Who are the Biggest “Winners”? The Effect of a New Metro Line Investment on Amsterdam Real Estate Market

By Juandra Alandi Irdham | 2701431

Supervised by Prof. Jos van Ommeren

This paper studies the consequences of a new transportation infrastructure project on the housing market. Between 1968 and 2018, the Netherlands’ government engaged in establishing a mass rapid transit system, called Metro Line 52 or the North/South Line (de Noord/Zuidlijn), which runs between the South and North parts of Amsterdam. The hedonic estimates generated from panel linear estimation strategies suggests that the overall “winners” from the metro establishment are the houses located within a 537-meter network distance from the nearest Metro Line 52 stations with an increase in house price by approximately 3.18% relative to the controlled areas. In addition, this paper also adopts a tree-based machine learning approach suggesting that the biggest “winners” from this investment goes to the properties with sizes of lower than 67 m^2 and located more than 2.2 km far away from the city centre.

The questions of whether investments in transportation infrastructure can promote an improvement in quality of living and urban development are the essential concerns for urban, transport, and real estate economists. These investments aim to improve accessibility and reduce traffic congestion in the areas where the infrastructure is placed, as some people’s commuting habits are strongly determined by their access to the nearest public transportation hub, and the others rely heavily on road traffic and automobiles usage. Therefore, transportation infrastructure can promote an economic gain which can be captured by the capitalization of housing prices located close to the affected areas. However, the transportation investment may also reduce housing value due to the negative externalities caused by increased noise, pollution, or high crimes. Understanding the positive and negative effects of transportation infrastructure development can provide a fruitful insight not only for the evaluation of the project but also for social cost-benefit analysis as well as designing and evaluating regional/urban/transportation policies and real estate investment portfolio.

Over the past few decades, the rapid transit system (RTS), also known as metro, subway, trams or heavy/light rail, has been commonly invested in major urban areas worldwide to deal with their urban challenges. An RTS can mitigate congestion problems and improve accessibility and connectivity to the city centre; thus, the city grows. Since the effects of transportation infrastructure development in a specific area are local by nature, it is expected that there are variations in values of houses between areas that are within and beyond the coverage area of the accessibility effect.

A large body of literature investigates the effect of rapid transit system (RTS) on housing price (see Debrezion et al., 2007; Higgins & Kanaroglou, 2016; Mohammad et al., 2013 for meta-analyses). However, these studies are prone to endogeneity problems. Some problems are related to measurement error and data accuracy, especially in distance measures and housing prices. However, measurement error in housing prices as the dependent variable in the hedonic studies poses no threat to the consistency of the standard hedonic regression. In contrast, the distance variable, mainly used as the primary independent variable in the hedonic regression, is likely to produce biased estimates due to measurement error. Moreover, some of these studies also encounter selection bias problems when determining the treated and untreated group in a quasi-experiment design due to some possible confounders that influence the treatment assignment and the housing price.

This paper studies the consequence of the investment in Amsterdam's newest metro line, also known as Metro Line 52 or the North/South Line (*de Noord/Zuidlijn)*, on the Amsterdam housing market. Using the geographical datasets from the municipality of Amsterdam, I linked the data of pedestrian and cycle route data and other locational point data with residence-level data points of Amsterdam housing prices from the most prominent real estate brokerage association in the Netherlands, i.e., the Dutch Association of Real Estate Brokers and Experts (NVM). The exposure of using pedestrian and cycle route data in the analysis is that I can implement a more realistic and accurate network analysis compared to the road or street data (see Diao, Leonard, et al., 2017 for an example) and correct the measurement error that is usually occurred in the Euclidean studies (see Baum-Snow & Kahn, 2000; Dubé et al., 2011). This paper also uses spatial innovation proposed by Linden & Rockoff (2008) and extensively used in Diao, Leonard, et al. (2017). Those studies use the local polynomial regression approach to identify the treatment zone of an event on housing prices and eventually use it as the basis of their treatment group assignment. Using 537-meter total-travel-cost network distance as the treatment zone based on local polynomial regression calculation, I estimate the average treatment effect of a metro line investment on Amsterdam housing prices using a linear hedonic pricing regression with a difference-in-differences approach controlling for spatial and time fixed effects. I find that the opening of the metro line 52 increases the value of houses in the treated areas located within a 537-meter total-travel-cost network distance from the nearest metro line 52 stations by approximately 3.18%, relative to other houses in the untreated areas.

In recent years, many policymakers, practitioners, and researchers have become increasingly interested in estimating a more personalized treatment effect and answer the question of who gains the most from treatment. Knowing for whom the treatment works best can provide a higher intuition to understand the true mechanism behind the treatment better. Building on a prominent paper in estimating heterogeneous treatment effect (see Athey & Imbens, 2016), I apply the causal tree approach that is based on a classification and regression trees (CART) method (see Breiman et al., 1983) to estimate the conditional average treatment effect (CATE) of the opening of the metro line on Amsterdam housing prices. This approach can produce clusters of treatment effects for the treated and untreated groups while allowing the effects to vary over their housing characteristics.

This research focuses on several questions. I first ask how Amsterdam residents value the change in accessibility caused by the Metro Line 52. This question can be addressed by estimating the elasticity of housing prices concerning the change in accessibility. Second, I ask whether there is an anticipation effect in the affected areas due to the announcement of the opening of metro line 52. The third question is about measuring how large the affected areas are. This question can be addressed by using the graphical result from local polynomial regression to find the distance where the reflection of a causal impact of the infrastructure development can be justified. The next question is about the commuting behaviour of Amsterdam residents. Is using Amsterdam pedestrian and cycle paths to go to the nearest metro station more realistic and accurate? Identifying whether using pedestrian and cycle path networks can correct measurement error that is primarily occurred in the Euclidean distance or not can answer this question. The next question is about how strong and dominant the effects are and how they can vary between different housing characteristics and among each leaf/subgroup of houses. Answering this question can help to answer the primary question in this paper which is to find who the biggest winners from the investment of the new metro line in Amsterdam are. Finally, I estimate the average treatment effects and then estimate its heterogeneous treatment effects to determine whether it has a positive or negative effect on housing price on average and on each subgroup of houses.

The remainder of this paper is organized as follows. In section I, I will discuss the literature review. Section II describes the quasi-experiment studied in this paper, i.e., the opening of Metro line 52 in Amsterdam on 22nd July 2018. Section III presents the data and empirical design. Section IV reports the empirical results, including a variety of robust and sensitivity analyses. Section V concludes this paper.

I. Literature Review

A. *Theoretical Framework in the Literature on the Economic Effects of Transportation System*

The effect of transportation system development has been extensively studied for a very long time in regional science, urban planning, and even economics, where the discussion focused more on the dynamic interplay between transportation improvements and urban land use. The study of the effect of transportation system development on urban land use involves the key urban economics concept called bid rent. This concept constitutes the maximum amount that a household or a firm is willing to pay, concerning its utility or profit levels, for a land.

Thünen (1875) suggests that the transportation savings determine the bid rent, and the land rents vary and depend on how close the land is to a central marketplace. Alonso (1964) argues that land rent does not vary in any way, except for areas close to the central business district (CBD) or employment centre where people who require shorter commutes can enjoy a cheaper commuting cost. However, to obtain such a benefit, people have to be willing to pay a premium which can be translated into an additional rent; thus the rent increases as people live closer to CBD. In contrast, other people, who require more lands instead of shorter commutes, locate themselves away further from the centre. Another extended concept has been presented where not only land rents changes, but also house price, building heights, and population density changes with respect to distance from the CBD (Muth, 1969; Richardson & Mills, 1973).

It is argued that the demand for a specific location increases when the location becomes more attractive due to some attractive urban amenities (Brueckner et al., 1999; Glaeser et al., 2001). In our case, a CBD's attractiveness leads to a higher demand for households to locate themselves close to the CBD. It also increases housing prices near CBD. However, the investment in transportation infrastructure, such as RTS, reduces this demand to some degree as it attracts households to locate themselves close to stations. Households living close to a station take the benefits of travel time and cost-saving from the investment. It also means that property or land values near CBD will decrease, whereas the house or land values near stations will increase. Some empirical studies have found that a new transportation system could flatten the bid rent curve of houses close to CBD and affect urban and even rural housing prices (see McMillen & McDonald, 2004; Zheng & Kahn, 2008). Higher accessibility (lower transportation costs) caused by the new transportation system would increase land value close to the station via the transportation system's capitalization effect.

The empirical estimation of RTS's effect on real estate price requires a statistical model to control other possible characteristics associated with real estate prices. This is where the hedonic price method proposed by Rosen (1974) comes in. This model offers an easy way to capture the effect by controlling attributes (e.g., property size, number of bathrooms, and distance to amenities) of the property observations. However, the standard version of this approach has some limitations, and how to deal with such limitations will be discussed further in the identification strategy section of this paper.

B. *Empirical Evidence in the Literature of the New Rapid Transit System Effect on Housing Market*

Academic literature documents an extensive study focusing on the effect of a new transportation system on property or land values[1]. This extensive literature produces a high degree of variability in

[1] Despite the study of the transportation improvement effect on housing market, there is a large literature studying the similar literature but focusing on other outcomes instead of land or property value. Those studies showed evidences that transportation improvement can affect population (see Deng et al., 2019; Zhao et al., 2017), economic growth (see Ahlfeldt & Feddersen, 2018; Meng et al., 2018), gentrification (see Jackson & Buckman, 2020; Lin & Xie, 2020),

capitalized value associated with the transportation system establishment. A large number of studies find a positive capitalization effect of accessibility to the nearest RTS on housing price (see Gibbons & Machin, 2005; Hess & Almeida, 2007; Higgins, 2019; Pilgram & West, 2018). Some studies find insignificant effects (see Gatzlaff & Smith, 1993; Landis et al., 1995). The other studies find a negative effect of transportation investment on housing prices (see Bowes & Ihlanfeldt, 2001; Diao et al., 2016).

Several meta-analyses have been done to date. Debrezion et al. (2007) conducted a meta-analysis to study this large variability. They collect 57 estimation results from previous studies and find that the residential property value is 2.3% higher than commercial property value every 250 m closer to a station. The variability can also be found in the type of transportation facility. They also find that commuter railway station has a consistently higher impact on property value than others. Mohammad et al. (2013) also conducted a similar meta-analysis based on 102 estimation results. They find that the average property value changes tend to be higher at distances between 500 and 805 m compared to distances more than half a mile away from a station. Lastly, Higgins & Kanaroglou (2016) perform a more extensive study involving 130 analyses (60 studies) and argue that the source of variability in this literature is the lack of proper empirical design, leading to a failure in capturing the overall capitalization effect and measuring other potential value.

Another essential part of the literature is that the change in land use usually occurs before the transportation system project is built and after the announcement of the project is announced, which is known as the anticipation effect. Arnott & McMillen (2007) propose a concept in which the changes of bid rent will exist after the transportation system operates, but the market values of houses or land will change before "opening day" (the first time the project is announced). Several empirical studies have been conducted to capture the anticipation effects (see Devaux et al., 2017; Diao, Fan, et al., 2017; Dubé et al., 2018; Gatzlaff & Smith, 1993; McMillen & McDonald, 2004). The idea behind the anticipation concept is that the housing unit is viewed as an asset class that generates income from the rent charged to the tenants. In this perspective, the housing price will be based on the expected present value of the total accumulation of the future rents. Thus, the housing price will increase when the expected rental price increases due to increased accessibility. I will discuss the anticipation effect of the Metro Line 52 project in the result section.

C. *The Heterogeneous Treatment Effect of an Opening of RTS on Housing Price*

In the study of the effects of transportation infrastructure development on housing prices, it is said that the effects may vary over different population groups. Levkovich et al. (2016) suggest that the effects of transportation infrastructure development on housing price vary between dwellings depending on their location, municipal affiliation, and distance to the site of the developed projects, and unobserved heterogeneity between properties. Duncan (2008) concludes that the capitalization effect is three times larger for multifamily than single-family housing. Furthermore, Li (2020) shows that the capitalization effects vary, depending on whether people live in a congested area or not. Kunimi & Seya (2021) suggest that it is fundamental to consider "who" benefits from the infrastructure project and where the benefits are located. Recently, there is extensive ongoing research in estimating heterogeneity in treatment effects using the machine learning method which will be discussed below.

The concepts of machine learning and causal inference are arguably originated from dissatisfaction with linear regression. Both were developed simultaneously, and both communities have shared and exchanged their views together, as they realized that the frameworks formulated by both fields could

employment (see Åslund et al., 2017; Jin & Kim, 2018; Mayer & Trevien, 2017). However, some empirical studies also find insignificant effect on the same indicator, e.g. population (see Mayer & Trevien, 2017), economic growth (see Garcia-Mila et al., 1996), and employment (see Jiwattanakulpaisarn et al., 2009).

provide a solution to each other field. One of the most intensively discussed issues regarding causal inference in machine learning these days is estimating the heterogeneous effect using machine learning, and the sample splitting approach. In most cases, the treatment effects vary across observations, meaning that the effect would be generalized to a population with different characteristics. Machine learning tools are deemed to be ideal for capturing the heterogeneity in treatment effects, assuming the assumption of confoundedness is held (Athey & Imbens, 2019).

Previous attempts try to solve this problem by proposing the regression-tree-based method, called causal tree (Athey & Imbens, 2016) and causal forest (Wager & Athey, 2018), which have become a popular innovation in recent years. Several major empirical studies have been carried out in detecting the heterogeneity in treatment effect using these approaches (see Athey et al., 2018, 2019; Athey & Imbens, 2016; Chin et al., 2020; Davis & Heller, 2020; Hussam et al., 2020; Miller, 2020).

One of the recent studies explores the effect of a new RTS on housing price using a machine learning approach. Chin et al. (2020) adopt Athey & Imbens (2016)'s causal tree with a difference-in-differences empirical design to estimate the conditional average treatment effects (CATE) of a new RTS across several dimensions of properties and its characteristics. Thus, this CATE estimation can break down the average treatment effects that correlate with the accessibility to the new RTS. They document a significant heterogeneity on observable dimensions in their study. They found that most apartment types in the treatment group experienced appreciation in property value, while several property types in the control group declined in value. The result also suggests that an apartment with three rooms, two bathrooms, construction age lower than five years, and located within 1 km from the nearest RTS station constitutes as the top 25% leaf that benefits the most from the treatment.

This paper applies approaches that are close to the work of Chin et al. (2020). The distinguishing features of this study can be seen from the use of network distance instead of Euclidean distance and the use of local polynomial regression as a method to find a more accurate distance for the treatment assignment instead of setting a discrete 1 km distance from the station as a basis of treatment assignment. Lastly, the major distinction is that this study uses the interactive terms of the treated areas and post event dummies as the binary treatment variable in the causal tree setting, while Chin et al. (2020) uses only treated areas as the binary treatment variable.

II. Quasi-experiment: Amsterdam Metro Line 52 (The North/South Line)[2]

This paper focuses on the city of Amsterdam, which is the capital and most populous city in the Netherlands and is located in the province of North Holland (see **Figure 1**). This urban area covers 21,949 hectares and accommodates almost 863,000 people[3]. The city provides a high level of pedestrian and cycle infrastructure; metro, tram, and bus system; ferry routes and canals; and city road. Amsterdam is also known as one of the most bicycle-friendly cities worldwide, as 48% of all home-to-work trips uses bicycle in Amsterdam[4]. Furthermore, CBS noted that Dutch people usually take their bikes or walk for trips within a five-kilometre distance[5]. The city has five metro lines, 14 tram lines, 43 bus lines, and ten ferry lines operated to date[6].

Recognizing the urban challenges such as the booming population and number of vehicles occurring in Amsterdam in the early 20th century, many feasibility studies had been published and sent to the Municipal Council of Amsterdam (Gemeente Amsterdam) in response to the possibility of developing

---

2 The terms of "the North/South Line" and "Metro Line 52" are used interchangeably in this paper.

3 Source: Statistics Netherlands (CBS) (2019)

4 Available at https://www.government.nl/documents/reports/2018/04/01/cycling-facts-2018

5 Source: Statistics Netherlands (CBS) (2016)

6 Source: Gemeente Vervoerbedrijf (2021)

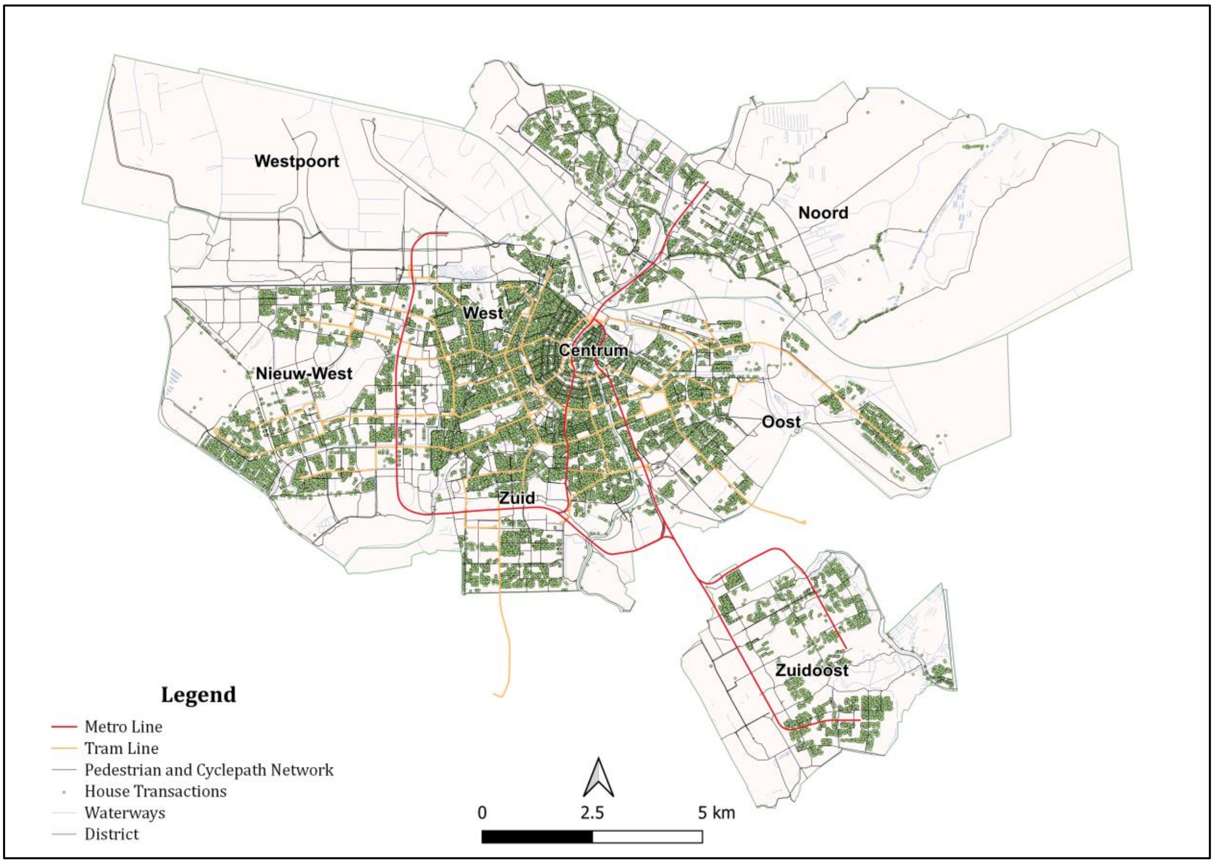

FIGURE 1. DETAILED MAP OF THE CITY OF AMSTERDAM IN 2020

public transport in Amsterdam. In 1968, Gemeente Amsterdam agreed with a strategic transport plan called the Plan Stadsspoor (Metroplan), which aims to create a metro system connecting all neighbourhood areas in Amsterdam. The original Metroplan suggested four different routes, comprising two lines running from the eastern part to the western part of Amsterdam, one circle line, and the last one, the North/South Line, has to be constructed in Amsterdam. However, some prolonged controversies related to the planning and construction of the metro system, especially for the North/South Line project (Mottee et al., 2020), will be further discussed below.

The initial phase of the metro system was started by constructing the East Line, which links the southeast part of Amsterdam (*Zuidoost*) with the city centre (*Centrum*) in 1970. Right after the line crosses the border of *Zuidoost,* the line is split into two branches. Those are the branches of Gaasperplas and Gein and later on are called Metro Line 53 and Metro Line 54, respectively. In 1975, the construction of the East Line led to community protests as it damaged the existing urban form. This also resulted in all projects being stalled and the loss of trust by the locals in the government. However, the continuation of the Metro Plan was reconsidered after a new and less damaging construction method was proposed to the Gemeente Amsterdam (van den Ende & van Marrewijk, 2019; Van Lohuizen, 1989). Several feasibility studies for the North/South Line project had been conducted. After considering the studies and the referendum in 1996, Gemeente Amsterdam decided to proceed with the plan and asked for funding from the Dutch National Government. The North/South Line proposal was considered promising and was approved on June 21st, 2000, by the Dutch National Government.

In October 2002, Gemeente Amsterdam made a final decision and decided to start constructing the Metro Line 52 in April 2003. **Figure 2** provides the detailed chronology of the journey of the North/South Line establishment. The construction of the North/South Line project is deemed to be controversial due to several geotechnical incidents, administrative problems, constantly increasing

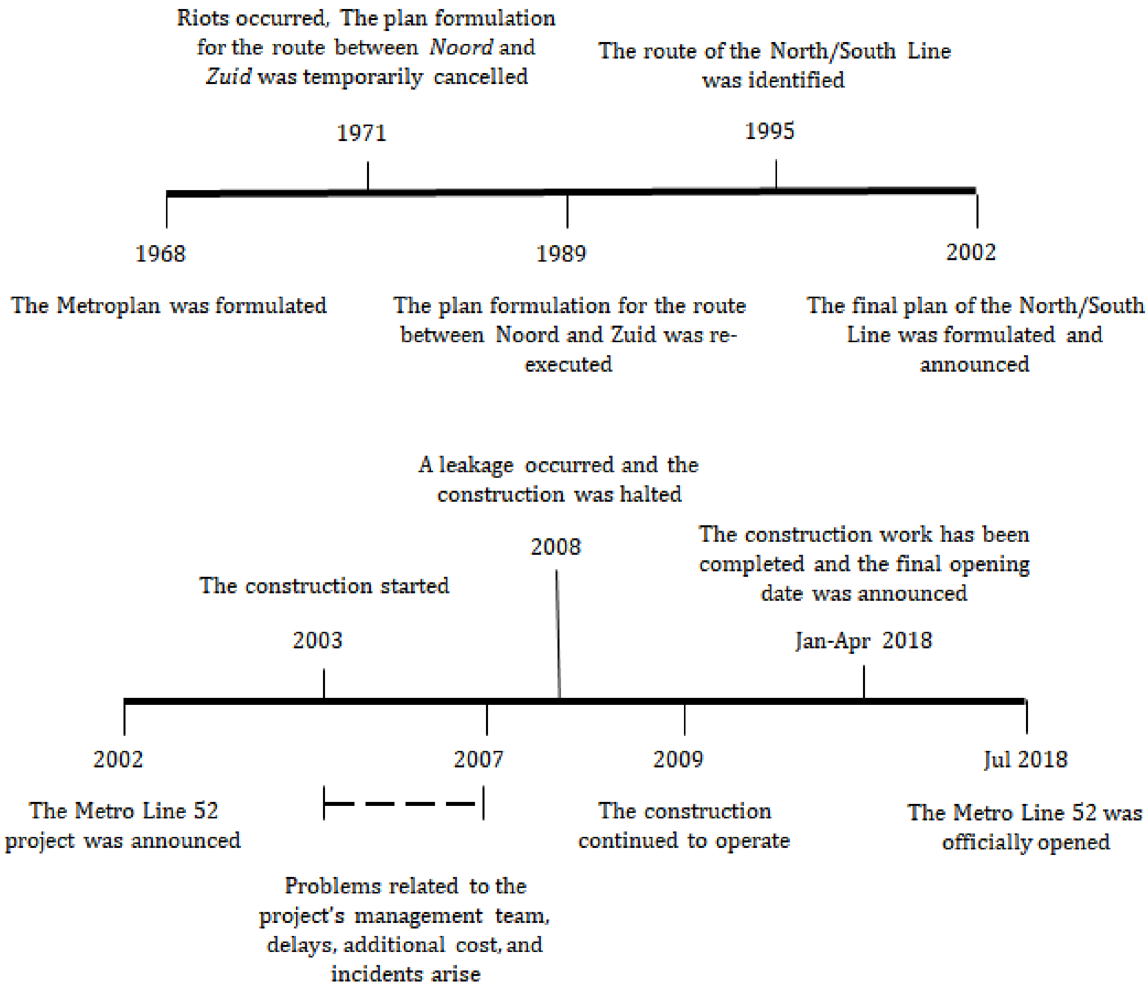

FIGURE 2. TIMELINE OF KEY EVENTS OF THE METRO LINE 52 PROJECT

estimated cost, several delays for the opening date, and severe loss of trust in the Gemeente Amsterdam (Mottee et al., 2020). However, Gemeente Amsterdam claimed that they have succeeded in completing the North/South Line. After 40 years of planning and 16 years of construction, the North/South Line project finally opened on July 22nd, 2018. The project was initially scheduled to be complete by 2007 at the cost of about 1.4 billion euros. In contrast, it was completed 11 years after the initial date and at the cost of approximately 3.1 billion euros.

The North/South Line connects the northern part of the Amsterdam (Noord) to the *Centrum* and the central business area in the southern part of Amsterdam (*Zuid*). The Metro Line 52 is a 9.7 kilometres metro link where 7.1 kilometres of the total length were built underground. This line is equipped with six new stations, including Europaplein, De Pijp, Vijzelgracht, Rokin, Noorderpark, Noord; and two existing stations, including the stations of Amsterdam Zuid and Amsterdam Central.

It has been argued that one of the challenges in studying the impact of transportation is that it requires a quasi-experiment setup causing enormous changes in accessibility that are mainly hard to be found (Banister & Berechman, 2001; Fernald, 1999). I argue that the Metro Line 52 project produce significant changes in accessibility as the new line's opening was accompanied by significant changes to the existing bus and tram systems to promote and support this new metro operation and remove any inefficient line. Later in this paper, I also provide graphical evidence showing that the opening of the Metro Line 52 can be a good case for quasi-experiment in this literature.

Empirical studies investigating the spatial impact of a transportation development project is prone to face simultaneity problem which makes it hard to investigate the flow of causality (Hoogendoorn et al., 2019). There has been argued that a transportation development project is placed to promote the economic growth of some particular areas; thus, the endogeneity problem arises. However, I argue that the Metro Line 52 establishment event can reduce the simultaneity bias in the study, as the project aims to reduce private vehicle usage and travel times, improve reliability and accessibility, and promote liveability [7] while not merely about improving the economic growth. Concerning dealing with endogeneity, this study assumes the moment of the opening of Metro Line 52 to be random.

III. Data and Identification Strategy

A. Data Sources

This paper uses micro-data on Amsterdam's housing prices provided by the Dutch cooperative association of real estate brokerage and appraisers in Netherlands (NVM/*de Nederlandse coöperatieve Vereniging van Makelaars en taxateurs*). This covers more than 75% of entire Dutch houses sold by NVM brokers[8]. The dataset covers 18 years of data collection from January 2002 until December 2020. Each property observation is geocoded at the precise residence level. The dataset contains detailed records of the transaction date, transaction price, six-digit postal code, and transactional and structural housing characteristics of the house transactions, such as the property size, age, sales condition, and size (see Appendix A for an overview of all variables used in this analysis). Table 1 presents the descriptive statistics which is divided into three primary columns, i.e., total sample, treatment group, and controlled group.

Since the goal of this paper is to capture the effects of Metro Line 52 on house prices and to prevent possible boundary discontinuity problems, I restrict the sample to the houses that are located within the radius of a 1,000-meter service area that is derived from the network analysis, from each Metro Line 52 station. **Figure 3** represents the locations of the sample housing transactions and the Metro Line 52 stations. I perform sensitivity analysis by expanding the radius to a 1,500-meter service area from each Metro Line 52 station. After performing several data selection procedures, it retains 23,668 observations in 2,161 6-digits postal codes. Appendix B describes another sample selection process that has been done in this study. The data have a panel structure and is surveyed daily, but a quarterly basis is used for the analysis.

Apart from the house location and its characteristics, I also link other locational characteristics such as administrative borders, road-related networks, locations of tram and bus stations, and all metro systems, including Metro Line 52, that Gemeente Amsterdam provides. Using QGIS, I measure the distance between the location of each house transaction and the nearest Metro Line 52 hubs and use this distance measure as the core variable in this analysis. I use total-travel-cost network distance[9] measured using QGIS Network Analysis Toolbox (QNEAT3), and it will be discussed further in the next section.

B. Total-Travel-Cost Network Analysis for Distance Measures

The model designed by Alonso (1964), Muth (1969), and Richardson & Mills (1973) regarding the resource allocation for a city, the so-called monocentric model, assumes that the land of a city has a plain and smooth terrain. However, I argue that ignoring spatial and topography obstacles can lead to a

---

[7] Source: Gemeente Amsterdam (2018) https://noordzuidlijn.wijnemenjemee.nl/lijnennet/over/

[8] Source: de Nederlandse Coöperatieve Vereniging van Makelaars (2021)

[9] The term "total-travel-cost network distance" is used interchangeably with the term "network distance" in this paper

Table 1.
Descriptive Statistics

	Total Sample		Treatment Group Total-Travel-Cost Network Distance ≤ 537 m		Controlled Group Total-Travel-Cost Network Distance > 537 m	
Observation	**23,668**		**7,379**		**16,289**	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Sales price (EUR)	439,372.07	322,450.16	398,019.76	279,223.03	458,104.87	338,579.41
Sales price/sqm (EUR)	4,574.632	1,902.184	4,622.731	1,691.31	4,552.843	1,990.039
Log (Sales price)	12.809	.581	12.729	.544	12.846	.593
Log (Sales price/sqm)	8.36	.369	8.38	.34	8.351	.381
Network distance to Metro Line 52 station (m)	656.258	221.306	385.468	107.818	778.927	132.781
Log (Network distance to Metro Line 52 station)	6.412	.424	5.901	.361	6.643	.175
Size of property (m^2)	95.74	50.25	86.823	47.313	99.779	51.016
Log (size of property)	4.45	.463	4.349	.465	4.495	.456
Property types						
Apartment	.9	.301	.94	.238	.881	.323
Terraced	.08	.271	.047	.212	.094	.293
Semi detached	.017	.128	.01	.099	.02	.139
Detached	.004	.065	.003	.057	.005	.068
Parking	.07	.255	.084	.278	.064	.244
Number of bathrooms	.965	.553	.94	.527	.976	.564
Number of kitchens	.803	.496	.798	.489	.806	.499
Number of balconies	.475	.544	.483	.547	.471	.542
Number of roof terraces	.169	.39	.155	.377	.176	.396
Owned private office	0	.009	0	.016	0	0
Maintenance score of the Outside	.791	.118	.798	.123	.788	.115
Maintenance score of the Inside	.788	.157	.801	.154	.781	.157
Good maintenance	.887	.317	.9	.301	.881	.324
Number of insulation types	1.337	1.659	1.535	1.769	1.247	1.598
Central heating	.877	.329	.868	.339	.881	.324
Listed property	.101	.301	.063	.243	.118	.323
Monumental property	.093	.291	.071	.256	.103	.305
Auction sales	.001	.038	.002	.044	.001	.036
Leasehold	.391	.488	.27	.444	.446	.497
Property is partly rented	.003	.057	.004	.06	.003	.055
Period of construction						
Before 1906	.367	.482	.444	.497	.333	.471
Between 1906 and 1930	.302	.459	.27	.444	.317	.465
Between 1931 and 1944	.076	.265	.033	.178	.096	.295
Between 1945 and 1959	.027	.161	.018	.133	.03	.172
Between 1960 and 1970	.031	.174	.02	.138	.037	.188
Between 1971 and 1980	.028	.165	.029	.167	.028	.165
Between 1981 and 1990	.039	.194	.038	.191	.039	.195
Between 1991 and 2000	.053	.225	.061	.239	.05	.219
2000 and later	.075	.264	.088	.283	.07	.255
Network Distance to other nearest metro station (m)	1,405.946	819.344	1,453.668	665.044	1,384.328	879.554
Network distance to the nearest tram station (m)	508.117	780.713	367.584	595.474	571.779	843.814
Distance to Centrum	2069.819	1066.49	1939.114	887.401	2129.028	1133.46
Distance to CBD	2638.69	1676.823	2383.23	1363.471	2754.415	1788.992
Log (network distance to other nearest metro station)	7.1	.558	7.168	.509	7.069	.577
Log (network distance to the nearest tram station)	5.724	.857	5.489	.725	5.83	.89
Log (distance to Centrum)	7.457	.676	7.428	.603	7.47	.706
Log (distance to CBD)	7.684	.655	7.637	.549	7.705	.697

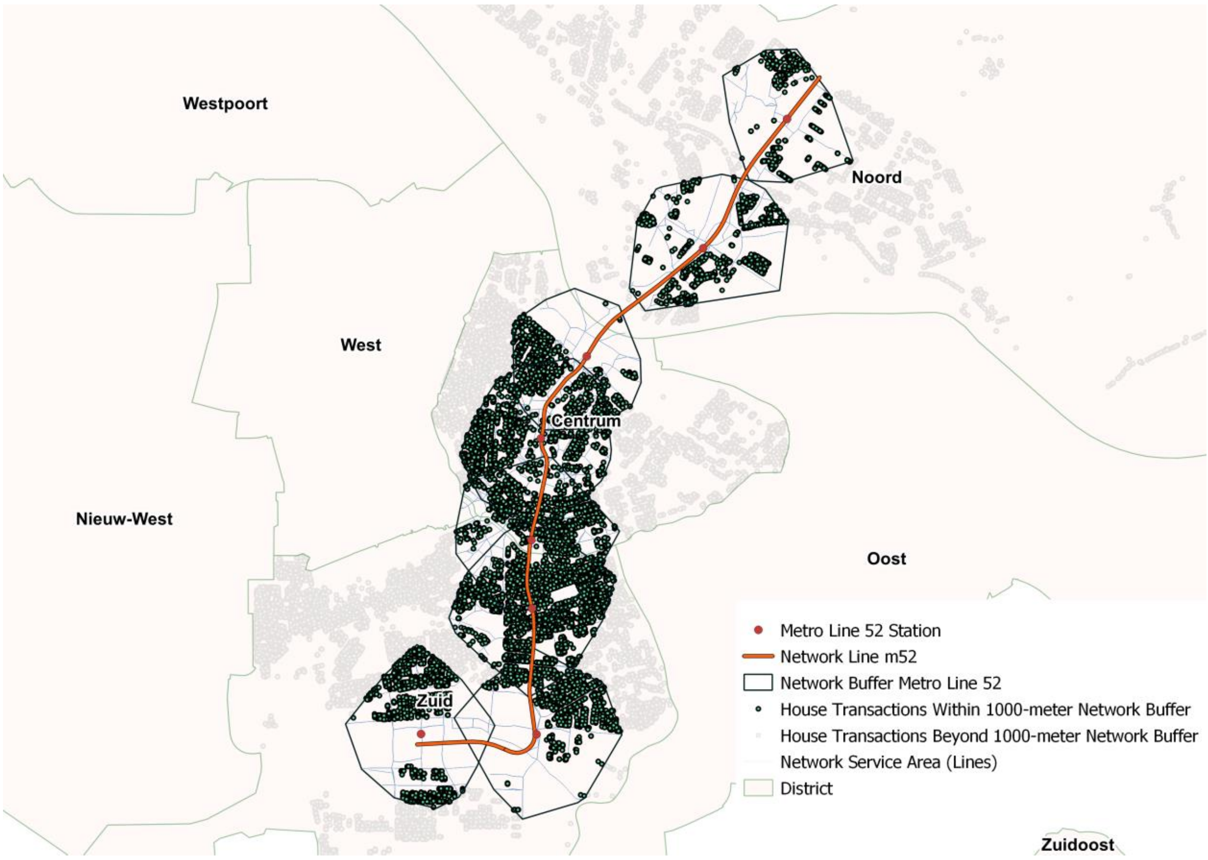

FIGURE 3. METRO LINE 52 AND HOUSE TRANSACTION WITHIN 1,000-METER NETWORK BUFFER

bias estimation in an analysis of the impact of infrastructure development. In practice, a government mainly uses zoning and land use planning to monitor and control the population density and the supply of land in a city. Furthermore, a zoning regulation is mainly formulated by considering the spatial and topography obstacles in a city, as Albert (2010) argues that spatial regulations are dependable according to the geography and topography of a city. Therefore, most Euclidean distance studies found in the literature of the infrastructure development impact are prone to produce bias results.

The use of Euclidean distance is widespread in this literature. The Euclidean distance can be an appropriate approach if the world is viewed as a two-dimension without considering topography obstacles. Euclidean distance seems to be an incorrect tool to measure an RTS station's distance to a house. RTS tends not to be constructed in a place where the surroundings are flat with straight topography lands. Instead, it is usually built in an area surrounded by natural obstacles such as rivers, inclines, ponds, trees; and infrastructure obstacles such as buildings, highways, and railroads.

A number of studies suggest using network analysis to handle the problem that arises from using the Euclidean approach as it considers geography and topography features (see Diao, Leonard, et al., 2017; Hess & Almeida, 2007; Higgins, 2019). However, the distances based on the length of a network are likely to produce a biased estimation. In practice, standard network analysis tends to face incomplete datasets, which cause imprecise measurement of network distance due to some unobserved networks that are not considered in the analysis. Spatial analysts tend to forcibly join the nodes of the house observations and the RTS stations to the existing network line; thus, any unobserved network between the house observation or the RTS station and the existing network line is not taken into account.

In this paper, I propose to use both the Euclidean and network analysis approaches to measure Euclidean distances from the start point to the existing network and from the existing network to the endpoint; and the walking distance in the network, respectively. This approach hereafter is referred to

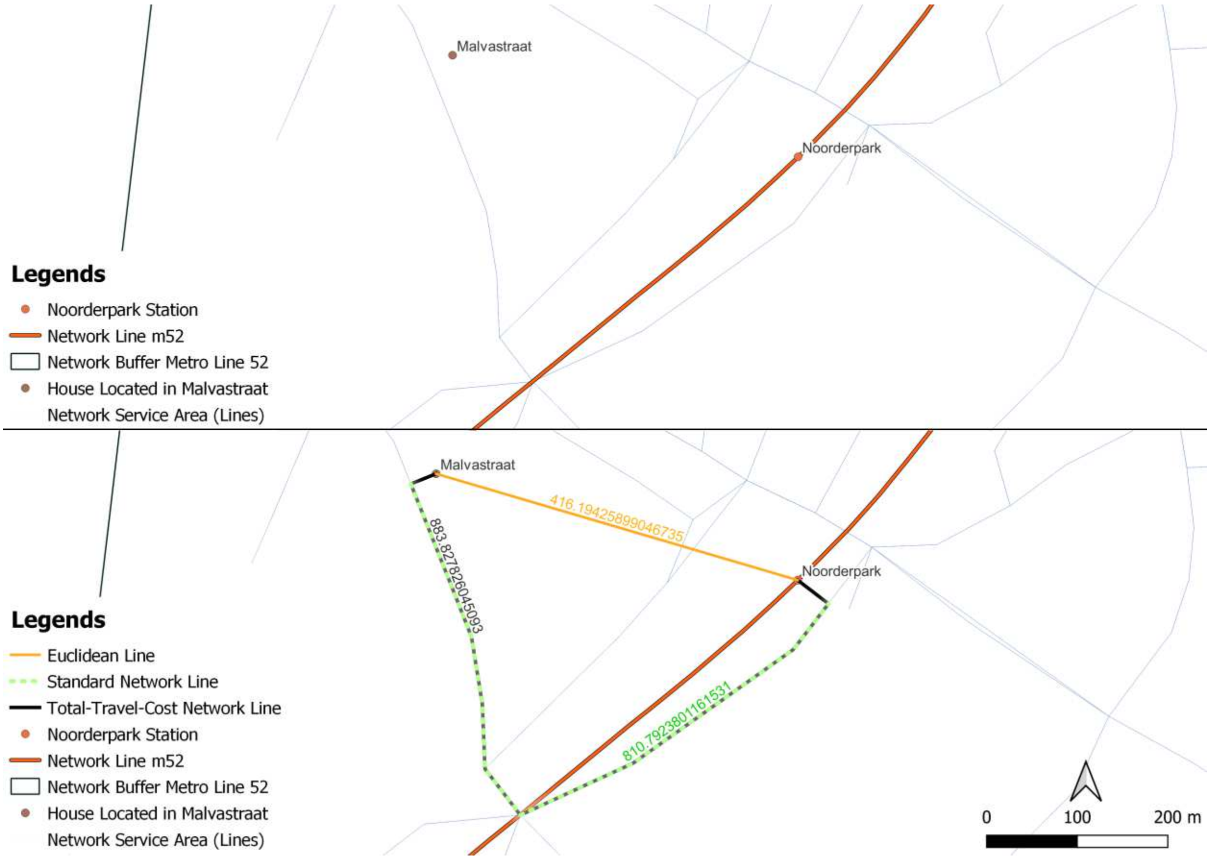

FIGURE 4. THE DIFFERENCE BETWEEN TOTAL-TRAVEL-COST NETWORK ANALYSIS, STANDARD NETWORK ANALYSIS, AND EUCLIDEAN APPROACH

as the total-travel-cost network analysis approach. **Figure 4** illustrates the difference between the total-travel-cost network analysis, the standard network analysis, and the Euclidean approaches. **Figure 4** shows that the house owner in the street of Malvastraat in Amsterdam requires to walk approximately 883 meters from the exit door of his house to the nearest metro line 52 station, which is the Noorderpark station. In contrast, the distance generated from standard network analysis is around 60 meters shorter than the total-travel cost network distance. The difference between both analyses is that the total-travel-cost network analysis considers the distance gap between the origin point and the network line that is usually unobserved. The total-travel-cost analysis uses a straight line towards the nearest existing network line to handle the unobserved line and correct the measurement error due to unobserved additional distance. Furthermore, the gap is much larger when the total-travel-cost network distance is compared with Euclidean distance. I will show that using total-travel-cost network distance is better as it can correct the measurement error commonly occurred in Euclidean and network analysis studies.

This paper uses pedestrian and cycleway networks data, as it represents a more behaviorally relevant proxy for accessibility, especially in Amsterdam. I overlay the pedestrian and cycleway networks layer onto the base layer of spatial data of house transactions and Metro Line 52 stations. Using the QNEAT3 feature in QGIS, I measure the shortest total-travel-cost distance for all sample homeowners to walk to their nearest Metro Line 52 stations. To the best of my knowledge, this study is probably the only one study in the literature of the effects of RTS on house price that uses total-travel-cost network analysis. For a robustness check, I also add the result from using Euclidean distance and standard-network distance in the main analysis.

C. Local Polynomial Regression for Treatment Assignment

In designing a quasi-experiment in this literature, it is commonly required to set the treatment areas that can vigorously represent the affected area derived from the change in accessibility. Some studies propose to use a 1 km distance from the RTS station to be the catchment area of the study (see Bartholomew & Ewing, 2011; Chin et al., 2020). A few other studies determine a discrete buffer zone of 400 meters from the RTS station to set the spatial boundaries of the accessibility effect, as it is argued that a 400-meter distance is an acceptable distance for people to walk (Untermann, 1984).

However, a number of studies suggest using the local polynomial regression[10], or so-called Fan regression (Fan, 1992; Fan & Gijbels, 1994), to perform the treatment assignment in a quasi-experiment project related to the housing market. Linden & Rockoff (2008) propose to use Fan regression in a criminal quasi-experiment study. After the sex offenders arrive at the neighbourhood, they find a local negative effect occurring on the house prices located close to the area where the sex offenders move in. They also find that the effect dissipates quickly when the house's location is getting further away from the location of the sex offenders' arrival. They also find no effect on the houses located at a certain distance from the location of the sex offenders' arrival. They provide graphical evidence using local polynomial regression that shows large differences in house prices before and after the sex offenders arrive. The large differences can be seen in the houses located within 0.1 miles and the houses located between 0.1 and 0.3 miles from the sex offenders' arrival location. Furthermore, Diao, Leonard, et al. (2017) apply a similar approach to investigate whether there is any differential non-linear treatment effect on houses located close to RTS stations and identify the distance boundary where the discontinuity occurred.

Building on a similar approach, I use local polynomial regression to identify the certain cut-off distance where large differences in house prices before and after the opening of the RTS and the discontinuity in housing price in such cut-off distance can be clearly seen. The basic concept behind this method is that it can be a convenient way to describe the disproportional price responses due to the change in accessibility in a simpler way. It is also argued that house prices that are close to an RTS station have a nature of non-linearities due to the housing supply that tends to be inelastic. Since the local polynomial regression can allow for non-linearities in house prices, this approach is deemed to be a more appropriate method to correct such a problem and for treatment assignment. Later on in this paper, I will show that the treatment group is assigned based on whether the location of a house is inside the 537-meter network-distance buffer or not.

Even though the approach is very similar to the works of Diao, Leonard, et al. (2017), the difference can be found in how I perform the parameter tuning, especially for tuning the bandwidth (λ). It is not clear how Diao, Leonard, et al. (2017) and Linden & Rockoff (2008) set the bandwidth. However, the bandwidth in this analysis is set based on the standard deviation of the network distance variable from the sample, which is 221.30 (see Table 1). Hastie et al. (2009) argue that when the Gaussian Kernel is assumed, the bandwidth (λ) should be equal to the value of standard deviation.

D. Graphical Evidence

If the housing services close to the RTS station are really attractive, we can expect that the prices of the houses close to the RTS stations should go up after the RTS opens for service. We can also expect that the closer the house to the station, the higher the price of the house. **Figure 5A** illustrates the price gradients of housing in Amsterdam using local polynomial regression with respect to the network

---

[10] Local polynomial regression is a nonparametric regression method and one of the kernel smoothing methods where it estimates a correlation between two variables to illustrate a smoothing curve.

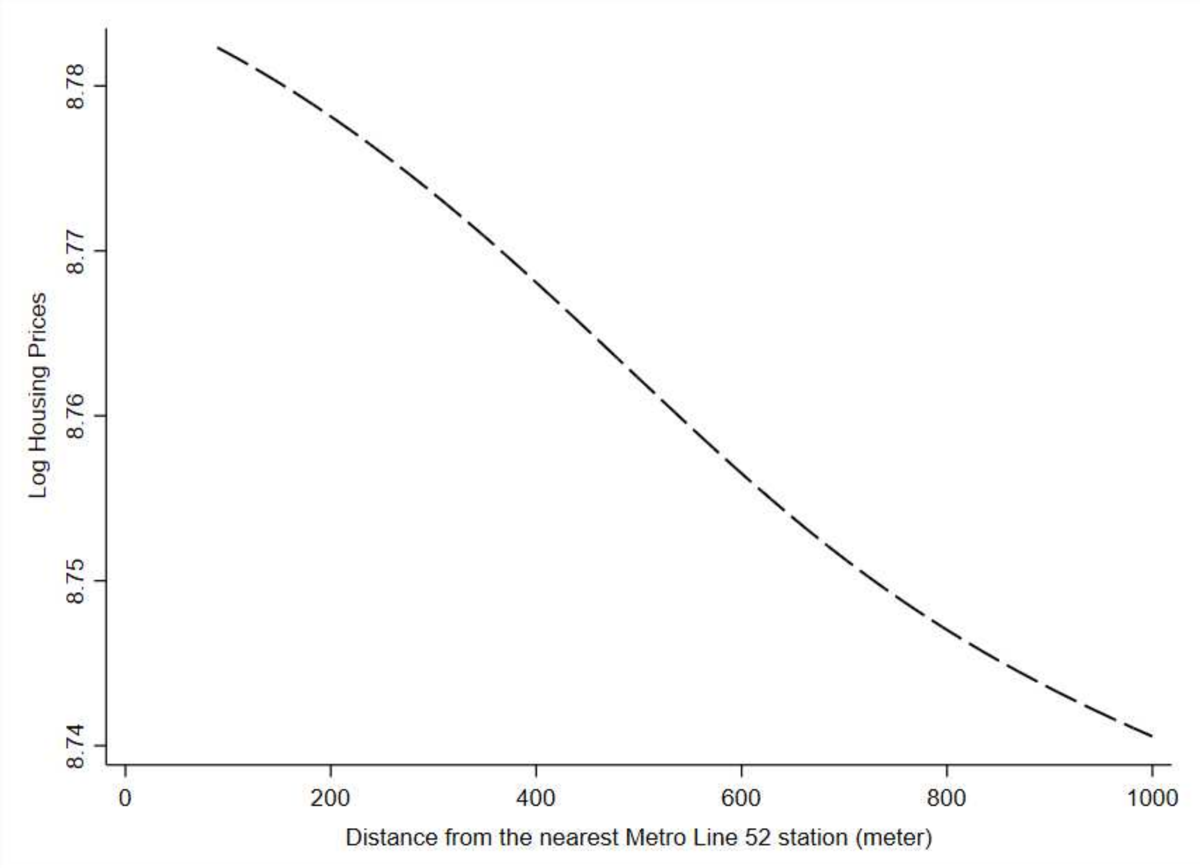

FIGURE 5A. PRICE GRADIENT OF NETWORK DISTANCE FROM METRO LINE 52 STATION
(Transactions during the year after opening)

Note: The result from local polynomial regressions with a bandwidth of 221 meters.

distance from house observations to the nearest Metro Line 52 stations after the Metro Line 52 opens for service. The one year (six months before and after the opening of Metro Line 52) is set as the time window for this analysis which is similar to the work of Linden & Rockoff (2008). This time window is set as a preventive measure against any other confounding factors that might influence the effect of the new metro line on housing price. The house prices located closest to the station have the highest price, and the prices are getting lower as the house's location moves away from the station. It seems that the result from this graphical evidence is aligned with the conceptual framework of the bid rent curve, where the prices are highest for houses that are closest to an attractive place and get smaller as the distance increases.

I also contrast the result with the price gradient of network distance to the station before the opening of Metro Line 52, which is illustrated in **Figure 5B.** The figure shows that both gradients are starting to resemble from the network distance of 400 meters to 650 meters and are tangent in the network distance of 537 meters. The diagram shows that the positive effects start to shrink from the distance of 400 m, and finds no effect after the location of a house crosses the distance of 537 meters.

It is argued that the causal impact can be conveyed for the above case if the increase of the prices is parallel with the opening of Metro Line 52, and after controlling any other factors (e.g. locational, structural, and transactional characteristics), there is no significant anticipation effect before the opening of the new metro. Both conditions can be observed by looking at the **Figure 6A** and **6B**. In addition, the latter condition will be further discussed in the next section. **Figures 6A** and **6B** represent the price gradients of the date of the house transactions concerning Metro Line 52. The measurement of the

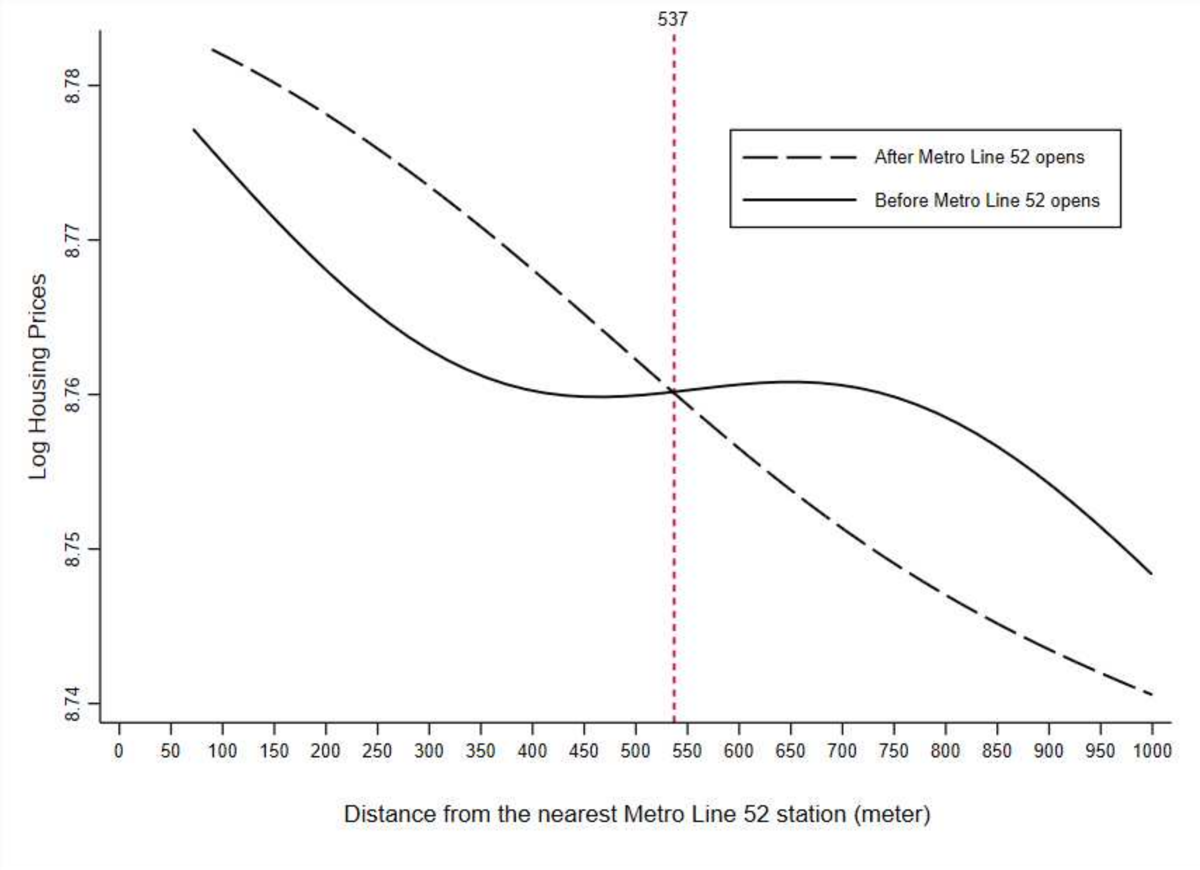

FIGURE 5B. PRICE GRADIENT OF NETWORK DISTANCE FROM METRO LINE 52 STATION
(Transactions during the year after opening)

Note: The result from local polynomial regressions with a bandwidth of 221 meters.

gradients is performed separately according to whether the houses that are located within 537 meters are transacted before or after the opening of Metro Line 52. The time window for this analysis is set to 2 years before and after the opening following Linden & Rockoff (2008).

If the anticipation effect does exist, we can expect that the increase in price occurs gradually during the time period before and after the opening of the metro system. As it turns out, the price trend witnesses a gradual increase before the opening of Metro Line 52 but witnesses a sudden huge increase afterwards. **Figure 6B** contrasts the price gradient of the date of the house sales located beyond 537 meters from the nearest station with the price gradient in **Figure 6A**. The difference between both gradients can be found in the curve of the houses located beyond 537 meters that continues, while the other curve is discontinuing and experiences a sudden increase after the Metro Line 52 opens. Generally, both groups of houses are still close to each other as they are all located within 1,000 meters from Metro Line 52 stations, as depicted in **Figure 3**. The question can be asked whether both groups of houses would have had the same price movement if the Metro Line 52 had not been established. The prices of the houses located beyond 537 meters are quite similar to the group of houses that are closer to the station prior to the opening of the metro. Thus, the group of houses that are slightly farther away from the affected areas can be used as the untreated group of house observations in this quasi-experiment.

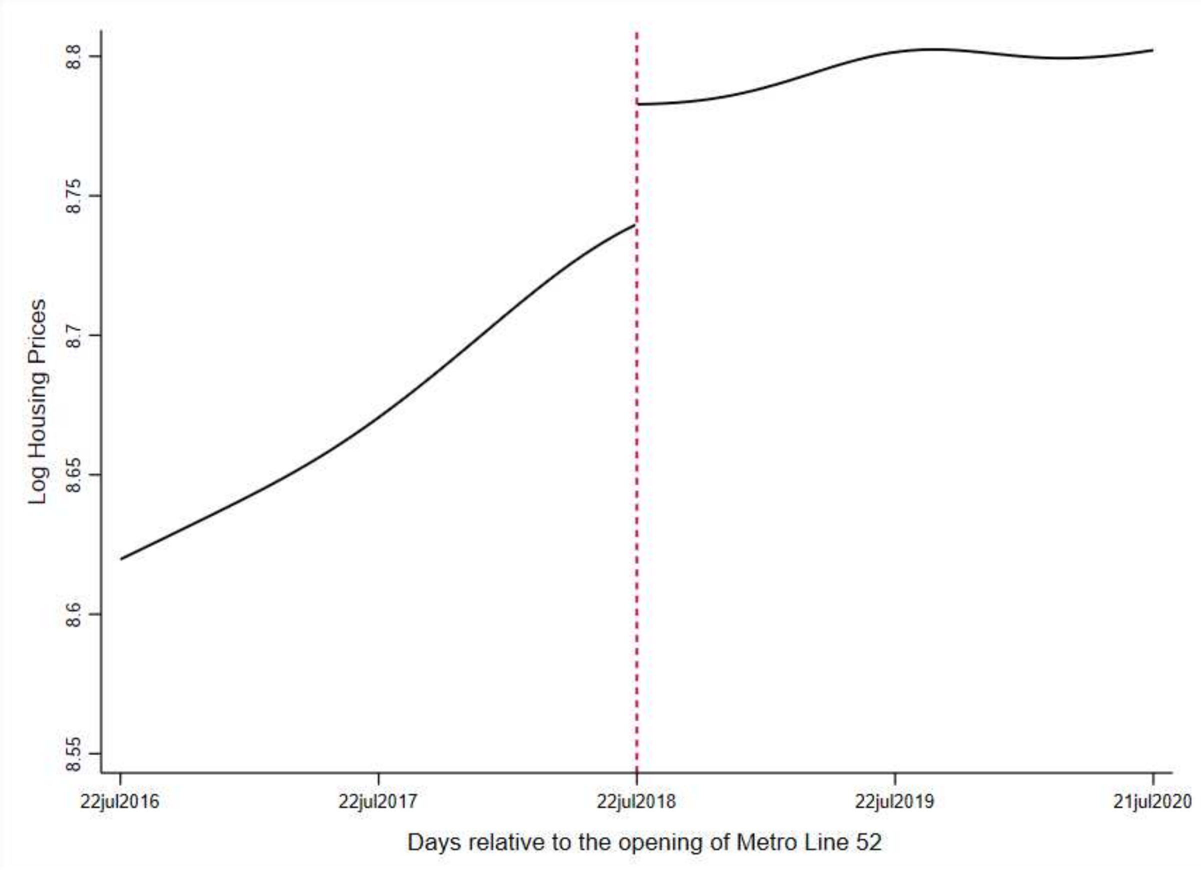

FIGURE 6A. PRICE TRENDS BEFORE AND AFTER THE OPENING OF METRO LINE 52
(House transactions within 537 meters from the nearest Metro Line 52 station)

Note: The result from local polynomial regressions with a bandwidth of 140 days

E. Empirical Design: The Linear Specification

The hedonic price model is by far the most popular model used in this mature topic. However, many extended versions, innovations, and new frameworks for identification strategy have been developed to correct endogeneity problems commonly occurred in conventional hedonic strategy. In earlier studies, the concept of "before-and-after" is used, where a set of dummy time-variables is added into the hedonic price specification, and cross-sectional real estate data pooled over time is used. The study by McDonald & Osuji (1995) was the first to use this method. The model can be specified as follows:

$$y_{it} = \alpha + \beta\, X_{it} + \gamma\, D_{it} + \delta\, H_{it} + \varepsilon_{it}$$

where y_{it} is a vector of log-transformed housing price; α is a constant term; X_{it} is a vector of variable(s) of interest; D_{it} is a vector of time dummies variable representing time fixed effects; H_{it} is a vector of housing attributes; ε_{it} is the error term; and β, γ, and δ are the parameters to be estimated. Housing attributes includes housing transactional, structural, and locational characteristics.

Furthermore, Gibbons & Machin (2005)'s work improved this concept by applying the difference-in-difference methodology to prevent the analysis from biases that are usually occurred in a cross-sectional study, for instance, some unobserved characteristics that are correlated with the accessibility and housing prices and the change of accessibility over time that is not accounted can lead to biased

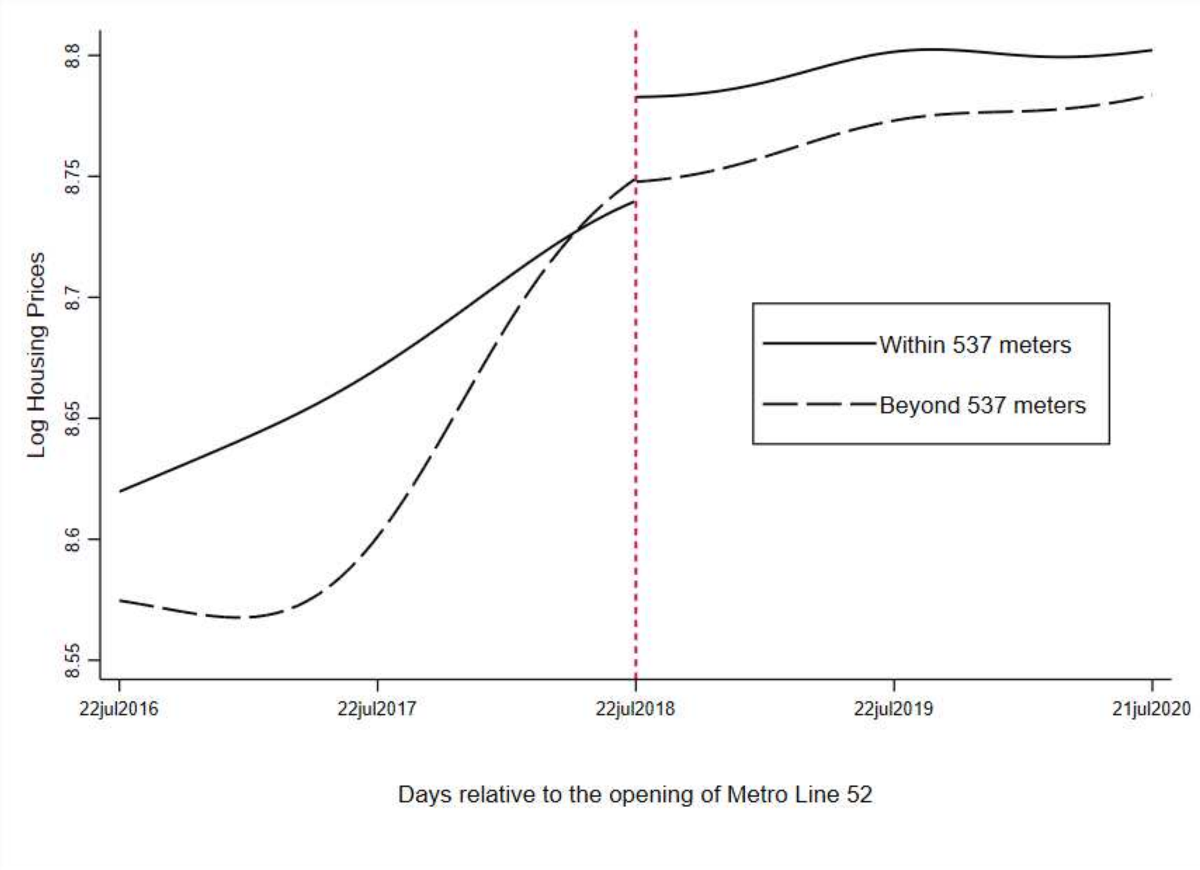

FIGURE 6B. PRICE TRENDS BEFORE AND AFTER THE OPENING OF METRO LINE 52
(Contrasting the house transactions within and beyond 537 meters from the nearest Metro Line 52 station)

Note: The result from local polynomial regressions with a bandwidth of 140 days

estimates. By adding time fixed effect accounting for the time temporal dynamics in the housing market, denoted as τ_t; and spatial fixed effects accounting for spatial features within a specific neighbourhood, denoted as φ_j; the specification can be upgraded as follows:

$$y_{it} = \alpha + \beta\, X_{ijt} + \delta\, H_{ijt} + \varphi_j + \tau_t + \varepsilon_{ijt}$$

The time fixed effect used in this analysis accounts for the quarter temporal dynamics in the Amsterdam housing market. Also, I use Amsterdam 6-digits postal code to represent the spatial features within the postal boundary. In the DID setting, causal effects can be determined as an estimation process of a counterfactual that changes over time for the treated group assuming the treatment does not occur. Athey & Imbens (2017) argue that this method assumes the change in outcomes over time for the control group is informative about the change in the treatment group in the absence of the treatment. Thus, the analyst should appoint each observation into the treated group and the control group. Most studies set the treatment area arbitrarily. However, this paper implements the local polynomial regression approach to determine whether the observation is assigned to the treatment group. Considering treated and untreated groups, thus, the specification for $\beta\, X$ would be:

$$\beta\, X_{ijt} = \beta_1 Treat_i + \beta_2 Post_t + \beta_3 (Treat_i \times Post_t)$$

where the variable $Treat_i$ is a dummy variable indicating whether or not the individual observation is in the treated group; the variable $Post_t$ is also a dummy variable indicating whether the house transaction has taken place after the RTS operates; β_1 and β_2 are the parameters to be estimated; and β_3 measures the average treatment effect. To specify the treatment area, I split the sample based on whether the network distances of the observations are within 537 meters. Thus, the houses located outside the 537-meter network-distance buffer are assigned to the controlled group. This robust empirical design has been considerably used in several papers related to the literature on the effect of an opening of RTS on housing price.

F. Empirical Design: Machine Learning Approach

This paper adopts the causal tree and honest approaches proposed by Athey and Imbens (2016). Causal tree approach was originally driven from the CART algorithm written by Breiman, Friedman, Olshen, and Stone (1983), where it focuses on estimating the heterogenous treatment effect and constructing valid confidence intervals for treatment. Similar to the standard regression tree, the causal tree is also grown based on a splitting rule which aims to minimize the risk function.

One of the distinguishing parts is that the causal tree estimates the average treatment effects within each leaf instead of the average of the dependent variable within each leaf. An analyst should specify a dummy treatment variable to tell the algorithm to take the difference between the sample average of the treated and untreated groups within the leaf. In machine learning, the selection of a model requires setting a specific tuning parameter to prevent from overfitting problems. This tuning parameter value is primarily chosen by performing cross-validation. In the case of tree-based machine learning, cross-validation is implemented by penalizing the number of nodes in the tree. In pruning the causal tree, the leaves are chosen based on the chosen risk function calculated while the tree is grown.

Another distinctive feature is that the causal tree is mainly performed along with the performance of honest estimation. This honest approach aims to estimate the treatment effect within the leaves of a tree by using an independent estimation dataset instead of using the dataset that has been previously used for building and pruning the tree. To perform a causal tree with an honest approach, the analyst first splits the datasets into two, i.e., training set and estimation set, and then uses the training set to build and prune the tree. Next, the pruned tree is used as the basis of the tree used for estimating the treatment effect using the estimation set. In the estimation process, the leaf estimates will be replaced by the new estimates calculated from using the estimation set.

The advantage of using these approaches is that it can answer one of the main questions asked in this paper, i.e., about finding the specific properties that have the highest treatment effect from the establishment of Metro Line 52. This analysis includes the same variables used in the linear estimation where the post and treated area dummies are used as the binary treatment variable, and other covariates (including the structural, transactional, and characteristics) used as the splitting variables in growing the causal tree. The dependent variable of interest is the log housing price per square meter. The honest causal tree functions are used as the splitting rule in growing the tree and used in pruning the tree in the cross-validation process. This analysis is based on a similar conceptual framework proposed by Chin et al. (2020) where the approach is to estimate the conditional average treatment effect (CATE) in the DID framework within each leaf using CausalTree R package[11] , which can be mathematically specified as follows:

$$
\begin{aligned}
CATE = &\{E[Y|X = x, D = 1, T = 1] - E[Y|X = x, D = 0, T = 1]\} \\
&-\{E[Y|X = x, D = 1, T = 0] - E[Y|X = x, D = 0, T = 0]\}
\end{aligned}
$$

[11] CausalTree R package is available at https://github.com/susanathey/causalTree

However, this analysis modifies the algorithm used by Chin et al. (2020) where instead of setting the treated area dummy as the binary treatment variable, this analysis sets the interactive terms of the treated areas and post-event dummies as the binary treatment variable in the causal tree setup as it can estimate the CATE better, especially in a single tree algorithm setup. I argue that setting only the treated area dummy as the binary treatment variable in the causal tree setup can lead to an incorrect specification in estimating CATE in a leaf as there is a high tendency that the regression tree will split the data based on post dummy, which is not a desirable case. As a result, the CATE will not represent the average treatment effect but only represent the housing price differences between the treated areas relative to the untreated areas either before or after the Metro Line 52 starts to operate.

IV. Empirical Result

A. The Pre-Treatment Trend

In conducting the analysis of the impact of a transportation development project using a difference-in-differences approach, it is essential to check whether there is an anticipation effect that occurs before the opening of the new transportation service. This can be done by checking whether the price trends of the treated groups are not statistically different from zero over time. Based on the graphical evidence produced by the application of local polynomial regression, I define the treated group as the group of houses located within 537 meters from Metro Line 52 stations and assign the rest of the observations into the untreated group. I estimate the temporal variations in the treatment effects using a specification as follows:

$$y_{ijt} = \alpha + \beta_1\, Within537_i + \sum_{z=2002/1}^{2020/2} \beta_z\, Within537_i \times HALFYEAR_z + \delta\, H_{ijt} + \varphi_j + \tau_{ijt} + \varepsilon_{ijt}$$

where $Within537_i$ is a dummy variable representing whether a housing observation is located within 537-meter network-distance from future Metro Line 52 stations or not; $HALFYEAR_z$ is a bi-annual time dummy variable; H_{ijt} is a set of housing characteristics; β_z is the interaction parameters between year dummy and treatment dummy; and τ and φ are quarterly time and 6-digits postcode fixed effects, respectively. **Figure 7** describes the interaction coefficients between the treatment dummy and year dummy, which represent the temporal variations in the treatment effect of the future opening of Metro Line 52. The figure shows a fluctuating trend during the periods before the opening event and a slight negative trend followed by positive trends, with some of it are statistically significant happening after the event of the metro opening. However, a more comprehensive study investigating the treatment effect after the opening of the metro line on housing prices within a 537-meter network distance will be discussed further in the next subsection.

If there is an anticipation effect, then it can be expected that the interaction coefficients before 2018 would be positive and statistically different from zero. However, the figure shows that from the year the Metro Line 52 project was approved and announced in 2002 until the opening, the effect of the future new metro line had not been statistically significant, while it becomes statistically significant at 1% in a period within the post-opening periods. I argue that the result shows no significant anticipation effect due to the negative sentiment perceived by the community about the controversies that had been going during the years prior to the opening along with the uncertainty of this project. This uncertainty causes no change in the expectations of the housing price, leading to no anticipation effect. Besides uncertainty, it is also probable that there is no significant effect on pre-trend analysis because almost all the construction works were taken place underground. Noises that are normally produced due to a

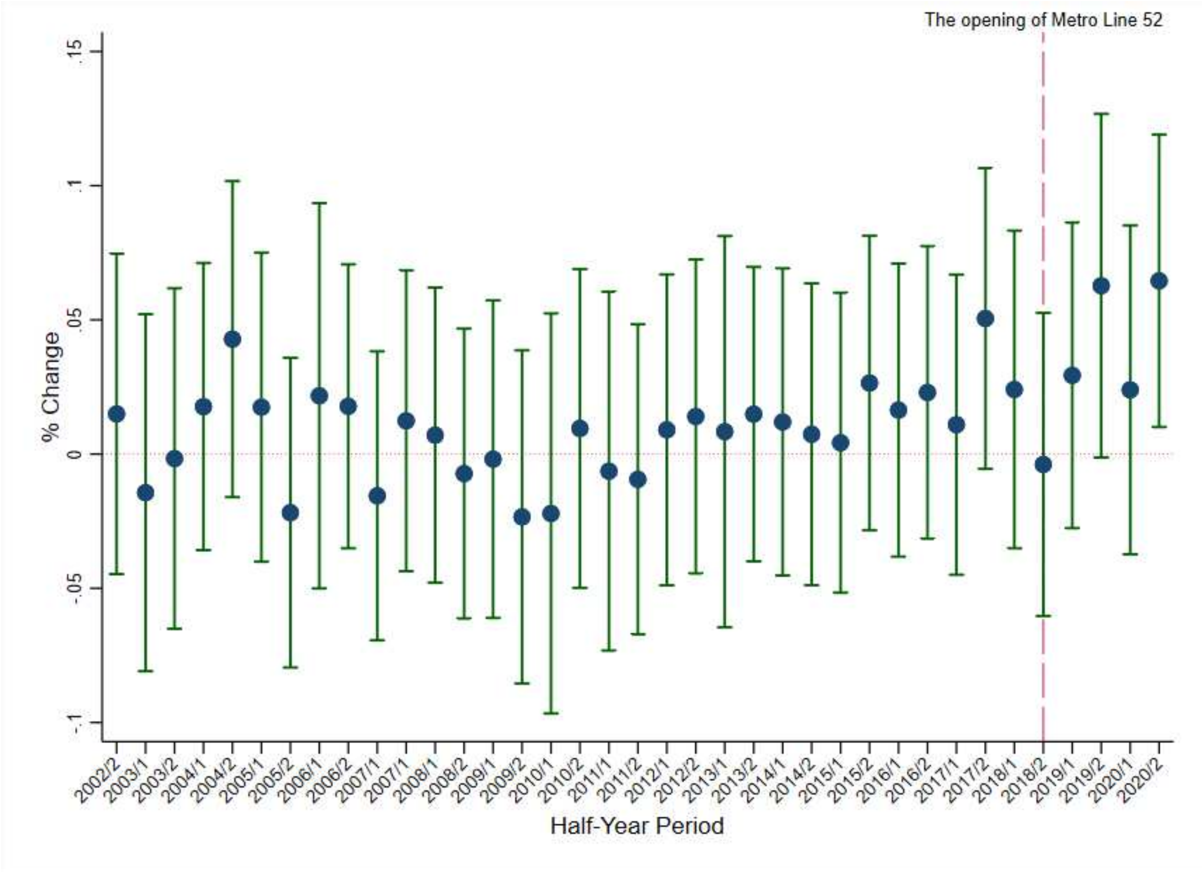

FIGURE 7. INTERACTION COEFFICIENTS REPRESENTING TREATMENT EFFECTS OVER TIME
(With 99% confidence intervals)

construction and a trigger of a decline in house price might not be the main driver in this analysis. Another possible reason is that the construction work requires a very long time to finish. Since the construction time is very long, even though there is an anticipation effect, it might not occur immediately after the plan was accepted and announced. In contrast, it is more likely to occur a few months or years before the opening.

The date of the Metro Line 52 opening was finally announced a few months prior to the opening via the local media. Therefore, there can be a case where there is a reaction in the housing market a few months or a year before the opening of the Metro Line 52. **Figure 7** also shows that there is an almost significant positive trend that emerged a year before the Metro Line 52 starts to operate while diminishing until the date of the opening. To check whether there is an anticipation effect before the opening, I slightly modify the baseline of the difference-in-difference approach and eventually add some other dummy variables representing whether a housing transaction is made during a year and a half year before the opening denoted as $Post_{t-12}$ and $Post_{t-6}$, respectively. I also add a dummy variable representing whether a transaction is made after the opening of Metro Line 52, denoted as $Post_{t=0}$, and a dummy variable representing the transactions made between a year and a half-year before the opening denoted as $Post_{(t-12)-(t-6)}$. **Table 2** summarized the result from the modified models.

The result from column 1 shows that the treatment effect for the one-year pre-opening is positive and statistically significant at 5%. I further add the post dummy into the model and find that the effect for the one-year pre-opening becomes more statistically and economically significant. This can be an indication that anticipation effects occur a year before the opening of the new metro line. Furthermore, I also test whether the anticipation effect occurs six months before the opening. Column 3 explains that

Table 2.
Sensitivity Tests: The analysis of anticipation effects within one year or a half year before the opening of Metro Line 52.

	(1) One year pre-opening	(2) One year pre-opening and post-opening	(3) A half-year pre-opening	(4) A half-year pre-opening and post-opening	(5) Biannual year pre-opening and post-opening
Within 537 meters	-.0141	-.0193	-.0127	-.0174	-.0194
	(.0124)	(.0125)	(.0124)	(.0125)	(.0125)
$Post_{t-12}$	.0053	.0038			
	(.0142)	(.0175)			
$Post_{(t-12)-(t-6)}$					-.0012
					(.0178)
$Post_{t-6}$			.0092	.0029	.0077
			(.0171)	(.0224)	(.0288)
$Post_{t=0}$		-.01		-.0172	-.0121
		(.0275)		(.0304)	(.0351)
Within 537 m × $Post_{t-12}$	.0221*	.0276**			
	(.0101)	(.0104)			
Within 537 m × $Post_{(t-12)-(t-6)}$					.0447**
					(.0141)
Within 537 m × $Post_{t-6}$			.0026	.0076	.01
			(.0137)	(.0139)	(.0139)
Within 537 m × $Post_{t=0}$		.0337***		.0321***	.0337***
		(.0086)		(.0085)	(.0086)
Observations	23,668	23,668	23,668	23,668	23,668
R-squared	.7202	.7205	.7201	.7204	.7205
Adj R2	.7189	.7191	.7188	.7191	.7191
Transactional characteristics	Yes	Yes	Yes	Yes	Yes
Structural characteristics	Yes	Yes	Yes	Yes	Yes
Locational characteristics	Yes	Yes	Yes	Yes	Yes
Quarterly fixed effect	Yes	Yes	Yes	Yes	Yes
Postcode fixed effect	Yes	Yes	Yes	Yes	Yes

Notes: The table displays the results of panel regressions with log housing price per square meter as the dependent variable and housing structural characteristics (e.g. log size, dummies representing property types, parking dummy, number of bathrooms, number of kitchens, number of balconies, number of roof terraces, office room dummy, maintenance score of the inside, maintenance score of the outside, good maintenance dummy, number of insulation types, central heating dummy, listed property dummy, and monumental property dummy), housing transactional characteristics (auction dummy, leasehold dummy, partly rented property dummy, and dummies representing years of construction) and housing locational characteristics (log network distance to the nearest tram station, and log network distance to the nearest metro station other than Metro Line 52 station) as the control variables. The quarterly time and postcode (6-digits) fixed effects are used in the estimations. The treatment estimations are defined by the total-travel-cost network distance of 537 meters from Metro Line 52 stations. $Post_{t=0}$ *is a dummy variable indicating the opening of Metro Line 52.* $Post_{t=6}$ *and* $Post_{t=12}$ *are dummy variables indicating whether a housing transaction is made one year and a half year, respectively, before the opening. The interaction between treatment variable and each of the time dummy indicates the treatment effects of the opening of Metro Line 52 on housing price for each time period. Robust Standard errors are in parentheses and the standard errors are clustered at the postcode (6-digits) level.*
**** p<.001, ** p<.01, * p<.05*

the treatment effect for the half-year pre-opening is also positive but not statistically significant. This result is still consistent when the post dummy is added (see column 4). There seems to be a diminishing trend starting from the beginning to the end of the one-year pre-opening period. Therefore, I test whether there is a heterogeneity in the anticipation effect within the period of one year before the opening by adding $Post_{(t-12)-(t-6)}$ into the model. The anticipation effect coming from the dummy $Post_{(t-12)-(t-6)}$, representing the house transactions made between July 22nd 2017 and January 22nd 2018, turns out to be economically and statistically significant at 1%, while I find no effect for houses transacted six months before the opening. This indicates that the anticipation effects are significant as early as one year prior to the opening, while the effects become insignificant the moment before the new metro line starts to operate.

These results show a strong indication that there is an anticipative action that occurred one year before opening the new metro line. However, these anticipation effects do not occur most of the time in the pre-opening period and even a few months prior to the opening of the Metro Line 52. Therefore, these pre-trend analyses can indicate that the design of the quasi-experiment applied in this paper and the difference-in-difference approach can still be assumed to be suitable for estimating the effect of a new metro line investment on the Amsterdam housing market.

B. Average Treatment Effect of Metro Line 52 on Housing Prices using OLS

Table 3 shows the results for the linear specifications described in the section of Empirical Design[12]. Building on McDonald & Osuji (1995), the first column describes the conventional standard hedonic pricing model in which the model controls for time fixed effects and housing characteristics using quarterly time dummies variable as well as other housing characteristics variables in a cross-sectional setting. Column 2 adds postcode dummies into the model to control for spatial fixed effect. Column 3 describes the base model, i.e., the panel estimates using the difference-in-differences approach. In this model, I cluster the unobserved spatial and time heterogeneities in the housing prices by adding postcode fixed effects (2,161 postal codes) and quarterly time fixed effects (76 quarters in the transaction date). Column 4 expands the study area into the houses located within 1,500-meter total-travel-cost network distance. Finally, column 5 modifies the base model by including the interaction terms between the average treatment effect and housing characteristics to test whether there is heterogeneity in the treatment effect across housing characteristics. All models aim to estimate the impact of the opening of Metro Line 52 on housing prices which is represented by an interactive term that is included in each model.

In columns 1 and 2, the coefficient of the interactive term between Log Distance to the Nearest Metro Line 52 Station and Post dummy represents the treatment effect in the cross-sectional design. Both models show a negative sign of the coefficient, which are as expected because the lower the distance is, the higher the house price will be. However, the cross-sectional model in column 2 shows a smaller magnitude compared to the previous model. This is an indication that there is still unobserved heterogeneity that is not considered in model 1. The result of model 2 shows that an extra 1% reduction in the distance from Metro Line 52 station is associated with a 2.85% increase in house price.

I further explore the result by including a dummy variable representing houses located within 537-meter total-travel-cost network distance to the nearest metro station and interact it with the post dummy variable to capture the average treatment effect in the panel estimation setup. All else equal, houses located within the treatment area witness a 3.18% capitalization relative to the houses located in the controlled area (see column 3). This result also implies that the new metro line in Amsterdam increases the willingness to pay of the buyers for houses that enjoy better accessibility.

The coefficients of the Post and Within537 dummies in column 3 also shows a negative sign. It indicates that there is a general downward price trend occurring after the opening event, and there is a price discount on average for houses located within 537-meter total-travel-cost network distance from Metro Line 52 station relative to other houses located outside the treatment area, respectively. It seems that the result in column 2 is overly estimated, as it shows a significant and strong positive trend after the opening of Metro Line 52, while the result in column 3 does not. Thus, column 3 affirm that there is still unobserved spatial heterogeneity that has not been controlled in column 2.

This result also reveals that the capitalization effects captured from the first two models are probably underestimated. It is assumed earlier that the capitalization effect is only received by the houses located within the treatment area. Later in the next subsection, I will show that there is a clear

[12] Table 3 only contains the estimates of the interest variables. The estimates of other control variables can be seen in Appendix A.

Table 3.
Panel linear estimates of the value of metro access

	(1) Standard Hedonic Controlling for Time Fixed Effect	(2) Standard Hedonic Controlling for Spatial & Time Fixed Effects	(3) Base Model	(4) Base Model	(5) Model to Test for Heterogeneity in Treatment Effect
Study Boundary	1,000 m	1,000 m	1,000 m	1,500 m	1,000 m
Post	.1845**	.1703**	-.0226	.0001	-.1907
	(.0589)	(.0529)	(.0217)	(.0138)	(.1516)
Log (Distance to the nearest Metro Line 52 Station in meter)	.0947***	.0092			
	(.0034)	(.0213)			
Log (Distance to the nearest Metro Line 52 Station in meter) × Post	-.033***	-.0285***			
	(.0084)	(.0076)			
Within 537 m			-.0171	-.0149	-.0895
			(.0124)	(.0123)	(.2338)
Within 537 m × Post			.0318***	.0231**	.1035
			(.0084)	(.0075)	(.2674)
Within 537 m × Post × Log Distance to CBD					-.049*
					(.0229)
Within 537 m × Post × Detached					-.2786*
					(.1162)
Within 537 m × Post × Parking					-.063*
					(.0309)
Within 537 m × Post × Partly Rented					.3251**
					(.0992)
Within 537 m × Post × Log Distance to the Nearest Tram Station					.0383**
					(.0134)
Observations	23,668	23,668	23,668	41,795	23,668
R-squared	.6879	.8316	.7204	.7268	.7271
Adj R^2	.6864	.8138	.7191	.7261	.7248
Transactional characteristics	Yes	Yes	Yes	Yes	Yes
Structural characteristics	Yes	Yes	Yes	Yes	Yes
Locational characteristics	Yes	Yes	Yes	Yes	Yes
Quarterly fixed effect	Yes	Yes	Yes	Yes	Yes
Postcode fixed effect	No	Yes	Yes	Yes	Yes

Notes: The table displays the results of cross-sectional (model 1 and 2) and panel regressions (models 3 to 5) with log housing price per square meter as the dependent variable and housing structural characteristics (e.g. log size, dummies representing property types, parking dummy, number of bathrooms, number of kitchens, number of balconies, number of roof terraces, office room dummy, maintenance score of the inside, maintenance score of the outside, good maintenance dummy, number of insulation types, central heating dummy, listed property dummy, and monumental property dummy), housing transactional characteristics (auction dummy, leasehold dummy, partly rented property dummy, and dummies representing years of construction) and housing locational characteristics (log network distance to the nearest tram station, and log network distance to the nearest metro station other than Metro Line 52 station) as the control variables. The quarterly time and postcode (6-digits) fixed effects are used in the estimations. The treatment estimations are defined by the total-travel-cost network distance of 537 meters from Metro Line 52 stations. Post takes 1 if a house transaction made after the opening event, while take 0 otherwise. Log distance is included in the model as the variable interest in cross-sectional study. All the interactive terms included in every model represent the average treatment effects. In estimating the model 5, I also include the interaction term between the average treatment effect and every housing characteristics while only display the result from the interaction terms displaying a significant effect. Standard errors are in parentheses
**** p<.001, ** p<.01, * p<.05*

capitalization effect with proximity to a Metro Line 52 station for houses located within the treatment area, while the results find no effect on houses located beyond the treatment area.

I also explore how much the realized gains in housing value are associated with the opening of Metro Line 52. I quantify the capitalization effect by considering the increases of accrued housing value within the treatment area. According to the result in column 3 and the average sales price per square meter for house transactions located within the treatment area which is EUR 4,622.73 per square meter (see **Table 1**), the Metro Line 52 average treatment effects of 3.18% can be translated into a price premium of EUR 147.003 per square meter.

I also perform a robustness check by expanding the study area boundary from 1,000-meter into a 1,500-meter total-travel-cost network distance from the nearest Metro Line 52 station. This approach increases the number of observations, almost doubling from 23,668 to 41,795 house transactions (see column 4). By setting a broader study area, I find a less significant ($p < 0.01$) and a weaker treatment effect accumulating to 2.31% compared to a highly significant ($p < 0.001$) treatment effect of 3.18% found in column 3. However, this extended-study-area model still shows a significant and positive effect; thus, the specification used for the base model can be concluded as robust.

Finally, I also aim to explore the result further and ask whether there is heterogeneity in the treatment effect. I again modify the baseline difference-in-differences model by interacting the coefficient previously representing the average treatment effect with all housing characteristics used in the main analysis. Column 5 documents that there is heterogeneity in the treatment effects across the distance to the nearest tram station and partly rented property dummy. This can also be an indication that there is heterogeneity in treatment effects across other housing characteristics. This heterogeneous treatment effect of the new metro line on housing price will be explored in the next few subsections and I will show that a machine learning approach can capture more personalized treatment effects with a valid confidence interval.

C. Spatial Variations in the Average Treatment Effects

To explore more on the spatial variation in the treatment effects, one of the approaches is to divide the treatment area into some sub-areas. This analysis divides the treatment area into 3 different zones: zone 1 constitutes the area within 0 to 200 m total-travel-cost network distance from the nearest Metro Line 52 station, and zone 2, as well as zone 3, constitute 200 to 537 m and 537 to 750 m, respectively. **Table 4** shows the results of the base model regression taking into account the three different zones that are previously created.

In column 1, we explore whether there is any difference in treatment effect for zone 1 and zone 2 within the treatment area with only using 7,379 observations. The result indicates that zone 1 has a higher trend than zone 2 following the metro opening. This also indicates that there seems to be a spatial variation in the impact of metro opening in the treatment zone. However, I also find no significance in the treatment effect between both zones. This may also indicate that there is no significant difference between both zones in terms of the treatment effect of the metro opening. Therefore, I argue that multiplying samples by expanding the study area would be necessary to further check for these spatial variations.

In column 2, I further expand the study area into the main study area in this study, i.e., within a 1000-meter network distance buffer with 23,668 observations. In this analysis, I also recheck the variations between both zones. The result shows that the impacts are 6.7% and 2.88% higher for houses located within zone 1 and zone 2, respectively, compared to the rest of the zones, and those are all statistically significant. This validates the spatial variations for zone 1 and zone 2 in the previous result, where zone 1 has a higher impact compared to zone 2 when the sample in the controlled area is added into the study. This also implies that the effect is stronger for the houses located in the zone closer to the station.

In column 3, I further add more spatial variation into the model by adding the zone 3 dummy indicating houses that are located within 537 – 750 m. The result is similar to the previous finding. It suggests that the impacts are still statistically significant and are 6.64 % and 2.82% higher in both inner zones (zone 1 and 2) compared to the reference zone. In contrast, the difference becomes no longer statistically different from zero when a house is located in the outer zone of zone 3. This complements

Table 4.
Spatial variations in the treatment effects

	(1)	(2)	(3)	(4)
Study Boundary	537_m	1,000_m	1,000_m	1,000_m
Post	.0291	-.023	-.0222	-.0219
	(.0342)	(.0216)	(.022)	(.022)
Within 0 – 200 m	.0515*	.0343	.0444	
	(.0242)	(.0284)	(.0312)	
Within 200 – 537 m		-.0166	-.0065	
		(.0124)	(.0178)	
Within 537 – 750 m			.0106	.0106
			(.0137)	(.0137)
Within 0 – 537 m				-.0071
				(.0178)
Within 0 – 200 m × Post	.0355	.067**	.0664**	
	(.0224)	(.0212)	(.0217)	
Within 200 – 537 m × Post		.0288***	.0282**	
		(.0087)	(.0099)	
Within 537 – 750 m × Post			-.0015	-.0015
			(.0103)	(.0103)
Within 0 – 537 m × Post				.0312**
				(.0096)
Observations	7,379	23,668	23,668	23,668
R-squared	.7726	.7206	.7206	.7204
Adj R^2	.7691	.7192	.7192	.7191
Transactional characteristics	Yes	Yes	Yes	Yes
Structural characteristics	Yes	Yes	Yes	Yes
Locational characteristics	Yes	Yes	Yes	Yes
Quarterly fixed effect	Yes	Yes	Yes	Yes
Postcode fixed effect	Yes	Yes	Yes	Yes

Notes: The table displays the results of panel regressions with log housing price per square meter as the dependent variable and housing structural characteristics (e.g. log size, dummies representing property types, parking dummy, number of bathrooms, number of kitchens, number of balconies, number of roof terraces, office room dummy, maintenance score of the inside, maintenance score of the outside, good maintenance dummy, number of insulation types, central heating dummy, listed property dummy, and monumental property dummy), housing transactional characteristics (auction dummy, leasehold dummy, partly rented property dummy, and dummies representing years of construction) and housing locational characteristics (log network distance to the nearest tram station, and log network distance to the nearest metro station other than Metro Line 52 station) as the control variables. The quarterly time and postcode (6-digits) fixed effects are used in the estimations. The treatment estimations are defined by the total-travel-cost network distance of 537 meters from Metro Line 52 stations. The treatment variables are based on whether an observation is located within 0-200 m, 200-537 m, 537-750 m, 750-1,000 m, 0-537 m, and 537-1,000 m of total-travel-cost distances from Metro Line 52 stations. Post dummy takes one if the house sales made after the opening and zero otherwise. The interactive variables estimate the treatment effect of Metro Line 52 on housing prices. Standard errors are in parentheses
**** p<.001, ** p<.01, * p<.05*

the previous interpretation where the strongest effects occurred for the houses located in the zone closest to the station while it gets smaller and smaller when a house located in the zone farther away from the metro station, and the treatment effects finally start to dissipate at a certain distance that is outside from the affected area. This can also be an indication that the treatment area calculated by the local polynomial regression might really represent the true affected area.

As a robustness test, I also expand the base model (see column 3 in Table 3) by adding an additional distance dummy representing zone 3. If the treatment area calculated by local polynomial regression can really represent the affected area, we can expect that the treatment effect on the house price for house transactions located in zone 3 will not be statistically different from zero. Column 4 confirms the statement where the result shows an insignificant 0.00% effect on house prices within zone 3. This confirms that the use of local polynomial regression is a robust approach for treatment assignment as the use of the treatment area assigned based on this approach can appropriately estimate the treatment effects in the study area within a 1000-meter network distance buffer from each Metro Line 52 station. This also affirms that the overall winners from the new metro line in Amsterdam are the houses that are located within 537-meter network distance from the Metro Line 52 stations.

Table 5.
Robustness tests: comparison between the base model and the model using different sources of distances

	(1) Euclidean distance ≤ 537 m	(2) Standard-network distance ≤ 537 m	(3) Total-travel-cost network distance ≤ 537 m
Post	.0063	-.0095	-.0226
	(.0156)	(.0192)	(.0217)
Within 537 m (Euclidean)	-.0283		
	(.0175)		
Within 537 m (Standard Network)		-.0215	
		(.0135)	
Within 537 m			-.0171
			(.0124)
Within 537 m (Euclidean) × Post	.0168*		
	(.0068)		
Within 537 m (Standard Network) × Post		.0267***	
		(.0078)	
Within 537 m × Post			.0318***
			(.0084)
Observations	35,672	26,488	23,668
R-squared	.7243	.7271	.7204
Adj R^2	.7234	.7259	.7191
Transactional characteristics	Yes	Yes	Yes
Structural characteristics	Yes	Yes	Yes
Locational characteristics	Yes	Yes	Yes
Quarterly fixed effect	Yes	Yes	Yes
Postcode fixed effect	Yes	Yes	Yes

Notes: The table displays the results of panel regressions with log housing price per square meter as the dependent variable and housing structural characteristics (e.g. log size, dummies representing property types, parking dummy, number of bathrooms, number of kitchens, number of balconies, number of roof terraces, office room dummy, maintenance score of the inside, maintenance score of the outside, good maintenance dummy, number of insulation types, central heating dummy, listed property dummy, and monumental property dummy), housing transactional characteristics (auction dummy, leasehold dummy, partly rented property dummy, and dummies representing years of construction) and housing locational characteristics (log network distance to the nearest tram station, and log network distance to the nearest metro station other than Metro Line 52 station) as the control variables. The quarterly time and postcode (6-digits) fixed effects are used in the estimations. The treatment estimations are defined by three different distance measurements, i.e., Euclidean, standard network analysis, and total-travel-cost network analysis. Standard errors are in parentheses
**** $p<.001$, ** $p<.01$, * $p<.05$*

D. Alternative Proximities for Measuring Distances

In this section, I will compare and discuss a few alternatives of distance measures as a proxy of accessibility that can be used in this literature. Table 5 shows the comparison between the results that uses a different measures of distance measurement techniques. Column 3 describes the main result of this paper which uses distance data from total-travel-cost network analysis technique. Columns 1 and 2 show the results using the same model but using distance data from Euclidean analysis and standard network analysis, respectively. The study area is within a 537-meter distance from the nearest Metro Line 52 station and the distance measurements differ depending on each measurement technique; Thus, all of the results are based on the number of different observations.

In the empirical analysis, it is essential to kick start the analysis by using the simplest technique to give some preliminary intuition about what the data is telling us about. I start the analysis by using the distance based on the simplest measurement method used in most studies of the impact of transportation infrastructure development projects on house price, i.e., the Euclidean distance. Since Euclidean distance use a linear distance from the station, the study area will always be larger than other distance measures. This study area contains 35,672 observations. The Euclidean estimation in column 1 suggest

Table 6.

Conditional Average Treatment Effects

	(1) CATE	(2) Standard Error	(3) T value	(3) Freq.	(5) Percent
Treatment x Leaf 1	.3006	0.0512	5.872	1,290	10.90
Treatment x Leaf 2	.3009	0.0517	5.820	1,339	11.31
Treatment x Leaf 3	.3561	0.0469	7.586	765	6.46
Treatment x Leaf 4	.3965	0.0623	6.360	877	7.41
Treatment x Leaf 5	.4452	0.0385	11.566	1,520	12..84
Treatment x Leaf 6	.4538	0.0602	7.534	1,134	9.58
Treatment x Leaf 7	.4847	0.0612	7.916	1,030	8.7
Treatment x Leaf 8	.5348	0.0496	10.773	1,311	11.08
Treatment x Leaf 9	.5855	0.0463	12.642	1,161	9.81
Treatment x Leaf 10	.6478	0.0436	14.849	1,408	11.90
Leaf 1-10	YES				
Observations				11,835	

Notes: The table displays the results of CATE estimation with log housing price per square meter as the dependent variable, the first until the tenth leaves as the dependent variable, and the interaction terms between the treatment variable and the leaves as the average treatment effect in each leaf. The CausalTree algorithm does not compute standard errors by default. Thus, the standard error in this table is computed by running the standard linear regression of the leaves and the interaction terms on log housing price per square meter.

that there is a positive trend in the interactive term representing the treatment effect of the metro opening within the treated area, but the effect is not significant. Note again that Euclidean measurement is prone to encounter a measurement error problem. It seems that some parts of the treatment effect are still unobserved in this Euclidean result.

In column 2, I further explore the analysis by using conventional network analysis to deal with the endogeneity problem commonly found in this literature. This approach shrinks the total sample of the study into 26,488 observations. The result shows that the treatment effect starts to be significant and increases by 0.99% to 2.67% compared to the earlier Euclidean study. It can be implied that this standard approach succeeds to absorb some parts of unobserved treatment effects that are used to be in the error term of the Euclidean estimator.

In practice, standard network analyses are still prone to measurement error as some unobserved networks, which are usually found in the studies, may still not be taken into account. I argue that the use of total-travel-cost network analysis can handle this issue. Although there is a very small shrinkage in terms of the total sample, the result of the estimation using total-travel-cost network analysis (see shows that the treatment effect increases from 2.67% to 3.18% compared to the estimation using the standard approach. It implies that this approach can correct the measurement error that occurred in the standard estimation. This increase of the treatment effect indicates that total-travel-cost network analysis approach can absorb the omitted coefficients that is used to be captured in the error term of the estimation using the standard network analysis approach.

E. Heterogeneous Treatment Effect of Metro Line 52 on Housing Prices using Causal Tree

In this section, I extend the main analysis of the linear panel estimation by introducing the causal tree approach into the framework to disaggregate the overall average treatment effects and estimating the treatment effect heterogeneity. This causal tree approach yields ten estimates of conditional average treatment effects (CATE) that differs between the leaves. Table 6 reports the result of the CATEs estimation[13]. I also explore how much the average level of each housing characteristics changes across leaves (see Appendix E).

[13] The CausalTree package does not include the standard error computation by default. To compute it, I estimate $Log(sales\ price/sqm) = \sum_{\ell} \alpha_{\ell} L_{\ell} + W \,.\, \beta_{\ell} L_{\ell}$, where L_{ℓ} is defined to indicate the assignment to leaf ℓ, and W is the term for the binary treatment variable. The interaction coefficients in this specification represents the average treatment effects in each leaf, as $E[Y|W=1, L=1] - E[Y|W=0, L=1] = (\alpha_1 + \beta_1) - (\alpha_1) = \beta_1$. The standard error around the coefficients of the interaction terms is equal to the standard error around the CATEs.

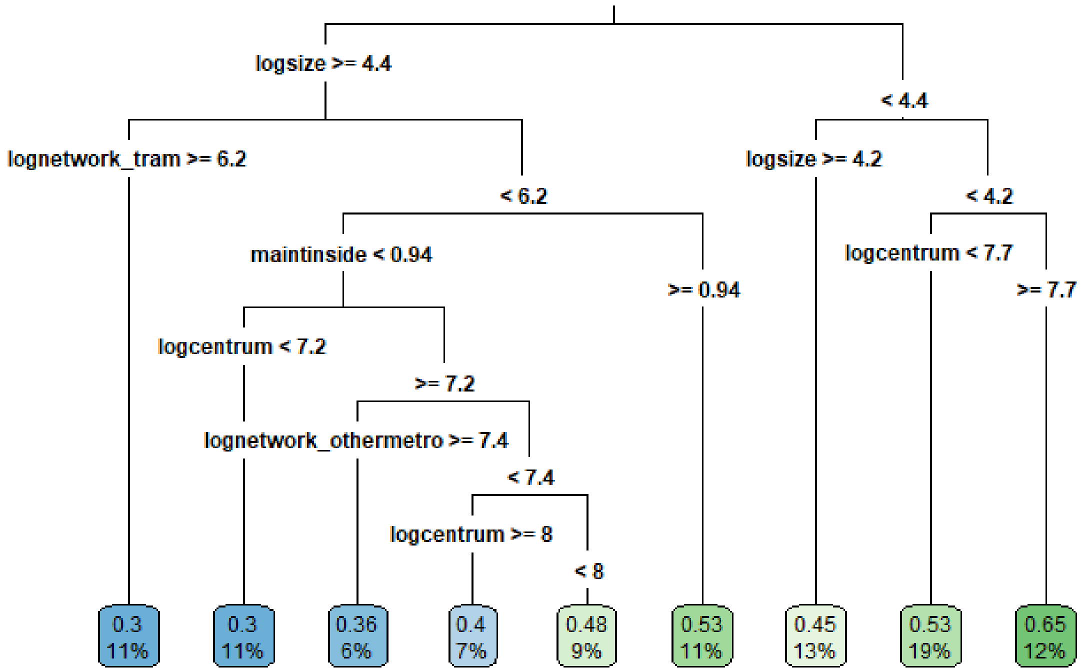

FIGURE 8. THE HONEST CAUSAL TREE RESULT USING THE INTERACTION BETWEEN TREATED AREAS AND POST DUMMIES AS THE BINARY TREATMENT VARIABLE

Note: the splitting process will stop until the leaf has at least 30 observations of either treated or untreated groups

The result shows that all CATEs display positive effects, and they are all statistically significant. Since all subpopulations that are located within the treated areas experience an uplift in housing price after the new metro line event relative to its subpopulation located beyond the treated areas, it seems that the binary treatment variable defined in this analysis works as intended. This may also make the robustness of the treatment assignment technique suggested in the previous section even stronger. Furthermore, by contrasting this result with the result from Chin et al. (2020) which finds 146 different estimates with 89 of which are positive and 53 of which are negative, and with the result from the implementation of their framework to the data used in this paper (see Figure 9) which also finds positive and negative estimates, it seems that the assignment of the interaction term between treated areas and post dummies works better in disaggregating the positive average treatment effect.

Furthermore, Figure 8 illustrates the regression tree diagram generated from the honest causal tree approach. In the figure, we can identify which subpopulation of houses has the highest and the lowest treatment effects. The properties benefiting the highest capitalization effects go to the properties with the sizes of lower than 67 m^2, i.e., equivalent to the log size of ln(4.2), and located more than 2.2 km, which is equivalent to the log distance to the city centre of ln(7.7), far away from the city centre of Amsterdam. This means that smaller size properties located far away from the city centre gain a higher capitalization than the other subpopulations. One explanation for this might be that smaller properties in such locations are more demanded as people are interested in a more affordable housing service as smaller properties tend to have a lower total nominal sale price while still benefiting from greater accessibility to the city centre the new metro line. In contrast, the properties with the size larger than 81 m^2, i.e., equivalent to log size of ln(4.4); located less than 493 m^2, i.e., equivalent to log network distance to tram station of ln(6.2), network distance from the nearest tram station; a maintenance score of the inside of lower than 0.94; as well as located less than 1.6 km, i.e., equivalent

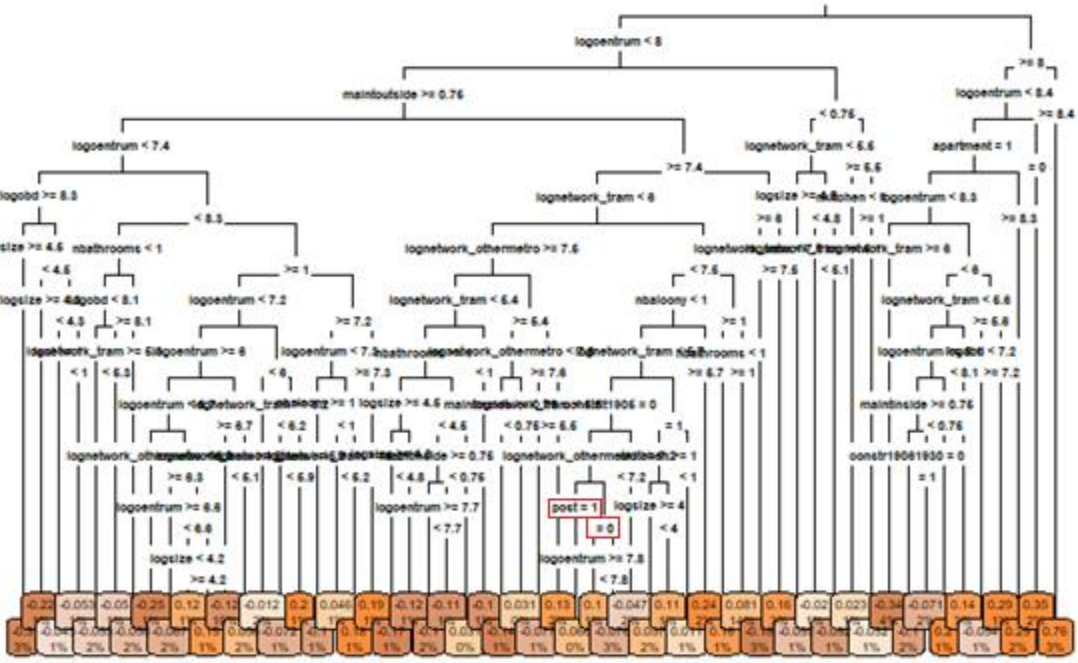

FIGURE 9. THE HONEST CAUSAL TREE ALGORTIHM USING ONLY THE TREATED AREAS DUMMY AS THE BINARY TREATMENT VARIABLE

Note: the splitting process will stop until the leaf has at least 30 observations of either treated or untreated groups and the red box indicates that the data is split based on the post dummy variable

to the log distance to the city centre of ln(7.2) far away from the city centre witness the lowest average treatment effect. It may indicate that a larger size of a house and a closer distance of a house to the city centre leads to a smaller treatment effect.

Lastly, I further add the regression tree result from using the treated area as the sole binary treatment variable (see Figure 9). It shows that the method leads to an incorrect CATE estimation. The method tells the algorithm to split the dataset based on the post dummy and assign it to the left leaf if the observation's post dummy is equal to 1 and the right leaf if the observation's post dummy is equal to 0. As a result, the left leaf only contains house transactions made after the new metro establishment, and the right leaf only contains hose transactions made before the event. In both leaves, the algorithm will only compute the price difference between the observations located within the treated areas and untreated areas without considering the house price difference between before and after the event. Consequently, this approach cancels out the DID framework in computing CATE within each leaf in the causal tree setting.

V. Conclusion

This paper argues that Amsterdam's Metro Line 52 establishment, a large rapid transit system infrastructure development project in Amsterdam from 1968 to 2018, constitutes a good case study to empirically study the effect of an RTS opening on the housing market. This event constitutes one of a few big transportation infrastructure projects invested in the urban area with a very dense transport network. Even though this quasi-experiment design might not be entirely random, I argue that this event can still be qualified as a rather exogenous event compared to other transportation development projects because this project has goals which are to reduce the car usage and travel time cost, improve reliability and accessibility, and create a more liveable urban area while not merely about improving the economic growth.

The estimates of the treatment effects over the time window from the pre-trend analysis presented in this paper suggest no significant anticipation effects prior to the opening of Metro Line 52. I argue that there might be a positive anticipation effect, but it is reduced by the negative sentiment perceived by society. Moreover, the result also does not show any significant negative effect that is probably driven by the noise because most of the construction work of this project occurred underground. However, I also found a significant anticipation effect that occurred as early as 1-year before the opening event, while the effect seems to disappear six months before the event.

To estimate the average treatment effects, this paper specifies the hedonic linear model with the difference in difference design and controlling for six-digits postcode and quarterly time fixed effects. The results show that the price of the houses located within 537-meter network distance to the nearest Metro Line 52 station increase by 3.18% relative to the houses located beyond the treated areas. The results are still robust even though the study boundary is extended from 1000 m to 1500 m. This result indicates that this paper's local polynomial regression approach can be an appropriate way to assign a house transaction observation into an accurate treated or controlled group. The result also shows that there are several significant treatment effects heterogeneity. This indicates that the willingness to pay of households vary depending on the heterogeneity in household characteristics.

I further check the spatial variations in the treatment effects by comparing treatment effect occurred in several different zones. The results confirm that the treatment effects are more substantial on the houses located in the first and second zones which are also located within the treated areas relative to the untreated areas. Furthermore, the result still shows a consistent significant positive effect for the first two zones after the third zone, which is located beyond the treated areas, included in the model, and it shows that there are no significant effects for houses in the third zone relative to the rest of the observations.

This paper also compares the treatment effects using another source of distance proxies of accessibility. I first argue that using Euclidean analysis would lead to measurement errors as it does not consider any spatial and topography obstacles while assuming that the distance can be drawn as a straight line between the origin and the destination. I also argue that using standard network analysis will also face the same measurement error problem as in most of the real practice cases; there is still unobserved distances that should be taken into account. The result confirms that the estimates applying the total-travel-cost network analysis approach are proven to correct the measurement error that can be found in the models applying standard network analysis or Euclidean approaches.

Apart from the main analysis, this paper also adopts the tree-based machine learning approach to estimate more personalized treatment effects by disaggregating the overall average treatment effect that is the typical object of interest in the previous analysis. The result shows that the highest treatment effect is received by the houses located within the treated areas, with a size of lower than 67 m2, and located 2.2 km away from the city centre. The framework of this approach modifies the method used in the earlier work by Chin et al. (2020), where this paper proposes to set the interaction term between the treated areas and post dummies as the binary treatment variable to estimate the CATE with a difference in difference correctly.

The results in this paper are important because they can provide a deep understanding of how accessibility from the improvement of transportation infrastructure is absorbed into the house price via the capitalization effect. This may be useful for regional or urban policy design to understand the mechanism of the effect works. Estimating a more personalized treatment effect may also be useful because knowing for whom the treatment works best can provide a better insight into knowing the true treatment mechanism. In addition, the use of machine learning in urban research provides a solution to compute a large number of dimensions of heterogeneity at a low cost.

All in all, I have investigated the overall winners from the large investment of a new metro line project in Amsterdam from 1968 to 2018 by exploring the effect of the opening of Metro Line 52 on the housing price in Amsterdam in this paper. I further extend the analysis to find the biggest winners among the overall winners by estimating the heterogeneous treatment effect of the Metro Line 52 opening on the housing market. This paper is one of a few papers in real estate economic literature that adopts machine learning techniques focusing on causal inference. In addition to the machine learning innovation, this paper also proposes several innovative approaches, such as the total-travel-cost network analysis to generate distances and local polynomial regression for treatment assignment, to contribute to the literature on the effects of urban transit infrastructure development.

Reference

Ahlfeldt, G. M., & Feddersen, A. (2018). From periphery to core: Measuring agglomeration effects using high-speed rail. *Journal of Economic Geography*. https://doi.org/10.1093/jeg/lbx005

Albert, S. (2010). The geographic determinants of housing supply. *Quarterly Journal of Economics*, *125*(3). https://doi.org/10.1162/qjec.2010.125.3.1253

Alonso, W. (1964). Location and land use. toward a general theory of land rent. Location and land use. Toward a general theory of land rent.nt. *Population Studies*.

Arnott, R. J., & McMillen, D. P. (2007). A Companion to Urban Economics. In *A Companion to Urban Economics*. https://doi.org/10.1002/9780470996225

Åslund, O., Blind, I., & Dahlberg, M. (2017). All aboard? Commuter train access and labor market outcomes. *Regional Science and Urban Economics*. https://doi.org/10.1016/j.regsciurbeco.2017.08.007

Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1510489113

Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*. https://doi.org/10.1257/jep.31.2.3

Athey, S., & Imbens, G. W. (2019). Machine Learning Methods That Economists Should Know about. *Annual Review of Economics*. https://doi.org/10.1146/annurev-economics-080217-053433

Athey, S., Imbens, G. W., & Wager, S. (2018). Approximate residual balancing: debiased inference of average treatment effects in high dimensions. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*. https://doi.org/10.1111/rssb.12268

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *Annals of Statistics*. https://doi.org/10.1214/18-AOS1709

Banister, D., & Berechman, Y. (2001). Transport investment and the promotion of economic growth. *Journal of Transport Geography*, *9*(3). https://doi.org/10.1016/S0966-6923(01)00013-8

Bartholomew, K., & Ewing, R. (2011). Hedonic price effects of pedestrian- and transit-oriented development. *Journal of Planning Literature*, *26*(1). https://doi.org/10.1177/0885412210386540

Baum-Snow, N., & Kahn, M. E. (2000). The effects of new public projects to expand urban rail transit. *Journal of Public Economics*. https://doi.org/10.1016/S0047-2727(99)00085-7

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

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). Classification and regression trees. In *Classification and Regression Trees*. https://doi.org/10.1201/9781315139470

Brueckner, J. K., Thisse, J. F., & Zenou, Y. (1999). Why is central Paris rich and downtown Detroit poor? An amenity-based theory. *European Economic Review*, *43*(1). https://doi.org/10.1016/S0014-2921(98)00019-1

Chin, S., Kahn, M. E., & Moon, H. R. (2020). Estimating the Gains from New Rail Transit Investment: A Machine Learning Tree Approach. *Real Estate Economics*. https://doi.org/10.1111/1540-6229.12249

Davis, J. M. V., & Heller, S. B. (2020). Rethinking the benefits of youth employment programs: The heterogeneous effects of summer jobs. *Review of Economics and Statistics*. https://doi.org/10.1162/rest_a_00850

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

Deng, T., Wang, D., Yang, Y., & Yang, H. (2019). Shrinking cities in growing China: Did high speed rail further aggravate urban shrinkage? *Cities*. https://doi.org/10.1016/j.cities.2018.09.017

Devaux, N., Dubé, J., & Apparicio, P. (2017). Anticipation and post-construction impact of a metro extension on residential values: The case of Laval (Canada), 1995–2013. *Journal of Transport Geography*. https://doi.org/10.1016/j.jtrangeo.2017.05.002

Diao, M., Fan, Y., & Sing, T. F. (2017). A new Mass Rapid Transit (MRT) line construction and

housing wealth: Evidence from the Circle Line. *Journal of Infrastructure, Policy and Development*. https://doi.org/10.24294/jipd.v1i1.22
Diao, M., Leonard, D., & Sing, T. F. (2017). Spatial-difference-in-differences models for impact of new mass rapid transit line on private housing values. *Regional Science and Urban Economics*. https://doi.org/10.1016/j.regsciurbeco.2017.08.006
Diao, M., Qin, Y., & Sing, T. F. (2016). Negative Externalities of Rail Noise and Housing Values: Evidence from the Cessation of Railway Operations in Singapore. *Real Estate Economics*. https://doi.org/10.1111/1540-6229.12123
Dubé, J., Legros, D., & Devaux, N. (2018). From bus to tramway: Is there an economic impact of substituting a rapid mass transit system? An empirical investigation accounting for anticipation effect. *Transportation Research Part A: Policy and Practice*. https://doi.org/10.1016/j.tra.2018.02.007
Dubé, J., Rosiers, F. Des, Thériault, M., & Dib, P. (2011). Economic impact of a supply change in mass transit in urban areas: A Canadian example. *Transportation Research Part A: Policy and Practice*. https://doi.org/10.1016/j.tra.2010.09.002
Duncan, M. (2008). Comparing rail transit capitalization benefits for single-family and condominium units in San Diego, California. *Transportation Research Record*, *2067*. https://doi.org/10.3141/2067-14
Fan, J., & Gijbels, I. (1994). Local Polynomial Modelling and its Applications. In *Local Polynomial Modelling and its Applications*. https://doi.org/10.1007/978-1-4899-3150-4
Fernald, J. G. (1999). Roads to prosperity? Assessing the link between public capital and productivity. *American Economic Review*, *89*(3). https://doi.org/10.1257/aer.89.3.619
Garcia-Mila, T., McGuire, T. J., & Porter, R. H. (1996). The Effect of Public Capital in State-Level Production Functions Reconsidered. *The Review of Economics and Statistics*. https://doi.org/10.2307/2109857
Gatzlaff, D. H., & Smith, M. T. (1993). The impact of the Miami Metrorail on the value of residences near station locations. *Land Economics*. https://doi.org/10.2307/3146278
Gibbons, S., & Machin, S. (2005). Valuing rail access using transport innovations. *Journal of Urban Economics*. https://doi.org/10.1016/j.jue.2004.10.002
Glaeser, E. L., Kolko, J., & Saiz, A. (2001). Consumer city. *Journal of Economic Geography*. https://doi.org/10.1093/jeg/1.1.27
Hastie, T., Tibshirani, R., & Friedman, J. (2009). Springer Series in Statistics The Elements of Statistical Learning - Data Mining, Inference, and Prediction. In *Springer* (Vol. 2nd).
Hess, D. B., & Almeida, T. M. (2007). Impact of proximity to light rail rapid transit on station-area property values in Buffalo, New York. *Urban Studies*. https://doi.org/10.1080/00420980701256005
Higgins, C. D. (2019). A 4D spatio-temporal approach to modelling land value uplift from rapid transit in high density and topographically-rich cities. *Landscape and Urban Planning*. https://doi.org/10.1016/j.landurbplan.2018.12.011
Higgins, C. D., & Kanaroglou, P. S. (2016). Forty years of modelling rapid transit's land value uplift in North America: moving beyond the tip of the iceberg. *Transport Reviews*. https://doi.org/10.1080/01441647.2016.1174748
Hoogendoorn, S., Van Gemeren, J., Verstraten, P., & Folmer, K. (2019). House prices and accessibility: Evidence from a quasi-experiment in transport infrastructure. *Journal of Economic Geography*, *19*(1). https://doi.org/10.1093/jeg/lbx027
Hussam, R., Rigol, N., & Roth, B. (2020). Targeting High Ability Entrepreneurs Using Community Information: Mechanism Design In The Field. *Working Paper*.
Jackson, S. L., & Buckman, J. (2020). Light rail development with or without gentrification?: Neighborhood perspectives on changing sense of place in Denver, Colorado. *Journal of Transport Geography*. https://doi.org/10.1016/j.jtrangeo.2020.102678
Jin, J., & Kim, D. (2018). Expansion of the subway network and spatial distribution of population and employment in the Seoul metropolitan area. *Urban Studies*. https://doi.org/10.1177/0042098017725476
Jiwattanakulpaisarn, P., Noland, R. B., Graham, D. J., & Polak, J. W. (2009). Highway infrastructure investment and county employment growth: A dynamic panel regression analysis. *Journal of*

Regional Science. https://doi.org/10.1111/j.1467-9787.2008.00579.x
Kunimi, T., & Seya, H. (2021). Identification of the geographical extent of an area benefiting from a transportation project: A generalized synthetic control. *Journal of Transport and Land Use*, *14*(1). https://doi.org/10.5198/jtlu.2021.1784
Landis, J., Guhathakurta, S., Huang, W., Zhang, M., Fukuji, B., & Sen, S. (1995). Rail transit investments, real estate values, and land use change: a comparative analysis of five California rail transit systems. *Monograph-Institute of Urban and Regional Development University of California*.
Levkovich, O., Rouwendal, J., & van Marwijk, R. (2016). The effects of highway development on housing prices. *Transportation*, *43*(2). https://doi.org/10.1007/s11116-015-9580-7
Li, T. (2020). The Value of Access to Rail Transit in a Congested City: Evidence from Housing Prices in Beijing. *Real Estate Economics*, *48*(2). https://doi.org/10.1111/1540-6229.12222
Lin, J. J., & Xie, Z. X. (2020). The associations of newly launched high-speed rail stations with industrial gentrification. *Journal of Transport Geography*. https://doi.org/10.1016/j.jtrangeo.2020.102662
Linden, L., & Rockoff, J. E. (2008). Estimates of the impact of crime risk on property values from megan's laws. *American Economic Review*. https://doi.org/10.1257/aer.98.3.1103
Mayer, T., & Trevien, C. (2017). The impact of urban public transportation evidence from the Paris region. *Journal of Urban Economics*. https://doi.org/10.1016/j.jue.2017.07.003
McDonald, J. F., & Osuji, C. I. (1995). The effect of anticipated transportation improvement on residential land values. *Regional Science and Urban Economics*. https://doi.org/10.1016/0166-0462(94)02085-U
McMillen, D. P., & McDonald, J. (2004). Reaction of house prices to a new rapid transit line: Chicago's midway line, 1983-1999. In *Real Estate Economics*. https://doi.org/10.1111/j.1080-8620.2004.00099.x
Meng, X., Lin, S., & Zhu, X. (2018). The resource redistribution effect of high-speed rail stations on the economic growth of neighbouring regions: Evidence from China. *Transport Policy*. https://doi.org/10.1016/j.tranpol.2018.05.006
Miller, S. (2020). Causal forest estimation of heterogeneous and time-varying environmental policy effects. *Journal of Environmental Economics and Management*, *103*. https://doi.org/10.1016/j.jeem.2020.102337
Mohammad, S. I., Graham, D. J., Melo, P. C., & Anderson, R. J. (2013). A meta-analysis of the impact of rail projects on land and property values. *Transportation Research Part A: Policy and Practice*. https://doi.org/10.1016/j.tra.2013.01.013
Mottee, L. K., Arts, J., Vanclay, F., Miller, F., & Howitt, R. (2020). Metro infrastructure planning in Amsterdam: how are social issues managed in the absence of environmental and social impact assessment? *Impact Assessment and Project Appraisal*, *38*(4). https://doi.org/10.1080/14615517.2020.1741918
Muth, R. F. (1969). Cities and housing; the spatial pattern of urban residential land use. In *American Journal of Agricultural Economics*.
Pilgram, C. A., & West, S. E. (2018). Fading premiums: The effect of light rail on residential property values in Minneapolis, Minnesota. *Regional Science and Urban Economics*, *69*. https://doi.org/10.1016/j.regsciurbeco.2017.12.008
Richardson, H. W., & Mills, E. S. (1973). Studies in the Structure of the Urban Economy. *The Economic Journal*. https://doi.org/10.2307/2231142
Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*. https://doi.org/10.1086/260169
Thünen, J. H. von. (1875). Der isolierte staat in Beziehung auf Landwirtschaft und Nationale Ökonomie. Band 1. In *Sammlung sozialwissenschaftlicher Meister*.
van den Ende, L., & van Marrewijk, A. (2019). Teargas, taboo and transformation: A neo-institutional study of community resistance and the struggle to legitimize subway projects in Amsterdam 1960–2018. *International Journal of Project Management*, *37*(2). https://doi.org/10.1016/j.ijproman.2018.07.003
Van Lohuizen, H. P. S. (1989). A new pre-metro line for Amsterdam? *Tunnelling and Underground Space Technology Incorporating Trenchless*, *4*(3). https://doi.org/10.1016/0886-7798(89)90079-

5
Wager, S., & Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, *113*(523). https://doi.org/10.1080/01621459.2017.1319839
Zhao, M., Liu, S., & Qi, W. (2017). Exploring the differential impacts of urban transit system on the spatial distribution of local and floating population in Beijing. *Journal of Geographical Sciences*. https://doi.org/10.1007/s11442-017-1403-7
Zheng, S. Q., & Kahn, M. E. (2008). Land and residential property markets in a booming economy: New evidence from Beijing. *Journal of Urban Economics*, *63*(2). https://doi.org/10.1016/j.jue.2007.04.010

Appendix

APPENDIX A. DESCRIPTIVE STATISTICS OF CONTROL VARIABLES AND THEIR EFFECTS ON LOG HOUSE PRICE USING HEDONIC APPROACH

Variable	Effect on log house prices
Auctioned sale	-.1767***
	(.0527)
Leasehold (erfpacht)	-.0206***
	(.006)
Property is partly rented	-.1566***
	(.04)
Period of construction	
Before 1906	(Reference)
Between 1906 and 1930	-.0068
	(.0043)
Between 1931 and 1944	-.0251***
	(.007)
Between 1945 and 1959	-.0233
	(.0153)
Between 1960 and 1970	-.0444
	(.0262)
Between 1971 and 1980	-.0529***
	(.0141)
Between 1981 and 1990	-.0438**
	(.0163)
Between 1991 and 2000	-.0046
	(.0112)
2000 and later	.0371**
	(.012)
Log (Size)	-.2087***
	(.0061)
Semidetached property	(Reference)
Apartment	-.1364***
	(.0162)
Terraced property	-.045**
	(.0145)
Detached property	-.0502
	(.0541)
Private parking space	.0935***
	(.0103)
Number of bathrooms	.016***
	(.0032)
Number of kitchens	-.0073*
	(.0035)
Number of balconies	-.0014
	(.0029)
Number of roof terraces	.0424***
	(.0037)
Room for (internal) office	-.0874*
	(.0393)
Maintenance score of the outside	.0864***
	(.0177)
Maintenance score of the inside	.2126***
	(.0137)
Maintenance state is good	.0441***
	(.0053)
Number of types of insulation	.007***
	(.0011)
Central heating	.0334***
	(.0055)
Listed building	.0316***
	(.0056)
Monumental	.0151**
	(.0053)

Variable	Effect on log house prices
Log (network distance to the nearest tram)	.0227*
	(.0088)
Log (network distance to other nearest metro station)	-.0022
	(.0317)
Log (distance to the CBD)	.2383***
	(.0713)
Log (distance to the city centre)	-.0035
	(.0691)
Observations	23668
R-squared	.7204
Adj R^2	.7191

Standard errors are in parentheses
**** p<.001, ** p<.01, * p<.05*

APPENDIX B. DATA SELECTION STRATEGY

No	Selection criteria	Number of observations
1	Initial dataset (housing transactions in Amsterdam from 2002 to 2020)	154,910
2	Remove cases with the network distance to the nearest Metro Line 52 stations higher than 1,000 m	23,954
3	Remove postal codes (4-digits) with less than 15 observations	23,941
4	Remove cases with unknown year of construction	23.668
5	Remove cases with zero size, unknown price, and unknown data in the other controlled variables	23.668

APPENDIX C. HOUSING CHARACTERISTICS IN EACH LEAF

	(1) Leaf 1	(2) Leaf 2	(3) Leaf 3	(4) Leaf 4	(5) Leaf 5	(6) Leaf 6	(7) Leaf 7	(8) Leaf 8	(9) Leaf 9	(10) Leaf 10
CATE	.3006	.3009	.3561	.3965	.4452	.4538	.4847	.5348	.5855	.6478
Structural Characteristics										
Log (size)	4.7460	4.8458	4.7723	4.7902	4.2759	3.9315	4.7296	4.7673	3.8845	3.9741
	(.0073)	(.0071)	(.0095)	(.0088)	(.0067)	(.0078)	(.0082)	(.0072)	(.0077)	.0070)
Property type										
Apartment	.6946	.7229	.9307	.9282	.9776	.9815	.9495	.9252	.9931	.9226
	(.0079)	(.0078)	(.0103)	(.0096)	(.0073)	(.0084)	(.0088)	(.0078)	(.0083)	(.0076)
Terraced	.2333	.2577	.0601	.0399	.0164	.0097	.0466	.0633	.0043	.0526
	(.0072)	(.0071)	(.0094)	(.0088)	(.0067)	(.0077)	(.0081)	(.0072)	(.0076)	(.0069)
Semidetached	.0628	.0157	.0092	.025	.0039	.0009	.0029	.0053	-9.6731e-18	.0234
	(.0034)	(.0033)	(.0044)	(.0041)	(.0031)	(.0036)	(.0038)	(.0034)	(3.5613e-03)	(.0032)
Detached	.0093	.0037	.0000	.0068	.0020	.0079	.0010	.0061	.0026	.0014
	(.0018)	(.0018)	(.0023)	(.0022)	(.0016)	(.0019)	(.0020)	(.0018)	(.0019)	(.0017)
Parking	.2473	.0515	.0510	.1197	.0362	.0088	.0126	.1053	.0224	.0142
	(.0067)	(.0066)	(.0087)	(.0081)	(.0061)	(.0071)	(.0075)	(.0066)	(.0070)	(.0064)
Number of bathrooms	.9364	1.1397	1.0327	1.0388	.8836	.8298	1.0311	1.0625	.8665	.8871
	(.0153)	(.0150)	(.0198)	(.0185)	(.0141)	(.0163)	(.0171)	(.0151)	(.0161)	(.0146)
Number of kitchens	.7860	.8962	.8667	.8062	.8026	.7698	.8680	.7963	.7339	.7685
	(.0137)	(.0134)	(.0178)	(.0166)	(.0126)	(.0146)	(.0153)	(.0136)	(.0144)	(.0131)
Number of balconies	.4178	.2323	.6562	.7252	.5013	.2751	.6058	.4729	.4005	.5724
	(.0146)	(.0143)	(.0189)	(.0177)	(.0134)	(.0156)	(.0163)	(.0145)	(.0154)	(.0140)
Number of roof terraces	.1659	.3413	.2444	.1186	.1164	.1208	.2097	.2906	.0922	.0575
	(.0107)	(.0105)	(.0139)	(.0130)	(.0099)	(.0114)	(.0120)	(.0106)	(.0113)	(.0102)
Private office	.0000	.0007	.0000	.0000	.0000	.0000	.0000	.0008	.0000	.0000
	(.0004)	(.0004)	(.0005)	(.0004)	(.0003)	(.0004)	(.0004)	(.0004)	(.0004)	(.0003)
Maintenance outside	.8110	.7429	.7498	.7519	.7896	.7937	.7374	.9216	.7916	.7749
	(.0029)	(.0029)	(.0038)	(.0036)	(.0027)	(.0031)	(.0033)	(.0029)	(.0031)	(.0028)
Maintenance inside	.7968	.7191	.7276	.7054	.7863	.7973	.7102	1.0000	.7996	.7666
	(.0037)	(.0036)	(.0048)	(.0045)	(.0034)	(.0039)	(.0041)	(.0037)	(.0039)	(.0035)
Good maintenance	.8946	.8536	.8575	.8039	.9013	.9030	.8165	1.0000	.8898	.8786
	(.0088)	(.0086)	(.0114)	(.0106)	(.0081)	(.0094)	(.0098)	(.0087)	(.0092)	(.0084)
Number of	1.6171	.7864	1.1333	.9031	1.4447	1.2434	1.0223	1.9085	1.6813	1.3359

	(1) Leaf 1	(2) Leaf 2	(3) Leaf 3	(4) Leaf 4	(5) Leaf 5	(6) Leaf 6	(7) Leaf 7	(8) Leaf 8	(9) Leaf 9	(10) Leaf 10
CATE	.3006	.3009	.3561	.3965	.4452	.4538	.4847	.5348	.5855	.6478
insulation types										
	(.0453)	(.0444)	(.0588)	(.0549)	(.0417)	(.0483)	(.0507)	(.0449)	(.0477)	(.0433)
Central heating	.7977	.8880	.8863	.9019	.9125	.8810	.8893	.9352	.8398	.8771
	(.0090)	(.0088)	(.0116)	(.0109)	(.0083)	(.0096)	(.0100)	(.0089)	(.0094)	(.0086)
Listed property	.1116	.3129	.0222	.0547	.0684	.1552	.0330	.1098	.0026	.0618
	(.0080)	(.0078)	(.0104)	(.0097)	(.0073)	(.0085)	(.0089)	(.0079)	(.0084)	(.0076)
Monumental	.1031	.2928	.0301	.0456	.0691	.1737	.0485	.1228	.0103	.0270
	(.0079)	(.0078)	(.0103)	(.0096)	(.0073)	(.0084)	(.0089)	(.0078)	(.0083)	.0076)
Transactional Characteristics										
Period of construction										
Before 1906	.2403	.7857	.4314	.0011	.2914	.5406	.3019	.4058	.4617	.1875
	(.0122)	(.0120)	(.0159)	(.0148)	(.0113)	(.0130)	(.0137)	(.0121)	(.0129)	.0117)
Between 1906 and 1930	.1101	.0642	.4431	.3877	.2928	.1182	.5204	.2990	.3445	.5355
	(.0120)	(.0117)	(.0155)	(.0145)	(.0110)	(.0128)	(.0134)	(.0119)	(.0126)	.0114)
Between 1931 and 1944	.0667	.0119	.0170	.2862	.0743	.0397	.0990	.0625	.0301	.1222
	(.0072)	(.0071)	(.0093)	(.0087)	(.0066)	(.0077)	(.0081)	(.0071)	(.0076)	.0069)
Between 1945 and 1959	.0504	.0052	.0013	.1642	.0158	.0132	.0019	.0198	.0017	.0185
	(.0043)	(.0042)	(.0056)	(.0052)	(.0040)	(.0046)	(.0048)	(.0043)	(.0045)	.0041)
Between 1960 and 1970	.0775	.0090	.0013	.0741	.0664	.0265	.0000	.0084	.0078	.0348
	(.0048)	(.0047)	(.0063)	(.0059)	(.0045)	(.0052)	(.0054)	(.0048)	(.0051)	.0046)
Between 1971 and 1980	.1016	.0157	.0013	.0000	.0487	.0362	.0029	.0031	.0138	.0192
	(.0044)	(.0043)	(.0057)	(.0054)	(.0041)	(.0047)	(.0050)	(.0044)	(.0047)	.0042)
Between 1981 and 1990	.0217	.0336	.0235	.0023	.0724	.0750	.0087	.0076	.0551	.0604
	(.0053)	(.0052)	(.0069)	(.0064)	(.0049)	(.0057)	(.0059)	(.0053)	(.0056)	.0051)
Between 1991 and 2000	.0899	.0642	.0431	.0011	.0651	.0917	.0563	.0625	.0172	.0064
	(.0064)	(.0060)	(.0079)	(.0074)	(.0056)	(.0065)	(.0068)	(.0060)	(.0064)	.0058)
2000 and later	.2419	.0105	.0379	.0832	.0730	.0591	.0087	.1312	.0680	.0156
	(.0071)	(.0069)	(.0092)	(.0086)	(.0065)	(.0076)	(.0079)	(.0070)	(.0075)	.0068)
Auction	.0031	.0015	.0052	.0000	.0013	.0009	.0029	.0000	.0017	.0014
	(.0011)	(.0011)	(.0015)	(.0014)	(.0011)	(.0012)	(.0013)	(.0011)	(.0012)	.0011)
Leasehold	.5450	.0530	.3242	.9225	.4191	.0802	.5029	.2952	.2679	.5874
	(.0119)	(.0117)	(.0155)	(.0145)	(.0110)	(.0127)	(.0133)	(.0118)	(.0126)	.0114)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Leaf 1	Leaf 2	Leaf 3	Leaf 4	Leaf 5	Leaf 6	Leaf 7	Leaf 8	Leaf 9	Leaf 10
CATE	.3006	.3009	.3561	.3965	.4452	.4538	.4847	.5348	.5855	.6478
Partly rented	.0023	.0075	.0131	.0000	.0000	.0000	.0126	.0000	.0009	.0000
	(.0015)	(.0015)	(.0020)	(.0019)	(.0014)	(.0017)	(.0017)	(.0015)	(.0016)	.0015)
Locational Characteristics										
Log (Distance to other nearest metro station)	7.2148	6.8410	7.5006	6.6669	7.1035	6.7506	7.1220	7.0275	7.2614	7.4712
	(.0138)	(.0135)	(.0179)	(.0167)	(.0127)	(.0147)	(.0154)	(.0136)	(.0145)	.0132)
Log (Distance to the closest tram station)	7.1017	5.5284	5.3930	5.3889	5.8164	5.5774	5.1728	5.3841	5.5317	5.8642
	(.0187)	(.0184)	(.0243)	(.0227)	(.0172)	(.0200)	(.0209)	(.0186)	(.0197)	.0179)
Log (Distance to CBD)	8.1507	8.0745	7.4124	6.5994	7.7532	8.0765	7.4824	7.5628	7.7074	7.5657
	(.0145)	(.0142)	(.0188)	(.0176)	(.0133)	(.0154)	(.0162)	(.0144)	(.0153)	.0139)
Log (Distance to Centrum)	7.4850	6.7931	7.6838	8.1449	7.4322	6.7130	7.7172	7.4250	7.5350	7.8627
	(.0150)	(.0147)	(.0195)	(.0182)	(.0138)	(.0160)	(.0168)	(.0149)	(.0158)	.0144)