22nd July 2008.[1] Construction works began on 15th May 2009, starting in Canary Wharf. Though originally planned to open in 2018, the project faced repeated delays, including for several months as a result of the COVID-19 pandemic.

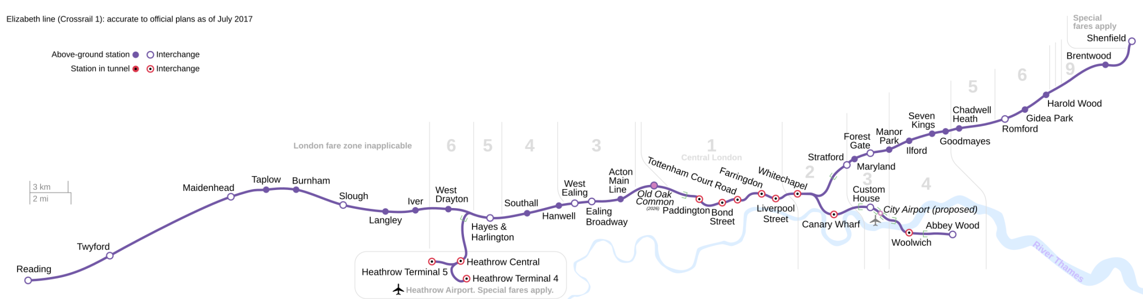

Figure 1: The Elizabeth Line Map

The line reached over 200 million trips annually in its second year of operation. Today, it carries one seventh of all trips by rail in the United Kingdom.

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

3 Estimation Strategy

For the baseline results, we start with a standard difference-in-differences model:

$$\ln(P) = \alpha + \beta_1 \text{Post} + \beta_2 \text{Treat} + \beta_3 (\text{Treat} \times \text{Post}) + H'\gamma + N'\theta + \varphi + \tau + \epsilon \qquad (1)$$

where $\ln(P)$ is log property price, $H'\gamma$ represents housing characteristics, $N'\theta$ represents neighbourhood characteristics, φ are spatial fixed effects, τ are time fixed effects and ε is the error term. The main coefficient of interest is β_3.

Let n be the number of spatial units. The general spatial autocorrelation model (SAC) is formally expressed as:

$$\begin{aligned} y &= \rho W y + X\beta + u \\ u &= \lambda W u + \varepsilon \\ \varepsilon &\sim \mathcal{N}(0, \sigma_\varepsilon^2) \end{aligned}$$

where y is an $n \times 1$ vector of cross-sectional dependent variables, W is the spatial matrix, X represents an $n \times k$ matrix of explanatory variables, ρ measures spatial lag and λ tests for spatial autocorrelation in the error term. This is an extension of an ordinary regression model that includes both a spatial lag term and a spatially correlated error term.

The spatial weight matrix, W, is an $n \times n$ positive symmetric and non-stochastic matrix with element w_{ij} at location i, j. The value of the weights measure the degree of spatial proximity for each pair of locations. By definition, the diagonal elements must equal 0. This formally incorporates spatial dependence into the model:

$$\begin{bmatrix} 0 & w_{1,2} & \cdots & w_{1,n} \\ w_{2,1} & 0 & \cdots & w_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n,1} & w_{n,2} & \cdots & 0 \end{bmatrix}$$

where values of the elements can be assigned in a number of ways, including

$$w_{ij} = \begin{cases} 1 & \text{if } i \neq j \text{ and } i \text{ and } j \text{ are contiguous,} \\ 0 & \text{otherwise.} \end{cases}$$

Combining the preceding DID regression with the SAC model produces the spatial difference-in-differences specification:

$$\begin{aligned} \ln(P) &= \rho W \ln(P) + X'\eta + u \\ u &= \lambda W u + \varepsilon \end{aligned} \tag{2}$$

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

4 Bibliography

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

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

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*, 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

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

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

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

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