The effect of tram expansion on housing prices in Greater Manchester (Case Study)

Introduction

This paper examines the impact of proximity to tram infrastructure on housing prices in Greater Manchester, utilizing a transaction-level dataset comprising over 25,000 geocoded housing property sales from 1995 to 2024. To achieve this, the research employs a staggered difference-in-differences design with an event study specification. This approach compares price changes for houses located within 1,000 meters of new tram stops (treated group) against those situated between 1,000 and 1,500 meters away (control group), examining differences over a ten-year window surrounding each tram stop opening.

Project Scale: 6-month project

Purpose: Master's thesis

Primary Stakeholder: Department of Economics, University of Barcelona

Tools Used:

- **RStudio** with geosphere, dplyr, and lubridate for geocoding and dataset construction.
- **Geospatial treatment assignment** based on 1,000–1,500m proximity to tram stops.
- **Econometric modeling in** fixest for a dynamic event study DiD with ring and year fixed effects.

Data Collection & Cleaning

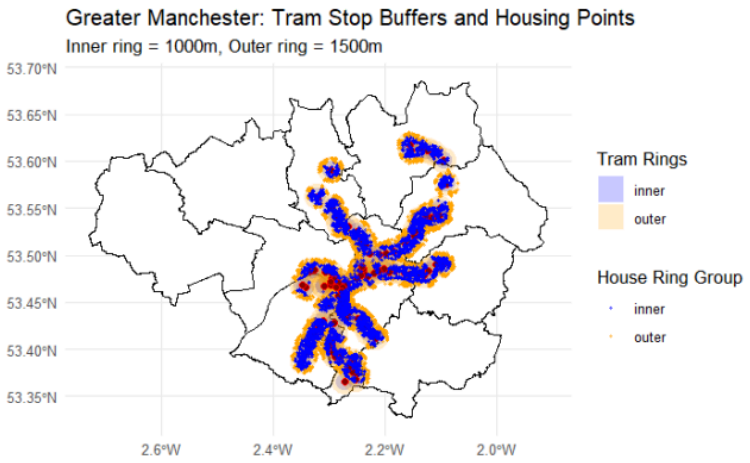
To build the dataset, I combined two main sources: housing transaction data from the UK Land Registry and tram stop data from Transport for Greater Manchester. The housing data initially contained around 100,000 transactions between 1995 and 2024, with details on address, sale price, property type, and ownership. After extensive cleaning and standardization of addresses, I geocoded the dataset to obtain precise latitude and longitude for each property. This step reduced the sample to about 80,000 usable transactions, as some addresses could not be matched.

Next, I merged this dataset with tram stop locations and manually assigned the opening date for each stop. By linking these datasets, I ensured that each property could be matched to its nearest tram stop in both space and time. After applying distance buffers and excluding cases with low sample sizes or overlapping catchments, the final analysis dataset contained around 30,000 transactions, of which approximately 6,000 fell within the ±5-year event windows used in the regressions

Geocoding & Treatment Assignment

Using the geosphere, dplyr, and lubridate packages in R, I calculated the Haversine distance between each property and its nearest tram stop. Properties within 1,000 meters of a stop were classified as treated, while those between 1,000 and 1,500 meters formed the control group. This distance-based approach ensured a credible quasi-experiment and avoided misclassification that often arises when treatment is defined using administrative boundaries.

A key decision was to discard tram stops that did not have enough sales in both inner and outer rings or that overlapped with nearby stops. This reduced the number of eligible stops to 83 but strengthened the reliability of the comparisons.

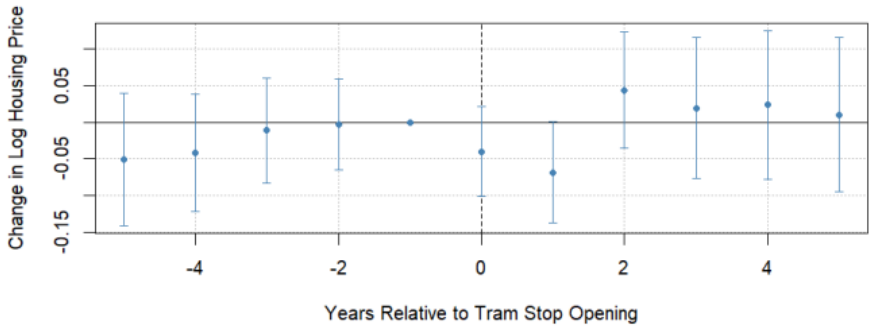


Econometric Analysis

To estimate the impact of tram access on housing prices, I applied a staggered Difference-in-Differences (DiD) design with an event study specification, using the fixest package in R. The model included both ring fixed effects and year fixed effects, with clustered standard errors at the ring level. This approach allowed me to test the parallel trends assumption directly and to trace how housing prices evolved each year before and after a tram stop opening.

The choice of an event study design was deliberate. A simple before-and-after comparison could have overstated the effect, but the event study allowed me to check for pre-treatment differences and to capture any dynamic effects around tram openings.

Figure 2: Event Study with Property-Level controls



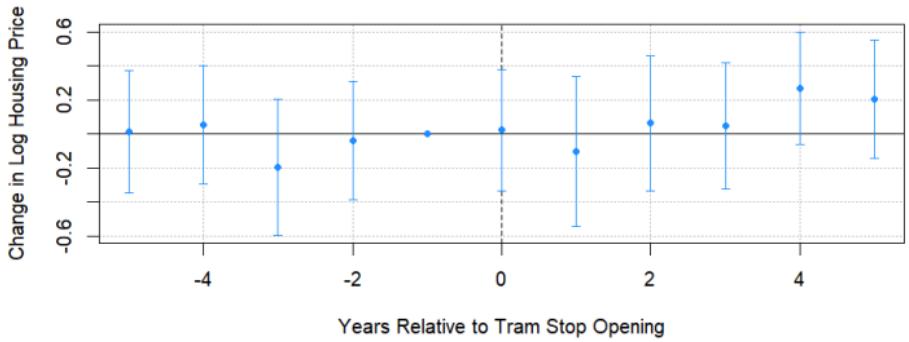
Robustness Checks & Heterogeneity

To ensure the results were not dependent on one specification, I carried out a series of robustness checks:

- Expanded the control group to 1,000–2,000 meters to test sensitivity to buffer choice.
- Added property-level controls such as type, new build status, and ownership.
- Split the sample into flats vs houses to explore heterogeneity by property type.
- Applied the Callaway & Sant'Anna (2021) estimator, which uses "not-yet-treated" properties as controls.

These checks confirmed the overall finding: tram stop openings did not lead to strong or sustained effects on nearby housing prices. The only consistent pattern was a small, short-term dip one year after opening, which did not persist.

Figure 5: Event Study: Flats Only



Challenges & Key Decisions

Several challenges shaped the project. First, geocoding led to significant data loss (from nearly 100,000 transactions down to about 6,000 usable cases). However, this data loss was necessary to maintain spatial precision. Second, I had to make defensible choices about treatment definitions. I used 1,000 meters for the treated group because it aligns with average walking distances to transit, but I tested wider buffers to check robustness. Finally, I balanced methodological rigor with data limitations: while the Callaway & Sant'Anna estimator provided a more advanced robustness check, I relied on the standard event study as my main specification due to sample constraints.

These decisions highlight the trade-offs between data availability, methodological precision, and interpretability.

Findings & Conclusion

The analysis showed no strong or lasting effects of tram stop openings on nearby housing prices. Apart from a small dip one year after opening, price changes were statistically insignificant across all specifications. Robustness checks confirmed this pattern, suggesting that in Greater Manchester's polycentric housing market, tram access alone does not systematically raise property values.

The most challenging part of the project was data cleaning and geocoding, which reduced the dataset from ~100,000 to ~6,000 usable transactions. This forced me to balance precision with statistical power. Another key decision was choosing a 1,000m buffer for treatment while testing alternatives for robustness.

From this project, I learned the importance of prioritizing data quality and designing transparent methods that can withstand robustness checks. A future iteration would include rental data and neighborhood characteristics to capture more nuanced effects.