The Residential Property Price Impact of Luas Investments

Brendan Connor

PricewaterhouseCoopers (PwC)

Abstract

This paper estimates a hedonic model of Dublin residential property prices to measure the price impact of developments to the Luas light-rail network in Dublin. The opening of a new Luas station had a substantial positive impact on within-walking-distance property sales prices. Also, the Cross-City Luas extension had a positive impact on property sales prices within walking distance of pre-existing Luas stations, particularly for Red Line stations. The model uses an innovative walking-time metric, created with Google's Distance Matrix API, to construct treatment and control groups for difference-in-differences tests of local price effects. The Irish Government has announced plans for €25 billion directed to the development of Dublin's transport infrastructure over the next 20 years with planning currently underway; this paper has important policy implications.

1. Introduction

Improved public transport infrastructure has an important role to play in creating a sustainable solution to Dublin's urban growth problems. Transport infrastructure developments have important environmental, social equity, land use, and property value implications. This paper examines the impact that the Dublin Luas light-rail line has on residential property prices.

The Luas is a light rail transit system operating in Dublin since 2004. There are two main lines with a current total of 67 stops, the Green Line (35 stops) and the Red Line (32 stops). The Luas has experienced a broadly consistent increase in per annum passenger numbers as well as an increase in its share of overall transport passengers over the past decade. Much of this increase in usage can be attributed to several line extensions added over the period 2009 to 2017. The Line B1 and Line A1 extensions were constructed in order to service outer suburban Dublin areas. The Cross-City extension was an extension that linked the Red and Green lines allowing for passenger interchange between these two previously separate lines. The Line B1, Line A1 and the Cross-City extensions provide the basis for the statistical analysis in this paper.

To provide the foundation for my test procedure, I estimate a geospatial-hedonic model of residential property prices in County Dublin. There are two versions of the pricing model, both versions assume that log sales prices are linearly dependent on the property's floor area, Building Energy Rating (BER), age, number of bedrooms, and quarterly seasonal dummies. The fully linear version of the model also includes a set of County Dublin area dummies based on Dublin postal codes and township designations. In the second version of the model, the area dummies are replaced by a nonparametric estimate of property sales price as a smooth two-variable function of north-south and east-west locations. The local pricing impact tests are then conducted using the pricing residuals from the pricing model. The two versions of the pricing model give similar test findings.

https://doi.org/10.32388/E9IEZG

The pricing model provides first-stage residuals corrected for the major observable sources of residential price variation. The paper then uses a difference-in-differences (DID) methodology applied to the pricing model residuals to estimate the local impact of Luas line stations on residential property prices. I distinguish between property sales within walking distance of the Luas stop and those reasonably close (within 3 kilometres) but not within walking distance. Walking distance is estimated using the Google Distance Matrix API; the upper bound for walking distance is set at 20 minutes' walk. Not-within-walking-distance is defined as more than 25 minutes of walking time. These two subsamples (properties within walking distance of a Luas and those reasonably close but not within walking distance) constitute the treatment and control subsamples for my DID procedure. This DID methodology corrects the pricing model residuals for unobservable differences in the features of the treatment and control samples, provided that these unobservable differences remain constant before and after the treatment event or inside/outside walking distance. The DID approach assumes that, in the absence of any developments in transportation infrastructure, any trends in the residential property price model residuals are the same for properties within walking distance of a Luas station (the treatment group) and within 3 kilometres but not within walking distance (the control group).

I perform two separate tests. The first test (Test 1) examines whether there is a price effect in the vicinity of new Luas stations after they open. The second test (Test 2) examines whether the Cross-City extension (which improved the overall connectivity of the rail network) of the Luas line in December 2017 impacted prices near pre-existing Luas stations. Both tests use the residuals from the hedonic pricing model in place of raw prices, and a DID methodology to control for other unobservable influences on prices. In Test 1, a new Luas station increased nearby residential property sales prices by an average of 12.6%. In Test 2, the Cross-Luas extension increased existing Luas-accessible neighbourhood property sales prices by an average of 9.8% but this increase was notably stronger for Red Line stations (11.3%) than for Green Line stations (5.1%).

The paper uses a comprehensive database of residential property sales which combines Property Price Register (PPR) sales price records, Building Energy Rating data from the official BER register, and Daft.ie property listing information. An innovation of the paper is its application of a walking-time variable based on the geocoordinates of each property sale and the geocoordinates of its closest Luas station generated using the Google Distance Matrix API. This walking-time variable has never been used before in the published literature on Irish property prices.

The residential property price increases associated with Luas extensions are not a measure of the extensions' overall economic value. First, the endogenous or non-random placement of transportation infrastructure makes it difficult to evaluate causal effects (Redding and Turner 2016). The Luas line extensions involved careful decision-making by Luas planners regarding the optimal positioning of stations. Thus, the associated increases in residential property prices may partly reflect the endogenous choice of planners about the best placement of line extensions. Second, the residential property price increases attributed to Luas line extensions reflect only part of their true economic value. These line extensions may also facilitate or impact business growth, tourism, commercial property prices, and environmental outcomes. By acting as a substitute for car travel the Luas may lower road congestion and associated urban air and noise pollution. The residential property price impacts reflect an important, but only partial, measure of Luas economic value.

Section 2 discusses related research literature; Section 3 describes the data and gives descriptive statistics; Section 4 introduces the pricing model and test procedures and gives the empirical results; Section 5 concludes.

2. Related Literature

This paper uses a hedonic model of house prices to isolate the effect of rail transport infrastructure on property prices. Rosen (1974) developed hedonic price theory; the basic idea is to divide the price of a good into the component prices of individual characteristics of the good. Using a hedonic model has effectively become the standard approach in the economics literature on house prices. In the context of real estate prices, the set of property and neighbourhood characteristic variables included in hedonic models varies from paper to paper and often depends largely on data availability. Commonly included variables include number of bedrooms, bathrooms, and the square footage of the property.

One of the key issues associated with hedonic price models is that not all characteristics are observable or quantifiable. If one of the unobservable variables affecting house prices is correlated with an included variable, then this leads to omitted variable bias. Hedonic house pricing models often have many explanatory variables since there are many observable features that can affect house prices. Heene, Coyne, Francis, Maguire and Maguire (2014) note that hedonic house pricing models can often have an excessive number of free parameters to estimate relative to the size of the dataset. This large number of observable characteristics can lead to an overfitting problem in regression.

Cropper, Deck, and McConnell (1998) illustrate how errors in measuring the marginal contribution of household characteristic variables vary with the functional form of the hedonic price index. They suggest that simpler functional forms tend to perform better empirically. They find that the marginal prices of some hedonic attributes are estimated more accurately than others. The hedonic attributes that are more important to consumer utility tend to be more accurately estimated than those that are less important to consumer utility. They compare various functional forms including quadratic, linear, and Box-Cox specifications. My model uses the semi-log form where the dependent variable is the log of sales price and the independent variables enter the regression model linearly. This semi-log specification is the most commonly used specification in the literature (Mayor, Lyons, Duffy and Tol 2012; Meha 2017; McMillen and Redfearn 2010).

Land value often constitutes a substantial proportion of a residential property's overall price. The determinants of residential property prices can be broadly assigned to five categories: physical characteristics, locational attributes, accessibility, financial factors, and inflation relative to market price (Miller 1982). It is reasonable to expect both the 'locational' and 'accessibility' determinants of residential property price to be captured in a property's land value. Bostic, Longhofer and Redfearn (2007) focus on land value and its role in determining movements in residential property prices. Bostic et al. argue that changes in the value of residential property will depend on a property's 'land leverage' ratio (ratio of land value to total value). Bostic et al. focus much of their analysis on the land leverage ratio, a concept that has considerable relevance to the Dublin application in my paper. The land value to total value ratio of Dublin residential properties is likely to differ substantially between properties in inner Dublin and those in Dublin's outer suburbs. Bostic et al. note that the price gradient (the decrease in property prices moving away from the city centre) is connected to the land leverage ratio. This implies a 'land leverage' gradient, which is the decrease in land value to the total value of properties in the outer suburbs compared to more central locations. In the context of the price impact from Luas access, there are two offsetting effects: on the one hand, more central properties with higher land leverage ratios will be more impacted by the land value increase of Luas accessibility as per the land leverage gradient discussed in Bostic et al.; on the other hand, more distant suburban properties may stand to benefit more due to an absence of transportation alternatives. These two opposing effects result in an ambiguous result as to whether the percentage price impact will increase/decrease with distance from the central business district.

Mayor, Lyons, Duffy and Tol (2012) perform a hedonic analysis of the value of rail transport in Dublin. Their research pre-dates the development of the Property Price Register in Ireland, and so relies instead on a database provided by the largest Irish real estate firm, Sherry Fitzgerald. Their use of log sales price as the dependent variable is the same as in my analysis, as well as their use of quarterly time dummies to adjust for trends in Dublin property prices over the course of their time sample (2001-2006). They find that nearby rail access has a significantly positive impact on property sales price, and that this effect is largest for light-rail (the Luas) followed by heavy rail and commuter transit. They find that locations extremely close to heavy rail transit lines suffer a price discount (presumably due to a noise-related negative public externality) but this effect is not evident for the light-rail Luas. Mayor et al. include three types of rail transport in their analysis, they find that the residential property price impact of transport infrastructure depends partly on the availability of transport alternatives. Related to this, Debrezion (2007) compares studies of rail station value which include variables for other accessibility factors (bus, highway, walking) and those that do not. He finds that the inclusion of these other accessibility factors in the model reduces the estimated impact of proximity to the rail station on property values. He finds that studies that accounted for highway accessibility estimated lower rail transit proximity effects on property values by an average of 4% relative to studies that did not account for highway accessibility.

The 'Announcement Effect' refers to the immediate impact on financial market prices when the news of a future event is announced rather than later when the event occurs. According to Fama's (1970) efficient markets theory of financial markets, financial asset prices should adjust to new information as soon as the new information is announced even if the news refers to future events (such as future dividend payments) which have not yet occurred. There is often not a substantial immediate price impact from the announcement of future public externalities in residential real estate sales prices. Residential householders do not have the ability to speculatively purchase houses prior to public externalities being in place, since they would need to live there in the interim period. This creates a very high "transaction cost" to attempting to earn profits from the announcement of future public externalities such as rail stations. This is relevant to this paper since the Luas Cross-City extension (one of the event variables in my study) was discussed for many years before being implemented. Dublin real-estate industry professionals observed that, in certain areas, price rises did not reveal themselves until a few months after the Cross-City service became fully operational (Hilliard 2018). Luas line extensions were in discussion for many years before they became a reality, which creates difficulty in defining a clear and reliable 'announcement' date.

Recently, preference has been given to a DID methodology to isolate the effect of transport infrastructure on residential property prices (Wardrip 2011). The DID methodology is favoured for two main reasons. Firstly, it potentially controls for variables not included in the hedonic pricing model. Secondly, it adjusts for nearby spatial effects which might be correlated with transport infrastructure changes (Dube and Legros 2011). Meha (2017) uses a difference-in-differences method to measure the price impact of a newly opened commuter rail line between Upsalla and Alvsjo in Sweden on nearby residential property prices. He takes the difference between average log sales prices within 2 kilometres of a new station and those beyond 2 kilometres and measures this difference both before and after the new line opens. I use an analogous difference in differences test methodology, but I use walking time rather than kilometre distance to distinguish the treatment and control subsamples.

McMillen and Redfearn (2010) discuss nonparametric and other nonlinear methods for estimating and testing house price models. They prefer the nearest neighbour approach over the smooth kernel approach since it adjusts for differences in the number of observations in different areas of the sample region. They use the natural log of sales price as their dependent variable. They consider a number of nonlinear estimation techniques including locally weighted regression, kernel regression and conditionally parametric regression. They argue that nonparametric methods are superior to spline or polynomial-based estimation methods. They suggest that nonparametric methods are particularly relevant in the case of spatial variables. The methodology of my study is consistent with this, since I use a combined model which has a linear component for the hedonic variables and a nonlinear nonparametric function for the spatial variables. McMillen and McDonald (2004) estimate the value of rail transport in the Chicago area. Their main specification is similar to mine; it is a partially linear semiparametric model. As in my model, they use a linear assumption for all the explanatory variables except the spatial variables where they use a nonparametric specification. One difference from my specification is that McMillen and McDonald have the sum of two separate nonparametric functions for the two geospatial coordinates (north-south and east-west) whereas I use a single, two-variable nonparametric function.

Anglin and Gencay (1996) use a partially linear semiparametric hedonic model of property prices to estimate a house price model for the Windsor, Canada region. Anglin et al. use a Gaussian kernel for their weighting method and apply the Robinson (1988) two-step method for the estimation of partially linear nonparametric models, which is also the method used in my study. Their paper differs from my study in that the nonparametric component of their model has four variables (four property characteristics) whereas mine only has two (north-south and east-west location variables). They use a set of dummy variables for neighbourhood values, and these location variables are in the linear component of their model rather than in the nonparametric component.

3. Data

This paper uses a large database generously provided by Daft Inc.. This database amalgamates information from three primary sources: the publicly available register of Building Energy Ratings (BER) for all non-exempt Irish properties, the Daft Inc. proprietary database of residential property sales, and the publicly available Irish Property Price Register dataset. The amalgamated database provides data for each of the 85,267 recorded residential property sales in County Dublin during the period January 1st 2010 to December 20th 2019. For each property sale, the database includes the date of sale, sales price, street address, longitude and latitude geocoordinates of the property, and a variety of house characteristics including floor area, number of bedrooms, number of bathrooms, year of construction, and Building Energy Rating (BER) letter rating on a fifteen-category scale from A1 to G. Following Lyons, Lyons and Stanley (2016) I use a one-for-one procedure to convert each BER rating into a quantitative value from one to fifteen. The database includes an individual calendar day for each sale, but for modelling simplicity I convert the sale dates into quarterly frequency from 1 to 40, covering the 40 quarters (ten calendar years) of the data.

In the raw data, there were a small number of extreme values. For example, the highest sales price recorded in the database is €139 million. I delete the three observations in the dataset with sales prices greater than €50 million. Although the database contains 85,267 observations, only 58,315 of these include data on the number of bedrooms. Since this variable is shown to be important in my hedonic pricing model, I restrict the statistical analysis to property sales which include this item. Excluding these observations lowers the average price in the sample by 1.09% so the sales without data on the number of bedrooms tend to be on average very slightly lower-priced properties.

In order to measure the geographic position of each property and the distances between these properties and Luas stops, I convert the longitude and latitude of each sale observation into a Cartesian grid, with north and east positive, and south and west negative. The arbitrarily chosen zero point of my grid, for convenience and to provide context, is the Dublin Spire on O'Connell Street, which has longitude-latitude coordinates of (53.3498° N, 6.2603° W). In my Cartesian grid, the Spire has coordinates of (0,0). Any property sale that has a longitude east of the Spire has a positive X coordinate; any property sale north of the Spire has a positive Y coordinate. South and west of the grid are negative values for the two coordinates. The grid (X,Y) values are measured in kilometres and give the distance from the longitude and latitude of the Spire to the longitude and latitude of the property sale.

I also add information on the Luas. In 2004 the Luas was constructed as two separate lines: the Green Line, which now has a total of 35 stations, and the Red Line, which has 32 stations. During my sample period, January 2010 to December 2019, the Luas Green and Red lines underwent three extension projects in total, adding 27 stations. Figure 1 graphs the two Luas lines and the three extensions of it which were completed during my sample period. The list of all Luas stations and their opening dates are shown in Table A1 in the Appendix. As with property prices, I convert the day of a Luas station opening to quarterly frequency. Luas stations opened before the first quarter of the sample, are denoted with a dash in Table A1. This quarterly variable is used to sort the data into "before and after" subsamples. I also convert Luas station longitude-latitude coordinates into Cartesian grid co-ordinates.

Figure 1: Luas Map with Line B1, Line A1, and Line BX (Cross-City) Extensions. Thejournal.ie (2017)

First, I compute the linear distance between each property and each Luas station; then, I find the Luas station that is closest to each property and its distance in kilometres. By doing this, each property sale has two new associated variables: the station number of the closest Luas station and the kilometre distance to this station. When used in my tests for Luas proximity value, I will screen these variables and focus on property sales within 3 kilometres of their closest Luas station. Properties beyond 3 kilometres from any Luas station play a role in estimating the hedonic pricing model, but their pricing model residuals are not used in any of the difference-in-differences tests.

I create a parallel variable associated with future Luas stations that have not yet opened at the time the property is sold. For each sale property in each quarter, I find the closest Luas station that has not yet opened as of that quarter. Each sale property (other than those in the last 8 quarters of the sample, after the last station has opened) is associated with the closest unopened Luas station and its distance in kilometres. For each property, I compare its closest open Luas station and its closest unopened (future) Luas station. Only the closer of these two stations is associated with that property sale. In this way, there is a single closest Luas station associated with each property sale. Finally, I take each property sale within 3 kilometres of the closest Luas and using the Google Distance Matrix API calculate the walking time of each property to the closest Luas.

3.1. Descriptive Statistics

Although it was not planned beforehand, the choice of the Dublin Spire evenly split the County Dublin property sales dataset. The "average" property sale was located 0.344 kilometres north of the Spire and 0.442 kilometres west.

Table 1: Descriptive Statistics of the Key Variables for Property Sales

	Mean	**Standard Deviation**	**Min.**	**Max.**	**Obs.**
Price (€)	366434.7	289412.5	5400	18150000	58315
Log Price (€)	12.63316	.586168	8.594154	16.71418	58315
Floor Area (m^2)	101.2076	48.69189	13.33	940.10	58315
Bedrooms	2.873412	1.016848	0	63	58315
BER	10.25863	2.9025	1	15	58315
Age	41.9688	33.3596	0	266	58315
North	.3445064	7.5383	-16.02395	30.84928	58315
East	-.4424752	5.8741	-17.15745	14.08818	58315

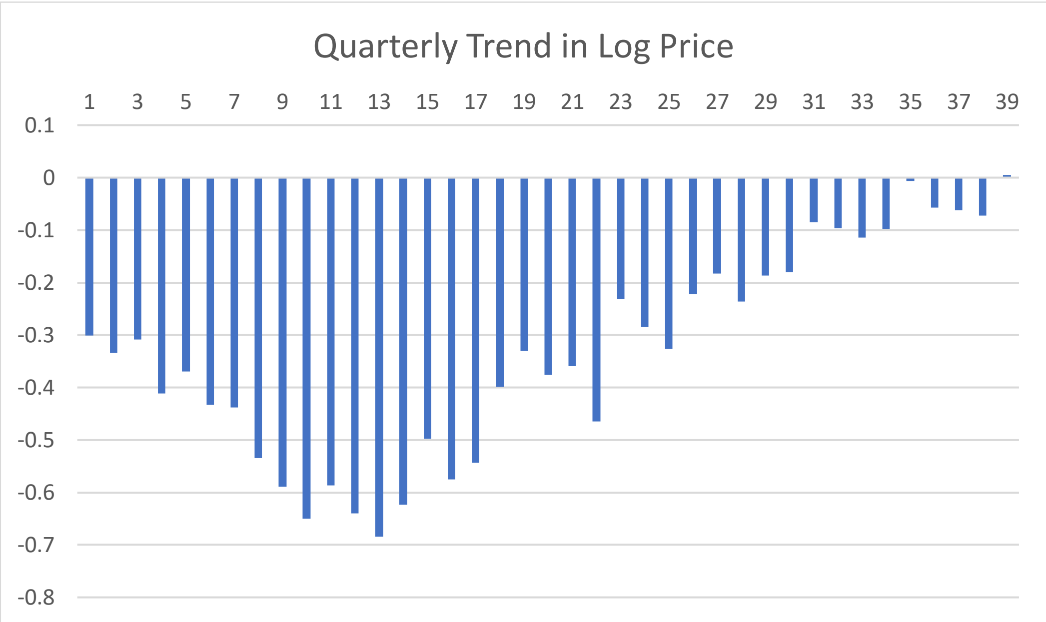
Table 2 examines average sales prices separated geographically. The first four rows consider sales prices located in the four "quadrants" of County Dublin using the Dublin Spire as (0,0). The next two rows consider property sales close to a Luas station (within 1 kilometre of a Luas station), between 1 and 3 kilometres to the nearest station, and the complementary set of property sales not within 3 kilometres of any Luas station.

Table 2: Descriptive Statistics of Sales Prices Sorted by Geographic Category

	Mean	Standard Deviation	Min.	Max.	Obs.
NE Price (€)	338213.3	209343.2	5400	3600000	15,876
SE Price (€)	541582	419241.4	5419	14000000	13,587
NW Price (€)	290712.7	165267.2	5900	3600000	11,910
SW Price (€)	305649	232416.7	5500		16,942
Price Near Luas (<1km) (€)	350452.1	252874.9	5900	4800000	13,924
Price Moderately Close to Luas (>1km,<3km) (€)	417004.6	353707.5	5419	14300000	17,909
Price Very Far from Luas (>3km) (€)	340639.4	252371.5	5400	18150000	26,482

Next, I looked at quarterly time trends in the key variables. I created forty quarterly dummy variables for all properties: for each property sale, each of the quarterly dummy variables is zero except for the quarter in which that property was sold, where it equals one. I regressed log prices on the 39 quarterly dummies (every quarterly dummy except the last one) and a constant. Since I did not include the final quarter in the set of dummy variables, the coefficients give the quarterly average log price levels relative to this final quarter (the fourth quarter of 2019). The results from this regression are shown in Figure 2 below. During my sample, average log property prices first dipped and then increased strongly until the second to last quarter, and then dipped very slightly. It is clear from this graph that a valid model of sales prices must include quarterly time dummies to account for this strong price trend (I will do this in my empirical work below).

Figure 2: Quarterly Time Trend in Log Price



This simple regression of log prices on quarterly dummies also allows me to construct an inflation-corrected version of log sales prices. Subtracting the relevant dummy coefficient from each log sales price gives the log sales price restated in units of the fourth quarter 2019 residential property price level. This inflation-corrected series is useful for comparing the prices of nearby properties when they sell at different times during the sample. This inflation-corrected series will be used later in the paper to create property price maps.

I decided to use four hedonic variables in my property price model: floor area, number of bedrooms, BER rating, and age (years since construction). The age of the property is calculated by subtracting the year of construction from the year of sale. To see whether these variables had time trends (for example if properties became smaller or more energy efficient over the sample period) I regressed each of them on the 39 quarterly dummies and a constant. The results are shown in Figures A2-A4 in the Appendix. None of these four variables have particularly strong time trends during the sample period, unlike log price as shown in Figure 2 above.

4. Empirical Analysis

In this section I first describe my empirical methodology, then the two versions of my pricing model for County Dublin property prices. I then use the unexplained residuals from the property pricing models to test for price effects of Luas station openings and the Cross-City extension.

4.1. Empirical Methodology

In order to test whether the Luas line has an impact on property prices I begin by estimating a hedonic model of log property prices. Then, using the residuals from this hedonic pricing model, I test whether there is evidence for the pricing impact of the Luas by looking for patterns in the model residuals.

My hedonic pricing model has the general form common in the literature reviewed in Section 2 above:

$$\mathrm{Log}(p_i) = c + b_1x_1 + \ldots b_kx_k + f(\mathrm{north}_i,\mathrm{east}_i) + e_i,\ i=1,\ldots,n$$

where c is a constant, $b_1,..,b_k$ are the linear coefficients associated with explanatory variables $x_1,\ldots,x_k$, and e_i is the unexplained residual. The explanatory variables include a set of quarterly dummy variables to capture property price trends, and property features such as square footage. I use two slightly different versions of the pricing model that differ in the specification of the location value function $f(\mathrm{north}_i,\mathrm{east}_i)$, either based on area dummies (the fully linear model) or based on weighted averages of nearby property prices (the partially linear semiparametric model). These two versions are discussed in detail later in this section.

Having estimated the pricing model, I use the estimated residuals e_i to test for the possible price impact of Luas access. These pricing residuals are sorted into before/after event treatment and control subsets. Using the pricing model residuals rather than the raw sales prices adjusts for possible differences in the average hedonic features of the collection of properties in the chosen subset of properties. Using the differences between treatment and control subsets adjusts for other nearby influences not captured by the hedonic model. The treatment subsets are properties within reasonable walking distance of the nearest Luas station (20 minutes or less walking time). The control subsets are properties close enough to a Luas station to share some of the local influences on property prices (within 3 kilometres), but too far from walking distance to be impacted by the practical advantages of Luas access (more than 25 minutes' walk). The properties within a walking time of 20 to 25 minutes are not included in either subset since their Luas accessibility is ambiguous. Beyond 25 minutes' walking time, most potential Luas users will choose another transport alternative.

Consider the pricing model residuals for property sales in the before-event treatment subset (e.g. near future Luas stations before they open), and an analogous set of residuals for property sales in the after-event treatment subset (e.g. near newly opened Luas stations after they open). Subtract from this the analogous difference for property prices in the control subsets. Under the null hypothesis of no price impact, the difference between the average differences should not differ significantly from zero. Under the alternative hypothesis that the event adds sales price value to nearby properties, the difference-in-differences will be positive. This can be tested using a t-test.

$$\text{t-test statistic} = ((m_{at} – m_{bt}) - (m_{ac} – m_{bc}))/(var_{at}/N_{at} + var_{bt}/N_{bt} + var_{ac}/N_{ac} + var_{bc}/N_{bc})\text{^}(1/2)$$

m_{at} = average pricing model residual in the after-event treatment subset
m_{bt} = average pricing model residual in the before-event treatment subset
m_{ac} = average pricing model residual in the after-event control subset
m_{bc} = average pricing model residual in the before-event control subset
var_{at} = variance of the pricing model residuals in the after-event treatment subset
var_{bt} = variance of the pricing model residuals in the before-event treatment subset
var_{ac} = variance of the pricing model residuals in the after-event control subset
var_{bc} = variance of the pricing model residuals in the before-event control subset
N_{at}, N_{bt}, N_{ac}, N_{bc} = number of observations in each of the four subsets.

One limitation of my approach is that this test statistic estimates a common mean difference for all relevant Luas stations and does not account for quality differences between stations. Stations can differ in their frequency of connections; for example, the Green Line stations beyond Sandyford (see Figure 1) have substantially lower service frequency than those closer in. Another limitation of my approach is that I do not attempt to identify the transport alternative set available in the vicinity of the various stations. I estimate a single average Luas access premium and do not attempt to differentiate it across stations.

In interpreting the estimated average impact, I make the simple assumption that the price impact begins in the quarter in which the Luas station opens. In the case of the Cross-City Luas extension, I additionally test using the quarter in which it received planning permission, as a test for an announcement effect. Measuring the announcement effects of the Line B1 and Line A1 extensions is outside the scope of the data used in my analysis. If announcement effects did occur for these line extensions, their estimated price impacts based on the opening dates may represent lower bounds for the total impact. Another important assumption I make is that the trends of residential property prices are the same for properties within walking distance of Luas stations and beyond walking distance, in the absence of any developments in transportation infrastructure. Note however that the DID methodology will mitigate any biases since it only uses properties reasonably close to the Luas station in calculating the difference in differences. The pricing model also includes quarterly dummies which will capture quarterly movements in Dublin-wide property price levels.

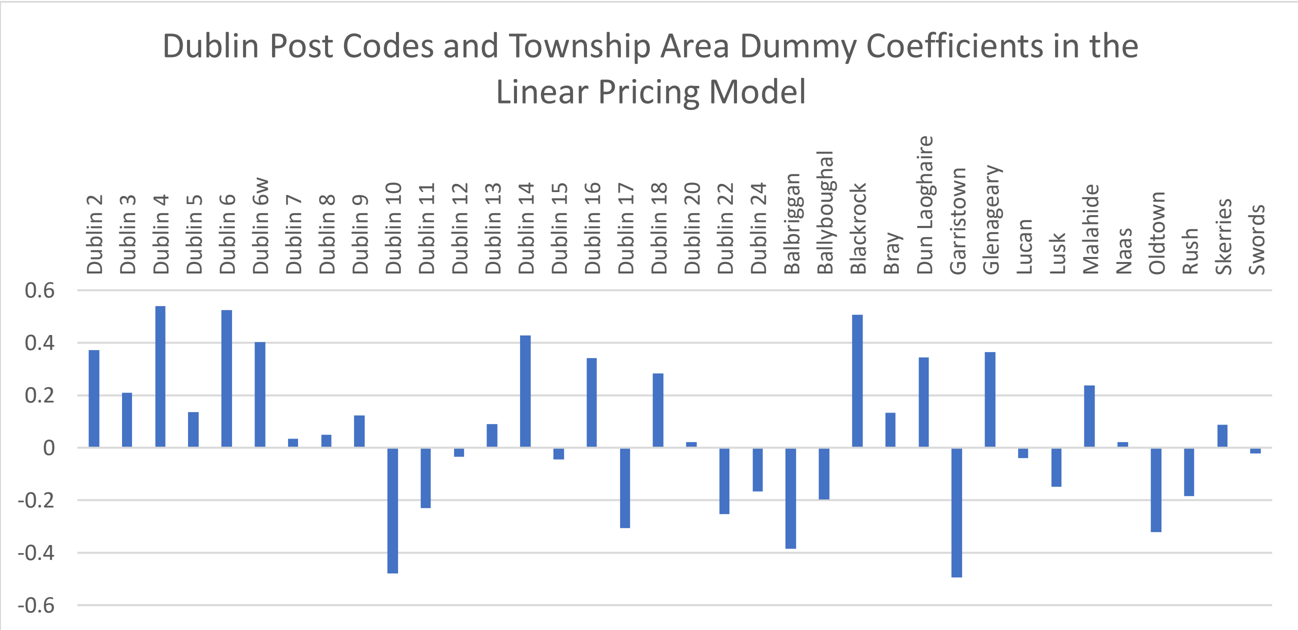
4.2. The Fully Linear Hedonic Pricing Model

The linear pricing model uses four hedonic variables, floor area, number of bedrooms, BER, and age, quarterly dummies to account for house price inflation, and local area dummies. In this version of the pricing model, the locational value function consists of a set of 36 dummy variables and their estimated coefficients based on Dublin postal codes and township designations:

$f(north_i, east_i) = h_1 d_1 + ... + h_{36} d_{36}$

where dummy variable d_j equals one if property i is in Dublin postal zone j and zero otherwise. The values $h_1,\ldots,h_{36}$ are the dummy variable coefficients which capture the relative price value of each postal zone. The thirty-seven local area dummies are based on Dublin area postcodes and County Dublin township designations (see Figure 3). I do not include a dummy variable for postal code area D1 (which is a central business district including the Dublin Spire) in order to avoid the dummy variable trap; hence the dummy coefficients represent the locational price value of each postal zone relative to postal zone D1.

Figure 3: Local Area Dummies



The fully linear model has the disadvantage that the spatial component of property prices is assumed to change abruptly when the postal district changes. Also, the postal code zones have no obvious link to locational values for property prices; they are designed to aid mail delivery rather than to differentiate property prices. Some postal codes have substantial variations in their property price locational characteristics across neighbourhoods within the same postal code.

4.3. The Partially Linear Semiparametric Hedonic Pricing Model

In order to have a smooth spatial component I consider the alternative of a partially linear semiparametric specification. In this version of the pricing model, the Dublin postal code dummy variables are replaced with smoothly varying local averages, using a kernel weighting scheme. I estimate f(north,east) at any value (north,east) by taking a weighted average of the log prices, giving more weight to nearby prices and less weight to those more geographically distant:

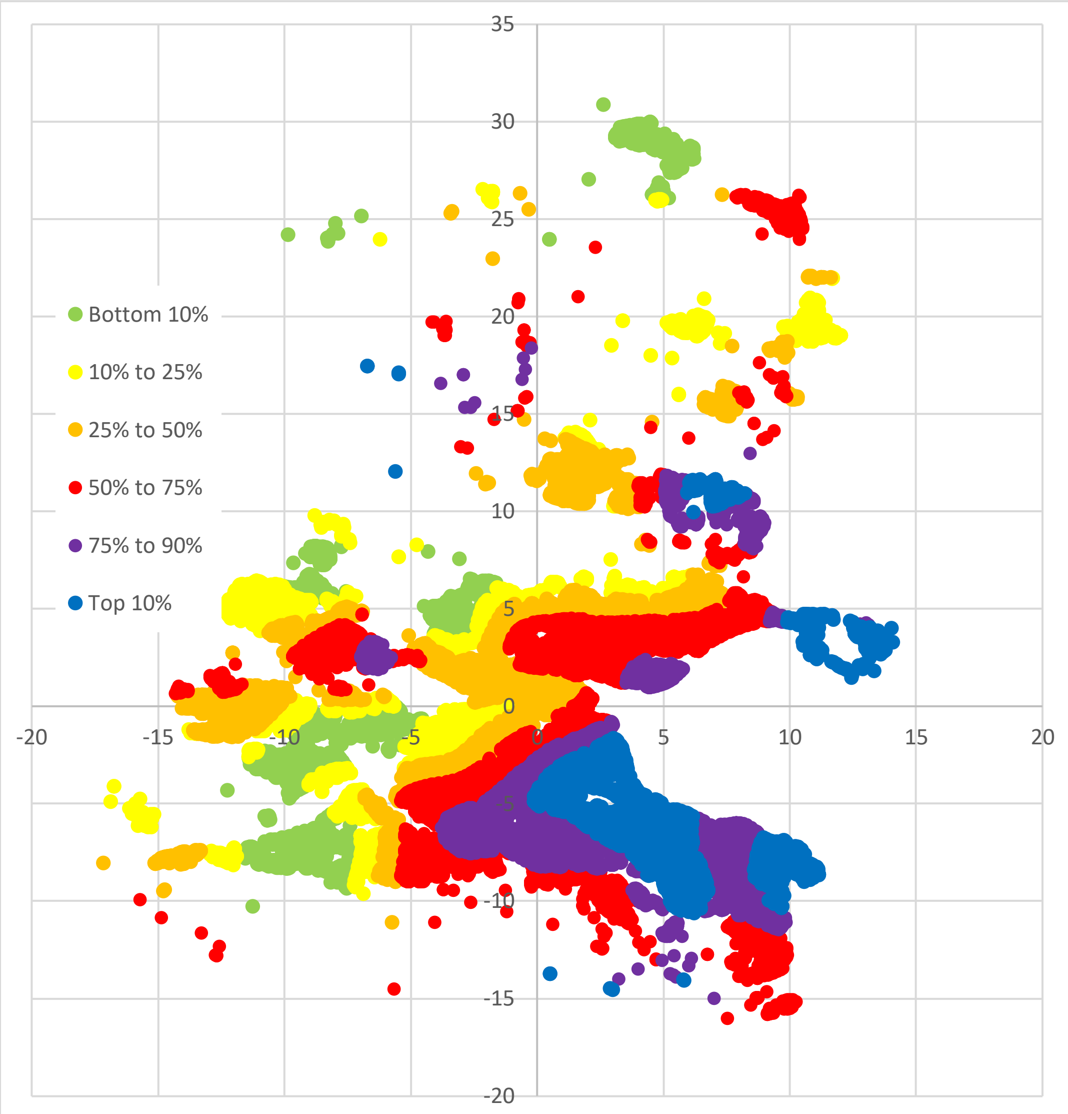
f(north,east) = weighted average of all log prices in the sample using weights w_i

$w_i = k(north_i, north, h_{north})*k(east_i, east, h_{east})/s$

where k() is a weighting kernel and s is the constant which ensures that the weights w_i sum to one. I use a Gaussian weighting kernel for k(), with smoothing parameters $h_{north} = 1.1418081$ and $h_{east} = 0.89576965$. These smoothing parameters are based on the rule-of-thumb method in Li and Racine (2007, p. 26).

Figure 4 shows the pricing map of County Dublin based on these local averages. Properties are colour-coded depending upon the category into which their local average log price falls. Blue denotes properties where the average price in the neighbourhood of that property is in the top 10% of all the local neighbourhood averages (above €625,712); purple is for properties where the local average price is in the top 25% to top 10% (between €518,378 and €625,712); red for those in the top half but below the top quarter (between €372,197 and €518,378); orange for those in the bottom half but above the bottom quarter (between €318,714 and €372,197); yellow for the bottom quarter but not the bottom 10% (€275400 and €318,714), and green for the bottom 10% (below €275,400). Local average property prices do not follow a simple east-west or north-south gradient in Dublin. There is a notable tendency toward high average prices in the southeast of the county, and along the east coast for its entire north-south range. The geographical pricing pattern is quite varied.

Figure 4: A Spatial Model of Dublin Residential Property Prices



Notes: For each property sale, the graph shows the average log sales price (detrended to the fourth quarter 2019 average price level) in the neighbourhood of the property, colour-coded according to its rank in the distribution of all Dublin neighbourhood average log prices. See the text for category breakpoints.

Next, I estimate the hedonic pricing model with the nonparametric pricing function in place of the postal zone dummy variables. I follow the Robinson estimation procedure for a partially linear parametric model as described in Li and Racine (2007, pp. 222-224). As in the figure above, I use time-detrended log prices so that all log prices are restated in units of fourth quarter 2019 price levels. First, for each observation, I subtract the local average log price from the log price variable to find the unexpected log price variable. I repeat this procedure for each of the four hedonic variables to find their unexpected components. Then, I do linear regression as in the case of the fully linear model, replacing the dependent variables and the four explanatory variables with their unexpected components.

The results from both versions of the pricing model are shown in Table 3. The coefficient estimates are similar in the two models, and the residuals from the two models have a correlation coefficient of 94.03%. The semiparametric version has a somewhat better overall fit, with a residual standard deviation of 0.308825, whereas the fully linear model has a residual standard deviation of 0.3286875. Interpreting the log-linear coefficients as percentage impacts, and using the partially linear model estimates, there is an approximately 0.5% increase in sales price for every extra square metre of floor space in the property. There is a 10% price premium per extra bedroom. Property age carries a price premium of 0.10% per year. Property prices fall by 0.36% for each unit increase in BER energy rating (note that zero is the “best” rating and 15 the “worst” hence the negative coefficient). The BER coefficient is somewhat smaller with the fully linear version. In general, the estimated coefficients using the partially linear semi-parametric model are very close to those from the fully linear model.

Table 3: Linear and Partially Linear Hedonic Pricing Models

	Floor Area	**Bedrooms**	**BER**	**Age**	**Dummy Variables**	**Residual Standard Error**	**R squared**
Fully Linear Model	.0051758 (126.11)	.1000468 (52.09)	-.001666 (-2.92)	.0012139 (22.15)	Quarterly and Area Dummies	0.328688	0.6851
Partially Linear Semiparametric Model	.0046667 (117.78)	.1047019 (57.40)	-.003625 (-6.64)	.000971 (18.49)	Quarterly Dummies	0.308825	.7220

Notes: The table shows the parameter estimates and their t-statistics (in parentheses) for the two versions of the geospatial-hedonic model of log sales prices. The R-squared for the fully linear model is adjusted for degrees of freedom; the partially linear model R-squared is not adjusted for degrees of freedom. Both models use the same 58,315 sales price observations.

4.4. Testing Price Impact on Properties Near Newly Opened Luas Stations

Having estimated the hedonic price model, I now use the partially linear pricing model residuals to test for price effects of Luas station openings (results using the fully linear model residuals are very similar and shown in the Appendix). In this subsection, I test whether opening a new Luas station increases nearby property prices. The test is obviously limited to the 27 Luas stations that opened during my sample period; see Table A1 for the list of Luas station opening dates.

In my test specification, a property sale is near a Luas station if it is within 20 minutes of walking distance. First, I find all property sales within 20 minutes walking distance of any future Luas station in any quarter before the station opens. This is the pre-event treatment group for my difference in differences test. I take the average of all the residuals from the pricing model for this pre-event treatment subset of property sales. Next, I find all property sales within a 25-minute walking distance but within 3km of any future Luas station before it opens. This is the pre-event control group. I take the average of the pricing model residuals from this control subset of properties. I find all property sales within a 20-minute walk of a Luas station after it opens (for the 27 stations that opened during the sample period). This is the post-event treatment group. I find all property sales more than a 25-minute walk from these 27 Luas stations after each one opens, but within 3 kilometres of it. This is the post-event control group. Taking the difference-in-differences gives the estimate of the price impact of Luas station openings on nearby property prices. Using the estimated variances gives the test statistic for whether this difference-in-differences is statistically significant (see Section 4.1 for the t-statistic calculation).

I also conduct the same tests on some designated types of Luas stations. I perform the test on only Green-line stations and only Red-line stations. Using local knowledge, I divide stations into "business district" stations and "non-business district" stations (see Table A1) and perform the test separately on them. The intuition for this test is that the "business district" stations are destinations rather than journey starting points, so residential properties near them might see less price impact from the station opening. The results are shown in Table 4 below. For all stations, there is an average positive price impact of 12.6% from the opening of a new nearby Luas station. The property price effect is slightly larger in non-Business District stations relative to Business District stations. For this test, it was not possible to separately estimate the price impact for the Red Line. The five new Red Line stations in my sample period opened 18 months (6 quarters) after the sample began and are all located in a relatively sparsely populated part of County Dublin. During the first six quarters of the sample, there were no recorded property sales in my database within walking distance of those five future stations. The two versions of the pricing model (fully linear and partially linear semiparametric) give very similar test findings so the results from the fully linear version are shown in the Appendix (Table A2).

Table 4: Property Price Impact of Newly Opened Luas Stations Using the Partially Linear Semiparametric Model

	Mean	**Before-Event Treatment Subsample Obs.**	**Before-Event Control Subsample Obs.**	**After-Event Treatment Subsample Obs.**	**After-Event Control Subsample Obs.**
All Stations	0.12593 (9.566)	1942	2117	2436	2840
Green Line	0.12252 (8.932)	1942	2117	1804	2493
Red Line	N/A N/A	0	0	632	347
Bus. Dist.	0.07455 (1.723)	504	136	205	65
Non Bus. Dist.	0.11558 (8.043)	1438	1981	2231	2775

Notes: The table shows difference in differences for treatment group (properties within walking distance) and control group (properties beyond walking distance) average price model residuals before and after new station openings. The t-statistics are shown in parentheses below the average difference in differences. The number of observations is shown for the before-event treatment group, before-event control group, after-event treatment group after event, and after-event control group.

Although the test statistics in Table 4 are significantly different from zero, it is not possible to interpret the mean differences as complete measures of the economic value of new Luas stations (see Section 1 above on endogenous placement and the interpretation of observed price impacts).

4.5. Testing Price Impact of the Cross-City Extension on Properties Near Existing Luas Stations

In December 2017, a Cross-City Extension was opened, linking the Green Line Luas with the Red Line Luas. This large infrastructure project greatly increased the potential usability of both lines since the travel range of each approximately doubled in length. In this subsection, I test whether the opening of the Cross-City extension of the Luas increased the sales prices of properties near existing stations.

I limit the subsample of this test to the 40 Luas stations which were open for the entire sample period. I find all property sales within 20 minutes walking distance of these Luas stations before the Cross-City extension was opened; these sales are the pre-event treatment group. Then I find all property sales outside 25 minutes of walking distance but within three kilometres of these Luas stations before the extension opened (the pre-event control group). I repeat the same procedure for sales after the Cross-City extension was opened, to obtain post-event treatment and post-event control groups of property sales. The results of the difference-in-differences test are shown below in Table 5. As in the last subsection, I also consider the test on Green-line only stations, Red-line only stations, business district stations, and non-business district stations.

The test finds a significantly positive effect, both overall and for the selected categories of stations. The magnitude of the pricing impact is larger for Red Line stations than for Green Line stations. The business district stations had a larger estimated pricing impact than the non-business district stations, going against my original intuition that the price impact would be smaller for "destination" stations in the business district than for those in more predominantly residential areas. In the case of the Cross-City Extension, there is the observable announcement date; August 2012 is when the project secured planning permission, corresponding to quarter 11 in my sample. I reran Table 5 using this announcement quarter in place of the quarter in which it became operational. All of the difference-in-differences test statistics become small and insignificant using this announcement date in place of the operational opening date (results not shown).

Table 5: Price Impact of the Cross-City Extension on Existing Stations Using the Partially Linear Semiparametric Model

	Mean	Before-Event Treatment Subsample Obs.	Before-Event Control Subsample Obs.	After-Event Treatment Subsample Obs.	After-Event Control Subsample Obs.
All Stations	0.09780 (10.866)	8893	7548	3705	2797
Green Line	0.05107 (3.576)	3112	2580	1385	1286
Red Line	0.11313 (9.660)	5781	4968	2320	1511
Bus. Dist.	0.10931 (5.637)	2630	1908	896	483
Non Bus. Dist.	0.07853 (7.780)	6263	5640	2809	2314

Notes: The table shows difference in differences for treatment group (properties within walking distance) and control group (properties beyond walking distance) average price model residuals before and after the Cross-City Luas extension for all existing Luas stations (not including new stations). The t-statistics are shown in parentheses below the average difference in differences. The number of observations is shown for the before-event treatment group, before-event control group, after-event treatment group after event, and after-event control group.

5. Conclusion

This paper tests whether improvements to the Luas light rail line impacted nearby residential property prices. I estimate a geospatial-hedonic model of property prices and use the price residuals from this model as the basis for the tests. I perform two tests. Firstly, I test whether the opening of new stations increased property prices in their vicinity, and secondly, whether the greater potential travel range and convenience provided by the Luas Cross-City extension increased property prices in the vicinity of existing stations. Both tests rely on a difference-in-differences methodology, where the treatment group consists of property sales price residuals within walking distance of a Luas station and the control group consists of property sales price residuals beyond walking distance but geographically not far (within a three-kilometre radius of the station). For both tests, I find statistically significant evidence for property price increases due to improved Luas access.

The geospatial-hedonic model of log property prices that I estimate combines a linear regression model based on quarterly time dummies and four hedonic variables (property floor area, number of bedrooms, energy rating, and age) and a model of locational value. In the partially linear semiparametric version of the pricing model, the nonparametric model of locational value is two-dimensional, based on the north-south and east-west locations of property sales. It is estimated by kernel regression methods. I also estimate a fully linear version of the pricing model in which the nonparametric locational value function is replaced with a set of linear dummy variables tied to Dublin postal codes and township designations. The estimation findings from the two versions of the model are quite similar, but the partially linear semiparametric version has greater theoretical and intuitive appeal.

The paper relies on a comprehensive database of residential property sales in Dublin for the ten-year period from January 2010 to December 2019. The database combines information from the Irish Residential Property Price Register, the Irish Registry of Building Energy Ratings, and the Daft Inc. property sales listing service. The database includes sales price and date for every registered property sale in the county, various property features such as floor area and number of bedrooms, and the geospatial coordinates of each property sold. The geospatial coordinates allow me to construct my nonparametric locational value function. Combined with the geocoordinates of all Luas stations, they also provide the necessary input to construct an innovative walking time variable, which gives the walking time from each property to its nearest Luas station. This variable is constructed using the Google Distance Matrix API.

Due to endogeneity issues and the public goods features of transportation infrastructure, the property price impacts that I measure are not unbiased measures of the true economic value-added from future Luas infrastructure developments. Nonetheless, the empirical and methodological contributions of this paper have considerable policy relevance. I confirm and extend the findings of Mayor et al. (2012) showing that new Luas stations have a positive price impact on nearby properties (without claiming that these price impacts entirely capture economic value). Although the empirical results using the partially linear semiparametric approach to Dublin price modelling do not differ strongly from those with a fully linear approach, this alternative methodology has considerable theoretical and intuitive appeal. The "walking time" metric that I implement in place of a linear-distance metric could also prove to be considerably useful for future policy analysis regarding Dublin's transportation.

Acknowledgements

I want to thank Professor Martina Kirchberger for her many helpful comments and insights on earlier drafts, Professor Ronan Lyons for sparking my interest in housing and transport infrastructure research, and Daft Inc. for generously providing access to their databases.

References

Anglin, P. M., & Gencay, R. (1996). Semiparametric estimation of a hedonic price function. *Journal of Applied Econometrics, 11*(6), 633-648.

Bostic, R. W., Longhofer, S. D., & Redfearn, C. L. (2007). Land leverage: Decomposing home price dynamics. *Real Estate Economics, 35*, 183-208.

Central Statistics Office. (2019). Regional population projections 2017-2036. Central Statistics Office, June 2019.

Cropper, M. L., Deck, L. B., & McConnell, K. E. (1998). On the choice of functional form for hedonic price functions. *The Review of Economics and Statistics, 70*(4), 668-675.

Debrezion, G., Pels, E., & Rietveld, P. (2007). The impact of railway stations on residential and commercial property value: A meta-analysis. *Journal of Real Estate Finance and Economics, 35*, 161-180.

Dube, J., & Legros, D. (2011). A spatial-temporal measure of spatial dependence: An example using real estate data. *Papers in Regional Science, 92*(1), 19-30.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance, 25*(2), 383-417.

Heene, M., Coyne, J., Francis, G., Maguire, P., & Maguire, R. (2014). Crisis in cognitive science? Rise of the undead theories. In: Proceedings of the Thirty-Sixth Annual Conference of the Cognitive Society, Cognitive Science Society, pp. 82-83.

Hilliard, M. (2018). Luas Cross City leads to property price jumps in Dublin 7. [Online] Irishtimes.com. Available at: https://www.irishtimes.com/news/ireland/irish-news/Luas-cross-city-leads-to-property-price-jumps-in-dublin-7-1.3726749 (Accessed 4th March 2020).

Li, Q., & Racine, J. S. (2007). Nonparametric econometrics. Princeton University Press.

Lyons, R. C., Lyons, S., & Stanley, S. (2016). The price effect of building energy ratings in the Dublin residential market. *Journal of Energy Efficiency, 9*, 875-885.

Mayor, K., Lyons, S., Duffy, D., & Tol, R. (2012). A hedonic analysis of the value of rail transport in the Greater Dublin Area. *Journal of Transport Economics and Policy, 46*(2), 239-261.

McMillen, D. P., & Redfearn, C. L. (2010). Estimation and hypothesis testing for nonparametric hedonic house price functions. *Journal of Regional Science, 50*(3), 712-733.

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.

Meha, B. (2017). The effect of transport innovation on property prices: A study on the new commuter line between Uppsala and Älvsjö. Uppsala University Working Paper, January 2017.

Miller, N. G. (1982). Residential property hedonic price models: A review. In C. F. Sirmans (Ed.), Urban Housing Markets and Property Valuation. Research in Real Estate, 2, 31-56.

National Transport Authority. (2019). Bus & rail statistics. Statistical Bulletin March 2019.

Redding, S. J., & Turner, M. A. (2016). Transportation costs and the spatial organization of economic activity. In *Handbook of Urban and Regional Economics*.

Robinson, P. M. (1988). Root-n consistent semiparametric regression. *Econometrica, 56*, 931-954.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *The Journal of Political Economy, 82*(1), 34-55.

Thejournal.ie. (2017). From December, this is what the Luas Map will look like. [Online] Available at: https://www.thejournal.ie/luas-cross-city-map-3334143-Apr2017/. (Accessed 13th March 2020).

Wardrip, K. (2011). Public transit's impact on housing costs: A review of the literature. Centre for Housing Policy, Insights from Housing Policy Research, August 2011.

Appendix

Table A1, Panel 1 – Red Line Luas Stops and Their Opening Dates

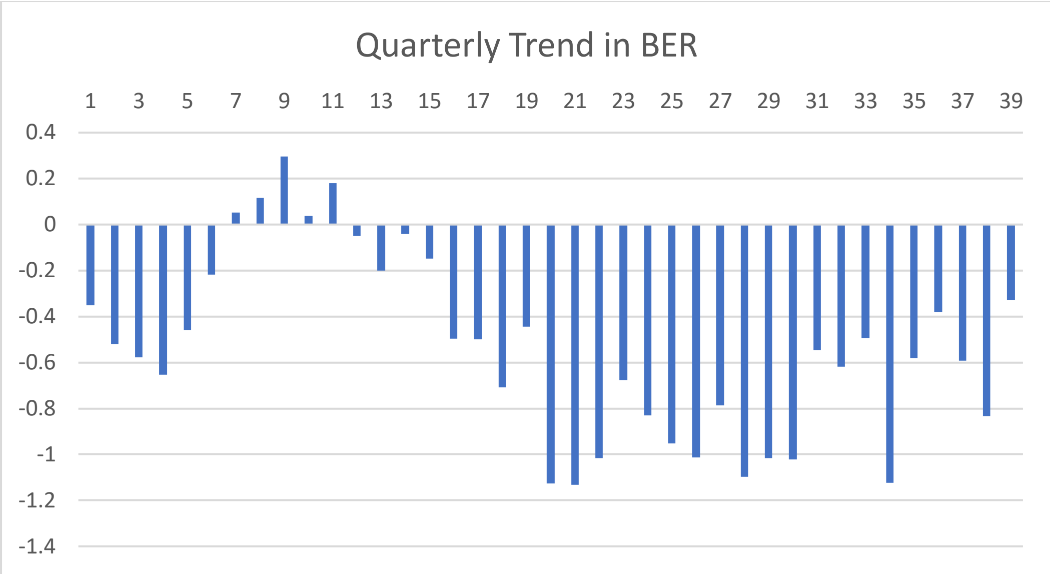
Luas Stop	Opening Date	Sample Quarter	Luas Stop	Opening Date	Sample Quarter
Saggart	02/07/2011	7	Suir Road	26/09/2004	-
Fortunestown	02/07/2011	7	Rialto	26/09/2004	-
Citywest Campus	02/07/2011	7	Fatima	26/09/2004	-
Cheeverstown	02/07/2011	7	James's	26/09/2004	-
Fettercairn	02/07/2011	7	Heuston	26/09/2004	-
Belgard	26/09/2004	-	Museum	26/09/2004	-
Tallaght	26/09/2004	-	Smithfield	26/09/2004	-
Hospital	26/09/2004	-	Four Courts (BD)	26/09/2004	-
Cookstown	26/09/2004	-	Jervis (BD)	26/09/2004	-
Kingswood	26/09/2004	-	Abbey Street (BD)	26/09/2004	-
Red Cow	26/09/2004	-	Busarus (BD)	26/09/2004	-
Kylemore	26/09/2004	-	Connolly (BD)	26/09/2004	-
Bluebell	26/09/2004	-	George's Dock (BD)	08/12/2009	-
Blackhorse	26/09/2004	-	Mayor Square NCI (BD)	08/12/2009	-
Drimnagh	26/09/2004	-	Spencer Dock (BD)	08/12/2009	-
Goldenbridge	26/09/2004	-	The Point (BD)	08/12/2009	-

Table A1, Panel 2 – Green Line Luas Stops and Their Opening Dates

Luas Stop	Opening Date	Sample Quarter	Luas Stop	Opening Date	Sample Quarter
Broombridge	09/12/2017	32	Cowper	30/06/2004	-
Cabra	09/12/2017	32	Milltown	30/06/2004	-
Phibsborough	09/12/2017	32	Windy Arbour	30/06/2004	-
Grangegorman	09/12/2017	32	Dundrum	30/06/2004	-
Broadstone	09/12/2017	32	Balally	30/06/2004	-
Dominick	09/12/2017	32	Kilmacud	30/06/2004	-
Parnell (BD)	09/12/2017	32	Stillorgan	30/06/2004	-
Marlborough (BD)	09/12/2017	32	Sandyford	30/06/2004	-
Trinity (BD)	09/12/2017	32	Central Park	16/10/2010	4
O'Connell Upper (BD)	09/12/2017	32	Glencairn	16/10/2010	4
O'Connell GPO (BD)	09/12/2017	32	The Gallops	16/10/2010	4
Westmoreland (BD)	09/12/2017	32	Leopardstown Valley	16/10/2010	4
Dawson (BD)	09/12/2017	32	Ballyogan Wood	16/10/2010	4
St Stephens Green (BD)	30/06/2004	-	Carrickmines	16/10/2010	4
Harcourt (BD)	30/06/2004	-	Laughanstown	16/10/2010	4
Charlemont	30/06/2004	-	Cherrywood	16/10/2010	4
Ranelagh	30/06/2004	-	Brides Glen	16/10/2010	4
Beechwood	30/06/2004	-			

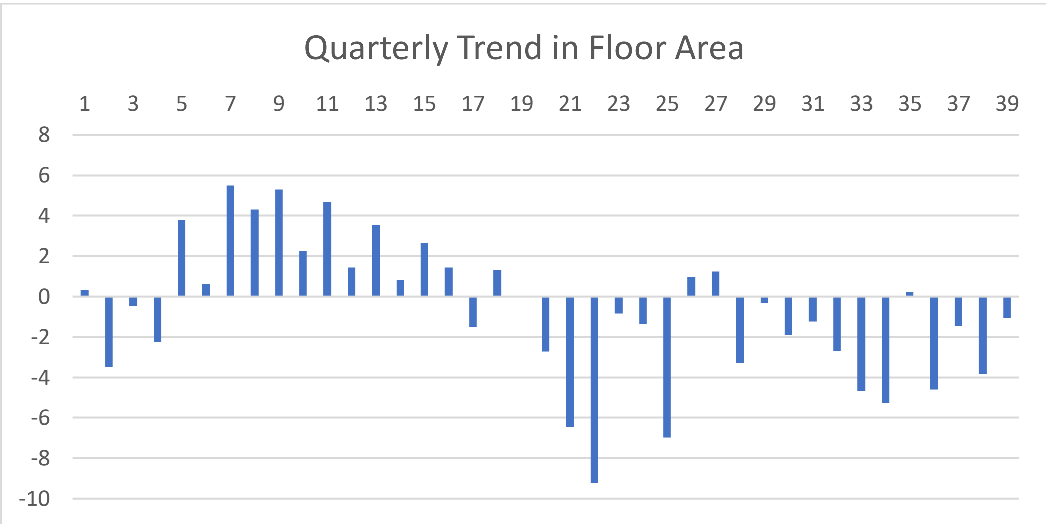
Notes: Stations Marked 'BD' are designated Business District locations.

Figure A1: Quarterly Trend in BER



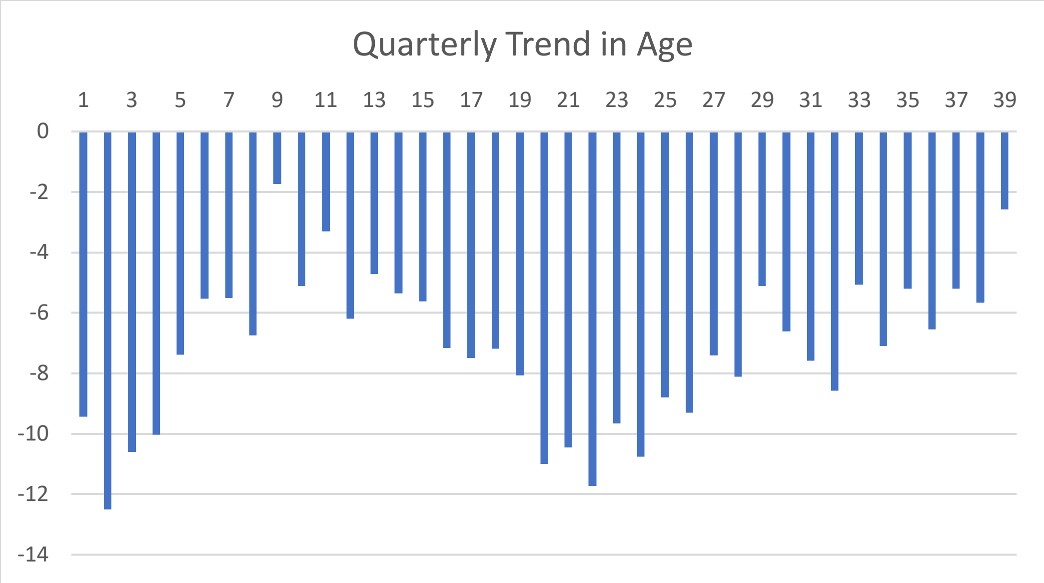
Notes: The figure shows the estimated coefficients from a regression of sale property BER (on a scale of 1 to 15) on a constant and a set of 39 dummy variables indicating the quarter in which the property was sold. The final quarter (the fourth quarter of 2019) is not included in the regression set of dummy variables.

Figure A2: Quarterly Trend in Floor Area



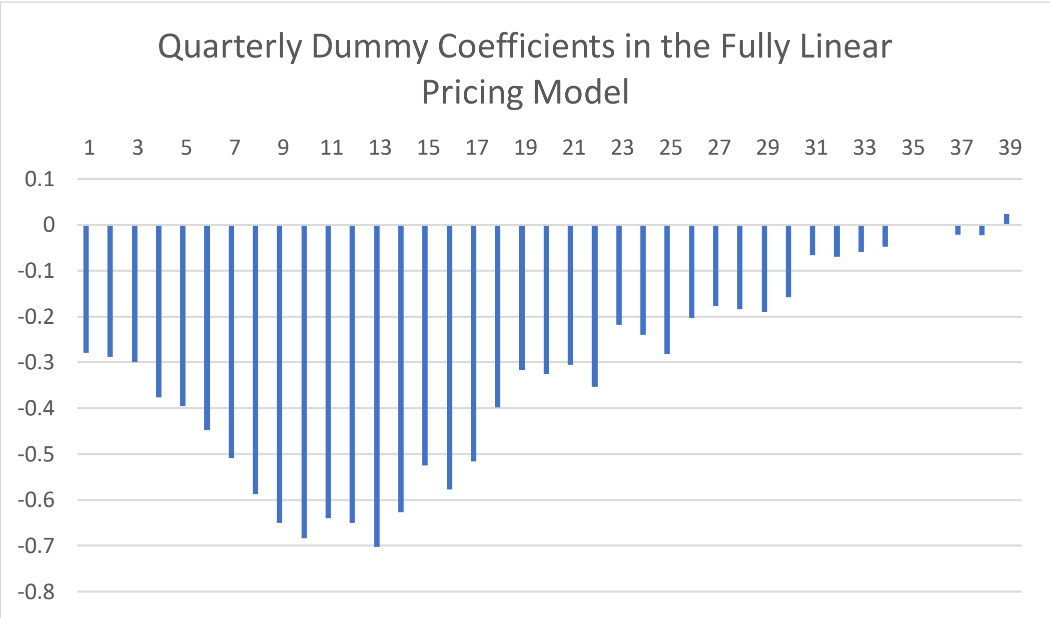
Notes: The figure shows the estimated coefficients from a regression of sale property floor area (in square metres) on a constant and a set of 39 dummy variables indicating the quarter in which the property was sold. The final quarter (the fourth quarter of 2019) is not included in the regression set of dummy variables.

Figure A3: Quarterly Trend in Age



Notes: The figure shows the estimated coefficients from a regression of sale property age (in years) on a constant and a set of 39 dummy variables indicating the quarter in which the property was sold. The final quarter (the fourth quarter of 2019) is not included in the regression set of dummy variables.

Figure A4: Quarterly Dummy Coefficients in the Fully Linear Pricing Model



Notes: The figure shows the estimated coefficients for the set of 39 quarterly dummy variables in the fully linear geospatial hedonic price model. The final quarter (the fourth quarter of 2019) is not included in the regression set of dummy variables.

Table A2: Property Price Impact of Newly Opened Luas Stations Using the Fully Linear Pricing Model

	Mean	Before-Event Treatment Subsample Obs.	Before-Event Control Subsample Obs.	After-Event Treatment Subsample Obs.	After-Event Control Subsample Obs.
All Stations	0.16638 (11.947)	1942	2117	2436	2840
Green Line	0.19032 (13.220)	1942	2117	1804	2493
Red Line	N/A N/A	0	0	632	347
Bus. Dist.	0.06031 (1.351)	504	136	205	65
Non Bus. Dist.	0.15462 (10.199)	1438	1981	2231	2775

Notes: The table shows difference in differences for treatment group (properties within walking distance) and control group (properties beyond walking distance) average log price residuals before and after new station openings. The t-statistics are shown in parentheses below the average difference in difference. The number of observations is shown for the before-event treatment group, before-event control group, after-event treatment group, and after-event control group. The pricing model residuals are from the fully-linear version of the pricing model.

Table A3: Price Impact of the Cross-City Extension on Existing Stations Using the Fully Linear Pricing Model

	Mean	Before-Event Treatment Subsample Obs.	Before-Event Control Subsample Obs.	After-Event Treatment Subsample Obs.	After-Event Control Subsample Obs.
All Stations	0.09358 (10.1315)	8893	7548	3705	2797
Green Line	0.05197 (3.564)	3112	2580	1385	1286
Red Line	0.09654 (8.057)	5781	4968	2320	1511
Bus. Dist.	0.09285 (4.521)	2630	1908	896	483
Non Bus. Dist.	0.07866 (7.695)	6263	5640	2809	2314

Notes: The table shows difference in differences for treatment group (properties within walking distance) and control group (properties beyond walking distance) average pricing model residuals before and after the Cross-City Luas extension for all existing Luas stations (not including new stations). The t-statistics are shown in parentheses below the average difference in difference. The number of observations is shown for the before-event treatment group, before-event control group, after-event treatment group, and after-event control group. The pricing residuals come from the fully linear version of the pricing model.