Transport for London and Department for Transport

Crossrail baseline evaluation

Evaluation of Crossrail pre-opening property impacts

Issued | 27 April 2022

Job number 249567-00

Ove Arup & Partners Ltd
13 Fitzroy Street
London
W1T 4BQ
United Kingdom
www.arup.com

ARUP

Contents

Executive summary

Our approach

- In 2016 Arup and Volterra were commissioned by Transport for London (TfL) and the Department for Transport (DfT) to conduct a Baseline Evaluation study of Crossrail. The announcement of Royal Assent for the Crossrail Bill was made in 2008. This signalled the formal go-ahead for the project. This report documents our work on an econometric evaluation of the pre-opening (i.e. announcement-related) impacts of Crossrail on property and planning outcomes. The report uses data to 2019, and so excludes the Covid-19 period.

- Our work is the most detailed evaluation of the pre-opening impacts of Crossrail to date. We used difference-in-difference econometric models with fixed effects, which are classified at Level 3 of the Maryland Scientific Methods Scale[1], and are considered to be a robust impact evaluation technique. We tested average impacts at the line-wide level, produced results for specific sections of the route, and analysed data around specific stations. **This method allowed us to calculate the direct impacts of the Crossrail announcement alone**, independent of other factors affecting market conditions. The same methodology can be used for a future post-opening evaluation (and an evaluation of the remaining pre-opening period).

Impact on residential properties

- Areas immediately around future Elizabeth line stations (0-500m away) experienced relatively high population growth and employment density growth over the pre-announcement period, compared to other parts of London. This suggests that **the Crossrail route was correctly targeted at providing transport capacity and connectivity to support areas with growing transport demand**.

- We found evidence that the Crossrail announcement **had a positive 2% (2.2%) impact on residential house prices in the areas closest to the stations (0 – 500m away) in the period between 2008 and 2019**. The largest growth was experienced along the Western London section (4%), and the lowest growth along the Eastern section (1.9%). Growth rates were similar to the average (at 1.9%) in deprived areas, and areas with wider regeneration or growth initiatives. The case for Crossrail was not predicated on residential price increases and although part of the funding for the line derived from developer contributions, the programme has not been funded on the expectation of any land value uplift captured directly from residential households.

[1] https://whatworksgrowth.org/resources/the-scientific-maryland-scale/

- **Factors other than the Crossrail announcement made a larger contribution to the overall increase in residential house prices between 2008 and 2019**. Residential housing unit prices per square metre in the whole study area (10 mile buffer area around all stations) increased from an average of £4,900 in 2008 to £6,000 in 2019, implying average growth of 21%. These rates were similar regardless of the distance to a Crossrail station, and all ranged from 20% to 33%.

Impacts on residential housing unit prices in the following areas	**500m to nearest Crossrail station**	**500-1000m to nearest Crossrail station**	**1-2 km to nearest Crossrail station**
Average impact on the whole route	2%	0	-2%
Western section outside London	3%	-3%	0
Western London section	4%	0	0
Central section	2.5%	2%	0
Southeastern section	3%	0	-6.5%
Eastern section	2%	-2%	-2%
Deprived areas	2%	-1%	-1%
Areas with associated major development	2.5%	0	-1.5%
Areas with wider regeneration or growth initiative	2%	0	-3%
Areas expected to benefit from additional labour supply	2%	0	-2%

- **We found evidence that the Crossrail announcement decreased the value of residential housing unit prices by around 2% for areas between 1 km and 2 km from the stations**, between 2008 and 2019. The current analysis is insufficient in explaining in detail what may be behind this result, although displacement may have been a factor.
- The Economy, Planning and Regeneration technical report found that **54,725 new homes were delivered within 1km of stations between 2008 and 2021 (two years longer than the baseline period ending in 2019)** which is around the same as the 57,000 estimated by the Crossrail Impact Study[2] in 2012 and lower than the 91,000 estimated by Future London[3] in 2016. These figures take into account both direct and indirect home-

[2] Crossrail Property Impact Study led by GVA: 08675 Crossrail Property Impact Study v27.indd

[3] Crossrail Property Impact and Regeneration Study led by GVA in 2016: gva_crossrail_property_impact_regeneration_study.pdf (rackcdn.com)

building, whereas the pre-opening announcement analysis presented in this report estimated only the direct impact on planning applications within a 2km buffer around stations. The original Business Case for Crossrail[4] did not include specific references to housing delivery targets, and was focused on supporting London's future economic development and employment. These figures take into account both direct and indirect home-building, whereas the pre-opening announcement analysis presented in this report estimated only the direct impact on planning applications within a 2km buffer around stations.

- **We found no evidence that the Crossrail announcement had a statistically significant impact on the number of planning applications within 1km of stations between 2008-2019,** which suggests that the announcement had little direct impact on building activity close to stations, compared to other areas of London. We were unable to quantify the overall (direct and indirect) impacts of the Crossrail announcement on the number of new homes delivered, through translation of the number of applications into the number of housing units.

Impact on commercial properties

- **We found evidence that the Crossrail announcement had a positive impact on office rent values in central London, with areas within 500m from stations gaining from subsequent increases of 3%.** This line-wide impact is mainly driven by the 3% increase in the Central section of the line since the majority of commercial units are located there. We found significant impact for none of the other sections. This finding is in line with previous evidence; a better connection between the regional centre and the smaller centres typically benefits businesses in the regional centre. There was no significant difference from the average 3% impact for areas with associated major development (3%), areas with wider regeneration or growth initiative (2.5%) and areas expected to benefit from additional labour supply (3%).
- Interestingly, offices located between 500m-1km from future Elizabeth line stations often had a higher rate of increase due the announcement than those closer to stations (a 7% line-wide increase and an 8% increase for the central section). **The highest levels of office increase due to the Crossrail announcement were in deprived areas (10%), areas with associated major development (7.5%), and areas expected to benefit from additional labour supply (7.2%).**
- Similar to residential housing unit prices, our findings suggest that **factors other than the Crossrail announcement made a larger contribution to the overall increase in London office rents between 2008 and 2019**. Average office rents in the whole study area stood at around £53 per square feet in 2019, an increase of around 30% since the end of 2008.

[4] https://learninglegacy.crossrail.co.uk/documents/appraisal-business-case-crossrail/

Impacts on office rent values in the following areas	500m to nearest Crossrail station	500-1000m to nearest Crossrail station	1-2 km to nearest Crossrail station
Average impact on the whole route	3%	7%	0
Western section outside London	0	0	0
Western London section	0	0	0
Central section	3%	8%	0
Southeastern section	0	6.5%	0
Eastern section	0	0	0
Deprived areas	0	10%	0
Areas with associated major development	3%	7.5%	0
Areas with wider regeneration or growth initiative	2.5%	6%	0
Areas expected to benefit from additional labour supply	3%	7%	0

Untested impacts

- We were not able to conduct a robust evaluation on the volume of office and retail development planning activity, and retail and industrial rents, due to the data constraints and methodology requirements. This does not mean that those impacts have not occurred.
- Our analysis did not use data on land availability and zoning around stations, due to challenges in gathering this data. We expect that using it would have provided useful contextual information on the differences in opportunities between sections and stations along the route.

Future uses of this research

- This evidence for changes in property values due to announcement effects alone may be factored into future economic cases and into funding plans for major projects. We expect further research on the post-opening impacts to add to the evidence base further.
- In the course of research we have explored an option of using a bespoke measure of accessibility to represent Crossrail impacts in the post-opening period. In addition to a binary radius approach, this continuous treatment approach may capture the impacts that Crossrail is expected to have on journey times and capacity across the rail network in London, and beyond.

Abbreviations

DiD	Difference in difference (method)
DfT	Department for Transport
FE	Fixed Effects
LSOA	Lower layer Super Output Area
TAG	Transport Appraisal Guidance
TfL	Transport for London
WWCLEG	What Works Centre for Local Economic Growth

1 Introduction

1.1 This commission

Arup and Volterra were contracted by Transport for London (TfL) and the Department for Transport (DfT) to deliver a baseline study on the pre-opening impacts of the Crossrail programme. Our work on this commission started in 2016, and is due to conclude in 2022, once the Elizabeth line is operational.

Within the overall context, this report provides a quantitative analysis of the baseline position and pre-opening impacts of Crossrail on selected property and planning indicators, seeking to answer the question ‘has the expected impact of the new line caused changes in the property market already, before the line has opened, over and above what would have happened anyway?’ Such effects are called ‘announcement effects’.

This analysis was updated in March 2020 and used data up to the end of 2019. As such it does not account for any impacts from Covid-19.

1.2 The Crossrail programme and the Elizabeth line

Throughout this report, we use Crossrail to refer to the construction project, and the Elizabeth line to refer to the railway once it is open. As our work focussed on the pre-opening impacts, we tend to use Crossrail more frequently, and later in the report, we use the shorthand of the “impact of Crossrail” for the impact of the announcement and implementation of the Crossrail programme.

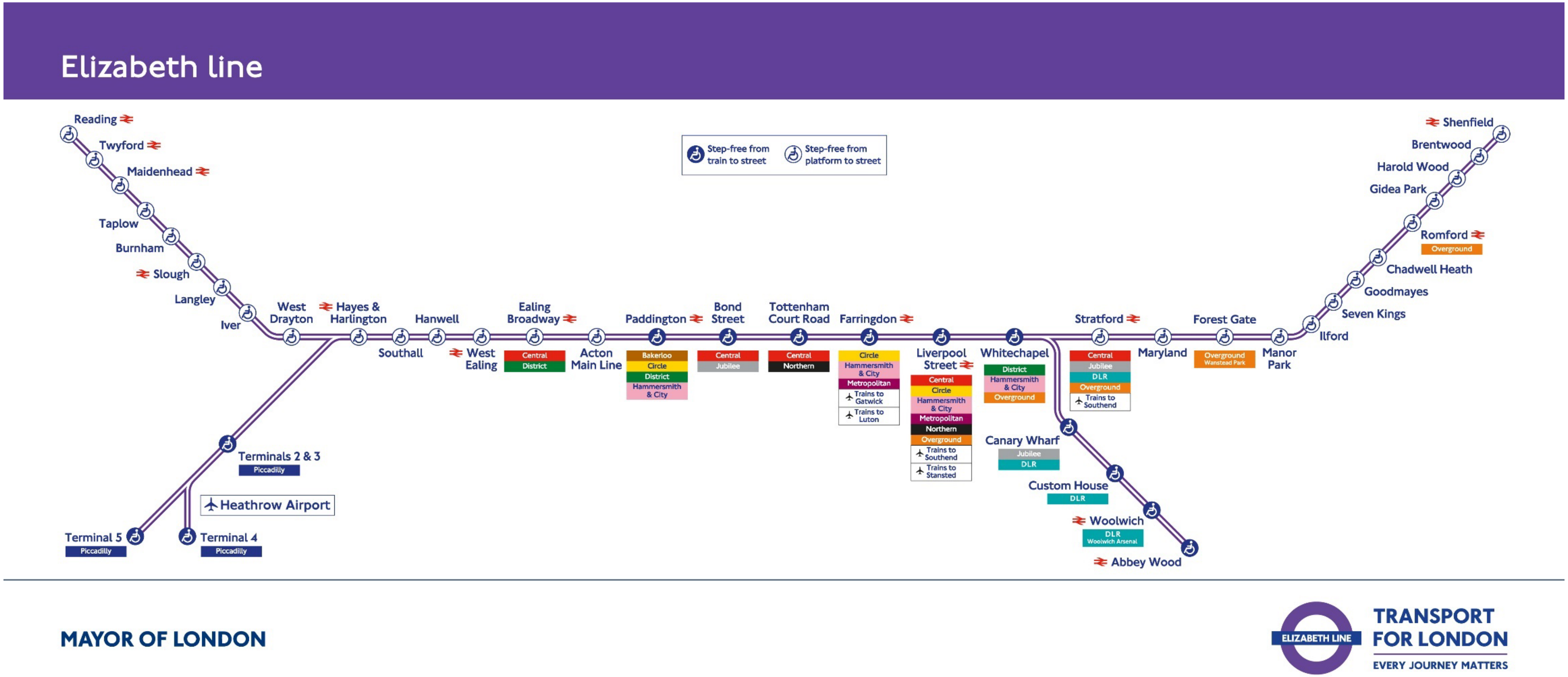
The Crossrail programme is constructing and commissioning a 118-km high frequency, high capacity railway line to serve London and the South East. It involves construction of 41 km of new tunnel under London that will be integrated with existing commuter railway lines to the east and west of London, allowing through services to travel the length of the route.

Once open, the Elizabeth line will provide a 10% increase to London’s rail network capacity, bringing an estimated 1.5 million additional people to within 45 minutes of central London and reducing journey times for existing rail users. Trains are 200 metres long with space for up to 1,500 passengers each.

The line will connect 41 stations between Reading and Heathrow in the west and Abbey Wood and Shenfield in the east. It includes 10 new stations and improvements to many existing station facilities. All stations will have step-free access and upgrades include new ticket halls, platform extensions, footbridges and provision of wayfinding information.

The Elizabeth line is planned to open in stages, beginning in 2019 with the transfer of existing national rail services currently operating from Liverpool Street to Shenfield and Paddington to Heathrow and Reading to operate as TfL Rail. This is to be followed by the opening of the central section of the Elizabeth line between Paddington and Abbey Wood in the first half of 2022. The final stage is the connecting of the eastern and western branches of the railway into the central tunnels and train running through the central section, later in 2022.

Figure 1 Elizabeth line route map



1.3 Supporting the growth of London

Once open, the Elizabeth line has the potential to support the growth of the capital, and to be a major stimulus to the economy. In a city that was growing, pre-Covid-19, by more than a million people every 10 years (Figure 2), and where total network capacity is already constrained, the line will add an estimated additional 10% to the network capacity.

Data presented in the London Infrastructure Plan 2050[5] show that observed trips in London were outstripping the forecasts on which the 2011 Mayor's Transport Strategy[6] was based. **Error! Reference source not found.** Figure 2 shows the growth in total trips over recent years compared to those forecasts. This outperformance is due largely to faster than expected population growth.

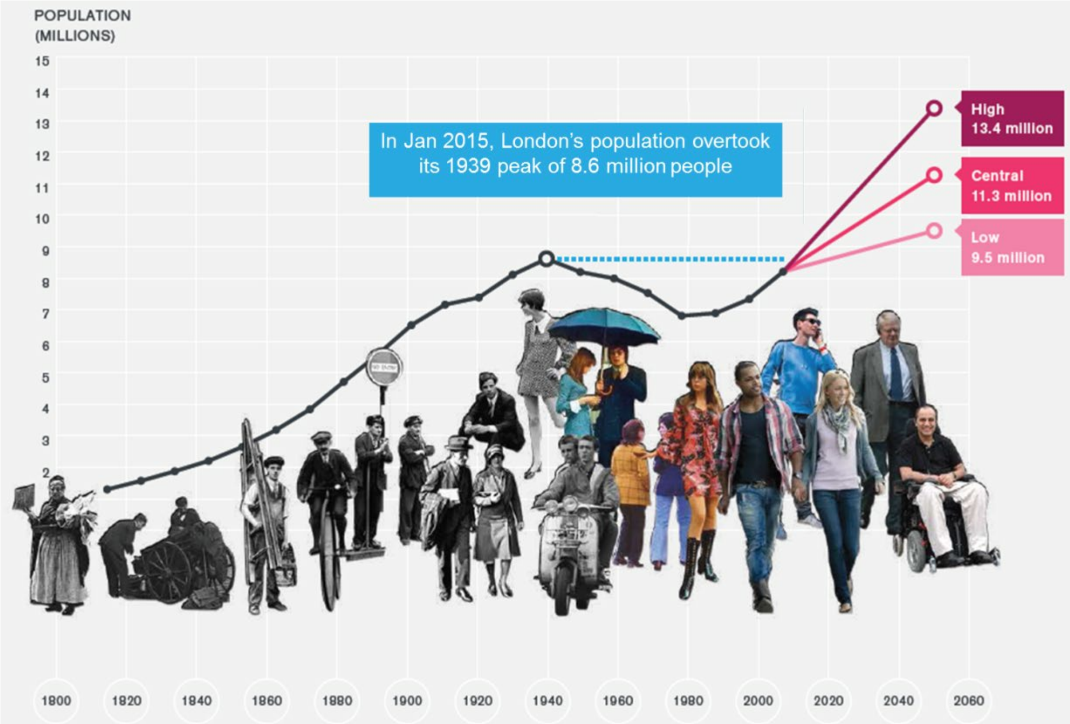
Within that, public transport trips are growing at a faster rate than total trips as the new population is more likely to use public transport. This is because younger individuals of working age (17-44), the demographic groups that comprise the majority of migration into London, have a higher propensity to use public transport than other age-based cohorts. It is also the case that inner London boroughs have seen higher growth in their populations and house building activity than elsewhere in the city (2008-18) and residents in these boroughs tend to have a higher propensity to use public transport than their outer London peers.

---

[5] www.london.gov.uk/what-we-do/business-and-economy/better-infrastructure/london-infrastructure-plan-2050

[6] https://www.london.gov.uk/what-we-do/transport/transport-publications/mayors-transport-strategy

Figure 2 London's population: past, present and future



Source: GLA, London Infrastructure Plan 2050, Transport Supporting Paper 2014; showing GLA Economics population projections (derived from ONS population projections)

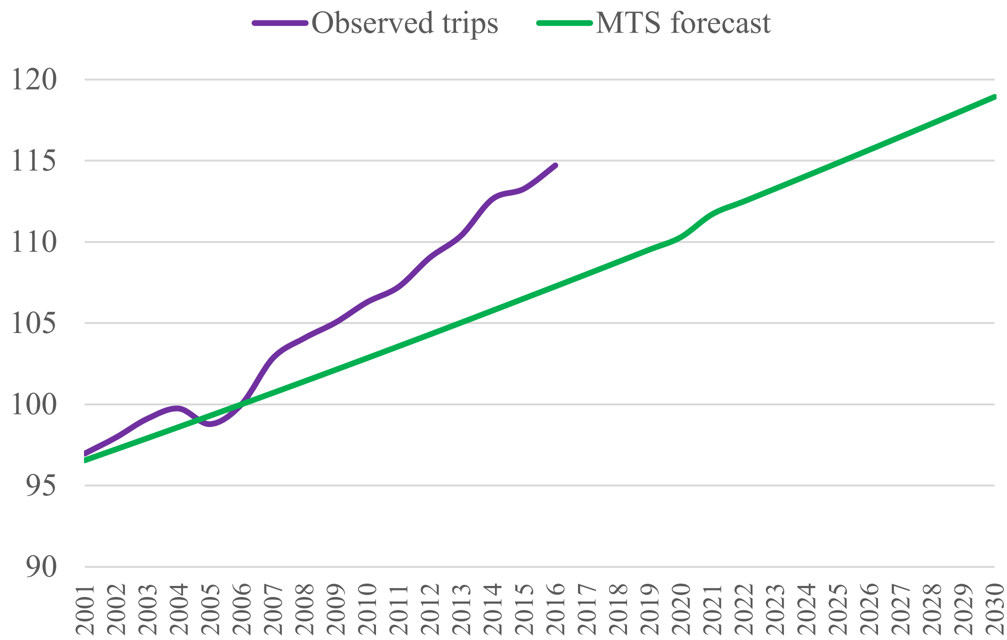
When proposals for Crossrail were initially developed, they were focused on reducing journey times and catering for expected growth by providing extra system capacity. The railway is also widely believed to have the potential to drive housing delivery and development, and those expected wider benefits played an important part in the final scheme business case. The analytical method used to make the case for those benefits was the first ever application of the Wider Economic Benefits (since then renamed to Wider Economic Impacts) methodology within the DfT Transport Appraisal Guidance framework[7]. The wider economic case was predicated on two hypothesised impacts:

- a positive relationship between travel-time accessibility and productivity, defined by agglomeration elasticities; and
- the role of the Elizabeth line in overcoming transport capacity constraints that would otherwise have prevented growth.

This pre-opening impact analysis is focused specifically on whether there is evidence of changes in the property market having occurred before the line begins operation. This means that the results are insufficient at assessing whether the estimated wider economic impacts were achieved.

[7] https://www.gov.uk/guidance/transport-analysis-guidance-tag

Figure 3 Observed growth in trips compared to forecast trips in London



Source: GLA, London Infrastructure Plan 2050, Transport Supporting Paper 2014; showing TfL trip data.

In a country such as the UK, which has a comparatively large economy and an existing mature transport network, there are two key aims of public spending on transport infrastructure:

- to respond to growing demand so that increased congestion and longer travel times do not increase costs and act as a constraint to growth (essentially, a preventative approach);
- to stimulate growth in local economies (rather than just respond to it) by generating agglomeration effects which increase productivity. This is based on the empirical observation that cities have higher levels of productivity than elsewhere and that, in general, the largest cities have the highest productivity levels.

The current evaluation evidence base on the impact of the wider economic impacts of transport infrastructure is summarised in a 2015 review by the What Works Centre for Local Economic Growth (WWCLEG)[8]. It broadly supports the hypothesis that transport can drive wider growth outcomes but notes that the evidence base for rail (in particular) is relatively sparse and focused on property impacts. The WWCLEG review finds no robust evidence of the impact of rail on employment and only one piece of evidence on business outcomes.

As part of the baseline evaluation this study has collected data and defined the methodology for post-opening evaluation(s) of the impact of the Elizabeth line.

[8] For a fuller discussion of the evidence base see ‘Evidence Review: Transport’, What Works Centre for Local Economic Growth, 2015.

1.4 The property market impacts of transport investment

This study is being undertaken in the context of an increasing focus on defining the land value uplift impacts of rail infrastructure. Evidence suggests that transport investment can lead to property value or land value uplift, although effects depend on distance from stations and can vary over time[9]. This uplift, if it can be captured, has the potential to be a source of funding for future projects but robust evidence from the UK on the scale of this impact is limited to one study[10].

Part of the funding package for the Crossrail programme relied upon the expectation of land value uplift and wider productivity impacts to businesses, which was captured through the traditional mechanisms (revenues, development contributions and general taxation) as well as through a new business rates supplement (BRS). The BRS was a two percentage point supplement above general business rates, levied on businesses in properties with a rateable value of more than £50,000. The BRS revenues were estimated to contribute £4.1 billion out of original £14.8 billion capital cost estimate (in 2008 prices, excluding rolling stock and Network Rail costs).

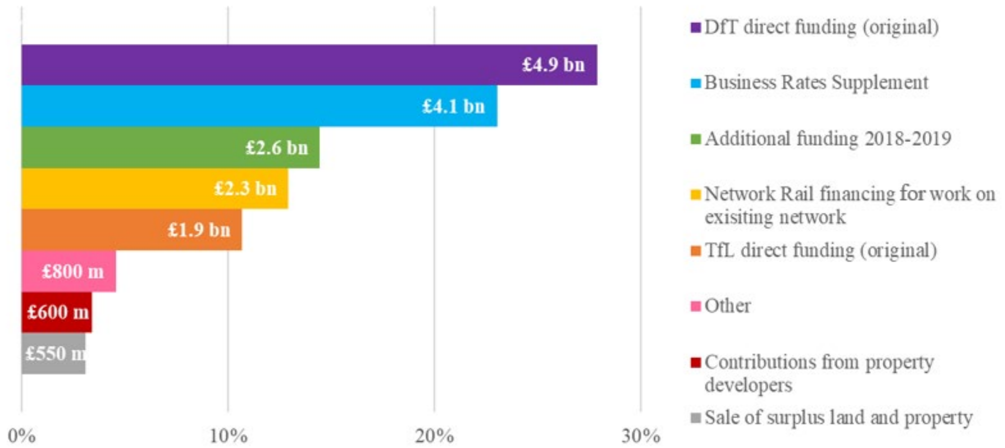
Also, drawing on the expected developments and land value uplift from Crossrail, the Mayoral Community Infrastructure Levy (MCIL) was introduced in 2012 to help finance the new line. MCIL is calculated based on the net additional floorspace for all planning permissions granted in central London, the northern part of the Isle of Dogs and within 1km of a Crossrail station for the rest of London with some exemptions (e.g. medical spaces, education or affordable housing). MCIL collections have so far contributed £865m in funding for Crossrail.

An annual update on Crossrail delivered to Parliament on 24th July 2018 confirmed an increase in the overall funding envelope for the delivery of the project from £14.8 to £15.4 billion, also stating that over 60% of the project's funding has been provided by Londoners and London businesses.[11] At the end of 2018 the total funding envelope increased to £17.6 bn[12].

Figure 5 below suggests that 60% is an approximation of London's contribution, given that some of the revenue used to finance TfL's investment comes from fares raised in the South East, and a proportion of infrastructure operators' contributions may come from central government or national enterprises.

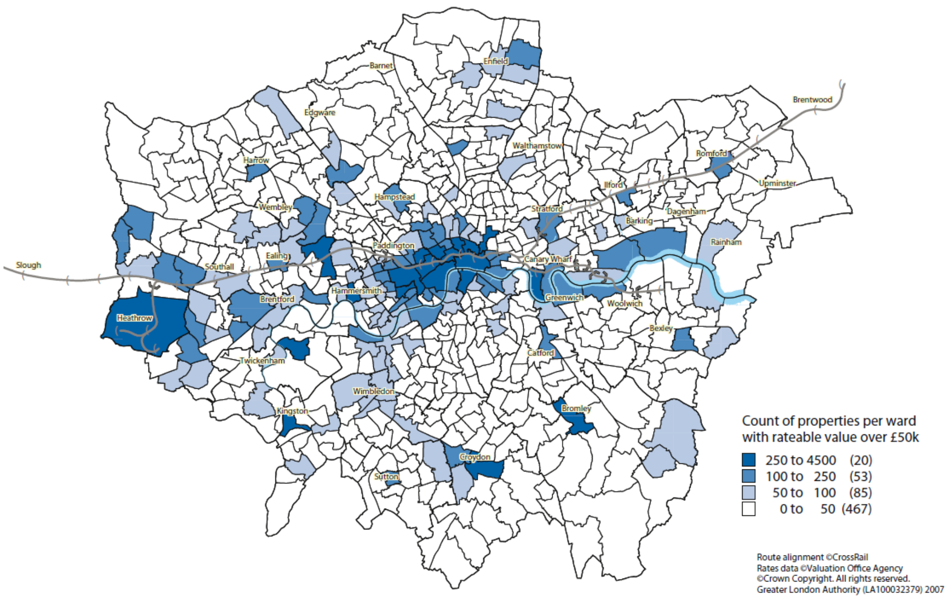
[9] WWCLEG (ibid).
[10] WWCLEG (ibid)
[11] Note that since then Crossrail's forecast outturn costs have risen significantly and its delivery schedule has been extended.
[12] https://www.gov.uk/government/speeches/crossrail-update-10-december-2018

Figure 4 Funding sources for the Crossrail delivery (July 2019)



Source: Arup graphic based on information available at https://www.crossrail.co.uk/about-us/funding

Figure 5 Businesses most likely to pay the Crossrail business rate supplement (based on 2005 ratings list – properties with rateable values above £50,000)



Source: GLA, 'Intention to levy a business rate supplement to finance the Greater London Authority's contribution to the Crossrail project: Final Prospectus', January 2010.

Better evidence on the existence, scale and timing of actual property value impacts could support the development of future funding packages for infrastructure across the UK, and will be of particular interest to major proposed schemes from Northern Powerhouse Rail and East West Rail to local metro schemes in cities across the UK.

1.5 This study in context

This study takes a quantitative approach to answering the question 'has the expected impact of the Elizabeth line caused changes in the property market already, before the line has opened, over and above what would have happened anyway?' Such effects are called 'announcement effects'. We refer to 'announcement impacts' or 'pre-opening impacts' interchangeably throughout the report, referring to all impacts that occur between the announcement and the opening of the line.

The analysis presented in this paper seeks evidence of the existence and timing of announcement effects by looking at the impact on property indicators. Other studies have found weak evidence in support of the existence of announcement effects of transport infrastructure on property and other indicators[13].

The study considers not whether the Elizabeth line *will* drive property impacts in future (when it is open) but whether we can find statistical evidence that it has done so to date. It should also be noted that the study adopts a quantitative approach to property impacts driven by the line announcement. It considers effects on planning activity, delivery and prices but does not consider in its scope effects on typologies, affordability, quality nor the activity during the opening delay period.

Other studies have attempted to establish the announcement impacts by comparing outcomes immediately around future Elizabeth line stations with the relevant area average (often at borough level). Among others, a study led by researchers from Leeds University[14] looked at land value uplift around rail investments and notably Crossrail. Accompanying this report are appendices which provide greater detail on individual model runs and compare the findings of this analysis with that of other published studies and where and why they may differ.

In 2018, Crossrail Ltd. published a study by the property company GVA[15]. It focuses primarily on predicting the future property market impacts of the Elizabeth line and it also uses property data to draw conclusions about the pre-opening impacts of Crossrail. The approach taken in the GVA study was to compare changing property outcomes in the areas around stations to the change in their respective area averages. An approach such as this does not control for the impact of other factors driving changing property prices, or the observation that areas in the immediate proximity of stations are often quite different from the wider surrounding area. For this reason, the impacts attributed to Crossrail may not be reliable. It also does nothing to address issues of selection bias (see section 2.3), whereby the locations that will be served by the Elizabeth line are chosen in a non-random way which reflects their characteristics such as their perceived growth potential.

It is impossible to be exact about the correct counterfactual because we can never observe this hypothetical alternative scenario in which the investment in the Crossrail programme had not been made. But we can use quantitative techniques to try to generate the best possible comparator.

[13] WWCLEG (ibid)

[14] Nellthorp, J., Ojeda Cabral, M., Johnson, D., Leahy, C. and Jiang, L. (2019). Land Value and Transport (Phase 2): Modelling and Appraisal. Final Report to TfN, WYCA and EPSRC. Leeds: Institute for Transport Studies, University of Leeds. https://doi.org/10.5518/100/18

[15] See, for example, GVA, 'Crossrail Property Impact & Regeneration Study 2012-2026', January 2018 and Hamptons International 'Linking housing markets: The effect of Crossrail on housing markets in London', 2016 – a summary of other studies is set out in an appendix.

This study uses econometric techniques (described in more detail in Chapter 2 of this report) to try to 'strip away' the impact of the many other factors that are known to affect these outcomes such as background economic changes, demographic and geographic factors, and other known developments, and estimate what proportion of the observed property market changes in areas around the stations are attributable to the announcement of the Crossrail programme.

The remainder of this report sets out the rationale for, and detail of, the methodology and data sources used for the pre-opening property impact analysis and its findings.

- Chapters 2 and 3 describe the methodology employed for the empirical analysis;
- Chapter 4 presents results of the analysis of Crossrail pre-opening impacts on the office property market;
- Chapter 5 presents results of the analysis of Crossrail pre-opening impacts on the retail property market;
- Chapter 6 presents results of the analysis of Crossrail pre-opening impacts on the industrial property market;
- Chapter 7 presents results of the analysis of Crossrail pre-opening impacts on the residential property market;
- Chapter 8 presents evidence of the pre-opening impact of the Elizabeth line on the volume of development planning activity; and
- Chapters 9 and 10 provide a summary of findings and final comments.

2 Key concepts

This chapter provides a high-level overview of the methodology and an introduction to some of the key concepts underpinning the approach to the econometric analysis. We explain how these general concepts apply to the Crossrail baseline evaluation.

Our evaluation approach is a difference-in-difference modelling with fixed effects. We explain the common trends assumption required for that technique to be used. We describe how that method enables us to tackle the selection bias, which is one of the biggest challenges when constructing a counterfactual in policy evaluation.

2.1 Time of treatment

This is a study of the impacts of the Crossrail announcement. For the purpose of the analysis we use the Crossrail Royal Assent date, July 2008, as the announcement date. We use annual data, so any observations from calendar year 2009 onwards are considered as having taken place after the announcement.

It is possible that there may be some overlap of these ‘announcement effects’ in the latter part of the pre-opening period with the treatment effect, due to the actual implementation in its earliest phases, which started in 2015. There has been a new train operator, TfL Rail, on the eastern section since May 2015, the phased introduction of new rolling stock since 2016, and Heathrow Connect services to Paddington were also taken over in 2018. However, property value impacts of the earliest phases of implementation might reasonably be expected to be minimal as there had been no new services in the periods covered by the analyses in this report (up to 2019).

A second type of announcement effect may have occurred from summer 2018 when it was announced that the opening of the Elizabeth line would be delayed from late 2018 to a later date. We account for this event by using year-fixed effects (see chapter 2.7), but we do not distinguish any potential impacts of the delay.

Furthermore, the announcement of Crossrail may have been anticipated for some time before July 2008. Proposals for the opening of the Elizabeth line, or a similar scheme, have been under development with more or less certainty for several decades. The Abercrombie Plan of 1947 makes reference to the need for cross-London rail lines, and the 1974 London Rail study was the first time the name ‘Crossrail’ was coined, in reference to a proposal to join the heavy rail lines to the west and east of London via a new line with stops at Paddington, Marble Arch, Bond Street and Liverpool Street, and a second tunnel to connect the southern rail network at Victoria and London Bridge, with the two lines intersecting at Leicester Square.

Similar schemes came forward in the 1980s and 1990s, with the latter resulting in a private bill to parliament that was ultimately rejected on affordability grounds, although the proposed route was safeguarded at that point. The ultimately successful Crossrail Bill was presented to Parliament in February 2005. As such, there is potential for pre-announcement impacts, which we do not investigate in detail in this report.

2.2 Defining the counterfactual

Defining the wider economic impact of public policy, and particularly place-based policies such as transport infrastructure investment, is challenging. Investment never takes place in a vacuum and there are always other factors to disentangle which may also affect economic outcomes. Some of these are difficult to observe in data or occur in parallel with transport investment and are therefore challenging to account for in quantitative analytical approaches.

To evaluate the impact of a project, we would ideally have an unambiguous understanding of what would have happened if the project had not been delivered. We could then compare that to what has actually happened. In evaluation research that “alternative state of the world” is called a counterfactual. Of course, it is not possible to observe the counterfactual. We cannot definitively know how a housing market, or businesses, would have performed in a world without the Elizabeth line.

In some policy contexts it is possible to create randomised experiments (similar to the approach taken in medical trials) which if well specified, can provide us with a high degree of confidence that we are accurately observing an alternative state of the world. When planning and evaluating railway lines it is neither possible nor desirable to do this.

Consequently, evaluation research concentrates on elaborating methods aiming at approximating the counterfactual as robustly as possible. This is predominantly achieved by identifying both:

- a treatment group – a group of subjects (people, businesses, locations) impacted by a project,
- a comparison (control) group [16] – a group of subjects which are not affected, whose performance serves as a counterfactual.

We can use this to infer the effect of the Crossrail announcement if we can demonstrate with reasonable confidence that they would otherwise have changed in the same way. Crucially, this means that the comparator areas should be similar in all relevant characteristics other than having received the investment, and therefore be expected to be a good proxy for how the treatment area would otherwise look.

If we are confident in the credibility of the comparison area(s) then we can be reasonably confident that the difference between the change in performance observed between the two groups, before and after the announcement of Crossrail or the opening of the Elizabeth line, forms our estimate of the impact. The

[16] Whilst this is often also referred to as ‘a control group’ in the literature, we tend to use ‘a comparison group’ throughout this study in order not to confuse the reader with a ‘control variables’ term.

performance of the comparison group is taken as a proxy for the ‘alternative state of the world’.

In line with standards of evidence adopted by the WWCLEG[17], we consider ‘robust’ evidence to mean as a minimum a before-and-after comparison (or continuous analysis during a period of change) of the changes observed in the area which benefits from the transport investment (the ‘treatment area’), with the changes observed in one or more comparison area(s).

2.3 Selection bias

Transport projects are usually planned and aligned to serve places considered to have the highest growth potential. Those places are not chosen randomly – there are reasons for them to be ‘selected for treatment’. If the Crossrail route alignment was planned to serve the places considered to have highest growth potential it is very difficult to be sure whether any observed change in outcomes is genuinely due to Crossrail, or due to the underlying growth potential that caused it to be selected as a Crossrail location in the first place. We mentioned in section 1.5 that some previous studies on Crossrail have not controlled for selection bias.

In addition, factors influencing decision-making can be difficult to observe and measure. These unobserved factors might include aspects such as the motivation or drive of individual business leaders or local politicians, the impact of other major transport investments, prevailing market conditions for businesses located in the area, changing fashions, tastes or political climates. Many of these cannot easily be represented by data points, and statistical analyses must select control variables carefully to include only the most directly relevant or risk losing statistical power (and thus, confidence in the results). These “unobservables” have the potential to over- or under-state estimates of impact where they are not controlled. Similar problems arise when we suspect that there may be reverse causality (for example, where the existence of strong growth pressure is a cause, rather than an effect of transport infrastructure investment).

[17] For more information on the relative strengths and weaknesses of different analytical approaches see the WWCLEG Guide to Scoring Evidence, available at http://www.whatworksgrowth.org/resources/scoring-guide/

To consider whether and to what extent selection bias is a problem for this study we look at whether areas around stations appear on average the same as, or different from, other parts of London. In our baseline analysis we found that areas immediately around future Elizabeth line stations (up to 500 metres from the stations):

- experienced relatively highest population growth over the baseline period between 2000 and 2008 (Figure 6);
- were already characterised by faster growing employment density before the Crossrail announcement (Figure 7).

These and other indicators suggest that the areas around future Elizabeth line stations were already economically different from the London average in some fundamental way prior to the announcement. In this case, the risk is that the Crossrail announcement could be wrongly identified as a cause of economic changes when they may be an effect of those underlying patterns (or both).

Because of these fundamental differences, to make a robust estimate of the impact of the Crossrail announcement it is not enough simply to compare growth along the line with background average growth rates and infer that the differential is explained only by the expectation of future Elizabeth line services. As a minimum, we must control for those socio-economic factors on which we can see that treatment and control areas are not the same. Controlling selection bias due to ‘unobservable’ factors is harder but possible to address using difference in differences analysis with fixed effects. Although it is unlikely that these techniques will have fully corrected for selection bias, they should have reduced selection (and omitted variable) bias to the extent practical within the parameters of the study. We describe this method in detail in chapter 2.7.

Figure 6. Population growth, bands around Crossrail stations and wider London benchmarks, based on ONS data (index 2002 = 100)

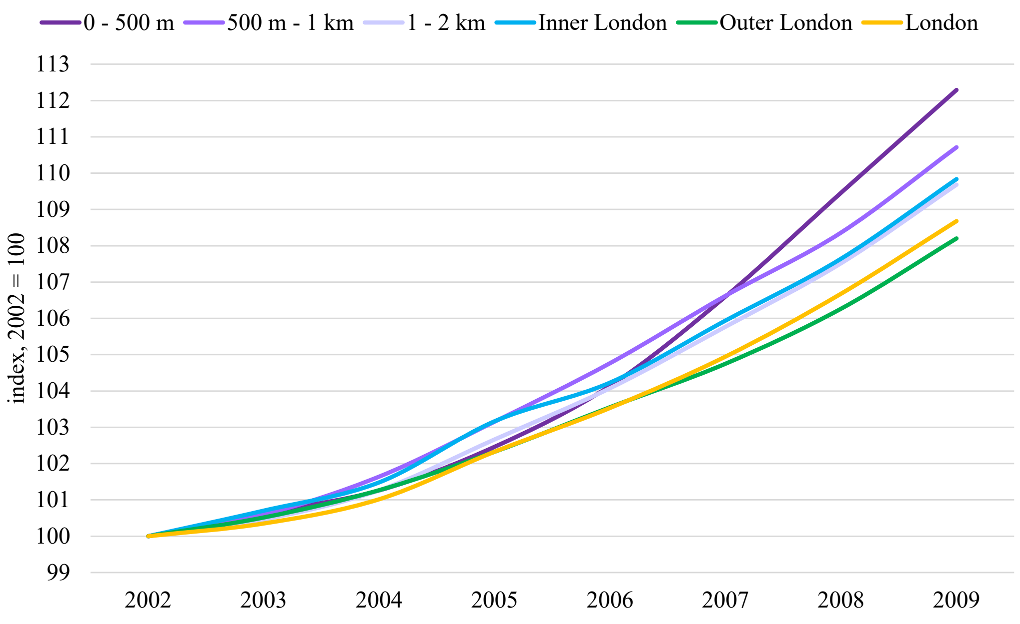
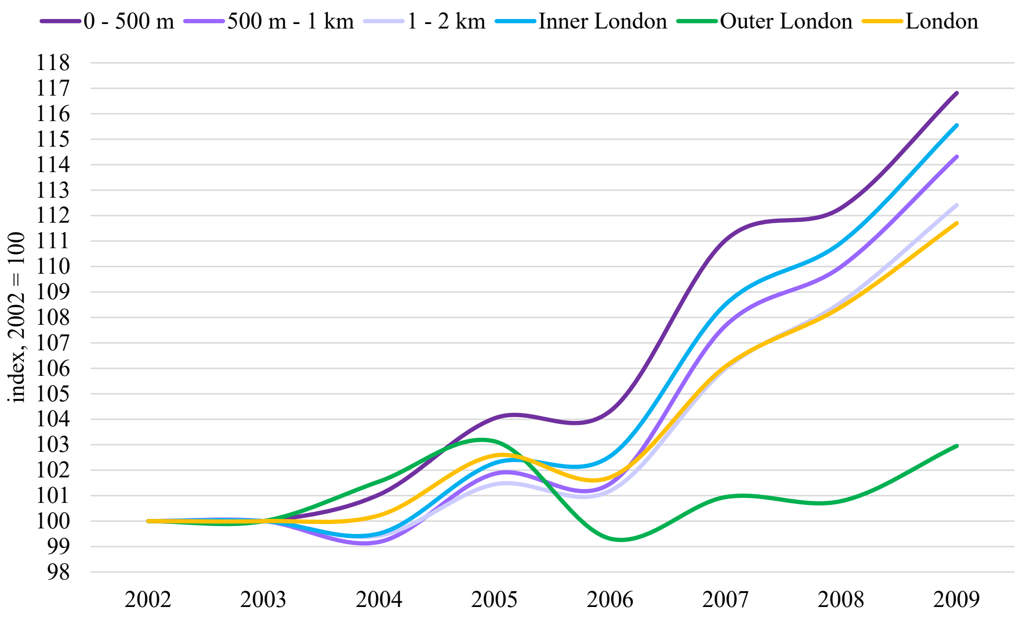


Figure 7. Employment density growth, bands around Crossrail stations and wider London benchmarks, based on ONS data (index 2002 = 100)



2.4 The difference in difference method

In a difference in difference (DiD) approach we compare changes in outcome values before and after an intervention for both treatment and comparison groups. The before and after change in the comparison group serves as a counterfactual for the treatment group. The difference in these differences is our estimate of the treatment effect. Table 1 presents the concept in a simple worked example.

Table 1 Example of Difference-in-Differences

Comparison group	Before (B)	After (A)	Difference (A) – (B)
Treated (T)	50	90	40
Comparison (C)	45	70	25
Difference (T) – (C)	5	20	**DiD effect= 15**

The same concept is used in a difference in differences regression analysis. Using regression techniques enables us to include additional variables in the model, which represent other factors that may have influenced the treatment and counterfactual groups' performance. That allows us to separate out their impacts and minimise the risk of misattributing the impact of those other factors to the Crossrail announcement.

Equation (1) shows specification for a simple difference in differences model:

$$(1)\ Y_{it} = \alpha + \beta * post_t + \gamma * treated_i + \delta * post_t * treated_i + \epsilon_{it}$$

where:

- $\boldsymbol{i}$ denotes an observation unit (e.g. a full postcode area);
- $\boldsymbol{t}$ denotes year;

- Y_{it} is a continuous outcome variable for each unit in each year;
- $Post_t$ equals one if for each year after the intervention happened (or zero for before);
- $Treated_i$ equals one for each unit *i* in the 'treatment group' (and zero for the 'counterfactual group'); and
- $Treated_i*post_t$ is an interaction between the two variables above, and equals one only for observations recorded in the treated area after the intervention took place.

The parameters represented by the Greek letters in the equation above (α, β, δ, γ), are the estimated coefficients. That is to say, the slope of the estimated line that shows the relationship between the outcome and the variable it sits in front of, such that

- the coefficient (β) estimated for the treatment period variable (*post*) represents the change observed in the comparison group over the same time (i.e. the end point for property values in our comparison areas);
- the coefficient (γ) estimated for the treatment variable (*treated*) represents the coefficient on treatment for treated observations before the treatment (i.e. the starting point for property values in our treated areas); and
- the coefficient (δ) estimated for the treatment effect variable (*post * treated*) shows a difference in coefficients for treated and comparison groups after the treatment. This is the coefficient in which we are most interested, because, as the name suggests, it is our estimate of the impact of the Crossrail announcement.
- the coefficient (α) represents a constant term in a regression model and it is not interpreted in this case.

As mentioned above, regression analysis allows us to include other explanatory variables into the basic specification above, in order to account for factors that affect the outcome of interest (see chapter 2.7). These would each have their own estimated coefficient.

Looking at the relative size of the coefficients on all the explanatory variables tells us about the relative importance of them driving observed changes in outcome, such that, for example, we may estimate that there has been a positive treatment effect, but that it is quite small in magnitude compared to other factors.

If we run a simple DiD regression on data used for the worked example above (Table 1), the estimated equation would look like the following:

$$(2)\ Y_{it} = \alpha + \mathbf{25} * post_t + \mathbf{5} * treated_i + \mathbf{15} * post_t * treated_i$$

2.5 Statistical significance

Having estimated the value of a treatment effect (or any coefficient) we must consider its statistical significance. With regression analysis we try to estimate relationships between variables based on pools of data which may be imperfect or inaccurate in some way. This means that any time we make an estimate, we will have to consider how confident we can be in its accuracy. To simplify, if we have a lot of variation in individual data points around our estimated ‘line of best fit’ then we are less confident in it. Or, if we have a small sample of data, we will generally be less confident in any relationships we estimate than if we have a very large sample. ‘Statistical significance’ means that we can be confident within a certain bound (usually 90%, 95% or 99%) that the effect we have estimated is accurate. For instance, if an estimate is statistically significant at 99% confidence level, it means that if we estimate a model many times on different data samples, in 99% of cases the value of the estimate will be within the estimated interval.

The lowest commonly acceptable level of confidence is 90%. Significance testing establishes the probability of finding these results in a world where the Crossrail announcement actually had no effect. When we talk about statistically significant findings of our models, we mean that they are significant at least a 90% confidence level. Regression tables further in the report show the level of statistical significance for each estimate. It should also be noted that we can be less confident in regression results where we are testing multiple different hypotheses, such as the effects at multiple different stations.

With the above in mind, it is important to look at all available evidence rather than focusing on singular findings which prove statistically significant. This is especially relevant for us when conducting analysis at section and station level.

In addition, it is a distinction of critical importance that the conclusion that ‘we have not been able to find statistically significant impacts’ is not the same as saying ‘we have estimated there is no impact’. An absence of statistically significant findings can sometimes be because the amount of data is insufficient, the quality of data is too low, or because there is too much variation in the data points. Importantly it could also happen because the relationship between the outcome and the treatment is not linear. The last instance is to some extent mitigated in our work by testing, for example, three distance bands separately (see section 3.2), as opposed to aiming to find one estimate of impact regardless of the distance from a station.

2.6 Common trends

For us to have confidence in a difference in difference analysis, we need what is called the ‘common trends assumption’ to be met. We should be confident that even if the treatment and comparison groups did not begin at the same level (in terms of property prices, or any other relevant outcome), their performance trend over the time period would have moved in parallel in the absence of the Crossrail announcement. This is an important check on the robustness of any defined counterfactual. The principal way of satisfying this is by comparing whether they were moving in parallel in the period leading up to treatment (rather than converging or diverging).

We can relax this assumption if we have a sufficiently large sample by constructing what are called area-year fixed effects. These would be parameters estimated for additional dummy variables, namely interactions between all-area and all-year dummy variables (excluding reference categories). By introducing these into a model we allow for different time trends in different places. This is explained in more detail in chapter 2.7, below, where we talk about the fixed effects technique.

2.7 Controlling for other factors and using fixed effects

The simplest form of evaluation would be a simple before and after comparison of the change observed around Crossrail programme stations with the change observed around comparable areas. The drawback of that approach is that it does not account for the impact of other factors that might also change outcomes. In terms of property prices, those other factors could include relative attractiveness of the neighbourhoods, transport connectivity, type of property stock in the area, availability of green spaces or quality schools, and many others.

To be sure that the difference between the change in performance in the treatment and comparison areas is only attributable to the Crossrail announcement, we must take into account other factors that may have influenced the treatment or comparison areas' performance and separate out their impacts, or we would risk misattributing the impact of other factors to Crossrail. In econometrics these factors are referred to as controls*,* as researchers need to 'control' for their impact on findings. In general, the extent to which credible controls have been used greatly affects our assessment of the robustness of the evaluation.

Control variables work well for factors that are measurable or somehow quantifiable (in econometrics language, 'observed'). These include things such as: proximity to roads or other non-rail transport modes; demographic characteristics of the population, characteristics of the existing property stock; range of local leisure, community and cultural amenities; amount and quality of open space nearby; and may also include negative factors such as proximity to 'bad neighbours' such as incinerators or heavy industrial sites; and high crime rates. Data representing these factors can be included in a difference in differences regression analysis.

However, it is possible to conceive of a range of other factors, which are not as easily observable but which can influence property market outcomes of areas (or businesses or individuals). As discussed in chapter 2.3, if they also affect the area's likelihood of being on the Crossrail route in the first place, we may also have a problem with selection bias.

One way to mitigate this is by using fixed effects. Fixed effects can be used for individual units (e.g. properties), for areas and for time (year by year). The term 'fixed effects' is used as a catch all for 'all the other unobserved factors for which we have not collected specific data'. In theory they should strip out any impacts due to non-Crossrail factors that we might otherwise wrongly attribute to the announcement. They include:

- area fixed effects (often called 'state fixed effects' in the literature) will strip out the effect of any factors that are specific to a neighbourhood at whatever statistical level they are defined. For example, some well-known reputational factor of a neighbourhood that is not well represented by the set of socio-economic and geographic control factors we have included. Area fixed effects can only control for locally specific factors that are constant over the time period;

- time fixed effects (in this case, the year) will strip out of the effect of time-specific factors that are not well represented by the other control variables. These are often used to strip out the effect of specific economic shocks like property market crashes, recessions, terrorist events, or Covid-19. Year fixed effects can only control for time-specific events that affect all places the same; and
- area-year fixed effects will control for unobserved factors that are specific to that area, in that year. For the example, the locally specific effects of a recession, or the closure of a major employer.

Equation (3) below represents a simple difference in differences specification with all three types of fixed effects.

$$(3)\ Y_{it} = \alpha + \beta * post_t + \gamma * treated_i + \delta * post_t * treated_i + FE_i + FE_t + FE_{it} + \epsilon_{it}$$

To implement fixed effects, we assume that each area of observation in a sample has its own unobservable specific set of factors. We introduce unit fixed effects in an econometric model by including dummy variables for each unit (except for a reference one). That means estimating a constant parameter for a time-series data for each unit separately. There are two technical assumptions we need to make:

- that these specific unobserved characteristics are constant across time; and
- that we have (outcome) data for at least two periods of time for both treatment and comparison groups of observations.[18]

If we are able to use area-year fixed effects in particular it allows us to be slightly more relaxed about the common trends assumption between treatment and comparison groups: by introducing these into the model we allow for different time trends in different places without biasing the estimate of the impact of the Crossrail announcement. In our analyses we introduce interactions between years and local authority fixed effects to account for any changes in policy or political factors.

2.8 The Scientific Maryland Scale

The Maryland Scientific Methods Scale (SMS) is a five-point scale ranging from one (for evaluations based on simple cross sectional comparisons'), to five (for randomised control trials). It aims to rank methods based on the extent to which they address selection bias.

The What Works Centre for Local Economic Growth uses a modified form of the SMS to rank evaluation evidence and considers the minimum standard of evidence

[18] This means either panel data (for the same property, from at least two years) or repeated cross-sectional data (average property values in a given area, for at least two years). The latter means not having repeated observations for each unit *i,* but for groups of units *s* (*s* might be postcode level, Lower Super Output Area (LSOA) level or Borough level for example). Fixed effects are then introduced at level *s*. In our case, while we have individual property observations, we introduce fixed effects at postcode level (everywhere except for planning analysis, where we have LSOA as units and LSOA-level fixed effects).

for inclusion in its systematic reviews to be SMS Level 3: a before and after comparison of treatment and comparison groups with justification provided that the comparison group selected is sufficiently similar to the treatment group.

This evaluation approach has been designed to achieve at least SMS Level 3 (defined as: a before and after comparison with a comparison group with reasonable attempts to control for selection on observable factors) and some elements may score SMS Level 4 (use of panel data and fixed effects which may to an extent be able to control for selection on unobservable factors).

2.9 Defining treatment

We use treatment variables to represent, in data form, the intervention that we are trying to evaluate. They can take two forms: binary or continuous. Binary treatment is the most common approach, whereby a binary variable takes the value of one for treated observations and zero for untreated observations. A continuous variable can take any value and allows us to represent a variation in the 'dose' of treatment (e.g. to allow some areas to have a big change associated with Crossrail and other areas a smaller change).

With a binary treatment variable, the definition of 'treatment' is usually being within a useful distance of a new station. This is constructed using a radius (distance band) around a station, the most common for urban areas (and the ones used in this report) being 500m, 1 km and 2 km (approximately a five, 10 or 20 minute walk).

This treatment definition was also used, among the others, in the land value impact evaluations conducted by GVA (2012 and 2017) and KPMG in collaboration with Savills (2016).

Including a binary treatment variable based on these distances is therefore not only a natural starting point but enables comparisons of our results with other studies based on a similar approach.

However, the Elizabeth line will only call at existing stations that are all already served by a rail-based mode of transport[19]. So those areas are already likely to have many of the characteristics typical of areas close to railway stations (including, but not limited to, existing better access to jobs and/or people). Defining treatment using a binary 'Elizabeth line station / no Elizabeth line station' approach may not be an accurate representation of the expected change that the Crossrail announcement will have brought to an area; it may oversimplify.

The case for the Elizabeth line was predicated on capacity and journey time-based accessibility. Places will not all gain (or lose) in the same way. They start from different baseline levels of service. Some origin-destination pairs may experience fairly limited or even negative journey time impacts, especially those in central London that are already well served by the existing network. Some will have longer journey times but fewer interchanges. Some will experience little change in journey

[19] Albeit some new station buildings are being constructed adjacent to existing stations.

time but will have increased or reduced service frequency compared to the current pattern.

Accessibility in developed urban areas rarely depends solely on one transport option – one rail connection, or one bus route. With dense transport networks such as in London, the accessibility of a single point is affected by the accessibility of other points, as people can conveniently change lines and modes at many locations. Areas not on the Elizabeth line will therefore still have improved accessibility when, for example, areas that used to require two interchanges can now be reached with just one. The observation that many more places would benefit from the Elizabeth line than just those along the alignment was central to the decision to levy the Crossrail Business Rate Supplement on all big businesses across London, and not just those close to a station.

We wanted our evaluation methodology to take that interconnectivity aspect into account, considering the impact of places directly served by the intervention but also those served indirectly. For this reason, we created a bespoke, continuous accessibility measure, to represent this complexity of treatment definition for different places (see Appendix C for more information). Using timetable and TfL Railplan inputs we estimated two types of accessibility[20] before and after the Elizabeth line opening for each rail station in the study area. Broadly speaking, by comparing the two we estimate the marginal change in accessibility that is represented by the Elizabeth line at any one station location. This anticipated net change can then be used as a continuous treatment variable. In the course of the study we discussed and agreed that the continuous treatment approach is likely to be more relevant for the post-opening evaluation.

In the pre-opening period people tend to focus their expectations of improvements in places close to the stations. That could have been seen for instance in numerous media articles on housing unit price levels around future Elizabeth line stations. It is likely that in the pre-opening stage people do not fully realise what the total network effects might be, especially in terms of capacity and journey times improvements for people starting their journeys in places not directly on the route.

Therefore, while change in accessibility is a more accurate representation of the relative impact the Elizabeth line will eventually have on the network, it is not necessarily true that in the pre-opening period individual workers, businesses or property investors will accurately perceive that change. Although the binary variable is a relatively simple measure of impact, it could be a better representation of how most people perceive the Elizabeth line in the period between the scheme being committed and services beginning to run, and it is the one that we have used in this study.

[20] One based on journey time (timetable), the second based on generalised journey time (Railplan).

3 Methodology details

We test pre-opening impacts of Crossrail in three distance bands – up to 500m, between 500m and 1 km, and from 1 km to 2 km from the stations. We evaluate average line-wide impacts and complement that analysis with looking at impacts in sections of the route and at specific locations where possible.

We run several regression models, with specifications varying with regard to the number and type of control variables include. All models include area and time fixed effects to control for unobservable factors. We also included a description of how to read simple regression outputs presented in the report.

3.1 Spatial levels of impacts

Our analysis uses several different geographies. The total study area is based on a 10 mile (16 kilometre) 'buffer' around the route alignment from which comparator observations are selected. Analysis is then broken down into sections and case study areas. We test for potential impacts at three spatial levels:

- average line-wide impacts (for everywhere within the 'buffer');
- impacts for different categories of places along the line (different sections within the 'buffer'); and
- impacts around individual stations.

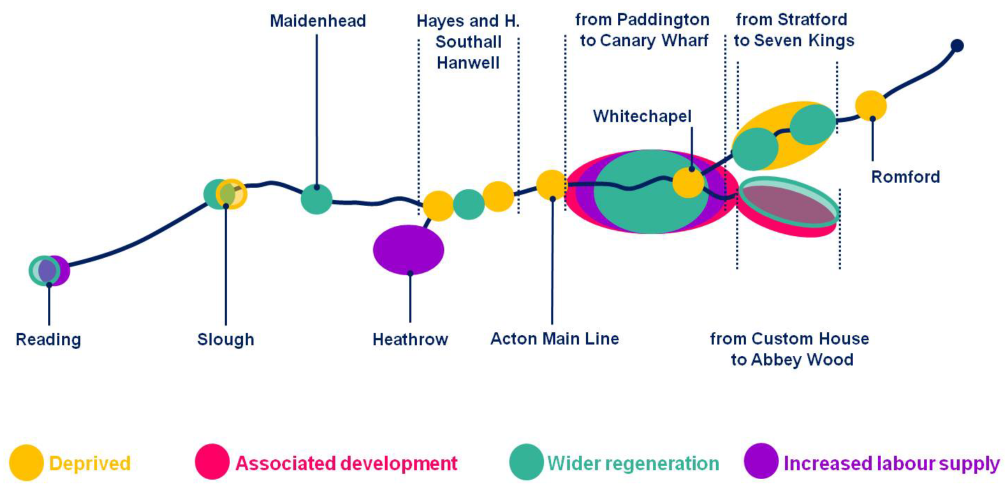
In the line-level analysis all observations within 2 km of any future Elizabeth line station are considered as treated. All the others make up a counterfactual group (except for outliers and any observations excluded from the analysis in line with our similarity approach described in chapter 3.6).

Line-wide analysis aims to estimate an average impact, robust to local variations and outliers, benefitting from the largest sample size. Section analysis aims to add more context and reveal potential spatial variation in impacts which would be flattened out in the line-wide approach. In addition to estimating average line-wide effects we test impacts specifically in these categories:

- geographic sections of the route;
- places with significant deprivation;
- areas with new stations or major development associated with the Crossrail programme;
- locations expected to benefit from increased labour supply; and
- areas of wider regeneration or growth initiatives.

Table 2 lists all the stations on the line and classifies them in accordance with the categories listed above, in addition to Figure 8. There are nine categories (five sections of the route and four additional ones) which means that in a section analysis we run our key model nine times, but each time on a slightly different dataset.

Figure 8. Map of categories of places for section analysis



In an analysis of a given category, only observations located within 2 km of the indicated stations are considered treated. All the other observations located within 2 km of any other Crossrail programme station are excluded from the dataset. Taking an analysis of the Outer Western impacts as an example, we see that this section of the line has eight stations, from Reading to Iver. Observations located within 2 km of any of these stations will be considered treated. Observations located within 2 km of any other station from West Drayton to Shenfield and Abbey Wood will be excluded from the analysis. All observations further than 2 km will be included as a counterfactual group, subject to an analysis of outliers and similarity analysis described in chapter 3.6.

An important contextual factor around the impact of the Crossrail's announcement on residential house prices – and on all property impacts in general as well – would be land availability around stations. Supply side limits can constrain growth even if the infrastructure investment creates the demand for new development. However, our analysis was unable to find suitable data for this analysis.

Table 2 Stations classified by route section and location category for section analysis

Station	Route section	Deprived	Associated development	Increased labour supply	Wider regeneration
Reading	Outer Western	No	No	Yes	Yes
Twyford	Outer Western	No	No	No	No
Maidenhead	Outer Western	No	No	No	Yes
Taplow	Outer Western	No	No	No	No
Burnham	Outer Western	No	No	No	No
Slough	Outer Western	Yes	No	No	Yes
Langley	Outer Western	No	No	No	No
Iver	Outer Western	No	No	No	No
West Drayton	Western	No	No	No	No

Station	Route section	Deprived	Associated development	Increased labour supply	Wider regeneration
Heathrow 123	Western	No	No	Yes	No
Heathrow 4	Western	No	No	Yes	No
Heathrow 5	Western	No	No	Yes	No
Hayes and Harlington	Western	Yes	No	No	No
Southall	Western	No	No	No	Yes
Hanwell	Western	Yes	No	No	No
West Ealing	Western	No	No	No	No
Ealing Broadway	Western	No	No	No	No
Acton Main Line	Western	Yes	No	No	No
Paddington	Central	No	Yes	Yes	Yes
Bond Street	Central	No	Yes	Yes	No
Tottenham Court Road	Central	No	Yes	Yes	Yes
Farringdon	Central	No	Yes	Yes	No
Liverpool Street	Central	No	Yes	Yes	No
Whitechapel	Central	Yes	Yes	No	Yes
Canary Wharf	Central	No	Yes	Yes	Yes
Custom House	Southeastern	No	Yes	Yes	Yes
Woolwich	Southeastern	No	Yes	No	Yes
Abbey Wood	Southeastern	No	Yes	No	Yes
Stratford	Eastern	Yes	No	No	Yes
Maryland	Eastern	Yes	No	No	Yes
Forest Gate	Eastern	Yes	No	No	No
Manor Park	Eastern	Yes	No	No	No
Ilford	Eastern	Yes	No	No	Yes
Seven Kings	Eastern	Yes	No	No	Yes
Goodmayes	Eastern	No	No	No	No
Chadwell Heath	Eastern	No	No	No	No
Romford	Eastern	Yes	No	No	No
Gidea Park	Eastern	No	No	No	No
Harold Wood	Eastern	No	No	No	No
Brentwood	Eastern	No	No	No	No
Shenfield	Eastern	No	No	No	No

3.2 Model specifications

We run a series of six models with slightly different specifications in order to test the sensitivity of the results. We present all of them in the case of line-wide analysis and then only the strongest model (4) for section and station discussion. All models are described below. Please do not confuse the notation of these models with the three equations presented in previous chapters. The equations present general concepts, whilst the models differ with regard to control variables as well.

Model 1 is a simple difference in differences regression. We have three treatment variables for the three distance bands of interest: 0-500 metres, 500-1,000 metres and 1-2 km. Each of those variables is interacted with a *post* variable. The model includes area- and year- fixed effects. The area fixed effects are estimated at a full postcode level (e.g. W1T 4BQ).

$$\text{(Model 1) } \mathrm{Y_{it}} = \\ = \alpha + \mathrm{post_t} + \mathrm{buffer\ 500_i} + \mathrm{buffer500\_1000_i} + \mathrm{buffer1000\_2000_i} + \mathrm{post_t} * \mathrm{buffer500_i} + \mathrm{post_t} * \mathrm{buffer500\_1000}_i + \mathrm{post_t} * \mathrm{buffer1000\_2000}_i + \mathrm{FE_i} + \mathrm{FE_t} + \epsilon_{\mathrm{it}}$$

where:

- *i* denotes postcode;
- *t* denotes years between 2000 and 2019;
- FE_i denotes area fixed effects estimated at postcode level; and
- FE_t denotes time fixed effects estimated at year level.

In model (2) we add several control variables (as listed below). These controls are either time-variant and area-variant or at local authority level which means they are not omitted from the model when estimating postcode- or LSOA-level fixed effects:

- *lease space:* Amount of space leased for each transaction (observation) in our dataset);
- *population density* at LSOA level, data for 2000-2018;
- *employment density* at LSOA level, data for 2003-2018; and
- local authority dummy variables.

$$
\begin{aligned}
(\text{Model 2})\ Y_{it} = \\
&= \alpha + \text{post}_t + \text{buffer } 500_i + \text{buffer500\_1000}_i \\
&+ \text{buffer1000\_2000}_i + \text{post}_t * \text{buffer500}_i + \text{post}_t \\
&* \text{buffer500\_1000}_i + \text{post}_t * \text{buffer1000\_2000}_i + FE_i + FE_t \\
&+ \mathbf{local\ authority_l} + \mathbf{leased\ space_k} \\
&+ \mathbf{employment\ density_{jt}} + \mathbf{population\ density_{jt}} + \epsilon_{it}
\end{aligned}
$$

where:

- *j* denotes LSOA; and
- *l* denotes local authority.

Model (3) includes an additional set of controls namely annual shares of sectoral employment per LSOAs described in chapter 3.3.

$$
\begin{aligned}
(\text{Model 3})\ Y_{it} = \\
&= \alpha + \text{post}_t + \text{buffer } 500_i + \text{buffer500\_1000}_i \\
&+ \text{buffer1000\_2000}_i + \text{post}_t * \text{buffer500}_i + \text{post}_t \\
&* \text{buffer500\_1000}_i + \text{post}_t * \text{buffer1000\_2000}_i + FE_i + FE_t \\
&+ \text{local authority}_l + \text{leased space}_k + \text{employment density}_{jt} \\
&+ \text{population density}_{jt} + \boldsymbol{employment\ share\ sector1_{jt}} \\
&+ \boldsymbol{employment\ share\ sector2_{jt}} \\
&+ \boldsymbol{employment\ share\ sector3_{jt}} \\
&+ \boldsymbol{employment\ share\ sector4_{jt}} \\
&+ \boldsymbol{employment\ share\ sector5_{jt}} + \epsilon_{it}
\end{aligned}
$$

In model (4) we include interactions between year dummy variables and local authority dummy variable to control for any wider changes impacting districts in a different way, for instance policy changes.

$$
\begin{aligned}
\text{(Model 4) } Y_{it} = \\
&= \alpha + \text{post}_t + \text{buffer } 500_i + \text{buffer500\_1000}_i \\
&+ \text{buffer1000\_2000}_i + \text{post}_t * \text{buffer500}_i + \text{post}_t \\
&* \text{buffer500\_1000}_i + \text{post}_t * \text{buffer1000\_2000}_i + \text{FE}_i + \text{FE}_t \\
&+ \text{local authority}_l + \text{leased space}_k + \text{employment density}_{jt} \\
&+ \text{population density}_{jt} + \mathit{employment\ share\ sector}1_{jt} \\
&+ \mathit{employment\ share\ sector}2_{jt} \\
&+ \mathit{employment\ share\ sector}3_{jt} \\
&+ \mathit{employment\ share\ sector}4_{jt} \\
&+ \mathit{employment\ share\ sector}5_{jt} + \boldsymbol{local\ authority_l} \\
&* \boldsymbol{year_t} + \epsilon_{it}
\end{aligned}
$$

In addition, we test potential findings using an incidental treatment approach. This is an approach used in evaluation literature to treat potential bias from reverse causality and selection bias. Reverse causality in transport evaluation presents itself when we attribute positive outcomes to the transport intervention, rather than to the fact the intervention was focused on locations which are already performing better than those which did not receive the intervention. There is no clear impact pattern in this situation, linked directly to the issues of ‘selection into treatment’ bias.

To tackle this challenge, the incidental treatment approach intends to introduce some hypothetical level of randomness in the evaluation design. It assumes that the key goal of a transport infrastructure scheme is to connect two places: where the scheme starts and where it finishes. All other places in between the two key places are assumed to have been added to the scheme without any particular reason, ‘just because they happened to be there’, so effectively randomly. Therefore, if we estimate a DiD model only for those areas which are ‘in-between centres’ we treat reverse causality arising from original non-random treatment assignment.

In our case we assume that the key goal of the Crossrail programme was to connect Reading, Heathrow, Shenfield and Abbey Wood to central London and Canary Wharf. These are the terminus stations. For this reason we exclude all those stations from the analysis (including all stations in a central section). This means that final data sample for the incidental approach does not include any observations located within 2 km of any of these stations. This approach tests impacts on all other in-between areas.

In our analysis we estimate two models using the incidental approach, using specification of model 3 and model 4. We then refer to them respectively as model 5 (without local authority-year fixed effects) and model 6 (with local authority-year fixed effects).

3.3 Control variables

We described in chapter 2.7 that using a fixed effect approach means that we include a dummy variable for each chosen unit of dimension (except for one). In our case the dimensions are years and spatial areas, usually postcode areas (everywhere except for planning impacts analysis where we use LSOAs).

In order for a regression model to be estimated there cannot be any collinearity between predictor variables. Collinearity occurs when dependent variables are highly correlated or when one independent variable can be expressed as a linear combination of the other(s).

This means that if we include dummy variables for each postcode area, we cannot include another variable, binary or continuous, which does vary within one postcode. Now most of the data available to us, and that we would normally use in a model explaining property values, comes from the census. The is census is only available for one year (2011), so it has the same value for each individual postcode. This means that any census data we include in a model would be collinear with our postcode and LSOA dummy variables. We can only include variables which are time- and area-variant, i.e. for a given area, they take a different value in at least two periods.

There are three time- and area-variant controls that we were able to use in the analysis of all outcomes:

- population density per LSOA and year;
- employment density per LSOA and year; and
- sectoral employment shares per LSOA and year.

Population estimates come from the ONS mid-year estimates. Employment estimates are provided through the Business Register and Employment Survey and Annual Business Inquiry for years before 2009. Absolute values in each year are then divided by the LSOA area to calculate densities. Sectoral employment shares are also calculated from the BRES and ABI data. The shares are included in an attempt to account for changes in sectoral and land use patterns in the area.

In addition, we used transition-level controls within separate outcome analyses:

- for all commercial transactions we use the amount of space leased in square feet;
- for residential transactions we use:
    - Type of a property as provided by land registry – detached, semi-detached, terraced or a flat; and
    - Market maturity – a property established on the market or a new one.

3.4 Presentation of results

Tables presenting regression results are simplified – they include estimates for the key variables only, with the addition of key model statistics such as R-Squared and Adjusted R-Squared.

The R-Squared is a measure of goodness-of-fit for regression models – it measures how much of the variation in the data is explained by the regression model. Also referred to as the coefficient of determination, it takes values from zero to one, which expresses what proportion of variation in the dependent variable is explained by the independent variables(s). The higher the value of the R-Squared, the better the fit of the model. An R-Squared of one means a ‘perfect fit’ so that the regression model was able to predict 100% of the variation in the data, whereas an R-Squared of zero means that the regression model was able to explain zero percent of the variation in the data.

However, the R-Squared is calculated in such a way that it increases with the number of independent variables included in the model (if other aspects are unchanged), regardless of the quality or appropriateness of those variables. The formula for the adjusted R-Squared accounts for the number of variables and enables us to compare fitness of models with different number of variables.

Under the goodness-of-fit measures we provide information on inclusion of various types of fixed effects variables, as explained in chapter 3.2.

We operate with three treatment definitions – for properties being located within 500 metres, within 500m to 1 km or within 1 to 2 km from a Crossrail programme station. That means that we report findings in three location bands. The estimate for the post-announcement variable is an estimate of the average difference in values between the periods before and after the announcement .

Each estimate is accompanied by the value of its standard error and information of its resulting statistical significance (see chapter 2.5 for more details). Standard errors can be used by the reader to test the significance level, but provided indicators (*, ** and ***) make it immediately clear:

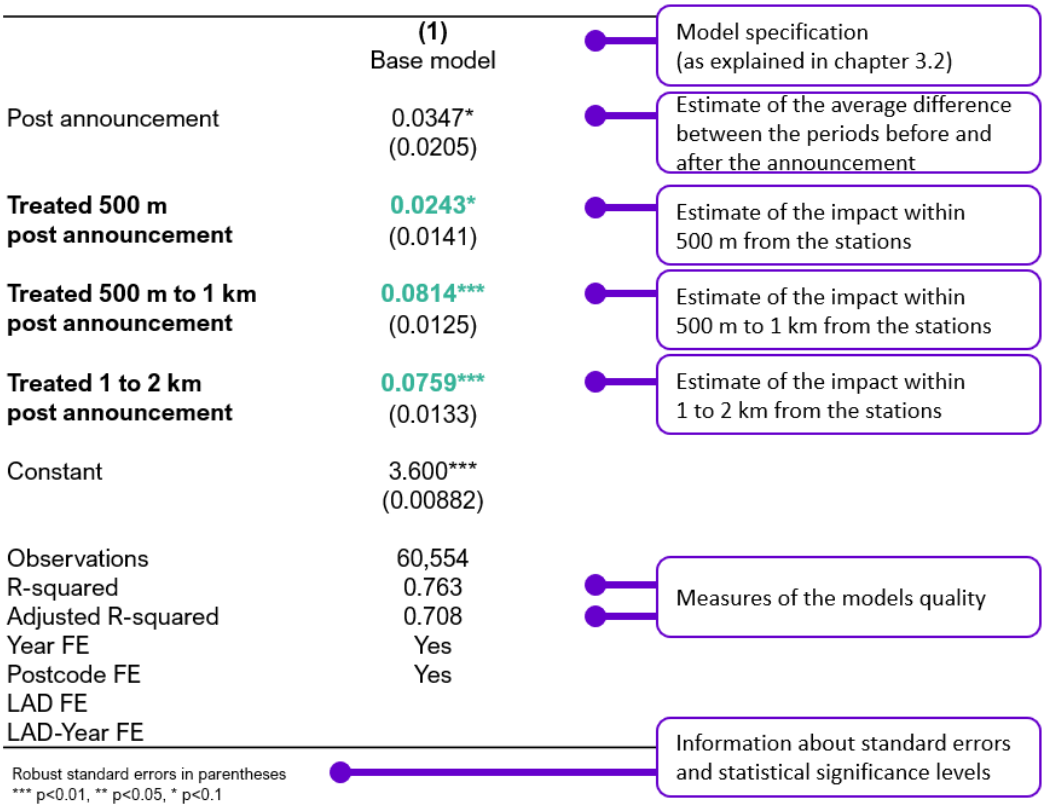
- *** means statistical significance at 99% confidence level;
- ** means statistical significance at 95% confidence level; and
- * means statistical significance at 90% confidence level.

Figure 9 shows an example of a regression output. Both R-Squared coefficients are relatively high, almost 80%, which suggests good fitness of the model. The first coefficient of the variable ‘post announcement’, 0.0347, means that office rents were on average 3.47% higher in the period after the announcement than before across the whole of the sample, regardless of their distance from a Crossrail station.

The standard error for this coefficient is 0.0205, which means that on a 90% confidence level (a 1.645 z-critical value) the coefficient is between 0.0009 (0.0347-0.0205*1.645=0.0009) and 0.068 (0.0347+0.0205*1.645=0.068) and it is significantly different (larger) than zero.

According to the next three estimates, there have been significant and positive Crossrail announcement impacts in all three distance bands: 2.43% in the first band (up to 500m), 8.14% in the second band (from 500m to 1 km around the stations), and 7.59% in the last distance band. The estimate of impact closest to the stations is significant at 90% confidence level, while the other two are significant at 99% confidence level. The constant (or ‘intercept’) estimate shows the average value of the time and area fixed effects included in the regression.

Figure 9 Presentation of regression outputs (an example to explain format)



	(1) Base model
Post announcement	0.0347* (0.0205)
Treated 500 m post announcement	0.0243* (0.0141)
Treated 500 m to 1 km post announcement	0.0814*** (0.0125)
Treated 1 to 2 km post announcement	0.0759*** (0.0133)
Constant	3.600*** (0.00882)
Observations	60,554
R-squared	0.763
Adjusted R-squared	0.708
Year FE	Yes
Postcode FE	Yes
LAD FE	
LAD-Year FE	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

3.5 Outliers analysis

We use the interquartile range technique, also known as the box-and-whisker plot approach, to eliminate outliers. Eliminating outliers is important as leaving data points in the sample which are the result of either inaccurate data or one-off events can lead to biased results. For each outcome variable we conduct an outlier analysis in sub-samples. For almost all outcomes that sub-sample is defined by a year and a local authority. In the analysis of office rents, we define sub-samples by office market areas instead of local authorities.

For a given sub-sample, defined by a year and a local authority or market area, we calculate two parameters of the outcome variable's distribution – the 25th and 75th percentiles, also known as first quartile and third quartile, Q1 and Q3. The difference between them is the interquartile range (IQ). The threshold values – i.e. the values allowing us to identify outliers – are then calculated in the following way:

- lower threshold = Q1 – 1.5*IQ; and
- upper threshold = Q3 + 1.5*IQ.

For each area in each year we then exclude observations with outcome variable values lower than the lower threshold or higher than the upper threshold.

3.6 Similarity analysis for the counterfactual

For each LSOA contained in a sample we calculate its probability of being treated. Treatment definition is based on the distance from a Crossrail programme station. If an LSOA contains a postcode located within 500m from any Crossrail programme station, then that LSOA is treated for the purpose of this analysis.

The probability of being treated is estimated through a logistic regression with the following explanatory variables:

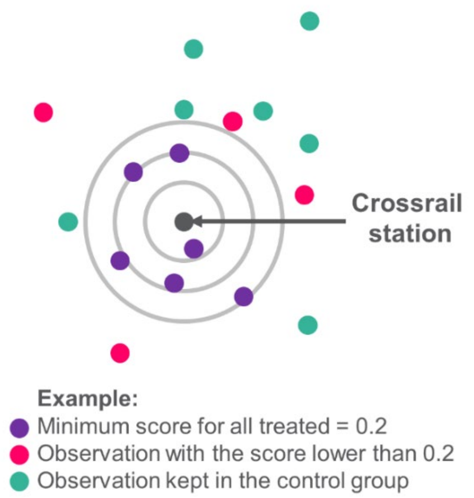
- rail accessibility in 2007;
- population density in 2011;
- employment density in 2011; and
- index of multiple deprivation in 2015 (which is based on 2011/2012 data).

Rail accessibility in 2007 is the same accessibility score estimated for potential continuous treatment approach and is described in the appendix. Since the accessibility score was estimated at station level and the logistic regression is run at LSOA level, we use an average of unweighted accessibility scores for all stations in a given LSOA.

We then analyse the distribution of the estimated probability for all the LSOAs that are considered treated. We identify a minimum (Pmin), a 25th percentile and a median value of the probability estimated across all treated LSOAs. We also set a rule that in order to be included in the counterfactual group of observations, or considered similar enough to treated observations, any postcode located further than 2 km from any Crossrail programme station needs to have an estimated probability of treatment at least at the minimum level estimated for the actually treated postcodes (i.e. Pmin).

If that rule excludes only a very small proportion of observations then we go one step further and set the threshold for the probability score to either the 25th percentile or, if the 25th does not exclude at least some observations either, to a median. We do this to test sensitivity of final estimates to the sample composition.

Figure 10. Representation of the similarity rule



4 Office rent impacts

We found consistent evidence of positive pre-opening impacts of Crossrail on office rents in central London. The average line-wide impact amounts to around 3% within 500m of Crossrail stations and around 7.5% between 500m and 1 km from the stations.

4.1 Introduction

This chapter presents findings of the various analyses relating to the impact of the Crossrail announcement on office rental values.

Whilst there is not necessarily an expectation that office rents will increase prior to the opening of the Elizabeth line, there is an expectation that the operational railway will bring agglomeration benefits to some parts of the economy (particularly knowledge-intensive and other service sectors) by increasing the size of the labour market catchment area and opportunities for knowledge spill-overs. This may make office locations close to future Elizabeth line stations relatively more desirable and hence valuable, especially where there has not been a significant increase in supply of floorspace. An estimation of wider economic benefits was included in the Crossrail Business Case assessment and contributed to increasing the benefit cost ratio from 1.45:1 to 2.54:1.[21]

The aim of this analysis is to test whether this likely impact has been reflected in values of office rents around the stations in anticipation of the Elizabeth line opening. We do not have a pre-existing assumption that it has been. Existing robust evidence on transport commercial property impacts is scarce and usually relates to post-opening impacts. The WWCLEG Review[22] in 2015 found only one relevant robust study on that matter published in English. It found no effects of proximity to stations[23].

4.2 Data source

Office rental data come from an individual transaction-level database provided by CoStar, a commercial property platform with a large proprietary database of commercial property data. The CoStar database holds commercial information and analytics on a web-based platform providing access to data on over 400,000 UK commercial properties. The data extracted for the analysis comprise all recorded office rental deals that have been transacted within the study area over the baseline period 2000 to 2019, and comprise over 60,000 observations.

[21] Before taking into account delay and cost increases that have subsequently arisen.

[22] Evidence Review: Transport', What Works Centre for Local Economic Growth, 2015.

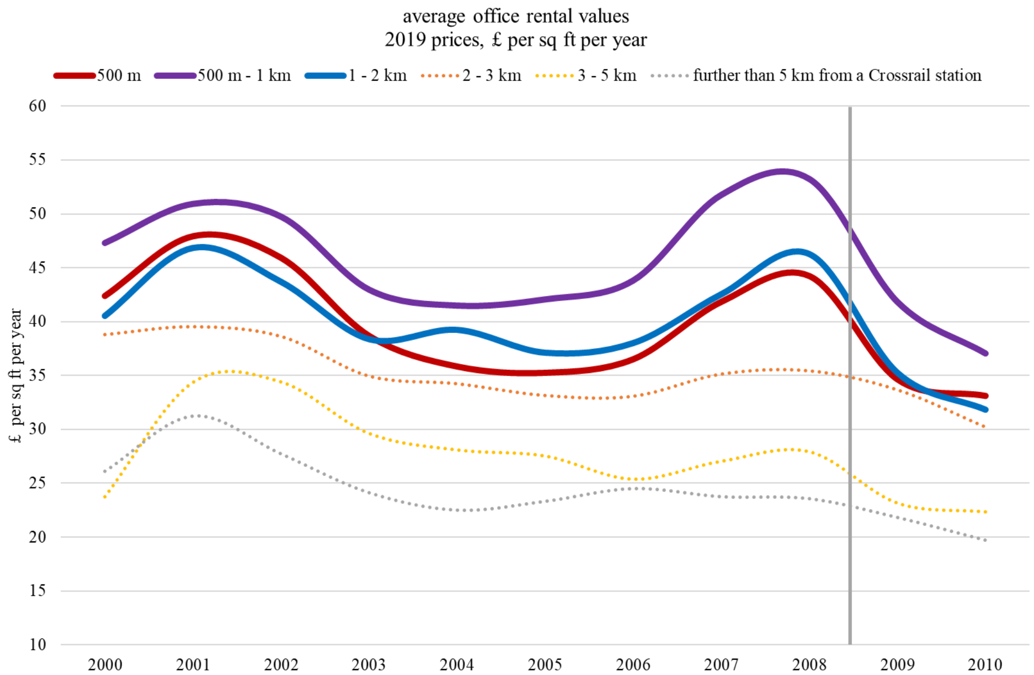
[23] 2011, Billings S.B., Estimating the value of a new transit option, Journal of Regional Science and Urban Economics

4.3 Pre-announcement trends

As we explain in chapter 2.6, there is a requirement that needs to be met to allow us to conduct a difference in differences analysis. Since we want treatment and comparison groups to be as similar as possible, we assume that they have been performing in the same way before the announcement of Crossrail. It does not require that they have had identical values of the outcome variables before the announcement, but that those value have been following similar trend.

Figure 11 presents changes in average office rents between 2000 and 2010 for treatment and control areas. Average values are lowest within 500m of Crossrail programme stations and highest within 500 – 1000 m, but in all those areas they have been moving in parallel. It means that the common trends assumption is met for the office rents analysis and we can proceed with modelling.

Figure 11. Average office rents, pre-announcement trends 2000-2010 for station distance bands



Source: CoStar data, sample excluding outliers and observations excluded based on the similarity test

Figure 12 below shows a high-level breakdown of the office rental dataset. Around 75% of transactions that occurred in the study area (50 miles around Charring Cross) took place in Inner London and 12% in Outer London. Around 7% are located in the West and the remainder from the South-West, North-East and North-West.

Figure 12. Number of office transactions for areas of London

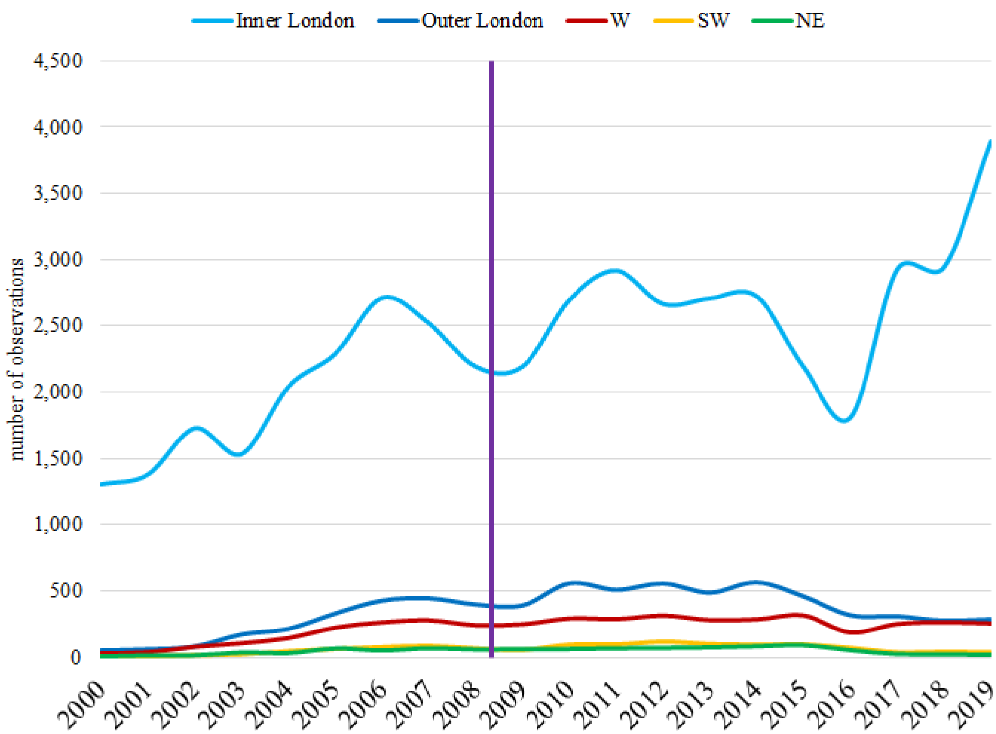
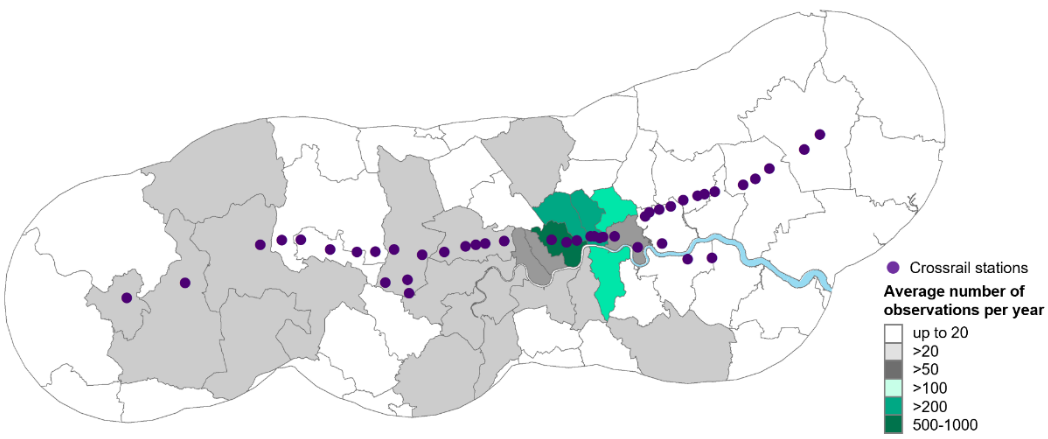


Figure 13 shows local authorities with the highest average annual number of transactions (average across all years) for the period 2000-2019. Inner London dominates the picture.

Figure 13. Office rental transactions – annual average per local authority 2000-2019



4.4 Understanding the data

Basic descriptive analysis of dataset composition and trends over time is a natural first and crucial step for any reporting based on data. In evaluation studies it is necessary in order to understand the broader picture of change and be able to validate our evaluation findings against it. In our case it provides valuable information for the future set-up of an ex-post evaluation study, identifying potential challenges.

Table 3 Office rental transactions – average yearly breakdown 2000-2019 (Arup analysis based on CoStar data)

Wider area	**Initial sample:** Average count	**Initial sample:** Average share	**Sample after excluding outliers:** Average count	**Sample after excluding outliers:** Average share	**Outliers**	**Outliers as % of initial sample**
London: Inner London	2,366	77%	2,308	77%	58	2%
London: Outer London	344	11%	324	11%	20	6%
Outside London: W	217	7%	209	7%	8	4%
Outside London: SW	63	2%	61	2%	2	4%
Outside London: NE	44	1%	44	1%	3	8%
Outside London: NW	23	1%	22	1%	3	12%
Outside London: E	12	0%	11	0%	2	16%
Outside London: S	4	0%	4	0%	2	43%
Outside London: N	2	0%	3	0%	1	44%
Sum	**3,076**	**100%**	**2,986**	**100%**	**99**	**3%**

Around 3% of the overall sample was excluded as outliers. When looking at data in more detail, a higher proportion of observations were excluded from areas outside London, although absolute numbers were much lower. In total we identified over 1,900 office transactions as outliers.

In addition to the outliers, almost 30% of the full dataset was excluded due to the similarity analysis (described in chapter 3.6) as being in locations dissimilar to the treated locations. Figure 17 maps all the excluded observations together – both outliers and the dissimilar ones.

Average office rents in the whole study area amounted to around £53 per square feet in 2019, an increase of around 30% since the end of 2008. Table 4 shows this change since 2008 for separate distance bands. Changes are higher when looking at median levels than when looking at averages. Generally, transactions with highest rents take place not in the direct vicinity of the station, but slightly further away, between 500m and 1 km (Figure 14).

Table 4 Change in office rents between 2008 and 2019 in areas around Crossrail programme stations (2019 prices, Arup analysis based on CoStar data)

Distance from Crossrail stations	Change in average rent	Change in median rent
Up to 500m	20%	31%
500m - 1 km	9%	20%
1-2 km	10%	26%
2-3 km	20%	22%
3-5 km	22%	26%
Further than 5 km	12%	28%

Figure 14. Median office rents around Crossrail programme stations, 2000 – 2019

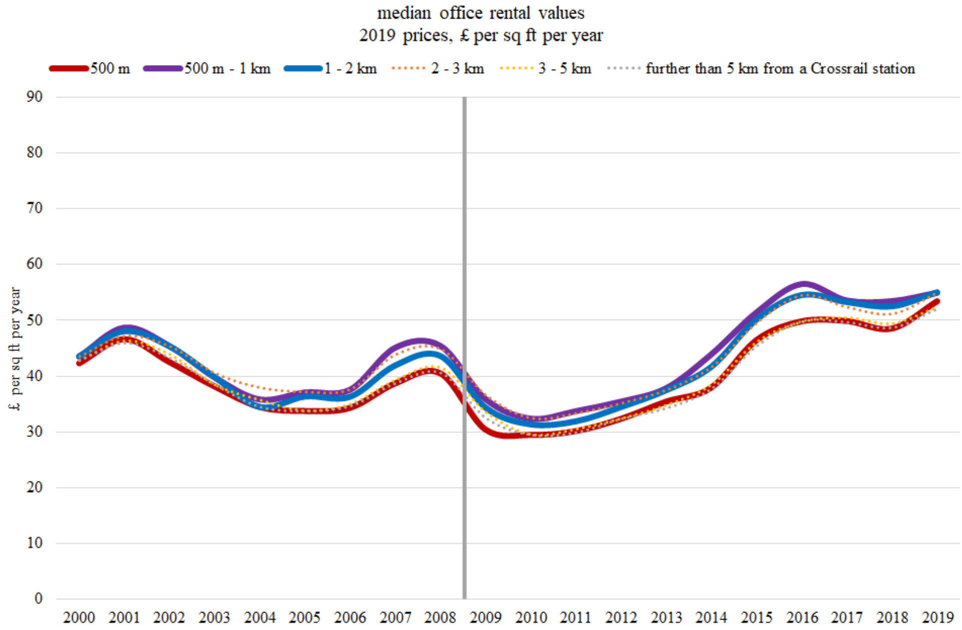


Figure 15. Median office rents around Crossrail programme stations, 2000 – 2019 index

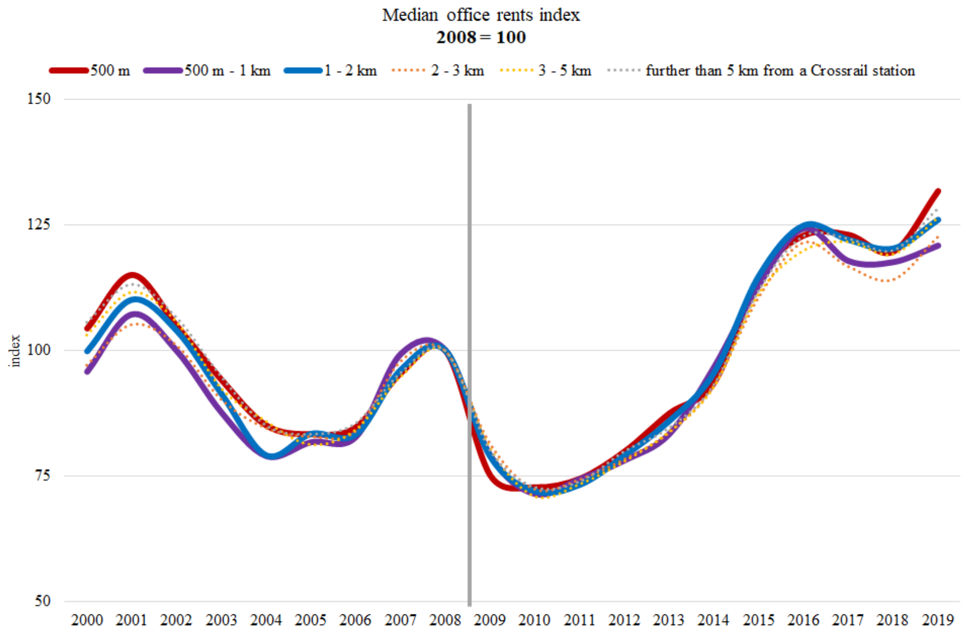


Figure 16. Office transactions without excluded records, 2000-2019

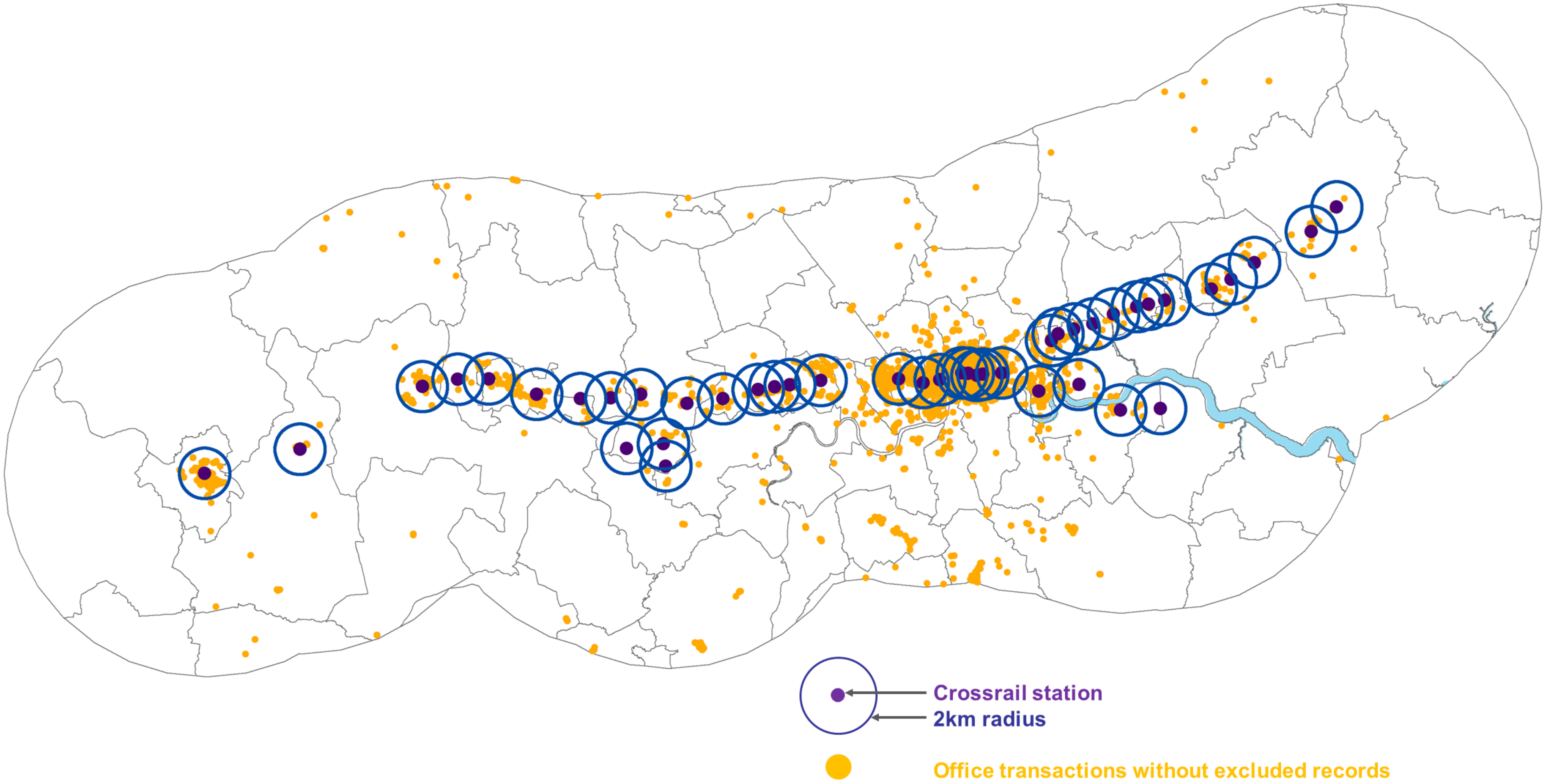
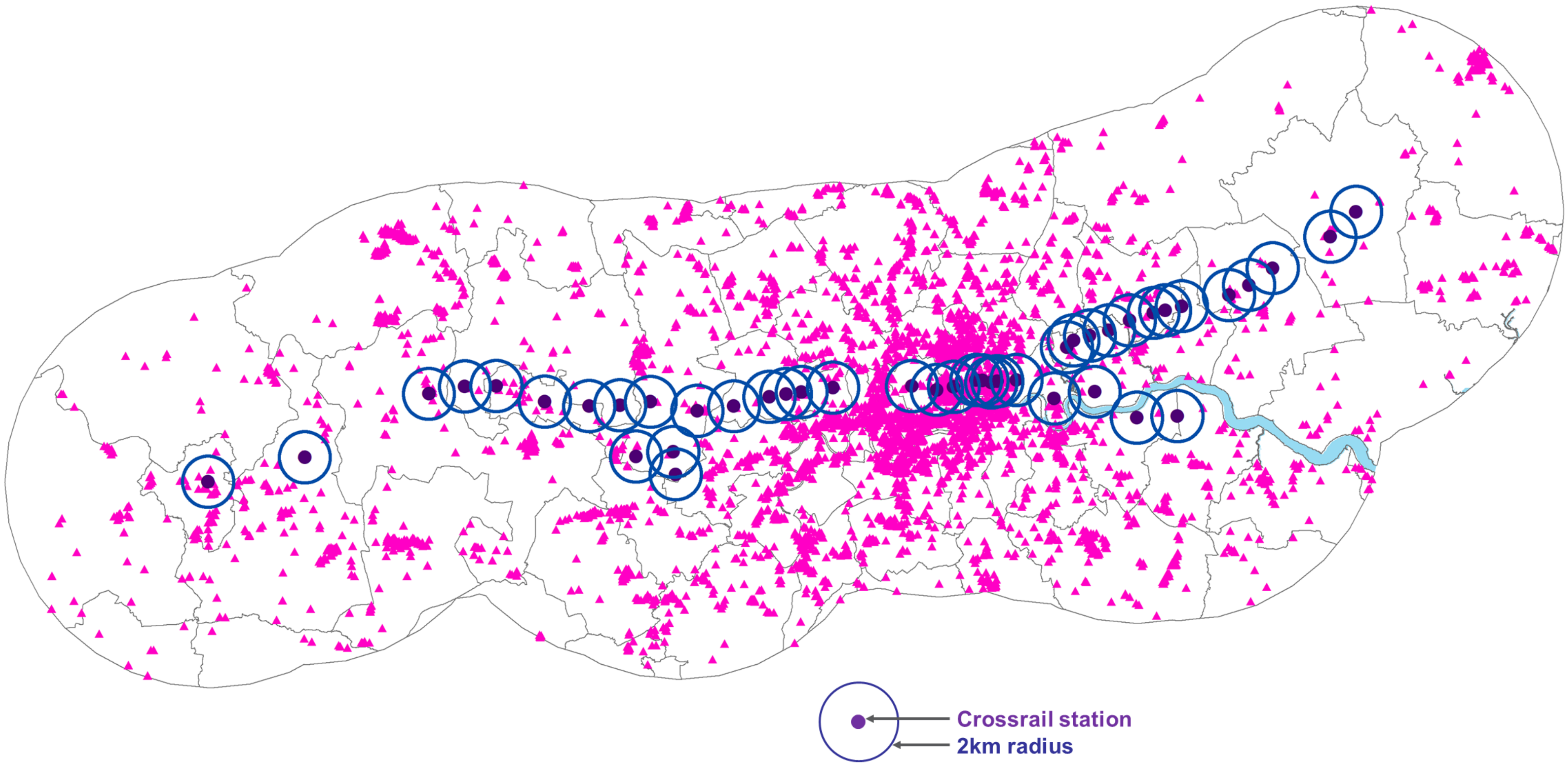


Figure 17. Office transactions: outliers and other excluded records, 2000-2019



4.5 Line-wide analysis

Column 1 in Table 5 below presents estimates of a simple difference in differences regression model 1 including three distance dummy variables interacted with the treatment variable and area- and year-fixed effects. The area fixed effects were estimated at full three-digit postcode level. This model suggests an overall pre-opening impact in three distance bands from Crossrail programme stations: 2.4%, 8.1% and 7.6% respectively in the 500m, 500m to 1 km and 1 to 2 km bands.

Estimates of model 2 and model 3 are very similar to the ones in model 1, but the estimate of impact in the 500m band is not statistically significant.

In model 4 we include interactions between year variables and local authority dummy (LAD) variables to control for any wider changes impacting districts in a different way, such as policy changes. The finding that the estimate for the last band impact becomes non-significant in model 4 is surprising, as it was consistently significant and at very similar level across all previous models (1-3). This suggests that adding interactions between local authorities and year variables controls for some part of the impact of the Crossrail announcement evident in previous three models. In addition, in model 4 the estimate at 'post announcement' becomes strongly significant and negative. This is counterintuitive, as we know from baseline analysis that on average rents have increased in the post announcement period, not decreased. Again, it seems that model 4 underestimates changes in property prices and Crossrail pre-opening impacts on them. Due to these findings, we don't consider the results of model 4 robust enough for interpretation.

Table 5. Regressions: Office rent impacts, for station distance bands full dataset, models 1-4

	(1)	**(2)**	**(3)**	**(4)**
	Base model	**with controls**	**with controls and employment shares**	**with LAD-Year interactions**
Post announcement	0.0347*	0.00413	-0.0104	-0.0726**
	(0.0205)	(0.0220)	(0.0224)	(0.0364)
Treated 500m post announcement	**0.0243***	0.0106	0.00886	**0.0281***
	(0.0141)	(0.0142)	(0.0142)	(0.0155)
Treated 500m to 1 km post announcement	**0.0814*****	**0.0806*****	**0.0796*****	**0.0781*****
	(0.0125)	(0.0122)	(0.0120)	(0.0149)
Treated 1 to 2 km post announcement	**0.0759*****	**0.0690*****	**0.0716*****	0.0108
	(0.0133)	(0.0133)	(0.0131)	(0.0152)
Constant	3.600***	3.461***	3.337***	3.575***
	(0.00882)	(0.0194)	(0.0516)	(0.0467)
Observations	60,554	60,554	60,553	60,553
R-squared	0.763	0.765	0.766	0.786
Adjusted R-squared	0.708	0.711	0.711	0.730
Year FE	Yes	Yes	Yes	Yes
Postcode FE	Yes	Yes	Yes	Yes
LAD FE		Yes	Yes	Yes

	(1) Base model	(2) with controls	(3) with controls and employment shares	(4) with LAD-Year interactions
LAD-Year FE				Yes

Column 6 in Table 6 shows the estimates of model 6, using an incidental treatment approach and local authority-time fixed effects (model 5 is very similar but does not include the authority-time fixed effects). None of the impact estimates are statistically significant. We found no statistically significant average line-wide impact on 'in-between areas'. This suggests that any line-wide impact found in previous models is in fact coming from the central section.

Table 6. Regressions: Office rent impacts, for station distance bands, full dataset, models 5-6

	(5)	(6)
	Base model with restricted observations	**With LAD-Year interactions and restricted observations**
Post announcement	0.0404	-0.150
	(0.0643)	(0.128)
Treated 500m post announcement	**-0.170*** **	-0.0775
	(0.0547)	(0.116)
Treated 500m to 1 km post announcement	0.0496	0.125
	(0.0561)	(0.109)
Treated 1 to 2 km post announcement	-0.0433	0.0403
	(0.0460)	(0.109)
Constant	3.103***	3.411***
	(0.117)	(0.115)
Observations	4,100	4,100
R-squared	0.818	0.870
Adjusted R-squared		
Year FE	Yes	Yes
Postcode FE	Yes	Yes
LAD FE		Yes
LAD-Year FE		Yes

Table 7 presents estimates of models 1-4, excluding a pool of observations as per the approach described in chapter 4. Modelling results are very similar to those reproduced in Table 5. This suggests that the results are robust to changes in the composition of the data sample and that our similarity analysis does not introduce any bias.

Table 7. Regressions: Office rent impacts, for station distance bands, records excluded, models 1-4

	(1) Base model	(2) with controls	(3) with controls and employment shares	(4) with LAD-Year interactions
Post announcement	0.0440**	0.0178	-0.0107	-0.0634**
	(0.0209)	(0.0213)	(0.0228)	(0.0301)
Treated 500m post announcement	0.0183	0.00604	0.00434	**0.0303****
	(0.0137)	(0.0136)	(0.0134)	(0.0142)
Treated 500m to 1 km post announcement"	**0.0656*****	**0.0668*****	**0.0706*****	**0.0768*****
	(0.0136)	(0.0130)	(0.0129)	(0.0149)
Treated 1 to 2 km post announcement	**0.0641*****	**0.0567*****	**0.0670*****	0.0209
	(0.0143)	(0.0147)	(0.0148)	(0.0165)
Constant	3.751***	3.622***	3.459***	3.696***
	(0.00959)	(0.0200)	(0.0663)	(0.0631)
Observations	42,079	42,079	42,078	42,078
R-squared	0.748	0.752	0.753	0.772
Adjusted R-squared	0.701	0.705	0.706	0.724
Year FE	Yes	Yes	Yes	Yes
Postcode FE	Yes	Yes	Yes	Yes
LAD FE		Yes	Yes	Yes
LAD-Year FE				Yes

4.6 Section analysis

Model 4 is consistently the strongest with regard to the goodness-of-fit measure (R-Squared and Adjusted R-Squared) and provides most conservative estimates of impact, which is why we focus on that specification when it comes to section and station analysis.

Model 4 for the central section (Table 8) shows an 8.0% impact within the 500m - 1 km buffer and 3.1% in the 500m radius. The results of analysis conducted on a limited dataset are very similar (Table 9), which means they are robust to the sample composition.

In addition, in both cases we find significant estimates for the 500m - 1 km band in the South Eastern section, although their values are quite different (3.8% vs. 6.5%).

Table 8 Regressions: Office rent impacts, for station distance bands and line sections, full dataset, model 4

	Outer Western	Western London	Central	South Eastern	Eastern
Post announcement	-0.245	-0.251	-0.0794**	-0.250	-0.250
	(0.182)	(0.183)	(0.0367)	(0.183)	(0.183)
Treated 500m post announcement	-0.0837	-0.103	**0.0315****	-0.00192	-0.00192
	(0.117)	(0.196)	(0.0159)	(0.0843)	(0.0843)
Treated 500m to 1 km post announcement"	-0.0779	0.143	**0.0805*****	**0.380*****	**0.380*****
	(0.151)	(0.125)	(0.0152)	(0.143)	(0.143)

	Outer Western	Western London	Central	South Eastern	Eastern
Treated 1 to 2 km post announcement	-0.147	0.0669	0.0146	-0.159	-0.159
	(0.116)	(0.0701)	(0.0161)	(0.223)	(0.223)
Constant	3.308***	3.336***	3.605***	3.326***	3.326***
	(0.0626)	(0.0646)	(0.0488)	(0.0647)	(0.0647)
Observations	21,288	20,054	57,363	19,791	20,248
R-squared	0.783	0.788	0.777	0.788	0.789
Adjusted R-squared	0.688	0.692	0.721	0.692	0.691
Year FE	Yes	Yes	Yes	Yes	Yes
Postcode FE	Yes	Yes	Yes	Yes	Yes
LAD FE	Yes	Yes	Yes	Yes	Yes
LAD-Year FE	Yes	Yes	Yes	Yes	Yes

Table 9. Regressions: Office rent impacts, for station distance bands and line sections, excluded records, model 4

	Outer Western	Western London	Central	South Eastern	Eastern
Post announcement	-0.148	-0.143	-0.0695**	-0.155	-0.149
	(0.124)	(0.130)	(0.0306)	(0.129)	(0.131)
Treated 500m post announcement	0.0423	-0.196	**0.0300****	0.0832	-0.182
	(0.287)	(0.660)	(0.0145)	(0.272)	(0.180)
Treated 500m to 1 km post announcement"	0.0731	-0.131	**0.0778*****	**0.654*****	-0.0313
	(0.291)	(0.119)	(0.0152)	(0.125)	(0.152)
Treated 1 to 2 km post announcement	-0.0765	0.0821	0.0239	0.164	0.0756
	(0.298)	(0.136)	(0.0172)	(0.250)	(0.0977)
Constant	3.420***	3.733***	3.745***	3.648***	3.628***
	(0.123)	(0.143)	(0.0682)	(0.145)	(0.142)
Observations	3,872	2,643	39,129	2,444	2,821
R-squared	0.857	0.887	0.731	0.878	0.893
Adjusted R-squared	0.785	0.817	0.679	0.805	0.822
Year FE	Yes	Yes	Yes	Yes	Yes
Postcode FE	Yes	Yes	Yes	Yes	Yes
LAD FE	Yes	Yes	Yes	Yes	Yes
LAD-Year FE	Yes	Yes	Yes	Yes	Yes

The central section is a strong component of most of the location categories we introduced in chapter 3.1. Unsurprisingly then, the analysis of these categories in Table 9 shows similar findings as for the central section: 2-3% significant impact within 500m and 6-7% impact within the second distance buffer. We also see a 10% impact within the second buffer for areas with significant levels of deprivation.

Table 10. Regressions: Office rent impacts, station distance bands and location categories, excluded records, model 4

	Deprived	With major associated development	Major regeneration area	Expected additional labour supply
Post announcement	-0.0402	0.000593	-0.0735	-0.0680

	Deprived	With major associated development	Major regeneration area	Expected additional labour supply
	(0.0843)	(0.0456)	(0.0647)	(0.0643)
Treated 500m post announcement	0.0202	**0.0287****	**0.0269***	**0.0323****
	(0.0338)	(0.0145)	(0.0158)	(0.0144)
Treated 500m to 1 km post announcement"	**0.107*****	**0.0754*****	**0.0609*****	**0.0722*****
	(0.0377)	(0.0162)	(0.0182)	(0.0161)
Treated 1 to 2 km post announcement	-0.103	0.0201	0.0113	0.0114
	(0.116)	(0.0192)	(0.0195)	(0.0189)
Constant	3.538***	3.727***	3.728***	3.709***
	(0.172)	(0.0683)	(0.0808)	(0.0681)
Observations	7,745	37,116	31,338	37,682
R-squared	0.746	0.714	0.754	0.723
Adjusted R-squared	0.694	0.662	0.706	0.673
Year FE	Yes	Yes	Yes	Yes
Postcode FE	Yes	Yes	Yes	Yes
LAD FE	Yes	Yes	Yes	Yes
LAD-Year FE	Yes	Yes	Yes	Yes

Table 11 summarises the impacts at section level, and Table 12 shows the numbers of treatment observations for this analysis.

Table 11. Impacts summary: Office rents, for station distance bands, section-level analysis

Section	500m	500-1000m	1-2 km
1 - Western outside	-	-	-
2 - Western London	-	-	-
3 - Central	3% **	8% ***	-
4 - South Eastern	-	6.5% ***	-
5 - Eastern	-	-	-
6 - Deprived	-	10% ***	-
7 - Areas with associated major development	3% **	7.5% ***	-
8 - Areas with wider regeneration or growth initiative	2.5% *	6% ***	-
9 - Areas expected to benefit from additional labour supply	3% *	7% ***	-

Table 12. Treatment data counts: Office rents, for station distance bands, section-level analysis

Section	500m	500-1000m	1-2 km	0-2 km
1 - Western outside	668	508	522	1,698
2 - Western London	115	185	336	636
3 - Central	5,699	17,101	7,491	30,291
4 - South Eastern	246	131	105	482
5 - Eastern	546	201	521	1,268
6 - Deprived	3,102	5,817	7,460	16,379
7 - Areas with associated major development	5,728	17,124	8,585	31,437
8 - Areas with wider regeneration or growth initiative	6248	15,648	4,566	26,462
9 - Areas expected to benefit from additional labour supply	5,941	17,398	14,345	37,684

4.7 Station analysis

Table 13 shows the number of office transactions within each distance band by station. Only the central stations (black ink) have enough observations to allow for a robust analysis. In Slough, Southall, Ealing Broadway, Custom House, Abbey Wood and Romford we observed on average only a few transactions per year (these stations are underlined in Table 13 for readers' convenience).

Please note that in the case of Bond Street and Tottenham Court Road stations the 1000-2000m buffer areas are overlapping. This means that some of the impact may be double-counted so the results for these two stations should not be added up.

Table 13. Number of office transactions by station and station distance bands

Station	500m	500-1000m	1000-2000m	0-2000m
1 - Reading	308	284	154	746
2 - Slough	77	82	39	198
3 - Southall	2	15	34	51
4 - Ealing Broadway	48	124	57	229
5 - Paddington	176	275	1,743	2,194
6 - Bond Street	1,929	6,719	7765	16,413
7 - TCR	2,974	7,708	10,789	21,471
8 - Farringdon	3,729	4,669	12,160	20,558
9 - Whitechapel	60	523	6,216	6,799
10 - Canary Wharf	287	501	253	1,041
11 - Custom House	16	112	96	224
12 - Abbey Wood	0	0	5	5

Station	500m	500-1000m	1000-2000m	0-2000m
13 - Romford	100	54	20	174

From Figure 19 to Figure 25 we present graphs of change in the average rent around the central stations and Reading. There are four cases where the common trends assumption is not exactly met and deviates significantly:

- Reading: all distance bands;
- Paddington: 0 – 500m;
- Whitechapel: 0 – 500m; and
- Canary Wharf: 1 – 2 km.

Table 14 presents Model 4 estimates for given stations (please refer to chapter 3.1 for an explanation of how we conduct station-level analyses). We present results from modelling on both the full and limited datasets, without data outliers.

We found a significant, positive and consistent impact across two types of datasets in the second distance band, between 500m – 1 km from the station, for all selected central stations, with the exception of Farringdon. As a side-check, we ran the analysis for the three cases mentioned above with unmet common trends assumptions. As expected, the model did not yield any significant estimates for those.

For Bond Street and Tottenham Court Road we found impacts for both the 500m and 500m – 1 km bands. However, the 500m and 500m – 1 km bands for these stations overlap, so the 0-500m impacts might be overestimated. Similarly, the 500m – 1 km estimate for Whitechapel might be overestimated as it is largely within the same distance from Liverpool Street station (which was not selected for an analysis as part of the wider project). Figure 18 shows the overlap between the distance bands for the central section.

Findings for stations are consistent with findings from the section analyses, where we found significant impacts of around 6 – 8% in 500m – 1 km for most section typologies.

Table 14. Impacts summary: variation of office rent impacts attributable to Crossrail announcement, station-level analysis for station distance bands, Model 4

Station	0-500m: All	0-500m: Excl.	500m-1km: All	500m-1km: Excl.	1km-2km: All	1km to 2km: Excl.
1 - Paddington	x	x	-	-	10%	17%
2 - Bond Street	10%	10%	11%	11%	-	-
3 - TCR	6%	5%	7%	5%	-	-
4 - Farringdon	-	-	4%	-	-	7%
5 - Whitechapel	x	x	9%	11%	-	-
6 - Canary Wharf	-	-	12%	18%	x	x

Figure 18. Office rents: overlaps of distance bands from Bond Street to Whitechapel

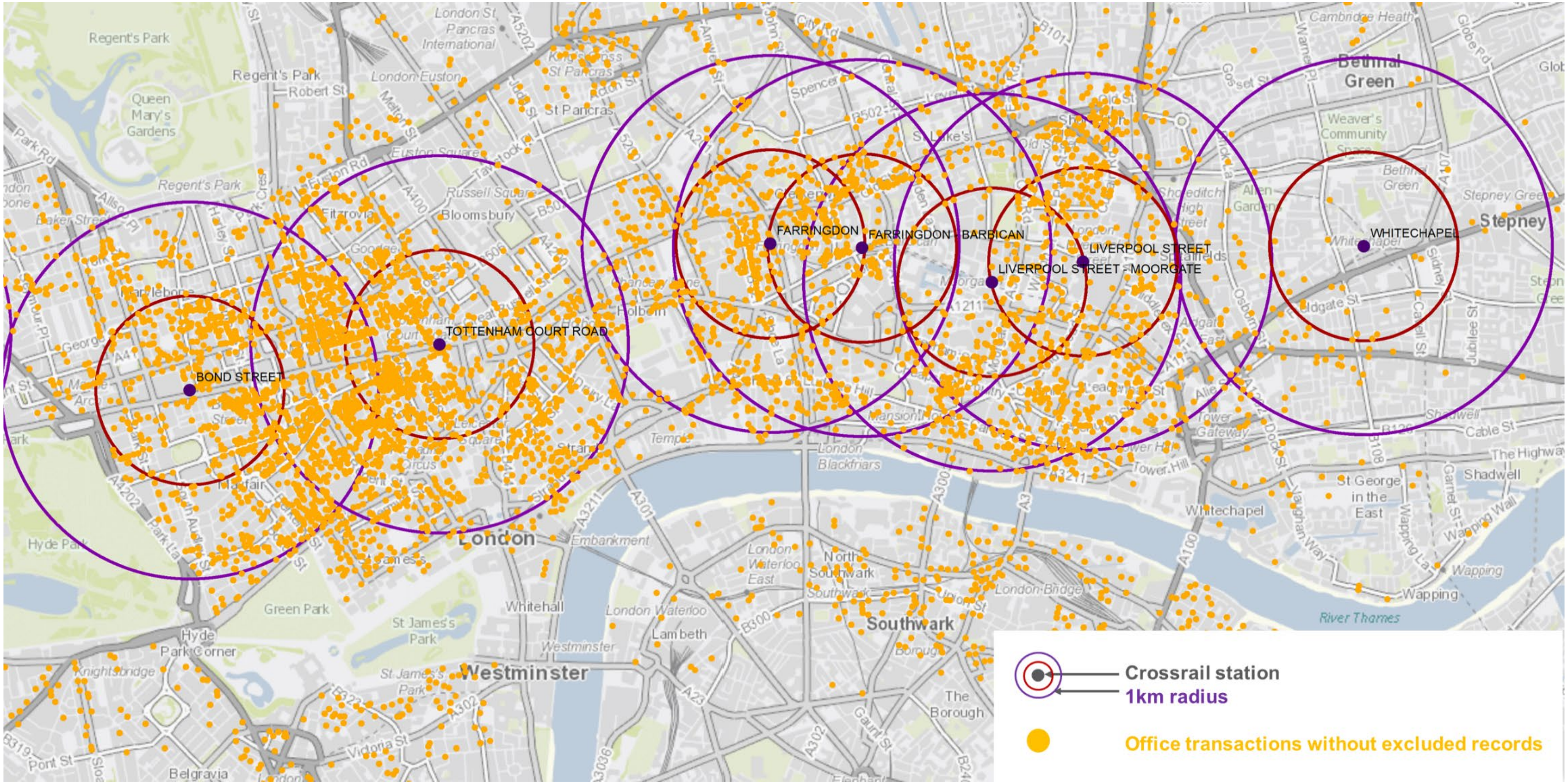


Figure 19. Average office rents for station distance bands around Reading station

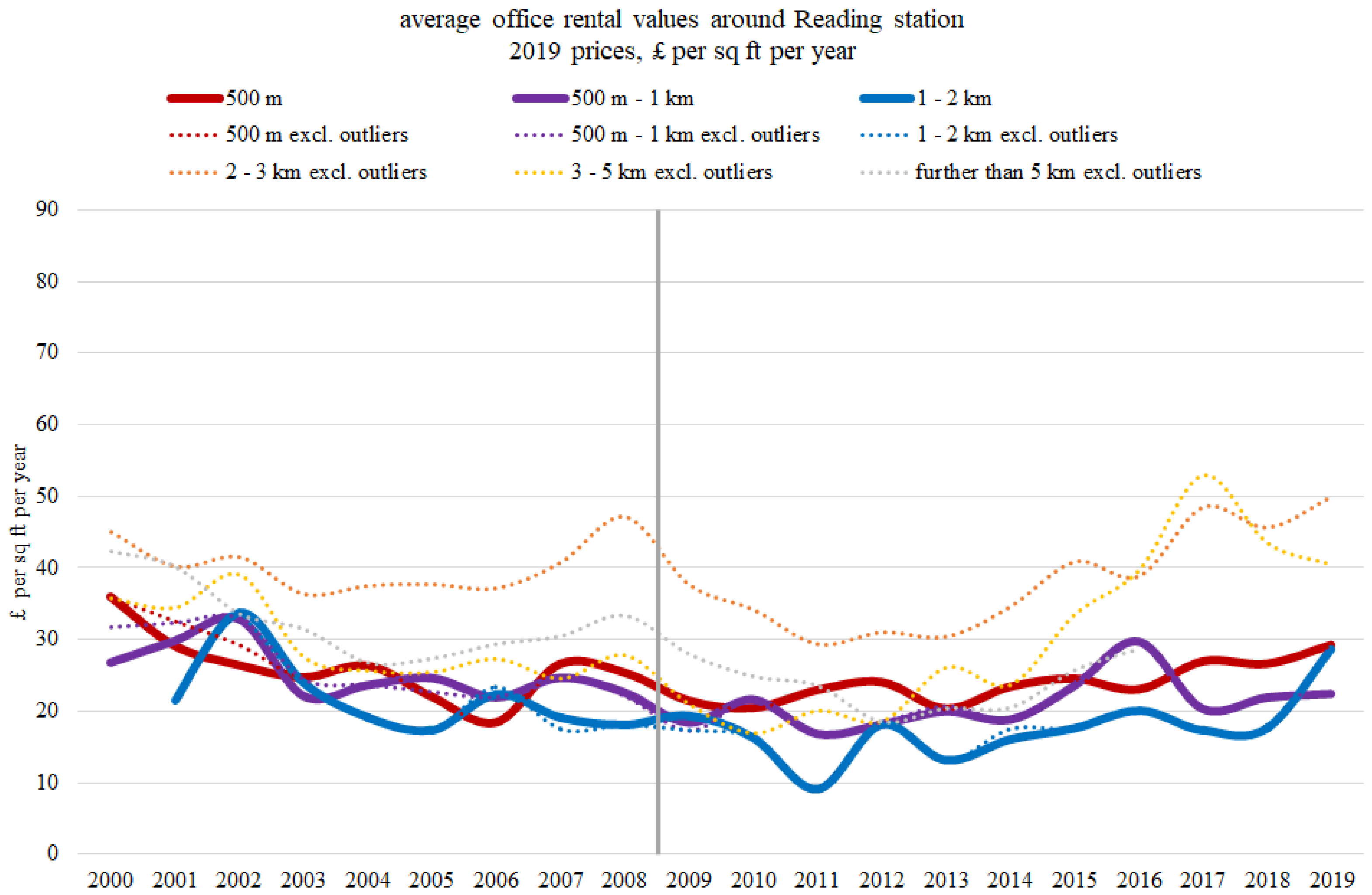


Figure 20. Average office rents for station distance bands around Paddington station

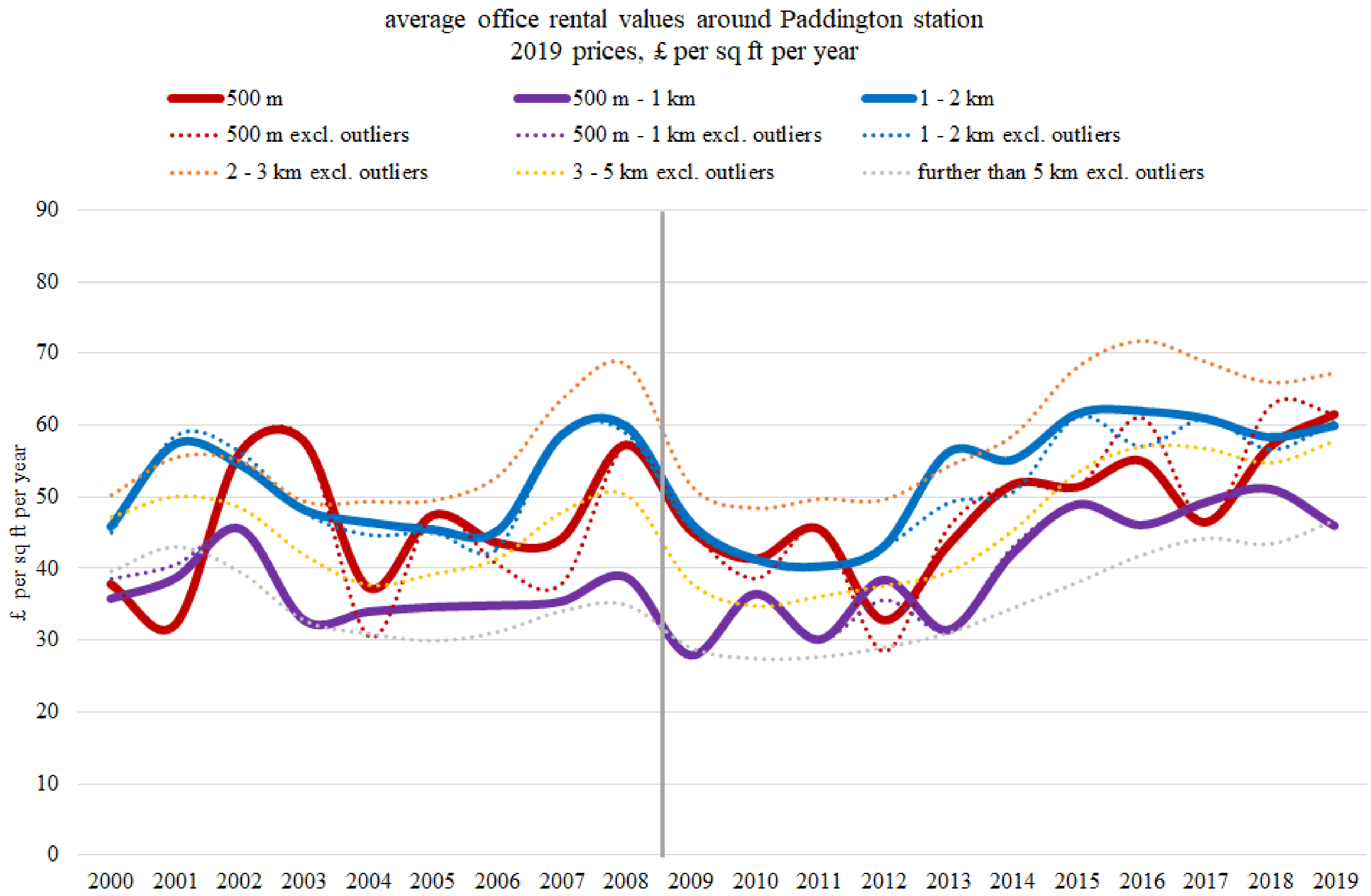


Figure 21. Average office rents for station distance bands around Bond Street station

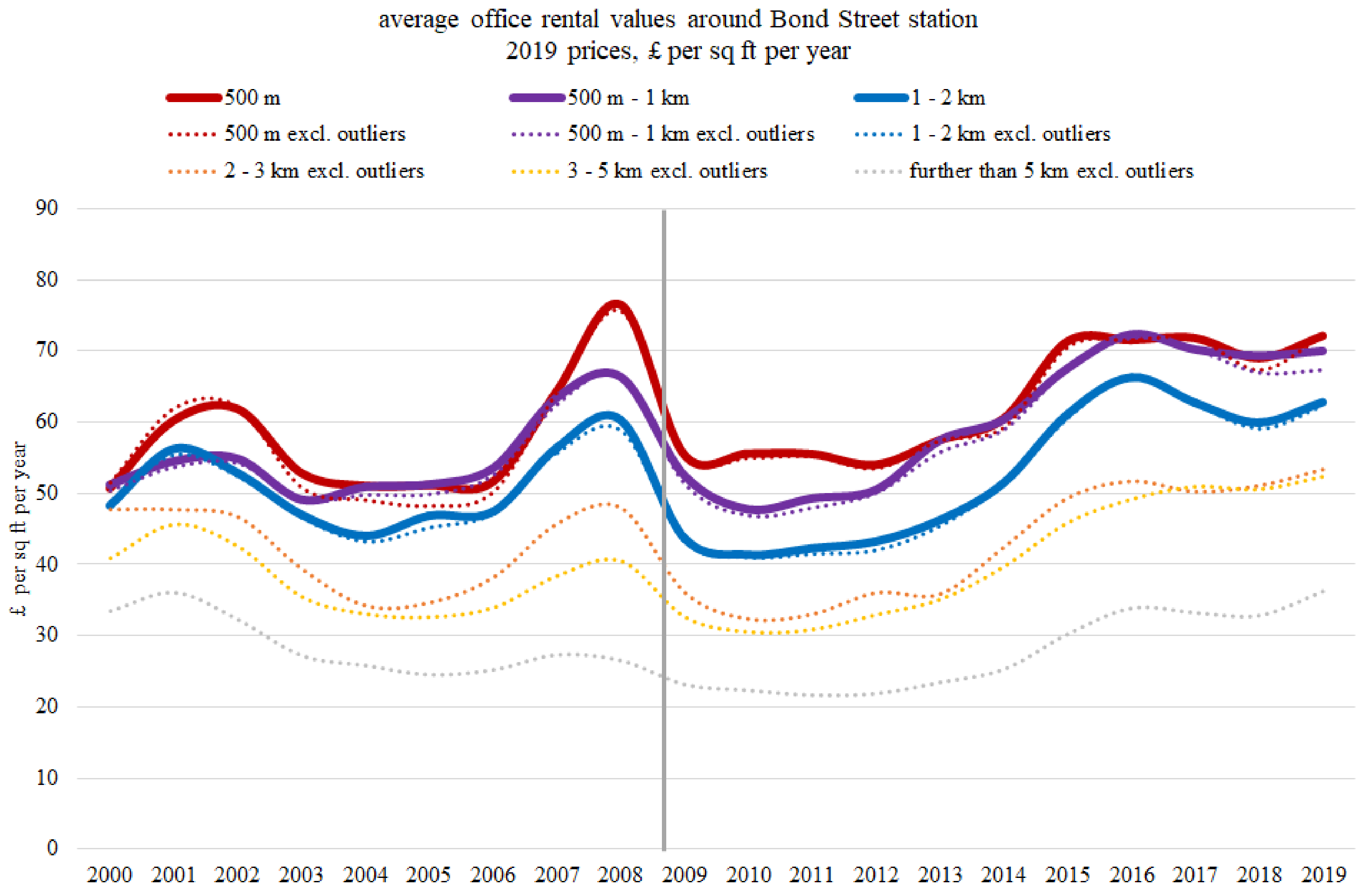


Figure 22. Average office rents for station distance bands around Tottenham Court Road station

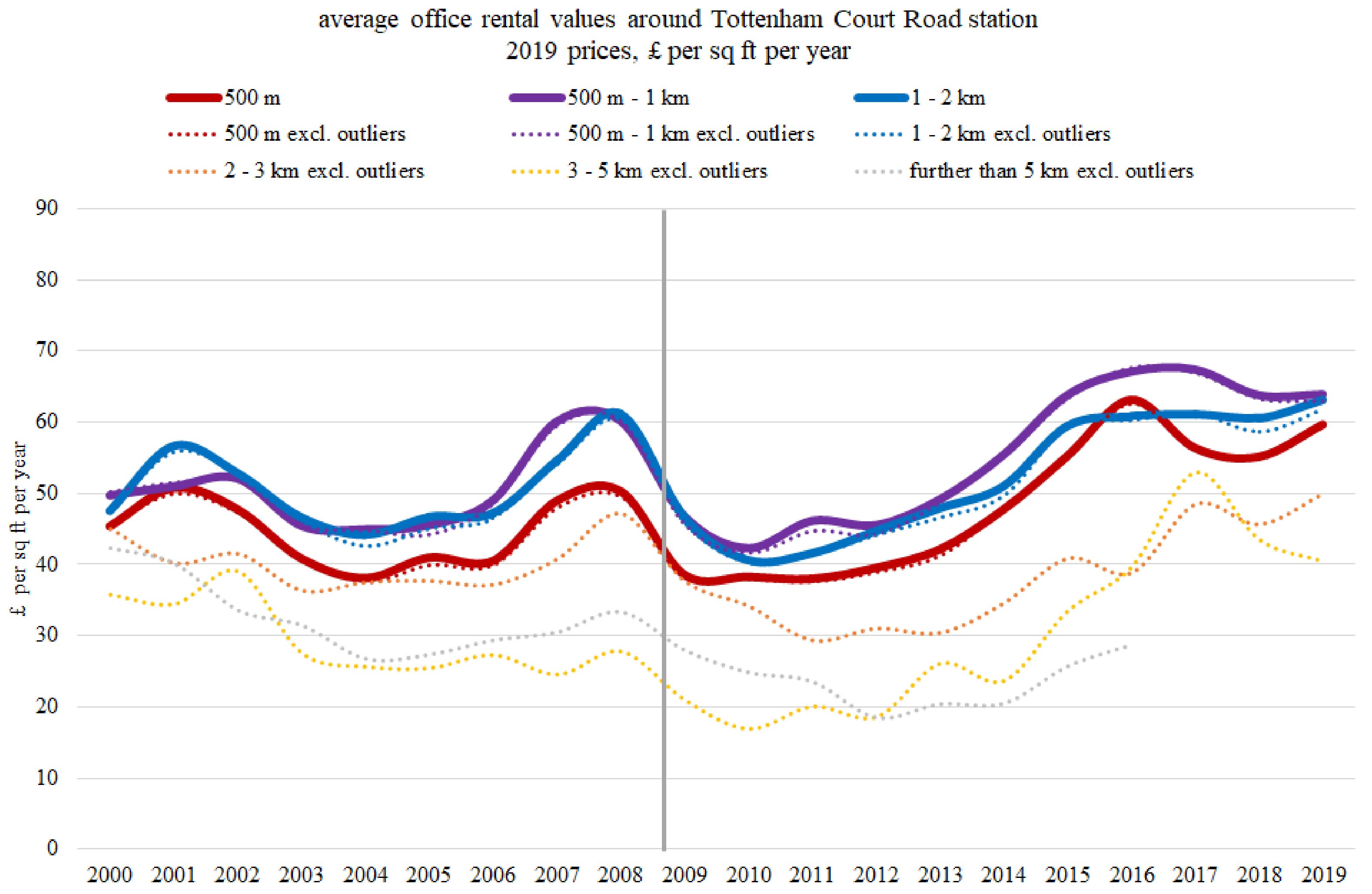


Figure 23. Average office rents for station distance bands around Farringdon station

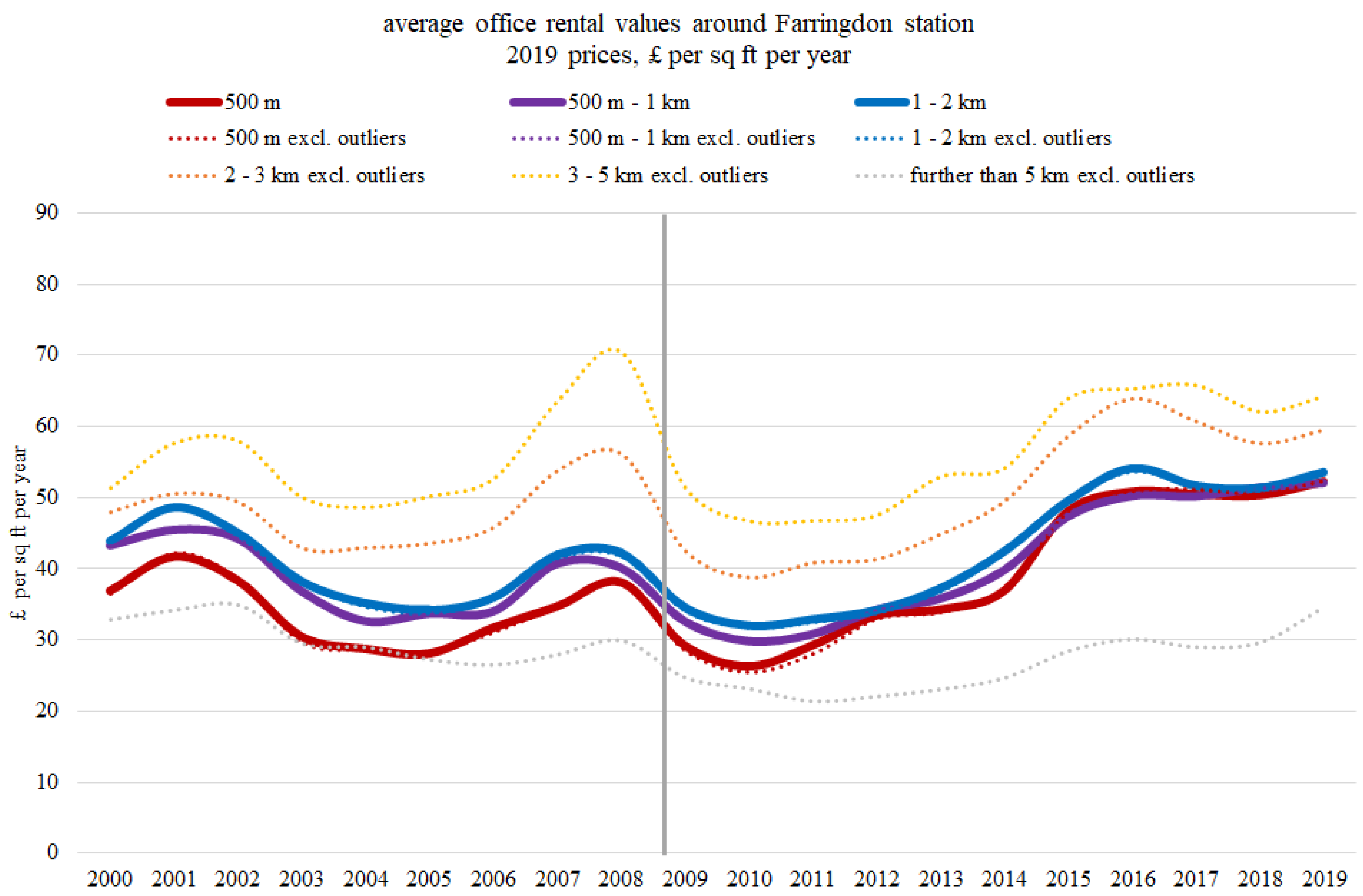


Figure 24. Average office rents for station distance bands around Whitechapel station (no transactions were completed for some years within 500m)

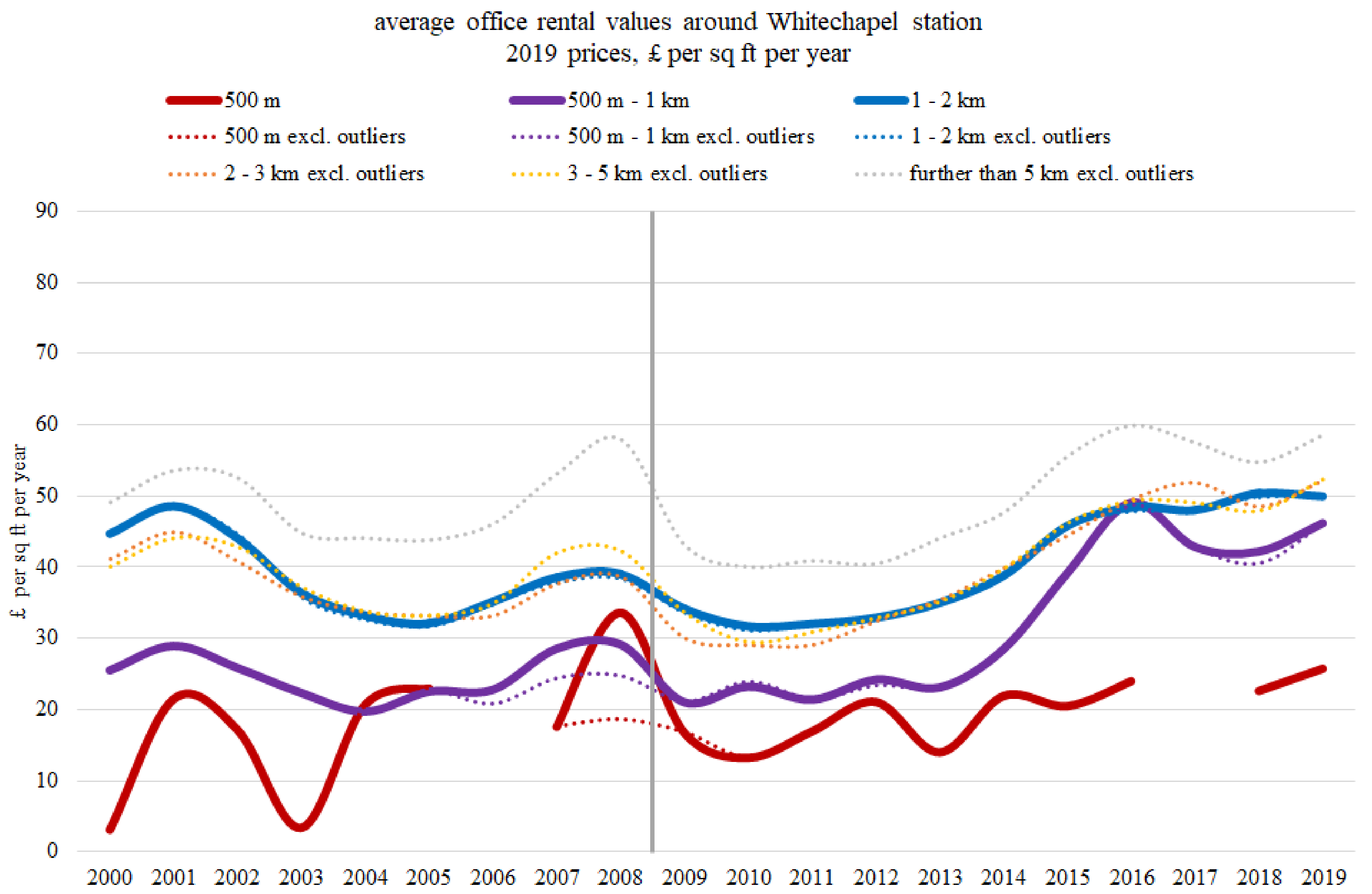
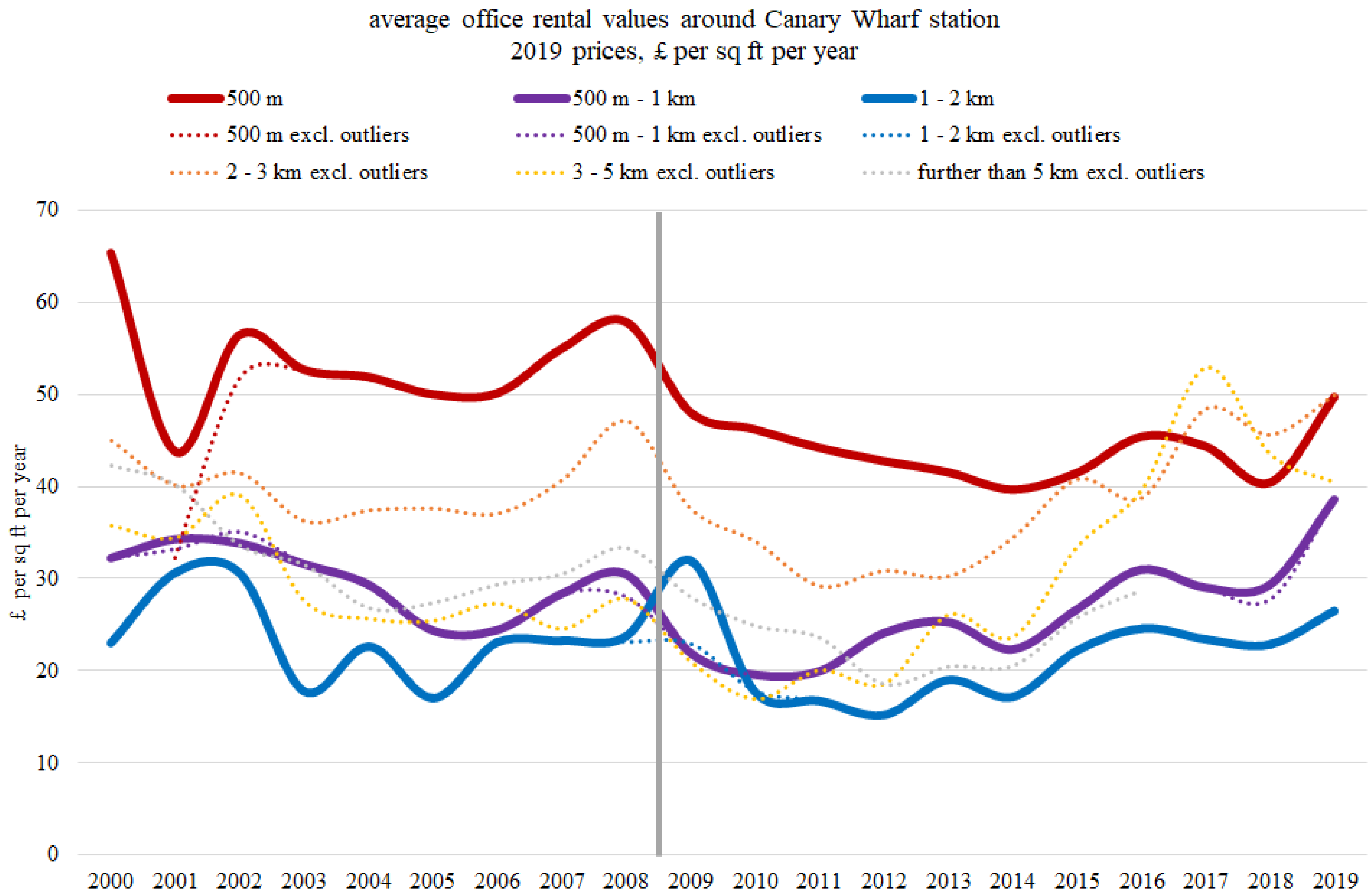


Figure 25. Average office rents for station distance bands around Canary Wharf station



4.8 Summary

Our hypothesis was that it would be possible to observe some announcement impacts on the value of office rents in areas of existing business markets, and in those expecting to experience the largest improvements due to the opening of the Elizabeth line.

We found an average line-wide announcement impact on office rental values of around 3% within 500m of Crossrail programme stations and around 7.5% between 500m and 1 km from the stations. However, using the incidental approach, we are able to confirm that these impacts are driven by changes in central London. This is also seen in the section analysis, where most typologies are skewed toward central stations.

We also found that areas considered as being deprived and those associated with major development have experienced higher than line-average impacts on their office rents. This is specifically true for areas between 500m and 1km of stations in areas deprived or with major associated development where the announcement had respectively an impact of 10.6% and 8.0% on the value of office rents. This confirms that these areas might experience the highest gains in attractivity for new businesses as a result of their enhanced connectivity to central London and of relatively low office rents compared to more central areas.

Table 15. Summary of office rent impacts by station distance band and category

Category			
Line-wide	3.0% **	7.5% ***	***No impact found***
Geographic section	Central: 3.0%	Central: 8.0%	***No impact found***
Deprived areas	***No impact found***	10.6% ***	***No impact found***
Areas of with major associated development	3.0% **	8.0% ***	***No impact found***
Areas of major regeneration or growth initiative	3.2%	7.0% ***	***No impact found***
Areas expecting increase in labour supply	3.4% **	7.6% ***	***No impact found***

Table 16 presents estimates of average impacts with general changes in average yearly rents between 2008 and 2019. The estimated average pre-opening Crossrail impact in the areas closest to the stations are 3% and 7.5% for areas between 500m and 1km. These estimated impacts are lower than the average of median rent increase which suggests that Crossrail was not the only factor which significantly affected rents during this period.

Table 16. Change in office rents between 2019 and 2008 around Crossrail stations

Distance	**Change in average rent**	**Change in median rent**	**Estimated average impact**
Up to 500m from the stations	20%	31%	3%
Between 500m and 1 km from the stations	9%	20%	7.5%
Between 1 km and 2 km from the stations	10%	26%	***no impact found***
Between 2 km and 3 km from the stations	20%	22%	***no impact found***
Between 3 km and 5 km from the stations	22%	26%	***no impact found***
Further than 5 km from the stations	12%	28%	***no impact found***

5 Retail rental value impacts

We were not able to test Crossrail pre-opening impacts on rental values because the common trend assumption could not be met. The data up to 2019 shows that it might be possible for that assumption to be met for the ex-post analysis.

5.1 Introduction

The interviews we carried out as part of the wider project show that many respondents expect retail catchments to change as a result of the Elizabeth line opening. This could lead to retail spaces around stations becoming more desirable for retailers than they would otherwise have been. Conversely, the Elizabeth line will only call at existing stations, which are all already served by a rail-based mode of transport and many of which are already established retail centres. It is therefore more likely that we can expect to see some local effects around specific stations as opposed to line-wide impacts. Nevertheless, many businesses may not be willing to take on higher rental costs before the line opens, and before it is clear that customer footfall has increased.

We would therefore not expect to see significant impacts on retail rents before the Elizabeth line opens.

5.2 Data source

Retail rental data come from the CoStar database, the same source that was used for office and industrial rent transactions. The data extracted for the analysis comprise all recorded retail rental deals that have been transacted within the study area over the baseline period. The study area we used comprises a 10-mile (16 km) corridor around the route of the Elizabeth line. We extracted data covering the period 2000 to 2019 and have over 23,000 observations in the full sample.

5.3 Pre-announcement trends

As we explain in chapter 2.6, there is a requirement that needs to be met to allow us to conduct a difference in differences analysis. Since we want treatment and comparison groups to be as similar as possible, we want to confirm that they performed in the same way before the Crossrail announcement. They are not required to have had identical values of the outcome variables before the announcement, but those values must have been following similar trends.

The common trends assumption is evidently not met before 2005 (see Figure 26). There is more parallel behaviour in the period between 2005 and 2009, but values around Crossrail programme stations increased significantly in 2006 and then dropped in 2009, while further than 2 km from Crossrail programme stations there was a gentle downward trend across this period.

Although the common trends assumption is not met for the full pre-announcement period, our main model includes interactions between local authority and year fixed effects, which to some extent might control for the differences in pre-announcement trends. For this reason, we decided to run the analyses, but not include the outputs. Regardless, we did not find any consistently significant impacts for retail rents.

5.4 Understanding the data for post-opening analysis

Figure 28 (below) shows the number of transactions recorded by geographic region. Just as with office transactions, most of the data comes from within London. Around 75% of all data come from Inner London and 12% from Outer London, with 9% from the Western part of the study area.

Figure 26 shows that the common trends assumption might not be met in post-opening analysis since retail rents in the areas closest to Crossrail programme stations fluctuate in a different way to those in other potential treatment areas. However, if we exclude the 500m treatment band, it appears that the common trends assumption is met from 2014 onwards for all other areas.

Figure 26. Average retail rents for station distance bands, 2000 – 2019

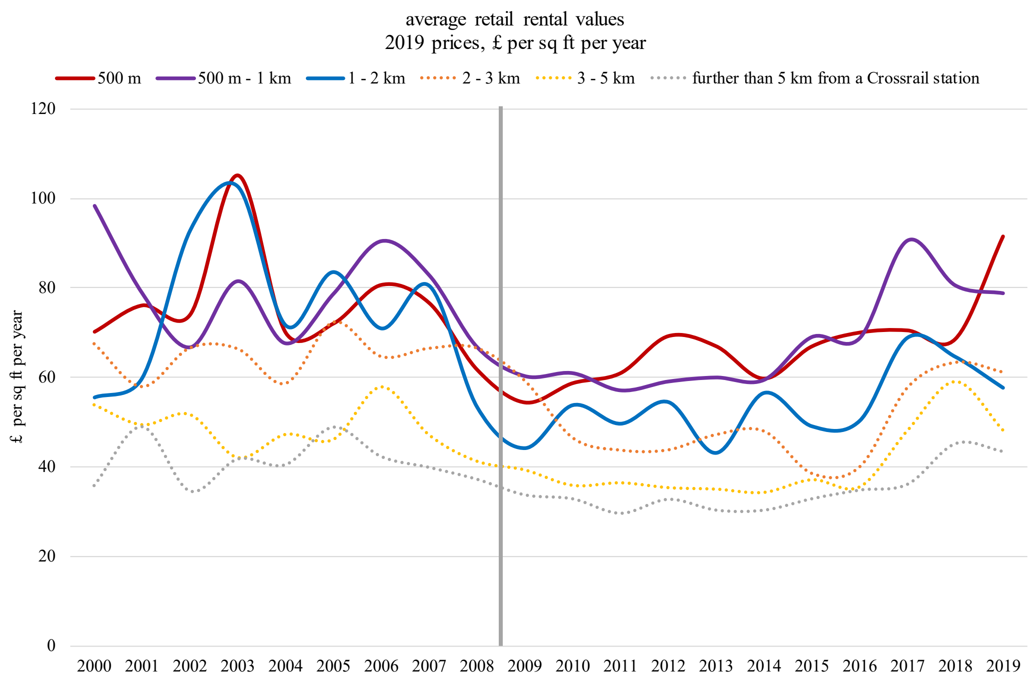


Figure 27. Average retail rents without outliers and other excluded for station distance bands, 2000 – 2019 index

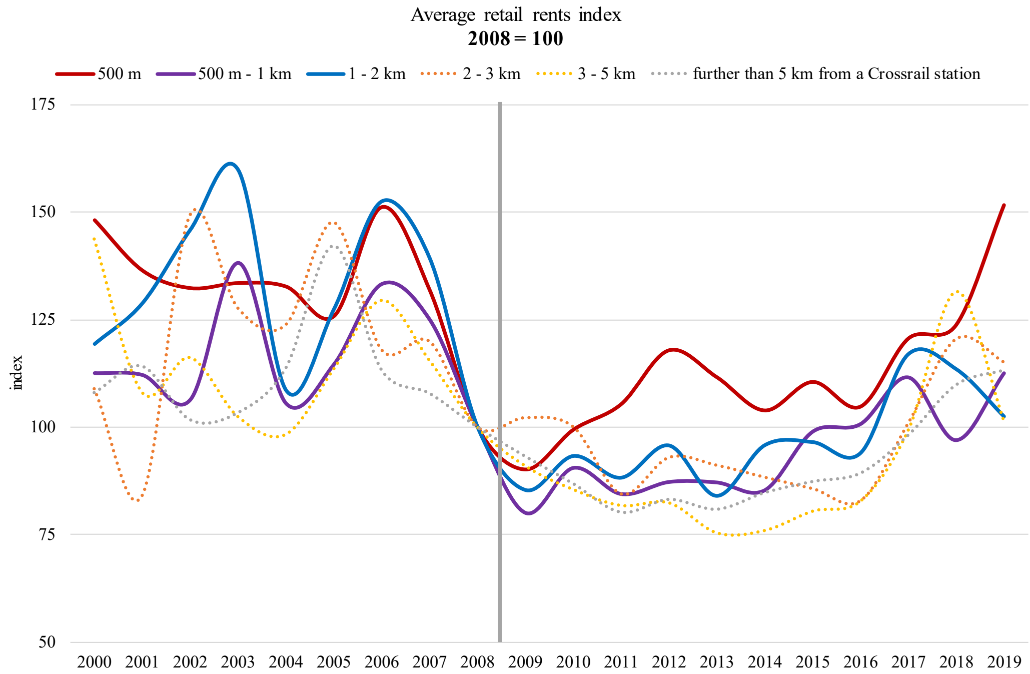
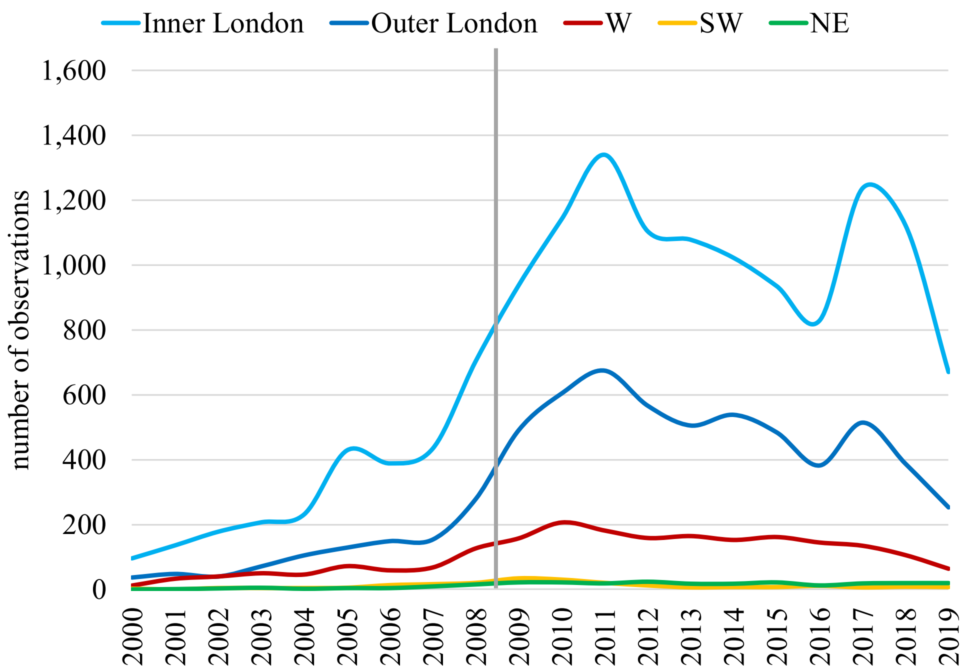


Figure 28. Number of retail transactions for areas of London



6 Industrial rental impacts

We were not able to test Crossrail pre-opening impacts on industrial values because the common trend assumption could not be met. The data up to 2019 shows that it might be possible for that assumption to be met for the ex-post analysis.

6.1 Introduction

We would expect industrial rents to be less likely to benefit from agglomeration than either office rents or retail rents. Evidence suggests that it is service sectors, particularly high-skilled services, which benefit most from agglomeration. For example, the effect of agglomeration is over three times as high for a sector such as finance and insurance as it is for manufacturing[24].

Nor is the effect necessarily consistent across all large metropolitan areas. Glaeser and Ressenger[25] (2010)) showed that for metropolitan areas in the United States, agglomeration impacts were far less pronounced for cities with a large proportion of low-skilled workers than for cities with a high proportion of high-skilled workers.

Accordingly, we would not expect to see significant pre-opening impacts on industrial rents.

6.2 Data source

The data source for industrial rental transactions is the CoStar database, as for the office and retail analysis.

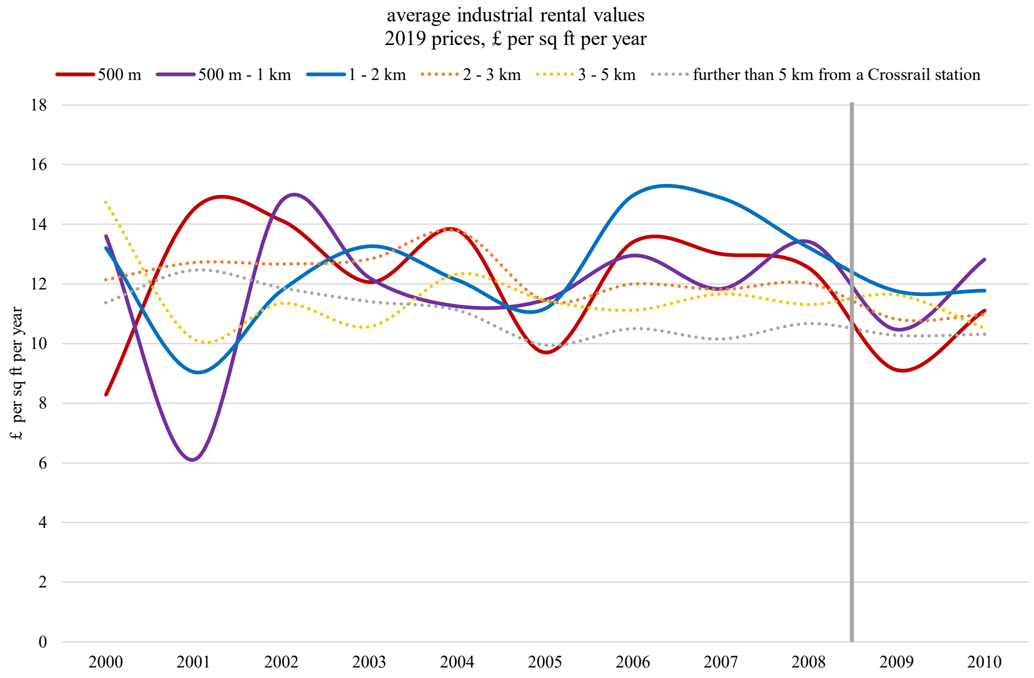
6.3 Pre-announcement trends

As we explain in chapter 2.6, to conduct a difference in differences analysis, we require the treatment and comparison groups to be as similar as possible before the announcement in 2008. It does not require that they have had identical values of the outcome variables before the announcement, only that those values followed similar trends.

Figure 29 shows that this assumption is not met for industrial rents. Rents in areas beyond 2 km from Crossrail programme stations are approximately stable across the whole period while rent values around the stations are much more volatile, especially before 2005.

[24]2014, Cheshire et al., "Urban Economics and Urban Policy: Challenging Conventional Policy"
[25] 2010, Glaeser and Resseger, "The complementarity between cities and skills", Journal of Regional Science

Figure 29. Average industrial rents for station distance bands, pre-announcement trends



6.4 Understanding the data for post-opening analysis

Figures 30 and 31 present average industrial values in different station distance bands between 2000 and 2019. It shows some common trends in the post-announcement period, notably in the increasing trend since 2018, suggesting, tentatively, that the common trends assumption may be met.

Figure 30. Average industrial rents for station distance bands, 2000 – 2019

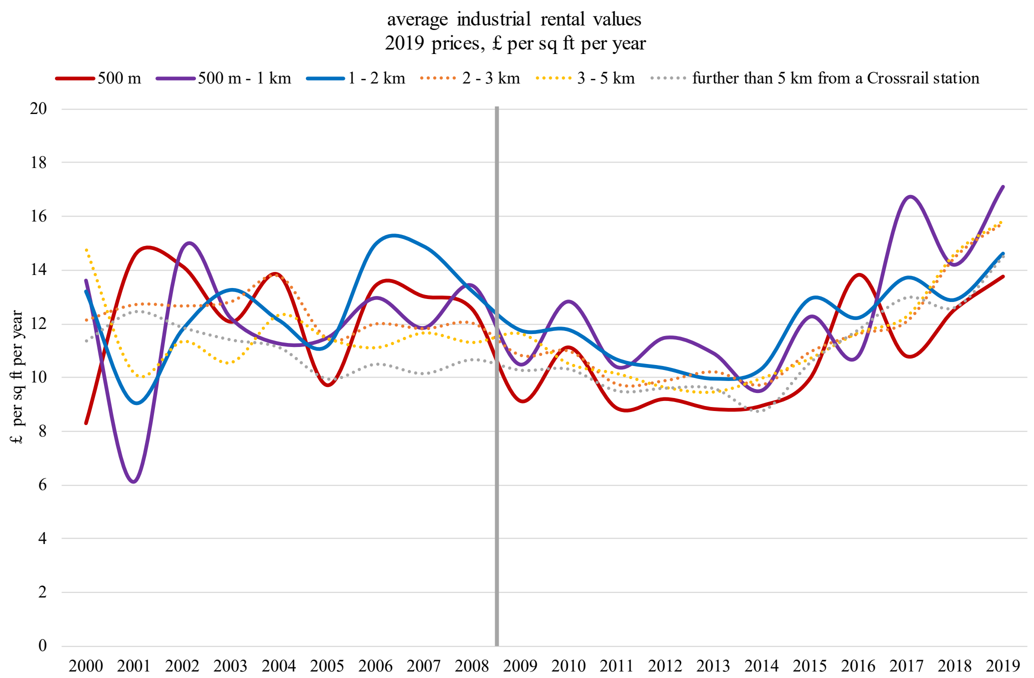
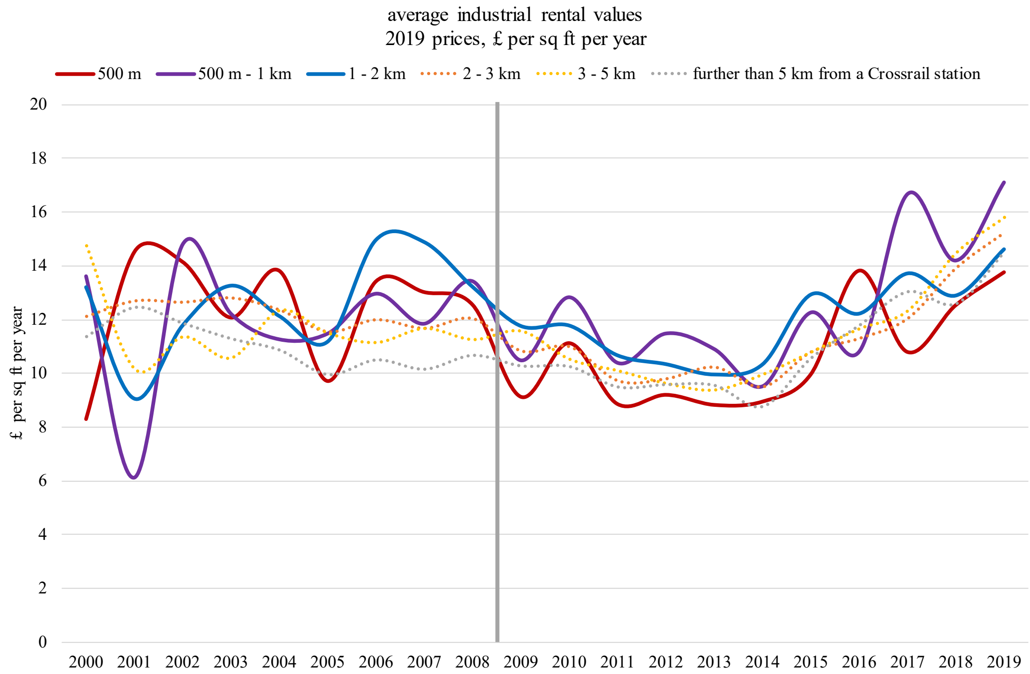


Figure 31. Average industrial rents for station distance bands without excluded, 2000 – 2019



7 Residential price impacts

We found that Crossrail had a positive impact on housing unit prices, but that this partly reflected some displacement effects in terms of demand for housing. Our analysis suggests that on the average line-wide level, prices in areas closest to the stations increased by around 2% due to Crossrail announcement effect, while properties in the areas from 1 km to 2 km from the stations experienced an equivalent slowdown in their value growth. This is observed particularly for locations in the eastern and south eastern section. We have not seen such displacement effects in central London.

7.1 Introduction

Residential properties located in well-connected, well-serviced, and accessible locations generally experience premiums in their values. The Elizabeth line is expected to improve the accessibility of many residential areas, particularly those outside central London. Buying a house is often a once-or-twice-in-a-lifetime investment, and so, more so than for other purchases, if house buyers expect an accessibility improvement, they may be prepared to make an investment years before the change comes about, in anticipation of a financial return.

Given housing affordability in London we might expect a pre-opening impact on residential property prices in some locations, particularly those outside London where the Elizabeth line will significantly decrease the commuting times into the central area. However, this is by no means certain, and housing unit prices have been increasing across the whole study area in recent years. In addition, the Crossrail programme announcement might have led to increased housing development, in which case additional supply could prevent values from increasing further. Therefore, we may not necessarily expect to find residential price impacts in our analysis.

Other studies have found that rail transport infrastructure projects have increased housing unit prices[26]. However, it is important to note that evidence is still scarce and usually considers only post-opening effects. In 2015, the WWCLEG's Transport Review[27] found seven studies exploring housing impacts, of which five 'found positive effects of proximity to stations, while two studies found no effect of proximity'.

The case for Crossrail was not predicated on housing unit price increases and although part of the funding derives from developer contributions, the programme has not been funded on the expectation of any land value uplift captured directly from residential households.

Future infrastructure schemes may attempt to make a case for capturing a proportion of residential housing unit price uplift to help fund the scheme. For that reason, an understanding of whether, when and how housing unit prices change because of the Crossrail announcement could be important in developing a robust funding case for such projects.

[26] What Works Centre for Local Economic Growth, 'Evidence Review: Transport', 2015
[27] Evidence Review: Transport', What Works Centre for Local Economic Growth, 2015.

7.2 Data sources

We used two datasets for an analysis of residential values: HM Land Registry Price Paid (LR)[28] and the Ministry of Housing, Communities and Local Government's Energy Performance of Buildings (EPB)[29].

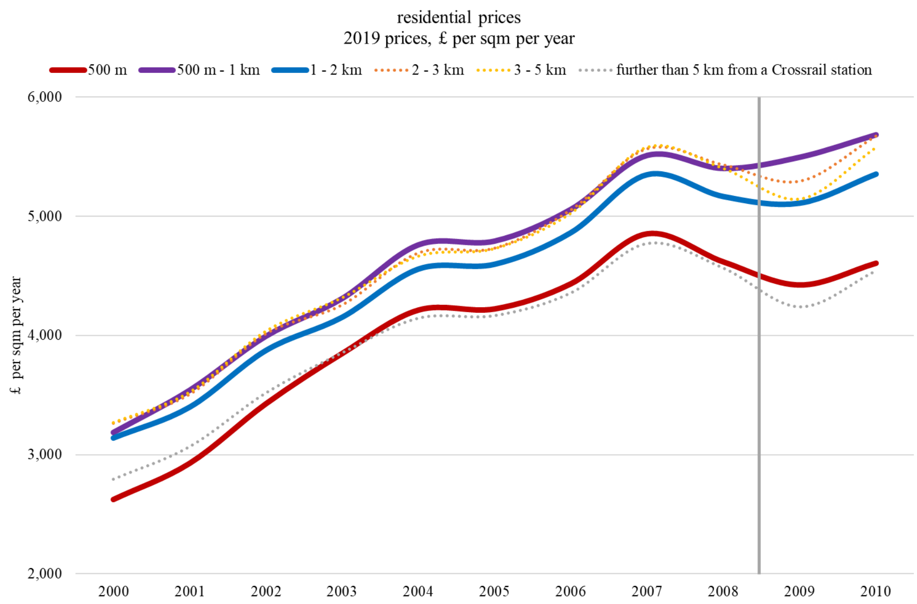
The LR data contains records of property sales in England and Wales submitted to the Land Registry for registration. The dataset includes information on sale price, transfer deed date, address, property type (detached, semi-detached, terraced, flats and other) and position on the market (new or established). However, it does not include data on the size of a subject property, which is one of a basic determinants of property prices.

The EPB dataset contains data drawn from Energy Performance Certificates (EPCs) issued for domestic and non-domestic buildings. All buildings constructed, sold or let since 2008 are required to have an EPC. In addition to data on energy efficiency parameters, EPCs contain property data, including total area. Combining Land Registry and EPB datasets by a property address allows us to obtain transaction prices per square metre which is a more appropriate outcome variable than just unit price.

7.3 Pre-announcement trends

Figure 32 shows that the common trends assumption is met for residential analysis.

Figure 32. Average housing unit prices for station distance bands, pre-announcement trends



28 https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads
29 https://epc.opendatacommunities.org

7.4 Understanding the data

Around 40% of our data comes from Outer London and 35% from Inner London. Around 20% in total are from West and South-West postcodes and 5% from wider North-East. Figure 33 shows numbers of transactions per high-level geographic area. It shows a clear dip in 2008-09 corresponding to the housing unit price crash and the beginning of the global recession. Demand returned to pre-crisis levels around 2014 and can be seen to have slowed substantially in recent years, especially in London.

Figure 34 shows the average number of transactions per local authority. A transaction is defined as a sale of a unique dwelling but that is not necessarily newly-built. Areas with the highest average numbers are Wandsworth, Bromley and Barnet. The fewest transactions took place in the City of London, South Buckinghamshire and Brentwood.

Figure 33. Number of residential transactions for areas of London

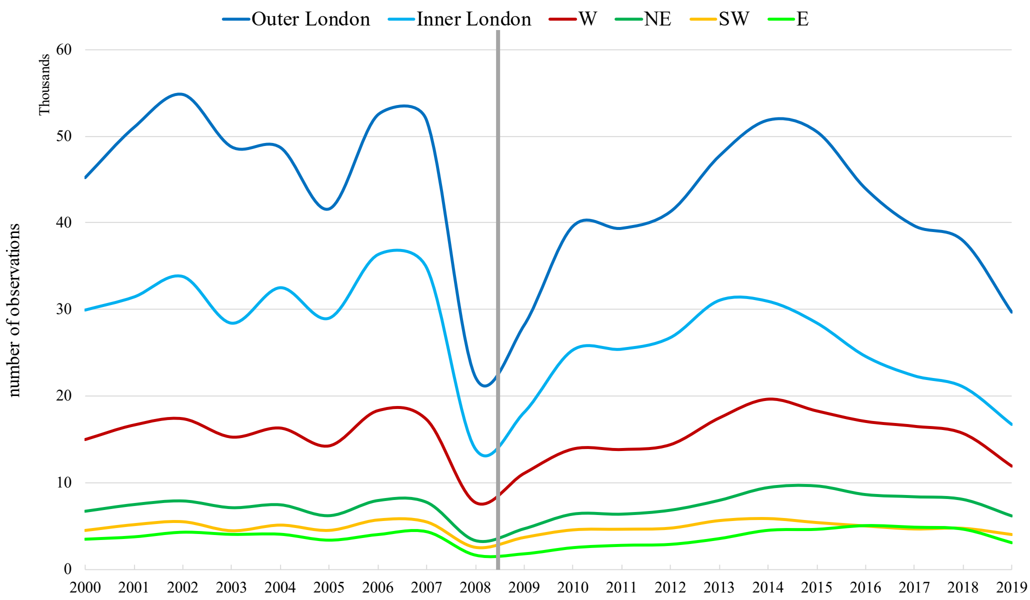


Figure 34. Residential transactions – annual average number per local authority

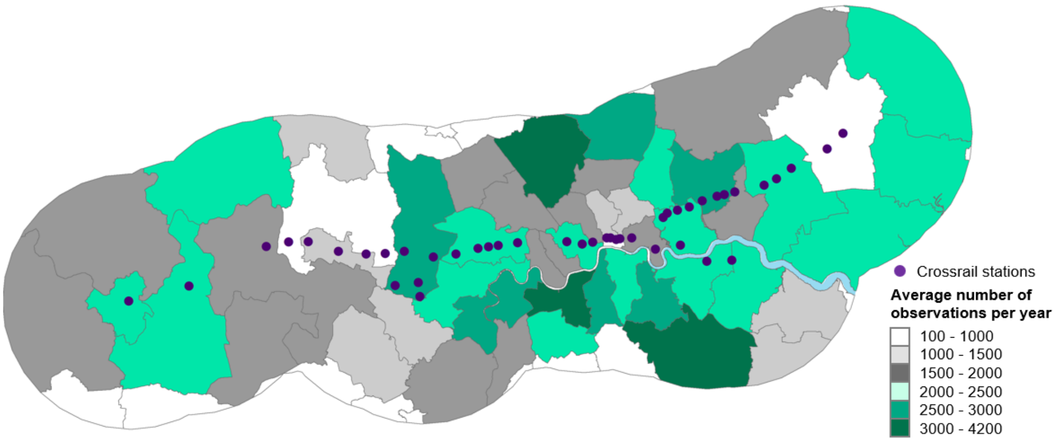


Figure 35 shows that the common trends assumption is also likely to be met for the post-opening analysis, because average prices in all areas moved in parallel up to 2019. Residential sale prices per square metre in the whole study area increased from an average of £4,900 in 2008 to £6,000 in 2019, implying an average growth of 21%. Growth rates have been broadly similar in all areas regardless of their distance to a Crossrail station, all ranging from 20% to 33%.

Figure 35. Residential sales prices for station distance bands, 2000 – 2019

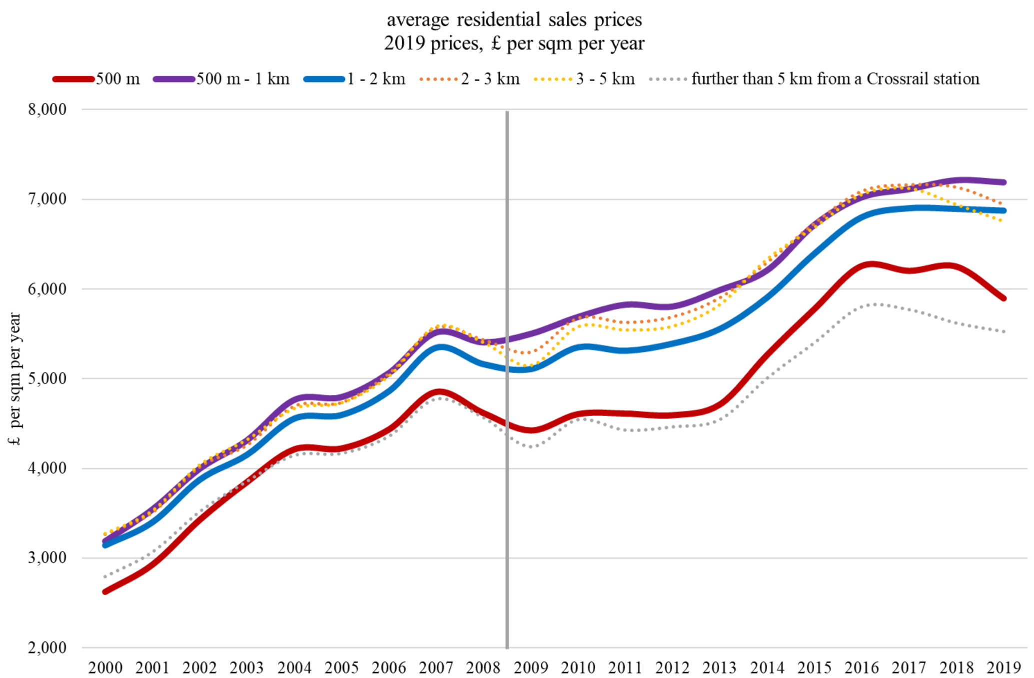
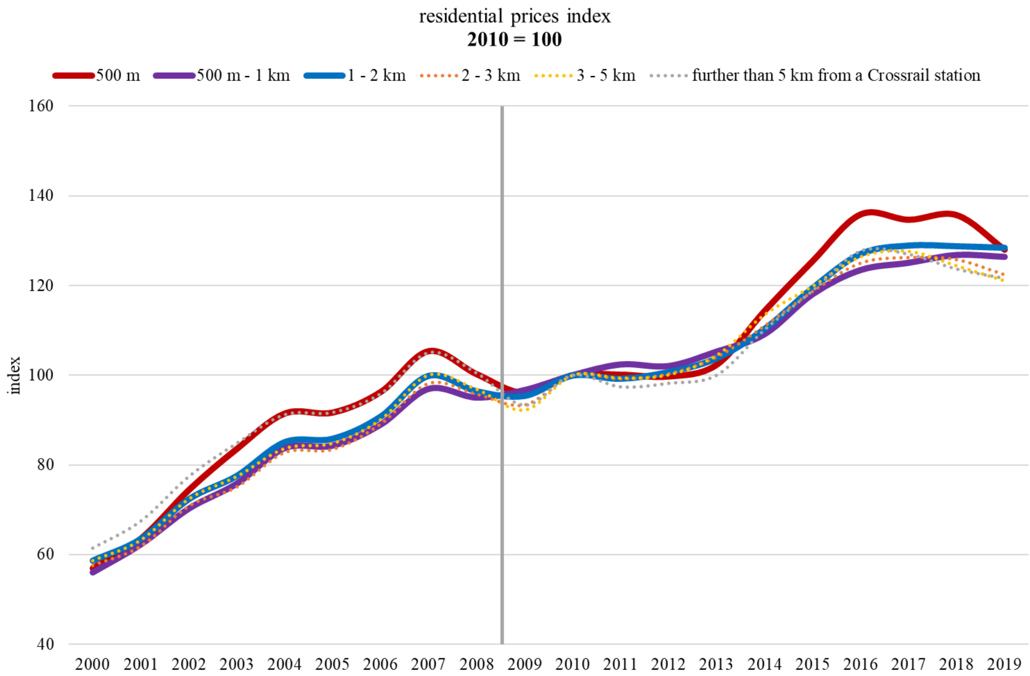


Figure 36. Residential sales prices for station distance bands, 2000 – 2019 index



In the analysis in Table 17, an average 4% of the original sample was excluded as outliers. There was a broadly similar proportion of outliers removed from all areas.

Table 17. Residential prices – average yearly breakdown

Wider area	**Initial sample:** Average count	**Initial sample:** Average share	**Sample after excluding outliers:** Average count	**Sample after excluding outliers:** Average share	**Outliers**	**Outliers as % of initial sample**
Outer London	43,377	42%	42,058	42%	1,685	4%
Inner London	27,047	26%	26,225	26%	1,027	4%
Outside London: W	15,351	15%	14,800	15%	672	4%
Outside London: NE	7,197	7%	7,000	7%	240	3%
Outside London: SW	4,798	5%	4,660	5%	178	4%
Outside London: E	3,658	4%	3,572	4%	101	3%
Outside London: S	1,268	1%	1,240	1%	36	3%
Outside London: NW	1,069	1%	1,037	1%	42	4%
Sum	**103,765**	**100%**	**100,591**	**100%**	**3,981**	**4%**

7.5 Line-wide analysis

Table 18 presents estimates from models 1 to 4 on a full dataset, without excluding any observations. As described in chapter 3.2, model 1 is a simple difference in differences analysis with postcode and year fixed effects, to which we then add controls across subsequent models.

The estimates of impacts for the full dataset stay consistent across all four models (Table 18). They show:

- a positive 1-2% impact within the 0 to 500m radius;
- zero impact within the 500m - 1 km band; and
- a significant negative impact of a similar scale, between -1% and -2% in the 1-2 km band (see figures highlighted in red).

However, when the strongest model (4) is estimated on a limited dataset (Table 19), the negative estimate is much smaller than in every other case and amounts to less than 0.1%. It should be noted that the models estimated on a limited dataset have higher R-Squared values. For example, the R-Squared values for model 4 are 0.84 on the limited dataset compared with 0.78 on the full dataset.

All four versions of the model show negative impacts on residential prices in the 1-2 km distance band. The nature of our methodology means that it is unable to assess the announcement’s impact on the overall level of prices. Instead, it compares different parts of the study area to each other. This means that it is unlikely that the negative impacts are caused by disbenefits from construction as they would be more likely for values closer to stations. It is more likely that those negative estimates suggest displacement impacts on the top of a study area-wide positive impact, especially given that majority of the data come from Outer London and outside London, where transport accessibility is lower and so being closer to a station matters more. The Crossrail announcement could have led to an increase in demand for dwellings close to the stations, and to a decrease in demand for dwellings further on.

Table 18. Regressions: Residential prices impacts for station distance bands, full dataset, models 1-4

	(1) Base model	(2) with controls	(3) with controls and employment shares	(4) with LAD-Year interactions
Post announcement	0.220***	0.225***	0.222***	0.179***
	(0.00181)	(0.00179)	(0.00182)	(0.0105)
Treated 500m post announcement	**0.0219*** **	**0.0111*** **	**0.0110*** **	**0.0207*** **
	(0.00244)	(0.00238)	(0.00238)	(0.00294)
Treated 500m to 1 km post announcement"	-0.00182	-0.00214	-0.00185	-0.00316
	(0.00428)	(0.00407)	(0.00407)	(0.00324)
Treated 1 to 2 km post announcement	**-0.0151*** **	**-0.0175*** **	**-0.0173*** **	**-0.0193*** **
	(0.00277)	(0.00271)	(0.00271)	(0.00238)
Constant	8.409***	8.377***	8.353***	8.458***
	(0.00142)	(0.00424)	(0.00458)	(0.00568)
Observations	2,073,687	1,966,972	1,966,151	1,966,151
R-squared	0.7660	0.7600	0.760	0.779
Adjusted R-squared	0.7474	0.7393	0.7394	0.7600
Year FE	Yes	Yes	Yes	Yes
Postcode FE	Yes	Yes	Yes	Yes
LAD FE		Yes	Yes	Yes
LAD-Year FE				Yes

Table 19. Regressions: Residential prices impacts for station distance bands, excluded records, models 1-4

	(1) Base model	(2) with controls	(3) with controls and employment shares	(4) with LAD-Year interactions
Post announcement	0.215***	0.220***	0.217***	0.210***
	(0.00241)	(0.00239)	(0.00241)	(0.0136)
Treated 500m post announcement	**-0.0220*** **	**-0.0299*** **	**-0.0293*** **	**0.0177*** **
	(0.00267)	(0.00261)	(0.00261)	(0.00300)
Treated 500m to 1 km post announcement"	**0.0105*** **	**0.0109*** **	**0.0112*** **	0.0000
	(0.00403)	(0.00386)	(0.00386)	(0.00294)
Treated 1 to 2 km post announcement	**-0.0697*** **	**-0.0684*** **	**-0.0678*** **	**-0.0117*** **
	(0.00376)	(0.00374)	(0.00373)	(0.00278)
Constant	8.463***	8.458***	8.429***	8.499***
	(0.00229)	(0.00722)	(0.00770)	(0.00873)
Observations	1,206,755	1,141,198	1,140,403	1,140,403
R-squared	0.813	0.807	0.807	0.828
Adjusted R-squared	0.797	0.789	0.789	0.812
Year FE	Yes	Yes	Yes	Yes
Postcode FE	Yes	Yes	Yes	Yes
LAD FE		Yes	Yes	Yes

	(1) Base model	(2) with controls	(3) with controls and employment shares	(4) with LAD-Year interactions
LAD-Year FE				Yes

7.6 Section analysis

In most sections of the route, the announcement of Crossrail seems to have impacted on the distribution of the demand for houses (Table 20). We see a 2-3% positive impact within 500m from Crossrail programme stations for all sections of the line. However, in most sections we see some negative impacts further out (in the 500m to 1 km and 1-2 km distance bands). As noted above, these patterns suggest displacement impacts.

Table 20. Regressions: Residential prices impacts, for station distance bands and line sections, excluded records, model 4

	Outer Western	Western London	Central	South Eastern	Eastern
Post announcement	0.179***	0.179***	0.179***	0.179***	0.179***
	(0.0105)	(0.0105)	(0.0106)	(0.0105)	(0.0105)
Treated 500m post announcement	**0.0312*** **	**0.0398*** **	**0.0255*** **	**0.0338*** **	**0.0187*** **
	(0.00358)	(0.00358)	(0.00353)	(0.00374)	(0.00343)
Treated 500m to 1 km post announcement”	**-0.0322*** **	-0.00216	**0.0181** **	0.0189	**-0.0146*** **
	(0.00699)	(0.00685)	(0.00767)	(0.0124)	(0.00531)
Treated 1 to 2 km post announcement	0.00355	-0.00657	-0.00577	**-0.0669*** **	**-0.0231*** **
	(0.00477)	(0.00601)	(0.00589)	(0.00657)	(0.00414)
Constant	8.445***	8.450***	8.470***	8.448***	8.439***
	(0.00464)	(0.00462)	(0.00564)	(0.00473)	(0.00460)
Observations	1,705,470	1,682,574	1,720,638	1,670,878	1,732,440
R-squared	0.767	0.768	0.783	0.770	0.767
Adjusted R-squared	0.747	0.747	0.764	0.750	0.747
Year FE	Yes	Yes	Yes	Yes	Yes
Postcode FE	Yes	Yes	Yes	Yes	Yes
LAD FE	Yes	Yes	Yes	Yes	Yes
LAD-Year FE	Yes	Yes	Yes	Yes	Yes

The patterns of impacts on residential prices according to location categories (Table 22) show little variation although it may be notable that the largest positive impact on within 500m of stations is seen in areas with major associated development. The largest negative impact is seen in the 1-2 km distance band for major regeneration areas.

Table 21. Regressions: Residential prices impacts, for station distance bands and location categories, excluded records, model 4

	Deprived	**With major associated development**	**Major regeneration area**	**Expected additional labour supply**
Post announcement	0.179***	0.179***	0.179***	0.179***
	(0.0105)	(0.0105)	(0.0105)	(0.0105)
Treated 500m post announcement	**0.0193*** **	**0.0266*** **	**0.0223*** **	**0.0210*** **
	(0.00339)	(0.00339)	(0.00337)	(0.00346)
Treated 500m to 1 km post announcement”	-0.00969*	-3.69e-06	0.00560	0.00578
	(0.00497)	(0.00487)	(0.00457)	(0.00587)
Treated 1 to 2 km post announcement	**-0.0121*** **	**-0.0158*** **	**-0.0339*** **	**-0.0227*** **
	(0.00404)	(0.00351)	(0.00345)	(0.00403)
Constant	8.447***	8.466***	8.457***	8.465***
	(0.00459)	(0.00574)	(0.00580)	(0.00568)
Observations	1,754,761	1,823,883	1,825,943	1,772,698
R-squared	0.766	0.782	0.779	0.782
Adjusted R-squared	0.746	0.763	0.759	0.763
Year FE	Yes	Yes	Yes	Yes
Postcode FE	Yes	Yes	Yes	Yes
LAD FE	Yes	Yes	Yes	Yes
LAD-Year FE	Yes	Yes	Yes	Yes

Table 22 shows number of treated observations in each distance band for each section analysis. The residential dataset is the largest we have at our disposal for this project.

Table 22. Treatment data counts: Residential, section-level analysis

Section	500m	500-1000m	1-2 km	0-2 km
1 - Western outside	125k	15k	47k	187k
2 - Western London	131k	6k	28k	165k
3 - Central	130k	20k	51k	201k
4 - South Eastern	125k	5k	25k	155k
5 - Eastern	164k	4k	53k	221k
6 - Deprived	171k	8k	68k	247k
7 - Areas with associated major development	131k	24k	76k	231k
8 - Areas with wider regeneration or growth initiative	158k	32k	120k	310k
9 - Areas expected to benefit from additional labour supply	132k	23k	74k	229k

Table 23. Impacts summary: Residential prices, section-level analysis

Section	500m	500-1000m	1-2 km
1 - Western outside	3%***	-3%***	0
2 - Western London	4%***	0	0
3 - Central	2.5%***	2%**	0
4 - Southeastern	3%***	0	-6.5%***
5 - Eastern	2%***	-2%***	-2%***
6 - Deprived	2%***	-1%*	-1%***
7 - Areas with associated major development	2.5%***	0	-1.5%***
8 - Areas with wider regeneration or growth initiative	2%***	0	-3%***
9 - Areas expected to benefit from additional labour supply	2%***	0	-2%***

7.7 Station level analysis

The main points to note from the station level analysis (Table 24) are:

- For all stations, we found 2-3% impact within 500m from a station. The consistency of this result is noteworthy;
- We found negative impact estimates for stations outside London and in the South East. The mix of positive and negative impacts in non-central areas suggests displacement effects; and
- We found no negative impacts for stations in central London.

Our station analysis includes six out of seven stations in Crossrail's central section. For all of them we found positive impacts closest to the stations, with more variety in the nature of impacts further afield.

- From Paddington to Tottenham Court Road we found significant positive impacts in most distance bands;
- In Farringdon and Whitechapel we found no impacts further than 500m away; and
- In Canary Wharf there seem to have been significant demand displacement effects in areas furthest from the station.

Please note that in the case of Bond Street and Tottenham Court Road stations the 1000-2000m buffer areas are overlapping. This means that some of the impact may be double-counted so the results for these two stations should not be added up.

Table 24. Impacts summary: Residential prices, station-level analysis by distance bands, model 4

Station	**0-500m:** All	**0-500m:** Excl.	**500m - 1km:** All	**500m - 1km:** Excl.	**1km - 2km:** All	**1km - 2km:** Excl.
Reading	**3%**	**2%**	**-4%**	**-2.5%**	0	**-3%**
Slough	**3%**	**2%**	**-7%**	**-6.5%**	0	0
Southall	**3%**	**3%**	0	0	**-6%**	**-5%**
Ealing Broadway	**3%**	**2.5%**	**-2%**	**-2%**	**7%**	**7%**
Paddington	**3%**	**2%**	**4%**	**3%**	0	**3%**
Bond Street	**3%**	**2%**	0	0	**10%**	**8.5%**
TCR	**3%**	**2.5%**	**4.5%**	-	**4%**	**5%**
Farringdon	**3%**	**2.5%**	0	0	0	0
Whitechapel	**3%**	**2%**	0	0	0	0
Canary Wharf	**3%**	**2%**	0	**3%**	**-10%**	**-11%**
Custom House	**3%**	**2.6%**	**4%**	**2.3%**	**-2%**	**-1.5%**
Abbey Wood	**3%**	**2.4%**	**3%**	**4.3%**	**-6%**	**-5%**
Romford		**2.3%**		**-2.7%**		**-2.7%**

7.8 Summary

We found consistent evidence across all spatial levels of analysis that there has been a 2-3% increase in housing unit prices around stations compared to what would have been expected without the announcement. Overall, for the period 2008-2019 prices around stations have risen by 50% in buffers 500m around stations (from £363,111 to £545,065 in 2019 real prices) and by 42% in buffers 500m – 1 km around stations (from £371,473 to £527,262 in 2019 real prices). This suggests that factors other than the Crossrail announcement made a larger contribution to the overall increase in residential house prices during the baseline period. The current analysis is insufficient in explaining in detail what may be behind this result, although displacement may have been a factor.

There seems to be a 2-2.5% impact on sale prices closest to central London stations and a slightly higher impact of 3%-8% further between 1 km and 2 km. The 'further' impacts were not detected around Farringdon and Whitechapel. Whitechapel is indeed a very different place from the West End, even if it is officially included as a central London station. In many ways it is more similar to areas such as Custom House and Abbey Wood on the south-east section of the line where we have also seen negative impacts in areas furthest from the stations. When looking at areas with significant levels of deprivation or places designated as major regeneration areas, they all seem to have experienced displacement effects..

Residential analysis shows that, similar to the analysis of commercial values, it is the urban centre that experiences increase in property values due to this transport investment.

There have been developments in close proximity to stations, such as the Berkley Homes development in Woolwich, where some 1,200 homes were completed and occupied above the new Woolwich station again in advance of the station becoming operational'[30]. The St Giles Circus project involves a £150m redevelopment of a central London site adjacent to Tottenham Court Road Station. The development includes leisure, retail, commercial, residential accommodation and a boutique hotel. In 2015 Centre Point, next to Tottenham Court Road, was converted from office space into luxury flats.

Another reason for higher demand for properties around stations might be the improvements in public realm which were commissioned for all 27 London stations and four stations outside London.

[30] Source: "Crossrail OSD collaboration and property value capture" by Ian Lindsay, published in 2018

Table 25. Summary of residential prices impacts

Distance band	Line-wide	Geographic sections	Deprived areas	Areas of with major associated development	Areas of major regeneration or growth initiative	Areas expecting increase in labour supply
	1% ***	All: 2-3.5% ***	2% ***	2.6% ***	2% ***	2% ***
	***impact not found***	Central: 2% ** Outer Western -3% *** Eastern -1.5% ***	***impact not found***	***impact not found***	***impact not found***	***impact not found***
	Very, very small negative impact ***	Southern Eastern -6% *** Eastern -2% ***	- 1% ***	-1.5% ***	-3.4% ***	-2% ***

8 Planning impacts

We tested Crossrail announcement impacts on the volume of residential development planning activity, but we have not found evidence of any impacts.

8.1 Introduction

In terms of pre-opening impacts, it is useful to understand whether Crossrail is having an impact on the volume of new development. Considering the number of applications overall is a relatively uninformative metric given that an application can be made for anything from one to many hundreds of housing units (and any amount of commercial floorspace). Even within the categories of minor and major, the number of units delivered can vary widely.

For this report, we used planning application data to calculate the number of new planning projects in each LSOA in each year. We estimate whether the announcement of Crossrail had any impact on the number of new development projects, regardless of their size or the number of related planning applications; therefore, we assume that the announcement had no impact on the size of average planning applications. We perform the analysis separately for office and housing developments.

8.2 Data source

The data source for planning application data (residential and commercial) is Glenigan, a proprietary commercial data provider that collects publicly available data on planning applications and places them into a usable database of all planning applications across the UK.

Each record in the Glenigan database is a planning application entry. However, multiple planning applications can relate to the same development, as different types of applications are submitted during the planning process – e.g. pre-planning submissions, outline applications, or detailed planning applications. In addition, both rejected and approved applications are reported in the database. Finally, planning applications relating to the same project are sometimes submitted separately to two or more planning authorities (where for example a project straddles local authority boundaries). All of these applications are reported separately in the database.

Each application has a type, whether it is a new development, refurbishment, extension, or a mix of the above. They are also assigned a property sector: housing, offices, industrial, education and others (Table 26). For the purpose of this analysis we looked at applications in the following three categories:

- Residential projects – applications categorised as Private Housing or Social Housing;
- Office projects – applications categorised as Offices/Commercial; and
- Retail projects – applications categorised as Retail.

Table 26. Count of planning applications by sector (2005 – 2019)

Application type	Count	Share
Civil (infrastructure)	6,971	3%
Civil (Utilities)	4,176	2%
Community and amenities	5,872	3%
Education	17,955	8%
Hotel and leisure	31,842	14%
Industrial	13,146	6%
Medical and scientific	7,858	3%
Offices/Commercial	28,765	12%
Private Housing	81,284	35%
Retail	31,335	13%
Social Housing	5,379	2%
Total	**234,551**	**100%**

In this iteration of the analysis it is the number of projects that we were interested in, not the number of planning applications. We counted each development project once in a year when the first related planning application was submitted. For simplicity, we only considered planning applications categorised as 'new' (Table 26) for new developments. We also used applications categorised as 'detail planning', which account for around 95% of office and retail applications for new development and 85% of housing applications, which means we were able to analyse the Crossrail impacts on a volume of new projects, and aimed not to include applications for small adjustments (Table 27).

Table 27. Count of planning applications by category (2005 – 2010)

Application type	Count	Share
Extension	22,102	9%
Extension/Refurb	24,504	10%
New	69,183	30%
New/Extension	2,055	1%
New/Refurb	4,170	2%
New/Refurbishment	733	>1%
New/Refurbishment/Extension	10	0
Refurb	61,979	26%
Refurbishment	44,577	19%
Refurbishment/Extension	5,238	2%
Total	**234,551**	**100%**

In the analysis of planning impacts, our units of analysis are LSOAs, for which we calculated numbers of new development planning projects in each year. For this reason, we did not have any application-level controls in this analysis, only the LSOA-level ones which were used in all previous analyses, namely:

- Population density (from model 2);
- Employment density (from model 2); and
- Employment shares (from model 3).

In addition, in the planning analysis we did not exclude observations as outliers. We excluded outliers in rental analyses as some transactions might achieve very specific rental values, while here in the planning analysis we looked at total numbers of new applications.

In this analysis we concentrated on detailed planning applications and omitting other categories shown in Table 28.

Table 28. Count of planning application type (2005 – 2019)

Application type	**Count**	**Share**
Approval of reserved matters	1900	>1%
Circular 18/84	11	>1%
Detailed planning	216,545	92%
Listed building consent	9,571	4%
Outline planning	5,191	2%
Planning not required	274	>1%
Pre-planning	1,083	>1%
Public Inquiry	9	>1%
Total	**234,551**	**100%**

8.3 Pre-announcement trends

The common trends assumption is met for an analysis of new development projects categorised as housing (Figure 37) and retail (Figure 39), although the retail dataset is too small to allow for a robust analysis.

For office data, the only treatment buffer behaving in parallel with the counterfactual group is the 1-2 km band (Figure 38), which means this is the only treatment area we could potentially use in further analysis.

Figure 37. Number of new housing planning projects regarding new development by distance bands

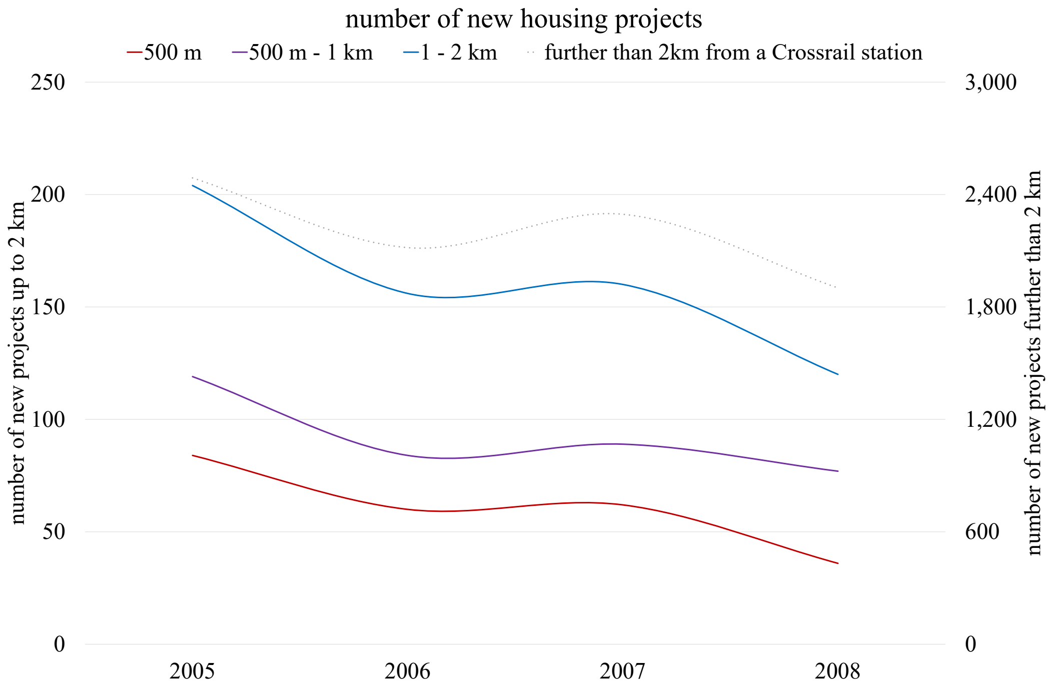


Figure 38. Number of all new office planning projects regarding new development by distance bands

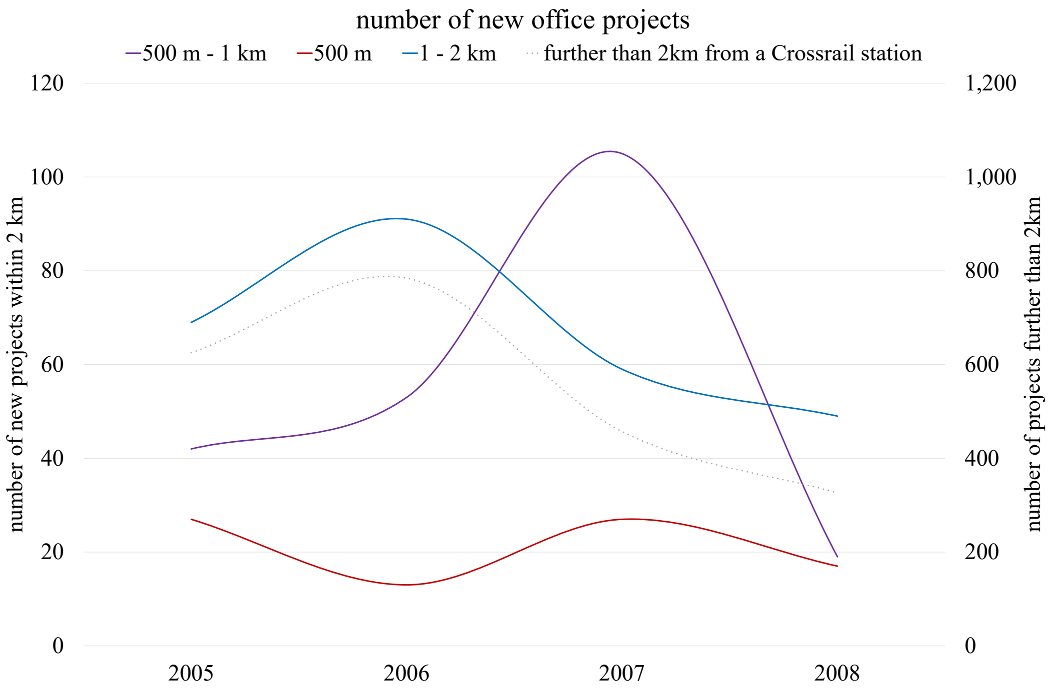
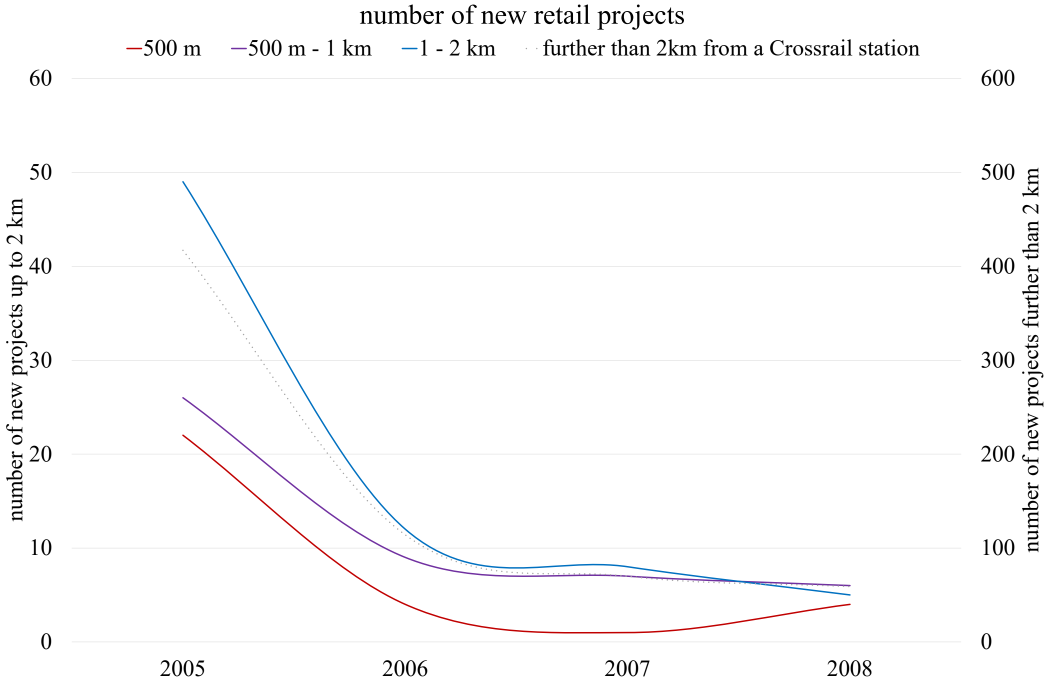


Figure 39. Number of all new retail planning projects regarding new development by distance bands



8.4 Understanding the data

Figure 40 shows total number of new housing planning projects across the region. In Inner and Outer London there were around 2,000 new projects started (i.e. with planning applications submitted) each year up to 2014, when the annual number increased significantly. In all major areas we see a significant drop around the time of the recession in 2008 and 2009, as expected. Figure 41 shows the same information but includes only projects categorised as new developments (i.e. excluding extensions, refurbishments etc.). The numbers are halved compared to the previous figure, but the dynamics of change are very similar across time.

Figure 40. Number of all new residential planning projects for areas of London

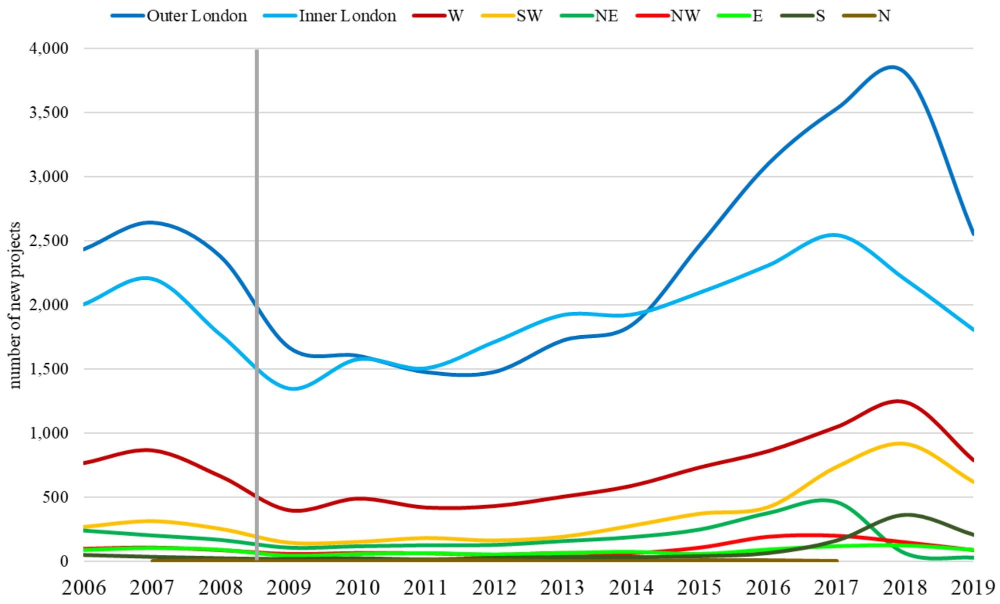


Figure 41. Number of all new residential planning projects regarding new development for areas of London

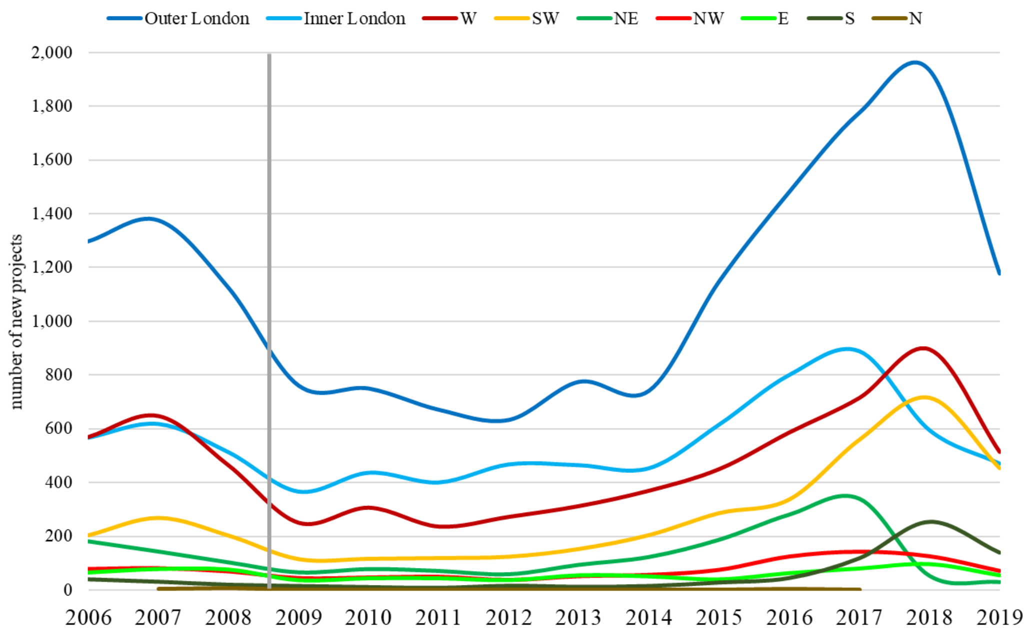


Figure 43shows the total number of office planning projects and Figure 44 shows only completely new developments. As expected, the recent numbers in Inner London are higher than in Outer London. There are significant decreases around the time of the recession in 2008 and 2009, following a similar pattern to housing projects.

Figure 42. Number of all new office planning projects for areas of London

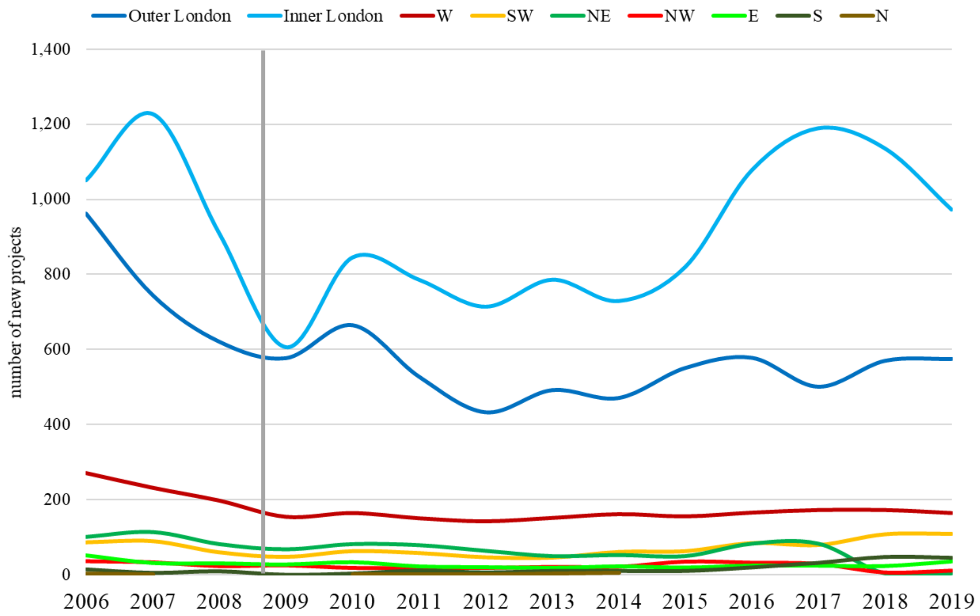


Figure 43. Number of all new office planning projects regarding new development for areas of London

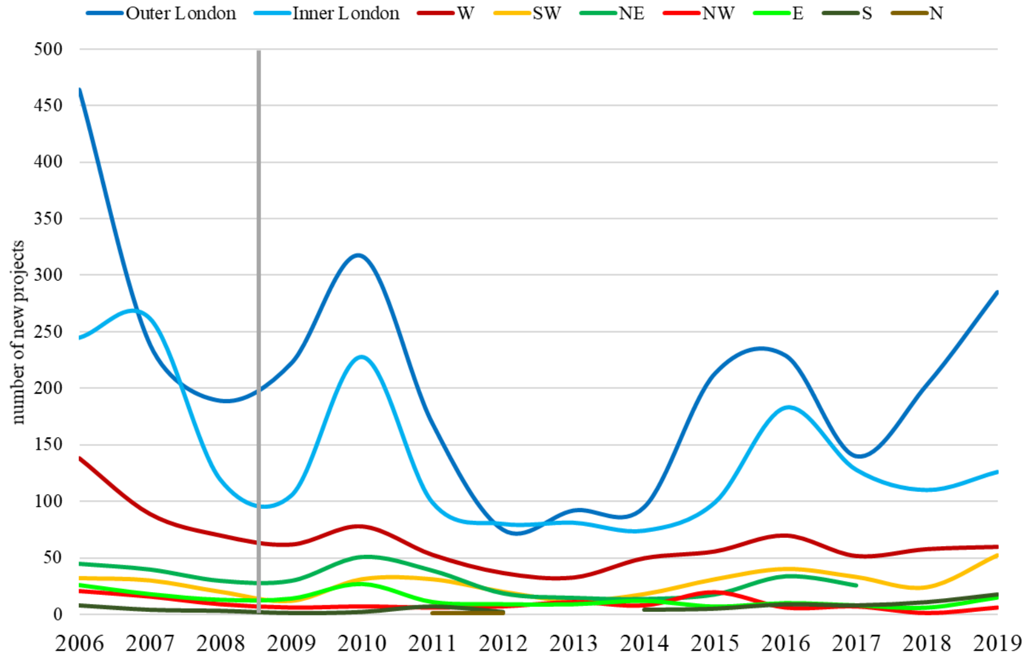


Figure 44 and Figure 45 show the numbers of new retail projects. As previously mentioned, the numbers of new development retail projects (Figure 45) are insufficient to allow us to obtain statistically robust estimates.

Figure 44. Number of all new retail planning projects for areas of London

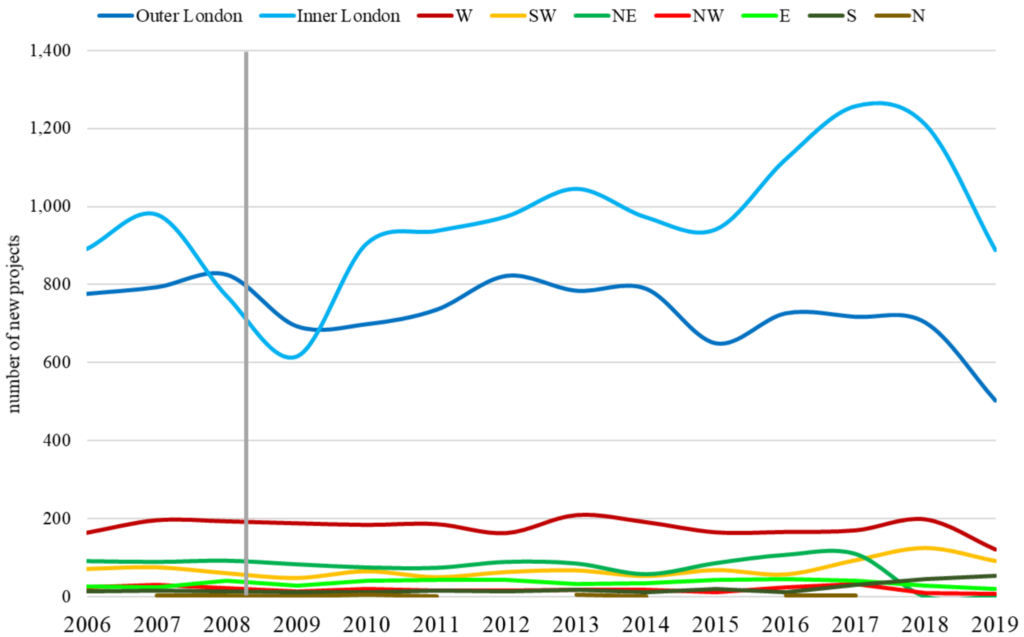
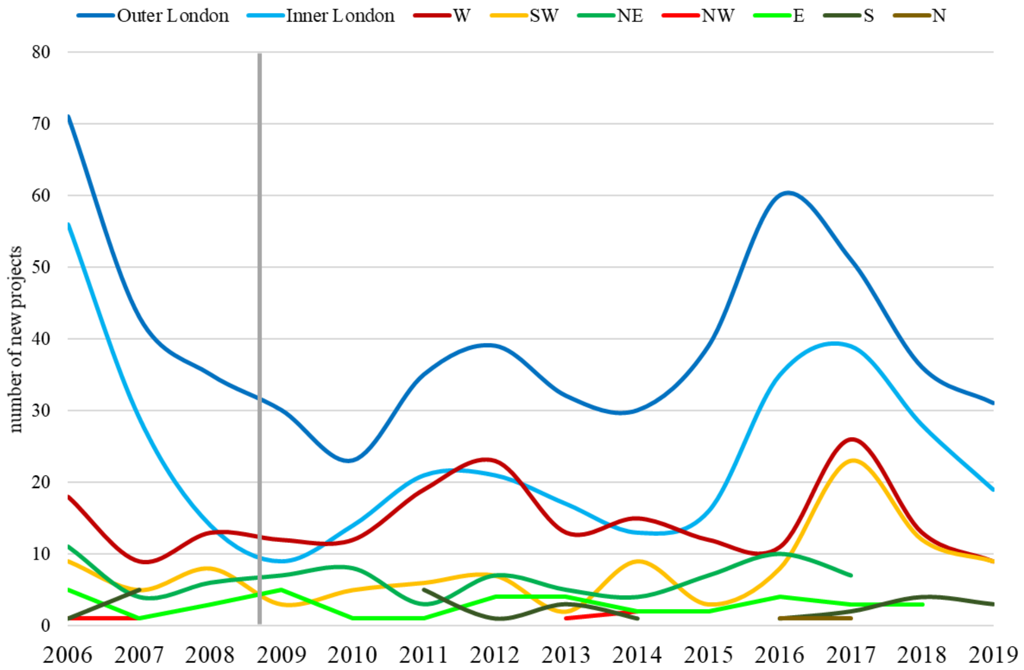


Figure 45. Number of all new retail planning projects regarding new development for areas of London



8.5 Housing projects

8.5.1 Line-wide

Using a full dataset (without excluding records from the counterfactual group), all models (1-4) suggest some negative housing impacts within the 500m buffer from Crossrail stations (Table 29) although this is only statistically significant for model 4. When we use a limited dataset, none of these estimates are statistically significant. Nevertheless, in both cases we observe low values of the R-Squared coefficient, which suggests that the model is not performing well in explaining the relationships in the data.

Table 29. Regressions: Planning impacts, housing, by distance bands, full dataset

	(1)	**(2)**	**(3)**	**(4)**
	Base model	with controls	with controls and employment shares	with LAD-Year interactions
Post announcement	0.0197	0.0492	0.0563	3.843***
	(0.0443)	(0.0451)	(0.0459)	(1.119)
Treated 500m post announcement	-0.135	-0.0975	-0.0933	**-0.232***
	(0.117)	(0.118)	(0.118)	(0.127)
Treated 500m to 1 km post announcement”	-0.0171	0.0256	0.0268	-0.156
	(0.0974)	(0.0981)	(0.0981)	(0.111)

	(1)	(2)	(3)	(4)
	Base model	with controls	with controls and employment shares	with LAD-Year interactions
Treated 1 to 2 km post announcement	-0.0293	-0.0136	-0.0154	-0.0906
	(0.0731)	(0.0732)	(0.0732)	(0.0840)
Constant	1.559***	1.793***	1.739***	1.574***
	(0.0333)	(0.0670)	(0.0880)	(0.0933)
Observations	21,980	21,899	21,896	21,896
R-squared	0.375	0.375	0.376	0.449
Adjusted R-squared	0.162	0.163	0.163	0.221
Year FE	Yes	Yes	Yes	Yes
LSOA FE	Yes	Yes	Yes	Yes
LAD FE		Yes	Yes	Yes
LAD-Year FE				Yes

Table 30. Regressions: Planning impacts, housing, by distance bands, data excluded

	(1) Base model	(2) with controls	(3) with controls and employment shares	(4) with LAD-Year interactions
Post announcement	0.0118	0.0445	0.0528	3.838***
	(0.0452)	(0.0460)	(0.0468)	(1.120)
Treated 500m post announcement	-0.132	-0.0791	-0.0742	-0.205
	(0.132)	(0.133)	(0.133)	(0.143)
Treated 500m to 1 km post announcement”	0.0526	0.118	0.115	-0.0203
	(0.111)	(0.112)	(0.112)	(0.130)
Treated 1 to 2 km post announcement	0.0481	0.0811	0.0767	0.0378
	(0.0912)	(0.0915)	(0.0915)	(0.105)
Constant	1.560***	1.821***	1.769***	1.581***
	(0.0336)	(0.0685)	(0.0903)	(0.0953)
Observations	21,045	20,964	20,961	20,961
R-squared	0.378	0.378	0.378	0.455
Adjusted R-squared	0.164	0.166	0.166	0.227
Year FE	Yes	Yes	Yes	Yes
LSOA FE	Yes	Yes	Yes	Yes
LAD FE		Yes	Yes	Yes
LAD-Year FE				Yes

8.5.2 Section analysis

Table 31 and Table 32 present results for the analysis of line section impacts. Again, the R-Squared indicators are low, which suggests low model quality. There is a single positive statistically significant estimate of impact for the outer western section within the 1-2km buffer Table 33. However, it is not repeated and therefore validated in the analysis of areas with wider regeneration or major associated development (Table 32).

Table 31 Regressions: Planning impacts, housing, by distance bands and line sections, limited dataset, model 4

	Outer Western	**Western London**	**Central**	**South Eastern**	**Eastern**
Post announcement	3.841***	3.841***	3.840***	3.840***	3.840***
	(1.149)	(1.147)	(1.143)	(1.147)	(1.141)
Treated 500m post announcement	-0.203	0.109	-0.413	0.00300	-0.212
	(0.389)	(0.432)	(0.277)	(0.478)	(0.213)
Treated 500m to 1 km post announcement”	0.185	-0.136	0.0765	0.0291	-0.0851
	(0.346)	(0.382)	(0.260)	(0.354)	(0.217)
Treated 1 to 2 km post announcement	0.339*	-0.0358	0.00823	0.0693	0.0901
	(0.199)	(0.213)	(0.181)	(0.205)	(0.164)
Constant	1.690***	1.715***	1.689***	1.636***	1.673***
	(0.112)	(0.112)	(0.113)	(0.101)	(0.109)

	Outer Western	Western London	Central	South Eastern	Eastern
Observations	19,088	19,109	19,396	19,117	19,570
R-squared	0.456	0.457	0.457	0.457	0.456
Adjusted R-squared	0.230	0.230	0.228	0.230	0.230
Year FE	Yes	Yes	Yes	Yes	Yes
Postcode FE	Yes	Yes	Yes	Yes	Yes
LAD FE	Yes	Yes	Yes	Yes	Yes
LAD-Year FE	Yes	Yes	Yes	Yes	Yes

Table 32. Regressions: Planning impacts, housing, by distance bands and location category, limited dataset, model 4

	Deprived	With major associated development	Major regeneration area	Expected additional labour supply
Post announcement	3.841***	3.839***	3.839***	3.838***
	(1.138)	(1.138)	(1.129)	(1.139)
Treated 500m post announcement	-0.256	-0.315	-0.116	-0.279
	(0.201)	(0.239)	(0.168)	(0.253)
Treated 500m to 1 km post announcement”	-0.0271	0.0409	0.0353	0.180
	(0.190)	(0.209)	(0.162)	(0.221)
Treated 1 to 2 km post announcement	0.123	-0.00783	0.111	0.0733
	(0.156)	(0.154)	(0.124)	(0.153)
Constant	1.668***	1.613***	1.580***	1.625***
	(0.110)	(0.0999)	(0.0973)	(0.101)
Observations	19,730	19,692	20,261	19,610
R-squared	0.456	0.457	0.456	0.457
Adjusted R-squared	0.229	0.228	0.228	0.228
Year FE	Yes	Yes	Yes	Yes
Postcode FE	Yes	Yes	Yes	Yes
LAD FE	Yes	Yes	Yes	Yes
LAD-Year FE	Yes	Yes	Yes	Yes

8.5.3 Station analysis

We did not find any statistically significant housing planning impacts at the station level of analysis, which is in line with the previous line-wide and section-level analysis. Similarly, the R-Squared indicators are consistently low in station-level models.

8.6 Summary

We have not found any consistent and statistically significant impacts from the Crossrail announcement on the number of new housing developments, as measured by planning applications.

We did not run the analysis for office and retail planning applications because of the unmet common trends assumption in the first case and the small amount of data available in the second case.

9 Summary of findings and conclusions

9.1 Findings

Our research shows that the announcement and further construction of Crossrail has had an impact on property values in the pre-opening period and that this was most visible in the residential and office market.

In our baseline analysis we found that areas immediately around future Elizabeth line stations (up to 500 metres from the stations) experienced relatively the highest population growth over the baseline period between 2000 and 2008 (compared to other parts of London) and were already characterised by faster growing employment density before the Crossrail announcement. This shows that the Elizabeth line will provide transport capacity and connectivity to support areas with growing transport demand.

An important contextual factor around the impact of the Crossrail's announcement on all property impacts would be land availability around stations. Supply side limits can constrain growth even if the infrastructure investment creates the demand for new development. However, our analysis was unable to find suitable data for this analysis.

Impact on residential properties

Our analysis on the impact of the announcement of new stations found significant, mostly positive results for residential housing unit prices, but found no significant impact on development around Crossrail stations.

We found that Crossrail had impact on housing unit prices, and that this may reflect some displacement effects. Our analysis suggests that on the average line-wide level, prices in areas closest to the stations increased by around 2% due to the Crossrail announcement, while properties in the areas from 1 km to 2 km from the stations have experienced an equivalent slowdown in their value growth.

Impacts on residential housing unit prices in the following areas:	500m to nearest Crossrail station	500-1000m to nearest Crossrail station	1-2 km to nearest Crossrail station
1 - Average impact on the whole route	2%	0	-2%
2 - Western section outside London	3%	-3%	0
3 - Western London section	4%	0	0
4 - Central section	2.5%	2%	0
5 - Southeastern section	3%	0	-6.5%
6 - Eastern section	2%	-2%	-2%
7 - Deprived areas	2%	-1%	-1%
8 - Areas with associated major development	2.5%	0	-1.5%

Impacts on residential housing unit prices in the following areas:	500m to nearest Crossrail station	500-1000m to nearest Crossrail station	1-2 km to nearest Crossrail station
9 - Areas with wider regeneration or growth initiative	2%	0	-3%
10 - Areas expected to benefit from additional labour supply	2%	0	-2%

This displacement pattern is particularly visible when looking at various route sections separately – in which case the displacement effects are visible in the Eastern and South Eastern sections only. The South East section seems to have seen the highest displacement effects – prices in the third distance band (1 km to 2 km from the stations) grew by 6% slower than they otherwise would have. A possible explanation is that public transport accessibility is currently relatively low in those areas, and so the transformation effects may lead to properties close to the station being most sought after. The analysis conducted at station level confirms those observations and displacement effects around non-central stations.

For some central stations such as Bond Street and Tottenham Court Road, the positive impacts were found in all three distance bands. However, we have not seen them in all central locations – the effect in the 500m radius is most robust.

Generally, the residential analysis shows that the Crossrail announcement has led to some additional price growth in the central location compared to what would have been seen otherwise, while in the non-central location the impacts are more commonly displacement effects between areas closer to and further from the future Elizabeth line stations.

We have not found significant evidence that Crossrail resulted in increased residential planning applications along the line. This might be because of the available data and the resulting quality of the models we used. It is also reasonable to think that because of the pre-pandemic demand for housing in London housing market, and the observation that all future Elizabeth line stations are already rail or Underground stations, most residential planning and development activities would have happened anyway.

Impact on commercial properties

Again, it is the central London market that has seen additional value growth due to Crossrail.

We found an average line-wide announcement impact on office rental values of around 3% within 500m of Crossrail stations and around 7.5% between 500m and 1 km from the stations. Additional analysis shows that majority of that line-wide impact comes from the central section, which dominates the office rent data set.

These findings described are in line with existing evidence: a better connection between the regional centre and the smaller centres mainly benefits the regional centre. Productive firms in the regional centre are now able to serve distant markets in smaller centres from their existing base.

Impacts on office rent values in the following areas:	500m to nearest Crossrail station	500-1000m to nearest Crossrail station	1-2 km to nearest Crossrail station
1 - Average impact on the whole route	3%	7%	0
2 - Western section outside London	0	0	0
3 - Western London section	0	0	0
4 - Central section	3%	8%	0
5 - Southeastern section	0	6.5%	0
6 - Eastern section	0	0	0
7 - Deprived areas	0	10%	0
8 - Areas with associated major development	3%	7.5%	0
9 - Areas with wider regeneration or growth initiative	2.6%	6%	0
10 - Areas expected to benefit from additional labour supply	3%	7.2%	0

Untested impacts

We were not able to test the announcement impacts on retail and industrial property values due to unmet methodology requirements. For industrial rents, we would not expect to see significant impacts, as existing evidence suggests that it is service sectors, particularly high-skilled services, which benefit most from agglomeration. We would not expect significant impacts before the line opens for retail property values. We suggest that it would be beneficial to evaluate that aspect in the ex-post analysis, bearing in mind any structural changes in the retail sector resulting from the Covid-19 crisis.

9.2 Future uses of this research

The methodology we designed and used for the evaluation of pre-opening impacts of Crossrail can be used for the post-opening evaluation as well. The difference in difference modelling with fixed effects is a robust framework which allows tackling of key evaluation challenges, namely selection bias and unobservable factors which might influence the outcomes. It is also an approach flexible enough to account for more control variables if needed, to construct a counterfactual group in a different way, or to introduce more distance bands to test the impact across them. While such changes would require adjustment to the underlying dataset, the analytical framework remains the same.

In the course of the research we have explored an option of using a continuous treatment approach instead of a binary one. The binary treatment is the most common approach. It is based on an expectation that a treatment does or does not take place, without any scaling of it, and requires an assumption on where it happens. The latter is usually defined through pre-selected distance bands around the stations, and we have done so in this research as well, using three distance bands within 2 km from the stations.

A continuous treatment approach allows for the impact to vary in scale (or in “a dose of treatment”). A property within 500 metres from a central London station might be impacted to a different extent than a property within the same distance from a station outside of London. Those impacts can be estimated if we have a continuous treatment measure. We have constructed such a bespoke measure for all rail stations in the study area, which indicates how much their journey-time-based accessibility will change once the Elizabeth line opens. We discussed and agreed that the continuous treatment approach should be tested in the ex-post evaluation, by which point people may have a better understanding of the scale of impacts that Crossrail is intended to bring, not only in station locations but across the whole transport network in London.

This pre-opening evaluation focused on property impacts, which perhaps more than other personal investments, depend on the perceived future value, therefore it is likely that major transport infrastructure projects or other interventions influence property values even before opening. Other wider economic impacts, such as changes in employment levels or compositions, are expected to take longer to realise, at least at a scale that would allow for an econometric analysis. Therefore, we would suggest conducting the ex-post evaluation for property values at least two years after the opening and waiting at least four years for employment and productivity impacts, bearing in mind that the full impact may be reached only decades after the opening.

The Covid-19 pandemic has severely impacted commuting and working patterns across London and other global cities, especially in knowledge intensive sectors where physical presence is usually not essential thanks to the existing digital technology. This in turn is impacting businesses in city centres in sectors such as retail, hospitality and culture. The discussion on how structural those behaviour and working changes will be is still ongoing, but it is unlikely to conclude before 2022. As with any impacts, years may pass before we can confidently assess how the pandemic has impacted London, its residents and businesses. Therefore we suggest to refrain from conducting the ex-post evaluation too early and waiting until markets have recovered from the pandemic.

Appendix A

Literature Review

A1 Literature Review

A1.1 Introduction

This appendix summarises existing *ex-ante* and *ex-post* estimates of the economic impact of the Elizabeth line, considering the methodologies utilised, their relative robustness, and the potential influence of these approaches upon the reported findings. The five studies considered are:

- Crossrail – Property Impact Study, GVA, 2012
- Transport for London – Land Value Capture, KPMG/Savills, 2016 (Unpublished)
- Crossrail effect puts London housing unit prices on the fast track, Lloyds bank, 2016
- The Overground Effect, Countrywide, 2016
- Proposed approach to baseline, pre-opening and post-opening impacts of Crossrail, Arup, 2016

A1.2 GVA (2012, 2018)

The purpose of the report published by GVA in 2012 and updated in 2018 was to understand the expected scale of property market benefits arising in terms of market activity, value uplift and how the Elizabeth line would support new opportunities. The basis for the assessment is *ex-ante* modelling of estimated increases in property values around future Elizabeth line stations over and above the modelled baseline. The study does not undertake any *ex-post* analysis although both versions of the report make use of data collected during the pre-opening period and conclusions are drawn about the extent to which development over that time is attributable to the Elizabeth line. The report is publicly available.

Methodology

The general methodological approach was the same for both versions of the report:

- Definition of Inner (0-500m), Outer (500-1000m) and Extended (1000m+) Zones of Influences. The latter varied for each station depending on the net improvement in journey times from or to Farringdon station resulting from the Elizabeth line service.
- Selection of baseline indicators relating to market activity, values, development capacity, development context and accessibility were collated for each station using consistent sources throughout.
- Choice of modelled baseline, which establishes predicated property values without consideration of infrastructure upgrades; for commercial property, data is drawn from VOA, Focus, ONS, IPD and Real Estate Forecasting Ltd, whilst for residential property Knight Frank's forecast from its Residential Research programme is used.

- Definition of “Crossrail Impact Factors”: multipliers to capture the potential additional impact in market value and performance generated by the Elizabeth line over the baseline, which have been developed on a bespoke basis for different sections of the route based on precedent and previous academic studies.

Values are set out on an annual basis from 2012-2021 (through the construction, testing and initial operation phases) but the authors add the caveat that results are not intended to provide time or location specific forecasts (and more illustrate the overall trend throughout the period).

Key Findings

The Study predicts that the Elizabeth line will have a considerable impact on the property market, in terms of delivery rates, occupier demand and the value of all forms of floorspace; commercial values around future Elizabeth line stations in central London will see an uplift of 10% above the baseline, whilst residential values will increase by some 25% in central London and 20% in the suburbs. It is forecast that between 2011 and 2021, the Elizabeth line will have a very significant impact on both residential and commercial property values. Furthermore, once the line is fully operational, it is likely that there will be a much more significant impact upon property values and transaction activity.

The general pattern is that the Elizabeth line will reinforce a number of the historically strongest performing areas, particularly in central London, though other markets may also strengthen. For example, residential values in the 500m zone of influence around Farringdon, Liverpool Street and Canary Wharf will move from underperforming the line section average to a stronger position. There are also opportunities to create new markets in areas which have historically underperformed when compared with the relevant local authority benchmark.

For some other areas benefits are predicted to be limited, while there may also be noticeable variation between impacts on the inner and outer impact areas. For example, whilst it is predicted that residential property values within the inner zone of influence at Southall will surpass the borough average by 2017, the impact upon the wider 1000m and extended zones of influence is expected to be more limited, potentially reflecting the low-density housing stock throughout much of the area. Another example is Abbey Wood, where residential values are predicated to remain 15% below the average for the eastern section of the line as a result of a variety of dampening factors.

For commercial office values, most of the benefits are likely to be felt in the Central section, with Bond Street, Paddington, Tottenham Court Road, Farringdon, and the Isle of Dogs cited as locations that are likely to see their performance improve. Meanwhile, outside of the Central section, it is generally judged that the Elizabeth line is likely to have a limited impact in terms of generating additional take-up of office space, or (particularly in the case of the eastern section) increasing values.

A1.3 KPMG/Savills (2016)

As part of a wider study commissioned by Transport for London (TfL) into the concept of 'value capture', *ex-post* analysis was undertaken by Savills on historic land value uplift around a series of live or completed TfL transport projects, including the Elizabeth line. A summary of this research was published by TfL in February 2017, although the full technical report was not published and has been supplied in confidence.

Methodology

The methodology used is a basic 'difference in difference' using panel data with average residential property values indexed and subsequently observed from one year before the start of construction to May 2016.[31] within a 500m radius of individual future Elizabeth line stations on a monthly basis (their 'zones of influence'). The change in values was compared with the corresponding change in properties falling in a 1-2km radius around the same station, with these areas serving as the 'controls'. The rationale for this is that the properties are effectively in the same neighbourhood, the defining difference between them being their proximity to the station. Any difference in the property price index between the 'treatment' and 'control' areas is assumed to be the land value impact of the transport investment. These uplifts were then averaged to provide an overall uplift for the Elizabeth line as a whole.

No additional controls are used to strip out the confounding factors. The analysis suggests that the findings may hold true in aggregate for the impact of the Elizabeth line but are less likely be a true reflection of individual station level outputs.

It is challenging to conclude the possible effect of the methodological choice upon the reported results given the range of possible external influences on property prices. Evidently, these influences may apply equally to the control areas as the zones of influence. It is also observed that, given the dense 'patchwork' nature of property values in London, the choice of a 1-2km buffer for the control may not provide an appropriate control for each station that is reflective of the local situation.

Key Findings

The assessment found that residential property values within the zones of influence remained broadly static between July 2007 and mid-2011, despite the commencement of Elizabeth line construction, with subsequent average values not deviating significantly from values within the control areas over the remainder of the study period (through to 2016). The exception is during the last three months, March-May 2016, which showed growth in values lagging behind those in the control areas.

Some small uplifts were noted at the individual station level, all of which were areas with adjacent high value markets from which demand was drawn, and a supply of period (pre-1930s) housing stock; even then, values grew just 0.4% faster pre-

[31] Analysis of commercial values is limited to quoting of secondary data from wider studies.

construction, and 0.1% since commencement of construction. Thus, overall effects are judged to be insignificant to negative.

A1.4 Lloyds Bank (2016)

In December 2016, Lloyds Bank released a short analytical public facing piece on the trend in the value of residential properties near future Elizabeth line stations compared with the surrounding local authority area.

Methodology

This analysis compared eight and two-year trend analysis of 12-month arithmetic Land Registry data for areas "around" future Elizabeth line stations, then it compared the average of these increases against the average increases for the overall local authority areas in which they sit. The article does not confirm the definition of "around" the stations, which makes it challenging to understand which areas are being compared. Some comparison is also made between housing unit values around future Elizabeth line stations and the wider local authority in which they sit, but at a fixed point in time using cross-sectional data.

The analysis undertaken is akin to a simple additionality assessment with no consideration of causality or controls. It is therefore not possible to attribute the observed increases in average values, or the difference in values versus the local authority area, solely to the Elizabeth line. This is because the impact of other factors is not considered. Furthermore, in some cases the logic for the selection of the local authority comparators is not clear; for example, property values around Iver are reported against average figures for Slough, which suffers particularly depressed values, but the station and the village associated with it are located within South Bucks. The rationale for this is unclear but makes it likely therefore that the "Crossrail effect" may have been overestimated for this location.

Key Findings

Housing unit prices near future Elizabeth line stations were reported to have risen 22% between 2014 and 2016, versus an average 14% growth for surrounding local authority areas and 13% growth for Greater London. Of 33 stations assessed, 28 saw average housing unit price growth for homes in the same postcode sector above that for the surrounding local authority (noting the discrepancy in defining 'surrounding local authority' mentioned above).

The increases around Abbey Wood, Forest Gate and West Drayton were noted as the highest, 47%, 46% and 46% respectively.

A1.5 Countrywide (2016)

Countrywide undertook analysis of residential property values around London Overground stations versus values in the wider TfL fare zones in which they sit. Separately, it also carried out an initial assessment of the proportion of investors, first time buyers and other owner occupiers purchasing properties around future Elizabeth line stations during the construction period. The research is unpublished, and we have not had sight of the methodology, only a presentation of headline findings. Observations on method have been received at second hand from TfL officers.

Methodology

The London Overground analysis compared residential property values using Land Registry data in fixed radius areas around London Overground stations, versus the average value of properties around stations across the equivalent TfL fare zone. This analysis was carried out against a blanket average ‘control’ and did not use a specific sample for the comparison. The assessment was carried out over the period 2010 to 2016, with the variation between the Overground stations and the ‘control’ calculated for each year and then across the whole period. No econometric analysis was undertaken to isolate the effects of Overground versus other transport improvements or wider factors, thus causality cannot be demonstrated.

The second piece of analysis again used Land Registry Data, comparing the relative proportions of homes purchased by investors, first time buyers and other owner occupiers for fixed radius areas around future Elizabeth line stations. This was presented graphically, and comparisons were also made between trends in western, central, and eastern sections.

Key Findings

For the property value analysis, Countrywide found the greatest impacts on values around Overground stations in Zone 4, with a 16.7% increase over values around other Zone 4 stations between 2010 and 2016. Values in Zone 3 were also notably higher (9.7%). Notably, aside from Willesden Junction, all of the strongest performing stations are located in outer north-east or east London.

Countrywide’s analysis of the Elizabeth line focused on the proportion of residential sales over time around future Elizabeth line stations completed by first time buyers, investors, and other owner occupiers. The analysis would appear to illustrate that a much higher proportion of residential sales during the construction years are completed by property investors, whilst first time buyers and other owner occupiers respond more slowly to change. However, notably, the effect was less pronounced in east London where first time buyer purchases continued to form a higher proportion of all sales than investors throughout the period (though the gap was closed somewhat).

A1.6 This study

Arup’s methodology described in detail in this report attempts to respond to the need for robust, *ex-post* evaluation of the Elizabeth line’s impact on wider economic, property and regeneration outcomes. It employs econometric techniques to attempt to isolate the extent to which observed property market changes are attributable to the Elizabeth line and control for wider observable factors which might also impact the outcomes.

Methodology

The analytical approach tests two different treatment variables (binary and continuous, based on expected accessibility changes) to attempt to fully capture the impact of the Elizabeth line, given the new infrastructure will primarily improve journey times and add additional capacity rather than provide new stations. Accessibility measure captures this more effectively, and furthermore allows the benefits generated by the Elizabeth line (as opposed to other transport interventions) to be isolated.

The modelling has been performed through linear egressions with control variable including fixed effects.

A1.7 Summary

In summary, The GVA study is modelled (predicted) rather than based on observed ('outturn') data and is therefore not a direct comparator to the Arup study. Both the Lloyds and Countrywide studies are based on simple comparison of property prices near a station to various average price changes in a form of analysis that does not allow for causal inference.

The KPMG/Savills work is the most methodologically robust (among the non-Arup studies), using a difference in difference analysis of panel data. However, it does not make use of fixed effects, nor include any other controls. Therefore, the validity of the analysis rests on the assumption that the 'treatment areas' (0-500m around a station) would develop in exactly the same way in the absence of the Elizabeth line as the 'control areas' (1-2km around the same station). It looks only at residential values not commercial values or new development and implicitly treats all areas served by the Elizabeth line as having the same impact, although the service patterns at different stations will vary.

The Arup approach includes area, year and area-year fixed effects and other control variable in the models. Including control variables attempts to account for observable differences between areas based on their economic, demographic, and geographic characteristics which might otherwise affect the estimate of the Elizabeth line's impact. The fixed effects should control for other, 'unobservable' differences between areas that are place or time specific. The 'treatment' variable is based on a bespoke accessibility matrix which estimates extent to which the Elizabeth line increases the accessibility of the station, expressed in terms of journey time. This should allow for a more nuanced estimate of land value impacts of the Elizabeth line, and a greater confidence that any association is causal.

The key points of methodology and the findings of each study are summarised for reference in the table overleaf.

Table 33 Summary of study methods and findings

Study	Method	Key Findings
GVA (2012, updated 2018)	- Economic model (predicted, not observed values) - Use of baseline which does not take into account the effect of infrastructure investment; - Application of Crossrail multipliers to estimate potential effects on residential and commercial values in Inner, Outer and Extended Zones of Influence; - Forecast and does not compare values over time.	- the Elizabeth line will have significant impacts upon both commercial and residential property values, completions and demand between 2011 and 2021; - Established markets, such as those in the Central section, will be further reinforced; effects more mixed on stations in East and West line sections; - Some positive impacts upon historically under-performing markets and opportunities to create new markets.
KPMG/Savills (2016)	- Before-and-after using panel data; - Comparison of residential values between one year prior to project completion and May 2016, looking at 500m buffers around stations versus control areas (1-2km buffers); - No use of controls for external influences on property values.	- Residential values increased in zones of influence broadly in line with control areas between mid-2007 and 2011 when construction commenced; - Small, localised uplifts in areas with stronger markets and period housing stock; - Overall effects judged to be insignificant to negative.
Lloyds Bank (2016)	- Residential property value trend analysis for areas around future Elizabeth line stations against average increases for wider local authority areas; - Assessment akin to simple additionality assessment with no consideration of causality or controls.	- 22% increases in average residential values around future Elizabeth line stations versus 14% for surrounding local authorities and 13% for Greater London; - Particularly substantial growth in some specific outer London locations.
Countrywide (2016)	- Time series analysis of residential property values for fixed radius areas around London Overground stations, with a comparison against the average for other stations within the equivalent fare zone; - Comparison against a blanket average as opposed to a tailored sample, and no controls used to isolate causality of Overground versus other factors.	- Zone 4 Overground stations experience the largest uplift in property values compared with other zone 4 stations; - Owner occupiers suffer imperfect information and respond more slowly to infrastructure investment construction/announcements than investors.

Study	Method	Key Findings
This study	- Fixed effects model to estimate impact of Elizabeth line-induced accessibility and capacity enhancements on a range of economic outcomes, including property impact; - Controls for both observable and unobservable factors that may impact property values (both residential and commercial) - Use of a matrix to isolate relative accessibility impacts of the Elizabeth line versus other infrastructure investments;	Please see the executive summary.

Appendix B

Journey time calculations

B1.1 Timetable based accessibility calculation

In TRACC software, a multi-modal transport accessibility tool, we added all relevant services' timetables, including the planned Elizabeth line timetable. Then we set all the stations' coordinates and network parameters – maximum walking distance and interchange penalty (an artificial time penalty, over and above the true time cost of changing trains, to represent the perceived inconvenience of interchange to travellers). The software produced a schematic network of public transport routes and calculated the shortest path between all origins and destinations (in this case stations). Our network does not include the bus system; however, it may include walking if two points on the route are within the maximum walking distance (set as a parameter).

The software algorithm works broadly in the following way:

1. For each origin point, find the nearest rail station.
2. Determine whether that stop is within a distance higher or lower than the maximum walking distance we set out.
3. If the stop is further than the maximum walking distance, assume you cannot make that journey and exclude the station from the analysis.
4. Repeat steps 1 – 3 for a destination point.
5. Find the shortest route between the origin and destination, taking into account a boarding penalty we set out.
6. Add the walking time at both ends to the time of journey on public transport.

In our case origin and destination points are already stations, so step six should have a marginal importance, unless it is more optimal to walk to another station and start a journey from there.

As the algorithm code is the intellectual property of the software producer, we cannot provide further detail on how it is structured, especially step five. We did not add any peak restrictions (in theory the software can calculate the shortest path available with an assumption of reaching the destination for instance by 9 am).

In the end we calculate the accessibility level for two scenarios:

(1) in 2016 without the Elizabeth line

(2) in 2016 with the Elizabeth line

The difference between those two is our "anticipated accessibility change due to the Elizabeth line". The relative change value (divided by the "2016 without the Elizabeth line" level) is our continuous treatment variable. This is a pure change due to the Elizabeth line and does not include impacts of other major transport schemes that open between 2016 and 2020 (e.g. Thameslink).

B1.2 Generalized Journey Times based accessibility calculation

The approach described above concentrates on changes in travel journey time resulting from distance and speed. However, improving journey times was not the only, or even the main purpose of the Elizabeth line. The Elizabeth line business case was strongly based on the need to add additional capacity to the London transport network (it offers around a 10% increase in total capacity). This will especially be important in analysing the Elizabeth line's impacts in central London, where areas already well served by the tube and rail network are overloaded and the ability to carry more people may be equally if not more important than reduced journey times.

Accessibility analysis based on standard journey times will not take into account capacity improvement. In order to capture this aspect, we also calculated an accessibility-measure based on Generalized Journey Times.

Generalized Journey Times are modelled through Railplan. Railplan is a strategic public transport model for London and the South-eastern. Railplan models the likely route and service choices of public transport users and the resulting levels of crowding on public transport networks in and around London[32]. It allocates the public transport trips generated by London Transportation Studies (LTS) model to all public transport modes: National Rail, Underground, Overground, Docklands Light Railway, London Bus and Tramlink.

The Railplan model allocates demand based on a generalised cost calculation built as a function of in-vehicle time, walk time, wait time and boarding penalties for every single combination of possible trips between two O-D pairs. In-vehicle time takes into account the greater costs associated with longer journey times while also including the additionally incurred costs that result from crowding. Walk time refers to the total time spent on walk links within the whole trip and takes the higher costs associated with having to walk to and from an origin/destination or to another public transport node. The model also includes wait time for services and boarding penalties which represent the variation in perceived costs associated with interchange by mode. Bus to bus interchange will incur a higher penalty than bus to rail interchange for example. The formula for the assignment procedure is as follows:

GC (Generalised cost) = IVT + (2* walk) + (2.5 * wait) + BP

- IVT = in vehicle time (including crowding factor)
- Walk = time spent on walk links within the whole trip
- Wait = time spent waiting for services
- BP = boarding penalty which are mode specific costs associated with having to change services

[32] http://content.tfl.gov.uk/londons-strategic-transport-models.pdf

This calculation produces the generalised costs between each Railplan zone pair (O-D) within the model. However, generalised cost calculations become significantly less accurate for areas outside of the Greater London boundary (GLB), as zone sizes become more aggregated, and the representation of public transport services becomes less detailed. That means that in many cases a high number of observations outside London might be mapped to the same Railplan zone and will therefore be assigned with the same level of accessibility improvement. That in turn leads to a loss of variation in accessibility level. This means that we are less likely to be able to pick up corresponding relationships with property values.

Appendix C

Journey-time accessibility note

Transport interventions aim at improving the accessibility of places and in some cases this accessibility improvement will lead to wider economic benefits over and above the direct benefit of reduced travel time costs. Journey time accessibility can mean various things for various agents. People considering where to live are interested in how easy it is from that location to reach places with jobs, health, education and entertainment services and so these factors are widely demonstrated to affect residential property prices. Companies deciding on where to locate an office are interested in how easy it is to reach employees, suppliers and other companies from the office. Knowledge intensive firms in particular often have an interest in the quality of work life that can be offered to current and prospective employees; places to get lunch, recreation areas and gyms, and shopping facilities, for example.

Accessibility in developed urban areas very rarely depends solely on one transport option – one rail connection, or one bus. With dense transport network such as in London, the accessibility of a single point is affected by the accessibility of other points, as people can conveniently change lines and modes at many locations.

Evaluation of transport interventions should therefore take that interconnectivity aspect into account, considering the impact of places directly served by the intervention but also those indirectly served (for example, places that would only have been accessible with two interchanges becoming accessible with one interchange). However, the prevailing approach in evaluation studies of new transport interventions so far is generally based on a simplifying assumption that impact of a transport project is felt only in areas directly served. The treatment is usually defined as reducing a distance to a new station. Binary treatment variables would then be defined based on being either within or outside of a radius (distance bands) around a station, the most common for urban areas being 500 m, 1 km and 2 km (approximately a 5, 10 or 20-minute walk).

Crossrail is also different from standard projects as it has been laid out entirely on the existing rail network. No new station locations have been identified for the project – although some stations are being substantially upgraded, all Crossrail stations have been accommodated at existing London Underground or National Rail stations: that is to say, they were already served by a rail based mode of transport. Defining treatment based on distance bands wouldn't therefore allow us to distinguish Crossrail's impact on the outcomes from other factors typical and already existing around rail stations. In addition to that, this approach would not allow to understand whether there are impacts of improved interconnectivity on areas not directly served.

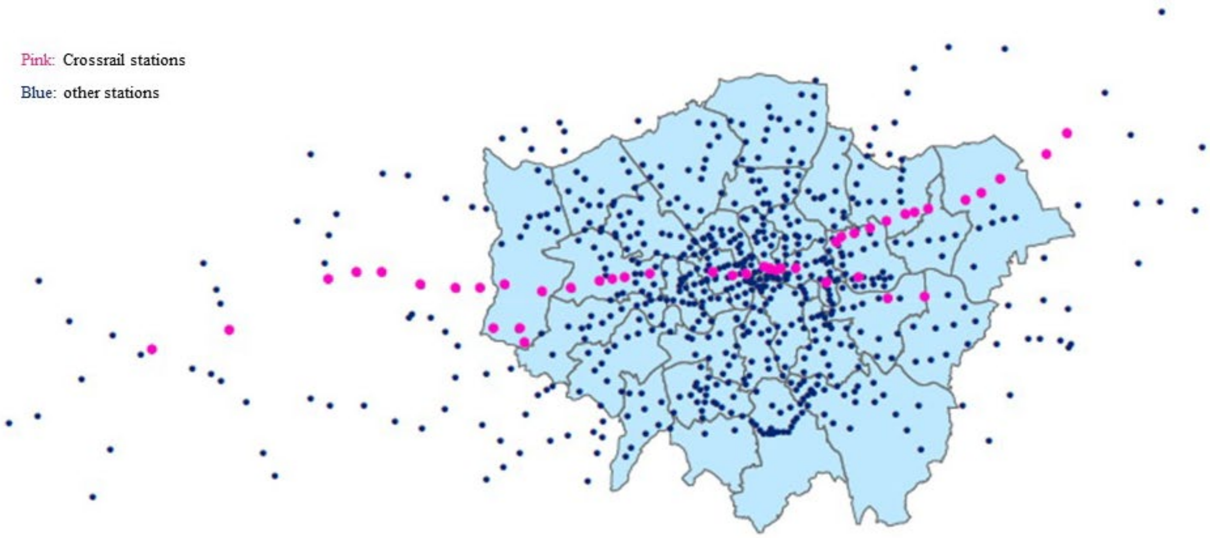
Therefore, for the purpose of this project we developed a bespoke accessibility measure. Two main aspects are discussed in the following section:

1. the algorithm,
2. the accessibility formula.

C1 The algorithm

Our study area is defined as a 10-mile corridor around the Crossrail route. Within that area we identified over 650 rail stations with all types of services – Underground, Overground, National Rail and Docklands Light Railway (see Figure 1 below). In TRACC software, a multi-modal transport accessibility tool, we added all relevant services' timetables, set all the stations' coordinates and network parameters – maximum walking distance and interchange penalty. The software then produces a schematic network of public transport routes and then allows us to calculate the shortest path between all origins and destinations (i.e. stations) [33].
Our network does not include the bus system, however it may include walking if two points on the route are within the maximum walking distance set as a parameter.

Figure C1: study area stations



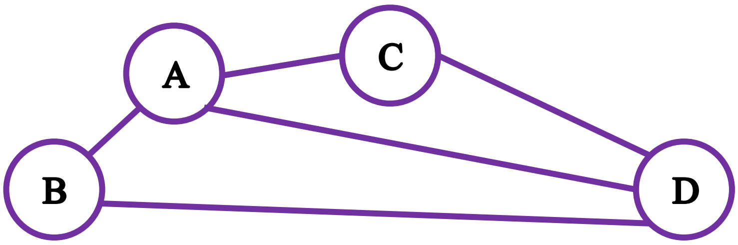
[33] As the algorithm code is the intellectual property of the software producer, we cannot provide further detail on how it is structured. We did not add any peak restrictions (in theory the software can calculate the shortest path available with an assumption of reaching the destination for instance by 9 am).

C2 The accessibility formula

The accessibility formula is essentially the closeness centrality formula, which has a recognised pedigree in the literature. For a given location, its accessibility is expressed as a sum of reciprocals of distances between that location and every other location in the network. Therefore for a hypothetical network of four locations A, B, C and D, the accessibility measure for A will be defined with the following formula:

$$Access_A = \frac{1}{JT_{AB}} + \frac{1}{JT_{AC}} + \frac{1}{JT_{AD}}$$

Distance can be defined either by geographical distance or by the time need to travel between locations (which of course depends to an extent on distance but also on service patterns), and the latter is what we use. For our hypothetical network drawn below, we assume the following:



Measure	Scenario 1 – Original journey times	Scenario 2 – Journey times reduced by 1 minutes each
From A to B	10	9
From A to C	100	99
Form A do D	1000	999

We experimented with an accessibility measure constructed simply as a sum of journey times and an average journey time between an origin and all destinations in the network. An advantage of this simple formula is an accessibility score measured in interpretable units, minutes. A disadvantage is a counterintuitive measure in which a higher accessibility score, is ‘worse’ (less accessible) rather than ‘better’ (more accessible). More importantly, in all modifications of this formula, a journey time reduction by a unit means the same, regardless if it refers to a reduction in journey time between places which are close or distant. A one minute reduction in journey time between Reading and Shenfield would therefore be valued equally to a one minute reduction between Paddington and Bond Street. Using a closeness centrality formula the one minute change between Paddington and Bond Street will have a greater impact on the total score improvement, because all changes impact the final score through their relation to the denominator.

Measure	**Formula 1a:** Accessibility measure = Sum of journey times (JT)	**Formula 1b:** Accessibility measure = average journey time	**Formula 1c:** Accessibility measure = Reciprocal of sum of JT	**Formula 2:** Accessibility measure = Sum of reciprocals of JT (Closeness centrality)
Scenario 1	10 + 100 + 1000 = 1110	(10 + 100 + 1000) / 3 = 370	1/(10 + 100 + 1000) = 0.000901	1/10 + 1/100 + 1/1000 = 0.111

Scenario 2	9 + 99 + 999 = 1107	(9 + 99 + 999) / 3 = 369	1/(9 + 99 + 999) = 0.000903	1/9 + 1/99 + 1/999 = 0.122
Accessibility ratio	1107/1110 = 0.997	369 / 370 = 0.997	0.000903/0.000901 = 1.0027	0.122/0.111 = 1.10
Accessibility improvement	Reduction in JT by 0.3%	Reduction in average JT by 0.3%	0.3%	10%

In essence, the closeness centrality formula puts greater weight on a unit of journey time change between places close to each other. In the case of Crossrail, this is not ideal either, because central London stations are already relatively well connected and relative journey time changes between them will obviously be smaller than between central London and Outer London, for instance. By putting greater emphasis on these changes and less on the fact of central London becoming more accessible from residential areas further along the route, we risk underestimating Crossrail's impact in central London. Similarly, when analysing accessibility in London suburbs, we might underestimate Crossrail's impact by putting slightly greater emphasis of them being better connected with places closer to them, and not taking into account the fact of them being also better connected to jobs and urban facilities in central London, even if central London is still relatively far from them. In order to mitigate that risk, three slightly different versions of accessibility score were constructed based on three types of weighting:

1. Equal weights – reciprocals' denominators are equal to 1,
2. Population weights – reciprocals' denominators are value of population within 1 km around a given destination,
3. Employment weights – reciprocals' denominators are value of employment within 1 km around a given destination.

Formulae for population- and employment-weighted accessibility score are the following:

$$Access\_Pop_{A} = \frac{Pop_B}{JT_{AB}} + \frac{Pop_C}{JT_{AC}} + \frac{Pop_D}{JT_{AD}}$$

or

$$Access\_Emp_{A} = \frac{Emp_B}{JT_{AB}} + \frac{Emp_C}{JT_{AC}} + \frac{Emp_D}{JT_{AD}}$$

In case of location A, its population-based accessibility improvement is slightly higher than the employment-based one. That means that due to our transport intervention location A becomes relatively better connected to locations with high population numbers than to locations with high employment. Also, the scale of improvement in both these cases in smaller than when looking at the unweighted score.

Table 34: Accessibility calculation example explanation

Measure	Scenario 3: **Journey time**	Scenario 4: **Journey time**	Both scenarios: **Population in destination**	Both scenarios: **Employment in destination**
From A to B	10	9	40	20
From A to C	20	19	50	75

Form A do D	30	29	85	5

This is why it is important to test all these different approaches through the modelling. We proceed from the assumption that for commercial property, the population-weighted measure is likely to have highest explanatory power, as employers (particularly in knowledge-intensive sectors) will be attracted to areas where they can reach the biggest possible pool of potential employees. Likewise, we expect the employment-weighted index to have the highest explanatory power for residential property where home buyers are likely to be more impacted by how easy it is to access areas with job opportunities.

Table 35 Accessibility calculation example calculation

Measure	Closeness centrality – equal weights	Closeness centrality – population weights	Closeness centrality – employment weights
Scenario 1	1/10 + 1/20 + 1/30 = 0.183	40/10 + 50/20 + 85/30 = 9.333	20/10 + 75/20 + 5/30 = 5.916
Scenario 2	1/9 + 1/19 + 1/29 = 0.198	40/9 + 50/19 + 85/29 = 10.007	20/9 + 75/19 + 5/29 = 6.342
Accessibility ratio	0.198/0.183 = 1.0823	10.007 / 9.333 = 1.072	6.342 / 5.916 = 1.071
Accessibility improvement	8.23%	7.2%	7.1%

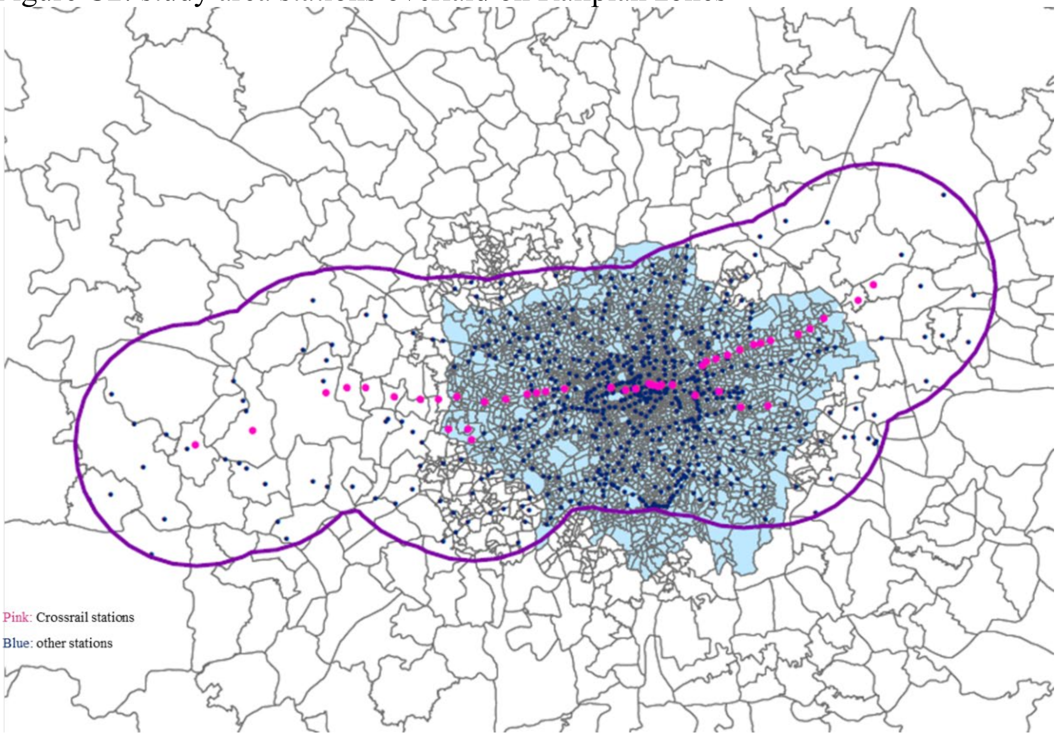
C2.1 Generalized Journey Times

Improving journey times was not the only purpose of Crossrail. In fact, the Crossrail business case was strongly based on the need to add additional capacity to the London transport network (it offers around a 10% upgrade in total capacity). This will especially be important in analysing Crossrail's impacts in central London, where areas already well served by the network are overloaded and the ability to carry more people may be equally if not more important than reduced journey times. Accessibility analysis based on Standard Journey Times (SJT) (as a result of distance and speed) will not take into account the capacity improvement. For this reason we also calculated a closeness-centrality-accessibility-measure based on Generalized Journey Times (GJT).

Generalized Journey Times are modelled through Railplan. Railplan is a strategic public transport model for London and the South-East. Railplan models the likely route and service choices of public transport users, and the resulting levels of crowding on public transport networks in and around London[34]. It allocates the public transport trips generated by LTS to all public transport modes: National Rail, Underground, Overground, Docklands Light Railway, London Bus and Tramlink.

[34] http://content.tfl.gov.uk/londons-strategic-transport-models.pdf

Figure C2: study area stations overlaid on Railplan zones



The Railplan model allocates demand based on a generalised cost calculation built as a function of in-vehicle time, walk time, wait time and boarding penalties for every single combination of possible trips between two O-D pairs. In-vehicle time takes into account the greater costs associated with longer journey times while also including the additionally incurred costs that result from crowding. Walk time refers to the total time spent on walk links within the whole trip and includes the higher costs associated with having to walk to and from an origin/destination or to another public transport node. The model also includes wait time for services and boarding penalties which represent the variation in perceived costs associated with interchange by mode. Bus to bus interchange will incur a higher penalty than bus to rail interchange, for example. The formula for the assignment procedure is as follows:

$$GC = IVT + (2 * walk) + (2.5 * wait) + BP$$

IVT = in vehicle time (including crowding factor)
Walk = time spent on walk links within the whole trip
Wait = time spent waiting for services
BP = boarding penalty which are mode specific costs associated with having to change services

This calculation produces the generalised costs between each Railplan zone pair (O-D) within the model. However, generalised cost calculations become significantly less accurate for areas outside of the Greater London boundary (GLB), as zone sizes become more aggregated and the representation of public transport services becomes less detailed (see Figure 2 above). That means that in many cases a lot of outside London observations will be mapped to the same

Railplan zone and will therefore be assigned with the same level of accessibility improvement. That in turn leads to losing variation in accessibility level and not being able to pick up their relationship with property values.

Since there is no perfect method, we are testing our modelling with 6 types of accessibility score – differing in use of SJT or GJT and one of three weighting types (2 x 3 = 6).

C3 Accessibility estimation outputs

Figures C3 and C4 below compare accessibility improvement calculated with an unweighted formula, with standard journey times and generalised journey times respectively. The colour of the hexagons represents average accessibility improvement in the relevant area. The darker the colour, the bigger the improvement. The first difference between two approaches is that the range of improvements based on unweighted formula is wider for SJT than for the improvements based on GJT. Maximum average is 22% for SJT-based and 11% for GJT-based calculations.

Figure C3: Average accessibility improvement (standard journey times, unweighted)

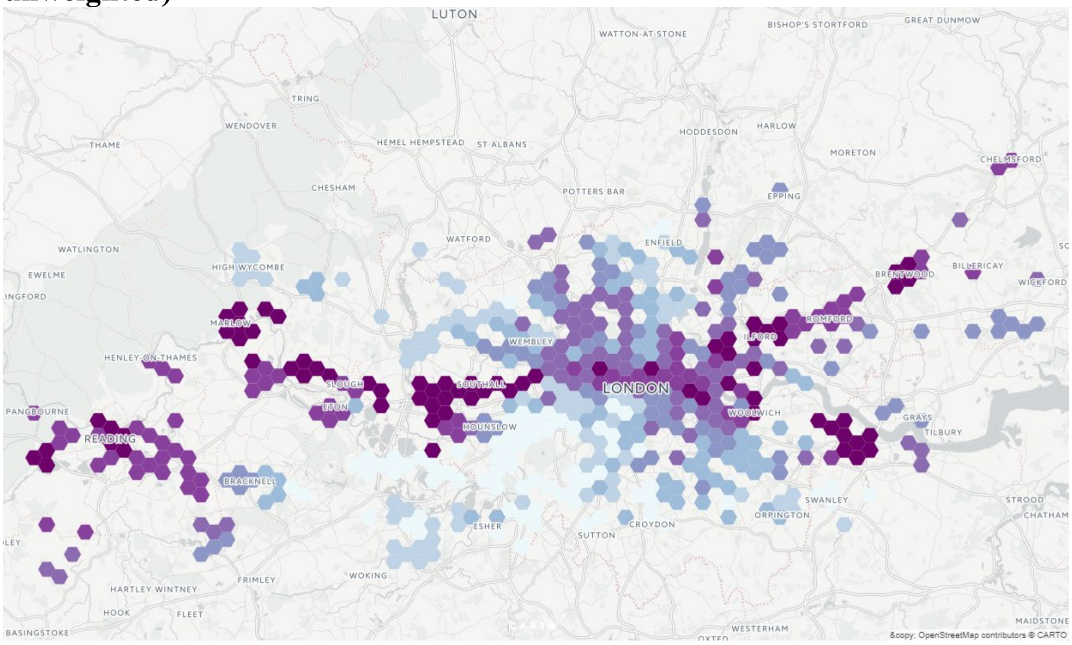
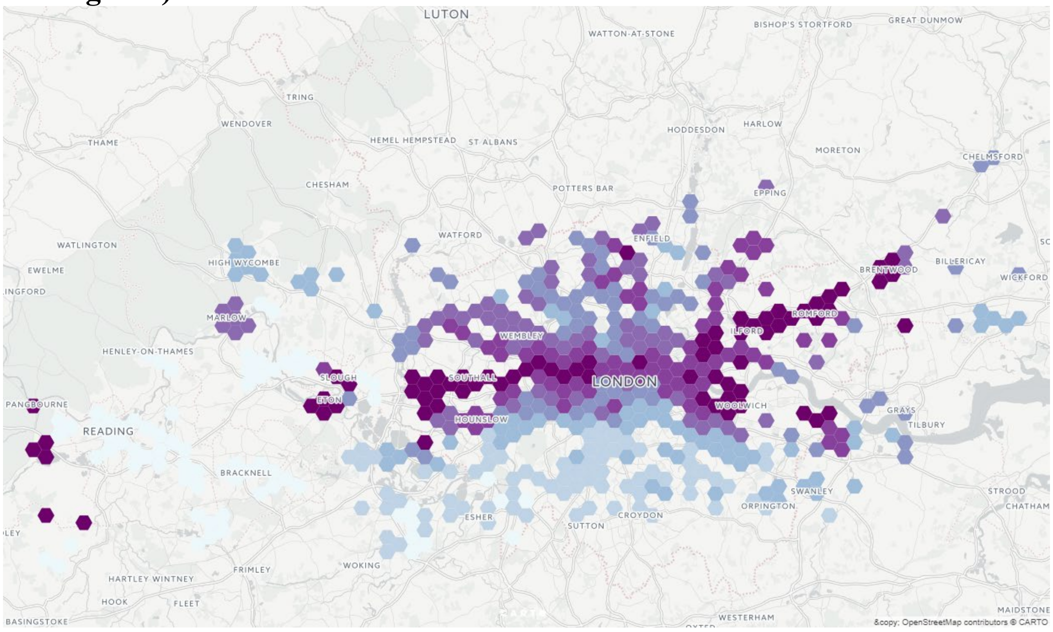
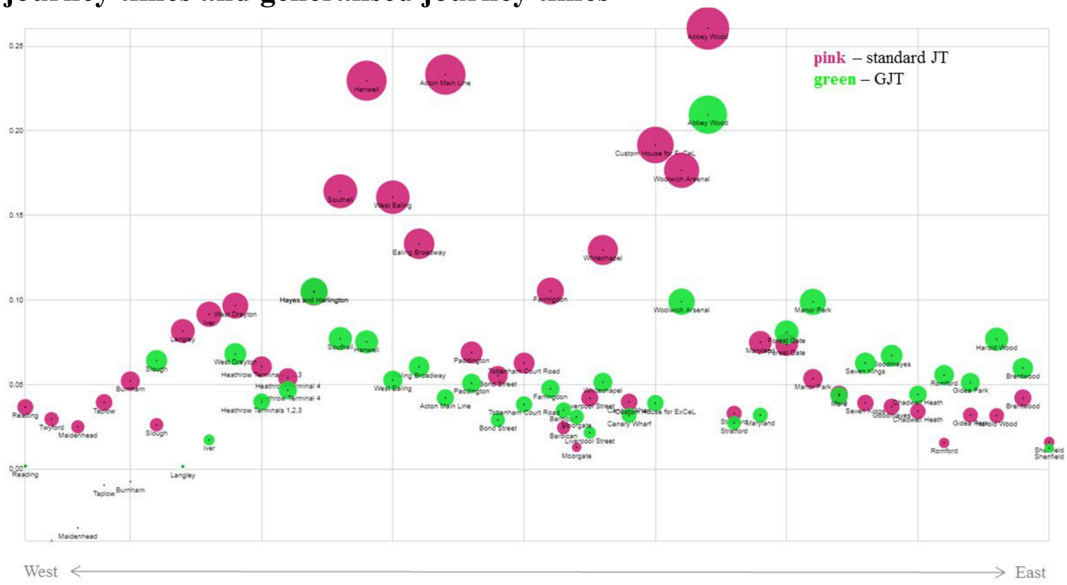


Figure C4: Average accessibility improvement (Generalised Journey Times, unweighted)



Secondly, the highest improvements based on SJT are observed along the Elizabeth line route, and in case of Western section also for locations around the line – especially strong around Southall, Marlow and Reading. London locations other than along the Crossrail route show relatively modest improvements. That changes however when we look at GJT-based scores. While Eastern sections stay strong (although the relative improvement is lower than in case of SJT), locations on the West show very, very modest improvements or even slightly negative impacts – Twyford, Maidenhead, Taplow and Burnham.

Figure C5: Accessibility improvement (unweighted: comparison for standard journey times and generalised journey times



The "negative improvement" result for a few stations is counterintuitive and requires special attention. In order to test that result:

1. We have analysed Railplan information. It is a fact that after Elizabeth line starts operation and some Great Western trains are discontinued, there will be less frequent rail service between Reading and Maidenhead (Figure C6, below). Journey time from some West stations to other stations on the network might also be longer, as Crossrail trains will stop on each station on the lines, while some Great Western trains that are going to be taken off currently omit some of the stations (Southall, Acton, West Ealing, Acton Main Line) – see Figure C9. On the other side however, total peak capacity is supposed to increase across the whole network, including the Reading – Maidenhead section (see slide 7). Similarly, crowding levels are modelled to decrease across almost all network, and definitively on the section from Reading to Paddington. To sum up, although deterioration of GJT-based accessibility on the West sounds plausible, we strongly believe this result requires a more detailed analysis of the Railplan data provided.

Figure C6: Railplan modelled changes in peak rail service frequency

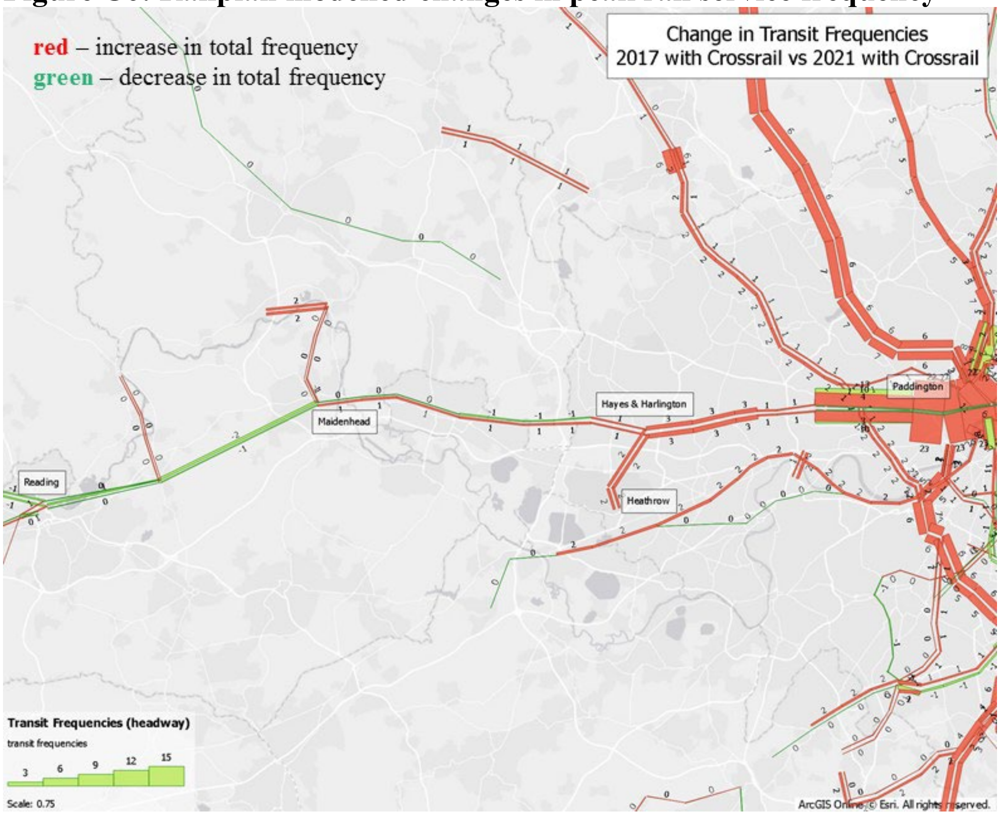


Figure C7: Railplan modelled change in peak rail service capacity

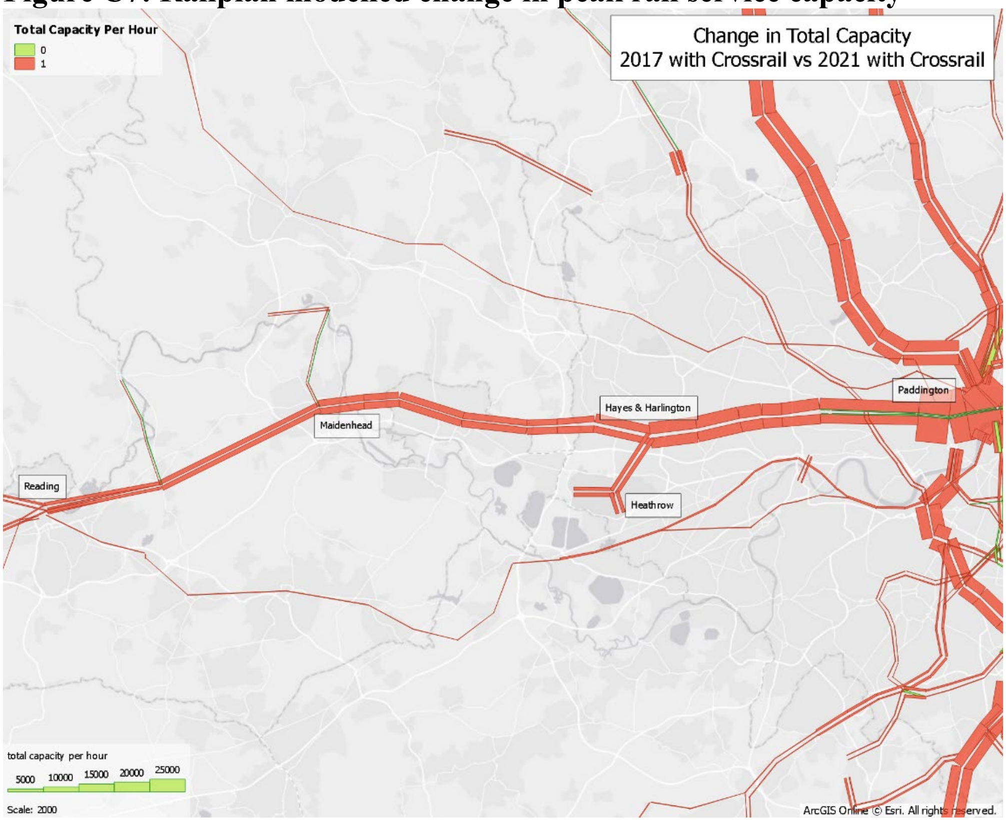


Figure C8: Railplan modelled change in peak rail service demand (passengers)

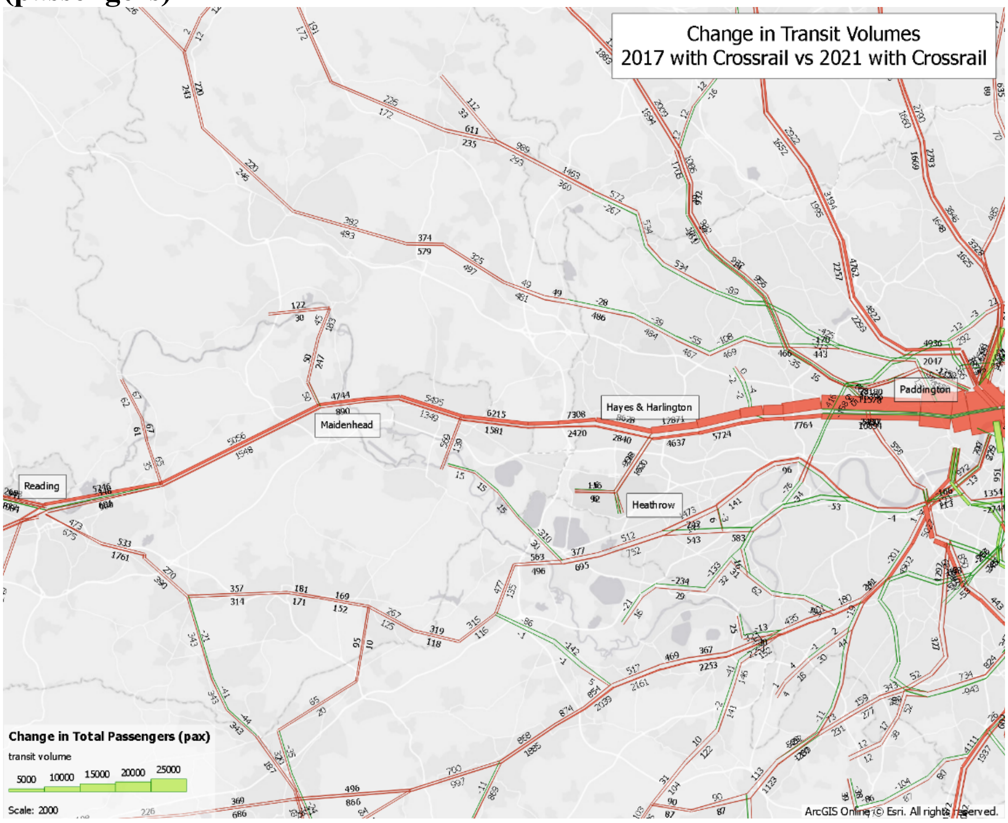
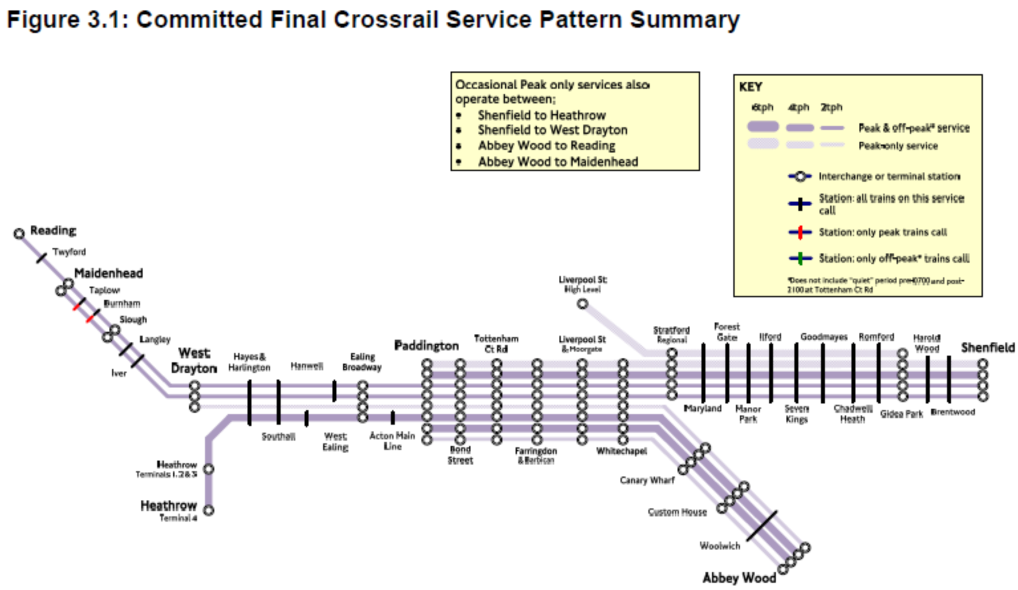


Figure C9: Railplan modelled change in crowding level



Source: Crossrail Business Case 2015

2. We analysed the synthetic score for the Burnham (as an example from the four stations mentioned above) in disaggregation for all components of the score – as the accessibility score is a sum of reciprocals of journey times between Burnham and all other destinations on the network.
   Figure C10 represents what the changes look like when the time is defined as standard journey time. According to this approach, due to Crossrail it will be quicker to get to stations in the centre and east of the study area, and there will be no change in time to get to the stations in the western part of the study area.
   Figure C11 shows what the changes look like when the time is defined as generalised journey time. With this approach, the connection between Burham and over half of the stations in the network will deteriorate – including a lot of central stations.

Figure C10: Simple Journey time changes from Burnham to other stations due to Crossrail

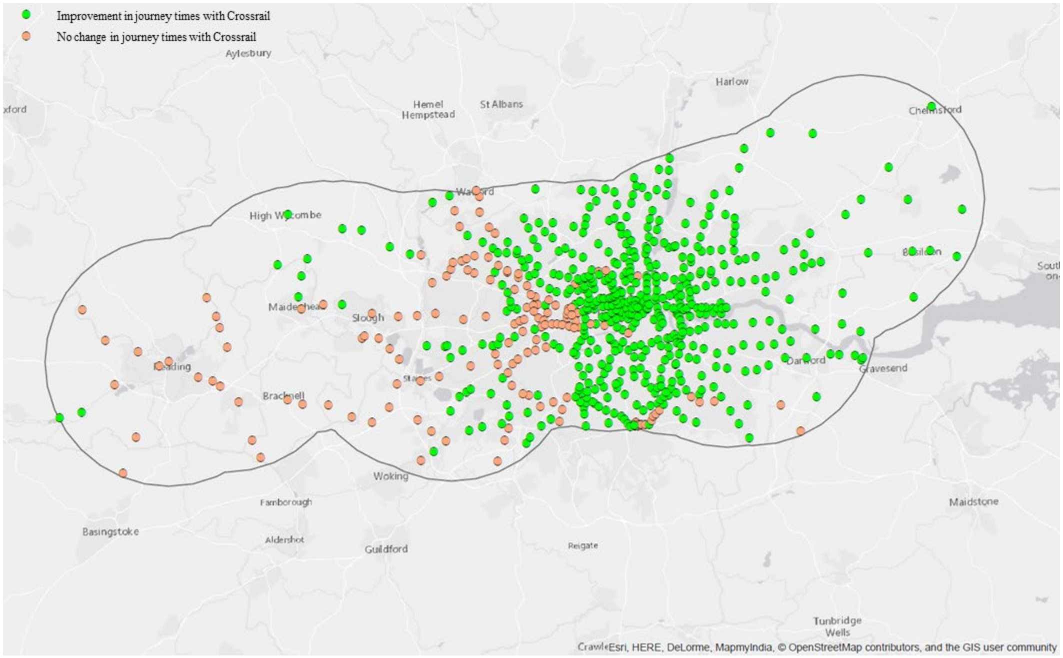
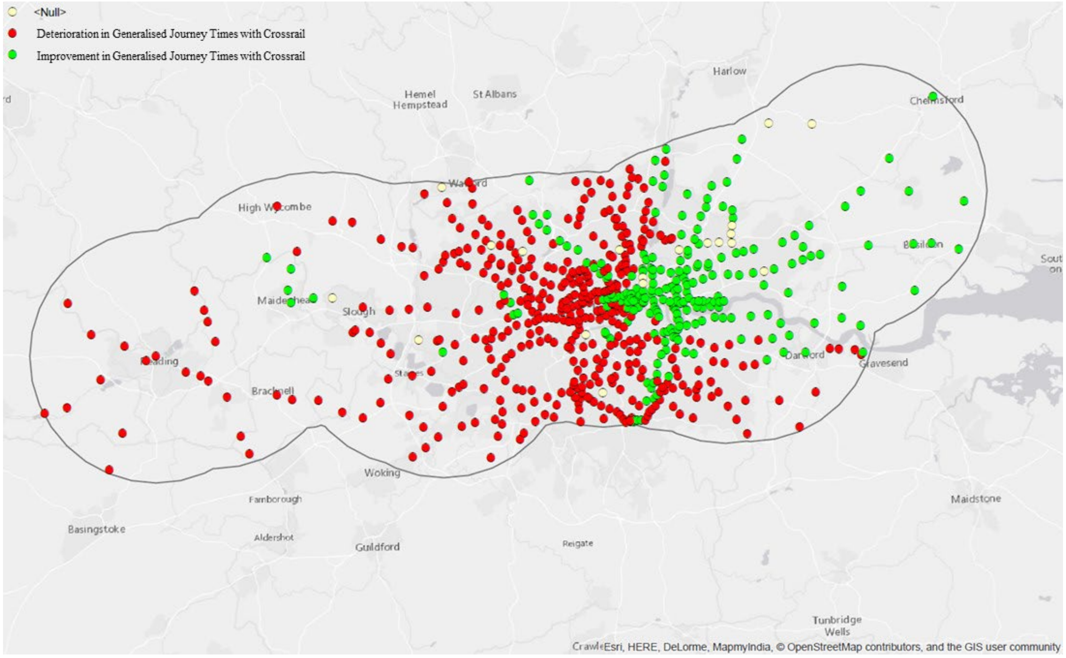


Figure C11: Generalised Journey Time changes from Burnham to other stations due to Crossrail



With the caveats described above, we estimated pre-opening impacts (in section 2.9) using changes in accessibility score – based on either normal Journey Times or Generalised Journey Times - as a treatment variable to take interconnectivity into account and have a more accurate representation of the relative impact of the Elizabeth line on the network. It also enabled us to conduct a similarity analysis for the counterfactual considering accessibility without the Elizabeth line (identified as the measured rail accessibility in 2007).