RUHR
ECONOMIC PAPERS

Joschka Flintz

The Value of Passenger Rail Access

RUHR UNIVERSITÄT BOCHUM RUB rwi

#1114

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #1114

Responsible Editor: Thomas Bauer

ISSN 1864-4872 (online) – ISBN 978-3-96973-292-2

Ruhr Economic Papers #1114

Joschka Flintz

The Value of Passenger Rail Access

Bibliografische Informationen der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at **http://dnb.dnb.de**

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

https://dx.doi.org/10.4419/96973292
ISSN 1864-4872 (online)
ISBN 978-3-96973-292-2

Joschka Flintz*

The Value of Passenger Rail Access

Abstract

This study quantifies the effect of train station openings on residential house prices across Germany over a 12-year period to provide generalizable evidence on the valuation of access to passenger rail by households. It is based on data on about 90 train station openings between 2009 and 2020 in Germany and a Difference-in-Differences model that uses three different control group variants to cover alternative assumptions about unobserved regional heterogeneity to mitigate problems arising from endogenous transport infrastructure provision. The results indicate that train station openings increase residential house prices on average by 5% (€18,000) within a distance of up to two to three kilometers. Notably, these positive effects are observed exclusively for properties without prior access to passenger rail services, and are significantly larger in more densely populated and urban areas.

JEL-Codes: C23, R21, R40

Keywords: Public transportation; regional passenger rail; hedonic price model; difference-in-differences; spatio-temporal analysis

October 2024

* *Joschka Flintz, RUB. – I am thankful for valuable feedback and comments by the participants of the 17th North American Meeting of the Urban Economics Association, the 13th ifo Dresden Workshop on Regional Economics, the 2024 Annual Conference of the International Transport Economics Association, the 3rd Causal Inference with Spatial Data Workshop, and the 62nd European Regional Science Association Congress, as well as the members of the research training group (RTG) 2484 – "Regional Disparities and Economic Policy". Special thanks go to Sven Werenbeck-Ueding, Johannes Gallé and Thomas Bauer for their suggestions and support. I gratefully acknowledge financial support from the German Research Foundation through the research training group (RTG) 2484 – "Regional Disparities and Economic Policy". – All correspondence to: Joschka Flintz, RUB, e-mail: joschka.flintz@ruhr-uni-bochum.de*

1 Introduction

Extensive and efficient public transportation infrastructure plays a pivotal role in the spatial distribution of economic activity and household location decisions by enhancing accessibility across regions (Mayer and Trevien 2017, Ahlfeldt and Feddersen 2018, Büchel and Kyburz 2020, Gibbons et al. 2024). This improved connectivity translates into reduced commuting times that ameliorate employment opportunities (Heuermann and Schmieder 2019, Tyndall 2021) and allows residents to meet their preferences for sustainable travel (Cao and Cao 2014), rendering locations near public transit stops increasingly attractive. In particular when it comes to connecting more rural areas with larger cities and facilitating commuting and travel over longer distances, the importance of an extensive regional passenger rail network becomes evident. As governments increasingly recognize the importance of shifting to more sustainable modes of transportation[1], access to rail services holds even greater appeal for individuals as policies are introduced to increase the attractiveness of public transportation, such as fare reduction programs (Cats et al. 2017, Liebensteiner et al. 2024). Consequently, residing near a train station becomes an increasingly attractive feature. This heightened appeal is expected to drive a surge in demand, creating an upswing in property prices that is indicative of households' appreciation of access to the public transport network (Rosen 1974).

This study analyzes the impact of train station openings in Germany between 2009 and 2020 on the prices of residential properties located near the stations taken into operation in order to estimate the value households place on living near a station that provides access to passenger rail services. The empirical analysis builds upon a novel, hand-collected data set on train station locations and openings for the period of 2009 to 2020 for Germany and a geocoded repeated cross-section of all residential advertisements on "Immobilienscout24", Germany's largest online platform for real estate. The train station commissionings build a quasi-experiment, which I exploit for identification in a staggered Difference-in-Differences setting using the fixed effects estimator proposed by Sun and Abraham (2021).

The results indicate that the commissioning of train stations that provide access to regional passenger rail services increases the prices of residential houses within up to two to three kilometers of the station by approximately 5%. Importantly, those effect sizes are observed exclusively for properties lacking prior access to passenger rail services. In contrast, houses already situated near another train station do not show any increase in value after the opening of a nearby station, indicating no utility gains associated with having access to a second or third train station. Furthermore, a heterogeneity analysis suggests substantially larger effects of train station openings on property prices in more urban areas compared to less densely populated or poorer neighborhoods. In absolute

[1] For instance, the German Federal Government aims to double rail passenger volumes by 2030 to meet emission targets (UBA 2024).

terms, the estimated effects translate to an average increase in house prices of €18,000 and a cumulative rise in property values of about €238 million for the entire sample, which corresponds to an average property price effect of €2.8 million per opened train station. While this figure already signifies a substantial appreciation for access to regional passenger rail services by households, it is crucial to recognize that this valuation forms a lower bound as not only households purchasing new homes benefit from enhanced accessibility, but also those who have perennially resided near the station that has recently commenced operations.

This study investigates the impact of commissioning regional passenger rail stations on property prices over a 10-year period, encompassing an entire country. In contrast, prevailing research concerned with the relationship between passenger rail infrastructure and real estate prices predominantly focuses on local rail services, such as subway and light rail transit, located in densely populated metropolitan areas (Gibbons and Machin 2005, Diao et al. 2017, Ahlfeldt et al. 2019, Welch et al. 2018, Ransom 2018, Brandt and Maennig 2012, Hess and Almeida 2007, Song et al. 2019). Moreover, most of the literature and especially recent studies, that leverage infrastructure openings or expansions for identification, consider only a single transportation network or project in their analysis. While a large body of research finds positive effects of public transport infrastructure commissionings on house prices (Gibbons and Machin 2005, Billings 2011, Dubé et al. 2013, Diao et al. 2017, Cohen and Brown 2017, Ke and Gkritza 2019, Rojas 2024), a number of studies report ambiguous and negative results (Ahlfeldt 2011, Chatman et al. 2012, Wagner et al. 2017, Devaux et al. 2017, Pilgram and West 2018, Ransom 2018). The mixed findings indicate that area-specific features, such as residents' travel preferences, the built environment and access to alternative transportation systems, can significantly affect the valuation of public transport access. Hence, results may be challenging to extrapolate and generalize, limiting their broader applicability. By analyzing the implications of multiple train station commencements distributed across an entire country over more than a decade, this study addresses the existing gap in generalizable findings and provides empirical evidence on a large scale, offering insights into households' appreciation for access to regional passenger rail services.

Moreover, to mitigate the problem of endogenous infrastructure placement in the identification of causal effects, and to enhance the results' robustness, this study introduces an extensive research design leveraging three distinctive control groups based on different regions, that serve as a counterfactual to areas in which a new train station is taken into operation. The three control group variants complement each other by covering alternative assumptions about the comparability of different locations and, thus, mitigate issues related to biased estimates through a poor selection of control group areas. For the different control group variants, I distinguish between properties within a certain distance band from the opened train stations, selected to be far enough away to be unaffected by possible accessibility improvements, but close enough to share similar location-specific characteristics, houses within proximity of stations that are similar in important characteristics to the train stations taken into operation, and properties located in villages or neighborhoods along former rail lines, where a major public transport industry association has recommended reactivating passenger rail service.

While earlier research predominantly employs cross-sectional hedonic pricing models to estimate the benefit of enhanced accessibility through transportation infrastructure (e.g. Bowes and Ihlanfeldt 2001) more recent studies use quasi-experimental settings in form of infrastructure openings or expansions in a Difference-in-Differences setting for the identification of causal effects (among others, Gibbons and Machin 2005, Billings 2011, Dubé et al. 2013, Diao et al. 2017, Wagner et al. 2017, Pilgram and West 2018, Fesselmeyer and Liu 2018) as it alleviates the problems of omitted variable bias, conventional cross-sectional hedonic price regressions fall prone to. However, it remains imperative to find a control group that closely mirrors the treatment group in order to obtain unbiased estimates. While Billings (2011) and Wagner et al. (2017) aim to establish a counterfactual scenario by selecting properties along an alternative/proposed transit route, the majority of studies either consider every dwelling that does not belong to the treatment group as part of the control group (e.g. Mohammad et al. 2017, Im and Hong 2018, Rojas 2024), or restrict the control group to houses outside the treatment area but within a certain threshold of the opened station (e.g. Diao et al. 2017, Fesselmeyer and Liu 2018, Ransom 2018, Yazdanifard et al. 2021). As a result, estimates may be biased due to lack of comparability between areas or due to spillover effects in the control group (Butts 2023), and sensitive to the selection of control group areas within a city (Pilgram and West 2018). In line with this notion, this analysis yields estimates that are somewhat comparable, yet vary substantially between the different control group variants, by up to more than four percentage points. This underscores the critical importance of a thoughtful and well-founded control group selection in this kind of research designs.

In the following, section 2 presents the data used in the analysis and section 4 explains the methodological approach. After that, section 3 and section 5 show descriptive statistics and the analysis' results, before section 6 concludes.

2 Data

The main data source for the empirical analysis is a repeated cross-section of detailed housing data at a monthly level obtained from the Research Data Center of the RWI - Institute for Economic Research. For the period from 2009 to 2020, the data contains all German advertisements for houses for sale on "Immobilienscout24", Germany's largest online real estate platform, facilitating a nationwide analysis of households' appreciation of access to passenger rail services (RWI and ImmobilienScout24 2022). The data holds detailed information on property characteristics, including offering price and geolocation, allowing the data to be spatially merged with various other spatial data sources to add information on socio-economic characteristics on the municipality and one square kilometer grid levels, as well as distances to highway entrances, railways, central places and train stations. As the results should represent households' valuation of access to passenger rail services, I exclude apartment buildings, that are likely purchased for investment purposes. Furthermore, I discard dilapidated buildings in need of renovation as well as properties with unusual characteristics, such as castles and houses constructed before 1700, with a minimum plot area of less than $50m^2$ or less than $30m^2$ living space. To increase representativeness, the data sample is restricted to properties in the range of 100,000€ to 1,500,000€ and houses in the highest 1% with respect to price per square meter have been dropped. The prices are inflation adjusted with 2015 as reference year.[2]

To add further control variables, I use commercial data on socio-economic neighborhood characteristics provided by the RWI's Research Data Center, which includes information on population, purchasing power, unemployment, building type composition, as well as gender and age distribution, among other things, at the one-square-kilometre grid level (RWI and microm 2022). Administrative borders as well as georeferenced information on the railway network are given by the federal agency for cartography and geodesy as of January 2020 (BKG), freeway entrance locations (2,382) and post code district borders are taken from OpenStreetMap as of April 2022, and information on central locations is provided by the federal office for building and regional planning (BBSR).[3] Tables A.1, A.2 and A.3 provide an overview of all control variables in the data set.

The georeferenced train station data is obtained from Germany's national rail provider "Deutsche Bahn" as a cross-section of April 2020 (DB Station&Service AG 2020). The data was aligned with information on train station locations from OpenStreetMap as of April 2022 and from public transport timetables for the first eight months of 2020 in order to perform validity checks. Only train stations that serve regular passenger transport are considered, i.e., train stations exclusively serving touristic purposes or freight transport are discarded from the sample. The train station cross-section was enriched with information

[2]For binary variables, I interpret missing values as zero since all of these variables are desirable features that users are likely to advertise (Schaffner 2020). Missing values in an advert's year of construction are recoded as the average construction year and the observations concerned flagged with an indicator variable. Moreover, observations that lack information on the house's category (e.g. single-family house or terraced house) as well as the object's condition I assign to a artificially created, separate category "Missing Information".

[3]The definition of central locations follows the German application of Walter Christaller's system of central places (Christaller 1980), that determines a location's significance based on its infrastructure compared to its surrounding environment's conditions. In Germany, municipalities are distinguished between higher-order centers, medium-order centers, lower-order centers and municipalities that only cover basic services, e.g. supermarkets or kindergardens. Lower-order centers provide services such as retail businesses, doctors and postal offices, medium-order centers contain institutions like cinemas and hospitals to fulfill periodic needs, while higher-order centers also provide access to establishments that offer specific services such as universities, museums and federal authorities.

on station openings[4] during the study period, obtained from "Allianz pro Schiene", an association of non-governmental organizations and companies in the railroad industry that is dedicated to the promotion and improvement of rail transport, from "Deutsche Bahn", and through own research efforts.

The train station data is augmented by accessibility measures that indicate how well a station is connected within the rail network (Table A.4). The measures are based on General Transit Feed Specification (GTFS) data from December 2019 to August 2020 that contains information on passenger rail time tables. The time tables are used to calculate travel time matrices that allow to determine the minimum travel duration between all train stations.[5] On the basis of these travel times, a number of different connectivity measures are calculated, which indicate the degree of accessibility provided by a station. For each station, I obtain the number of train stations and the largest station, in terms of population around the station, that can be reached without changing trains, as well as the travel time to the main station of the closest larger city, and the inversely travel time-weighted sum of accessible population. For the latter, I determine the average population within one kilometer of each train station over the study period, using population counts on the one square kilometer grid level. Then, for every train station, I sum up the population counts surrounding all accessible train stations, taking into account the inverse weight based on the time needed to reach each station. Table B.1 presents descriptive statistics on various station attributes.

A total of 6,433 train stations have been served by public passenger transport during the period of 2009 until 2020 with 6,305 stations being still active at the end of 2020. In the study period, 220 train stations have been taken into operation, either on a rail line that was already served by passenger rail, or in context of the (re-)activation of a complete railway segment.[6] The opened train stations are comparable to all stations in operation during the study period in terms of socio-economic neighborhood characteristics, but tend to be located in more central locations and exhibit poorer connectivity within the train station network. There is no correlation between the timing of station commencement and attributes (Table B.2). 353 train stations in Germany serve long distance rail, two of which were added during the study period. The spatial distribution of German train stations shows that the majority of stations is located in more densely populated areas

[4] The term "opening" is used for newly opened as well as reactivated train stations. Train station relocations or train stations that have been commissioned to replace an already existing station are not considered as "station openings".

[5] The travel time matrices are calculated using the R-package "tidytransit" and allowing for a maximum of two transfers (it is assumed that at least eight minutes are required to transfer), and ignoring the possibility of transferring, respectively.

[6] For example, seven new train stations were taken into operation in December 2013 as the rail segment "Heinsberg - Lindern" connecting Heinsberg with the German rail network was reactivated. In contrast, in September 2013 a new train station opened in Munich's district Freiham and was added to the already active rail line connecting Munich-Pasing with Herrsching in Munich's south-western hinterland. The respective train stations are highlighted as orange triangles on the right hand side map in Figure 1.

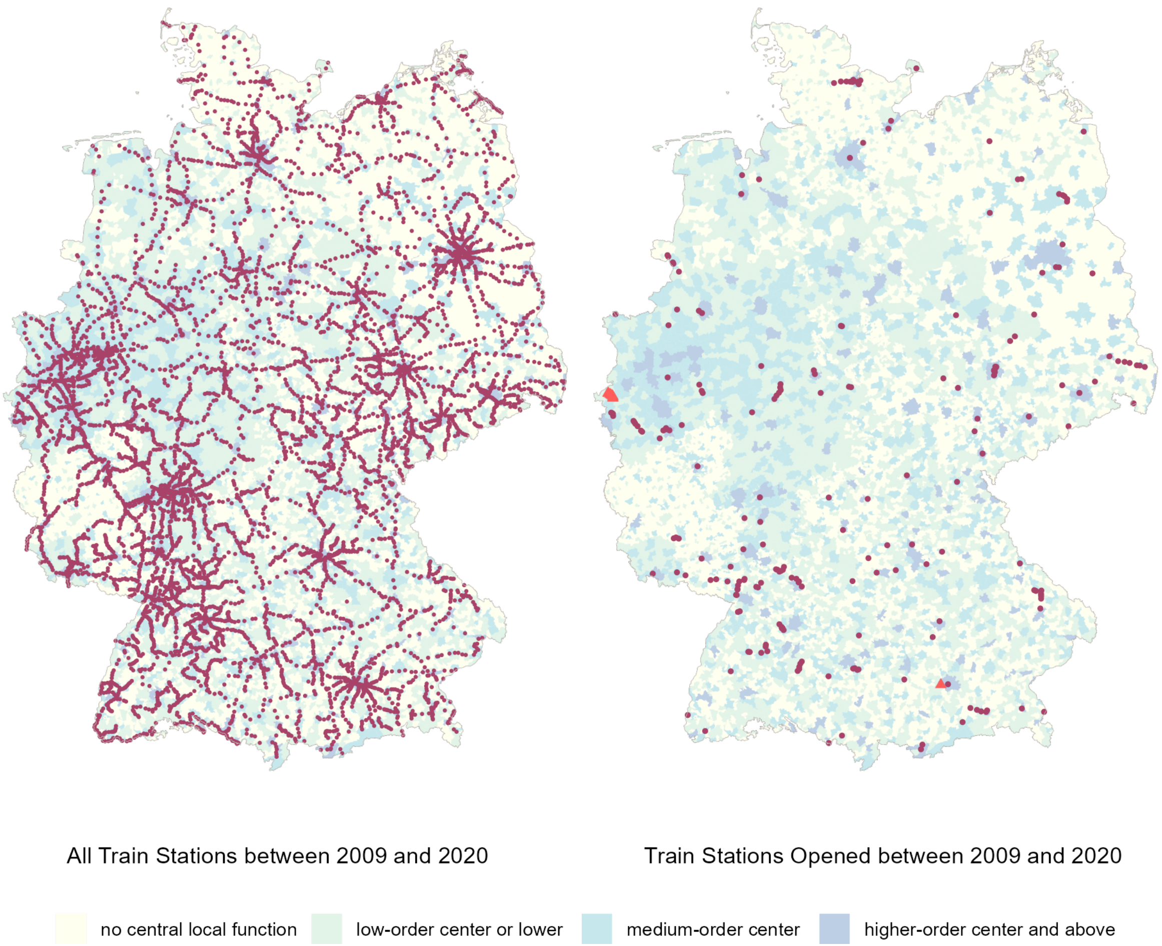

Figure 1: Train Stations in Germany

while more rural places often lack access to the German rail network (Figure 1). There is about one train station per 100 square kilometers in municipalities covering only basic services, 1.6 in lower-order centers and 2.5 in medium-order centers, while municipalities designated as higher-order centers have 6.3 train stations per 100 square kilometers.

3 Descriptive Statistics

The cleaned data set on residential houses for sale contains 5,053,385 observations across 12 years. Each year at least 261,286 (2019) houses are observed. The maximum number of observations per year is 556,965 and occurs in 2009. The houses are on average 3.5km located away from the nearest train station that is operational at the point of observation. 50% of the houses are located within 2.2km of a train station, of which about 1.3 million are within 1km. In contrast, residents of about 300,000 properties have to travel more than 10km to reach the nearest train station (Figure 2).

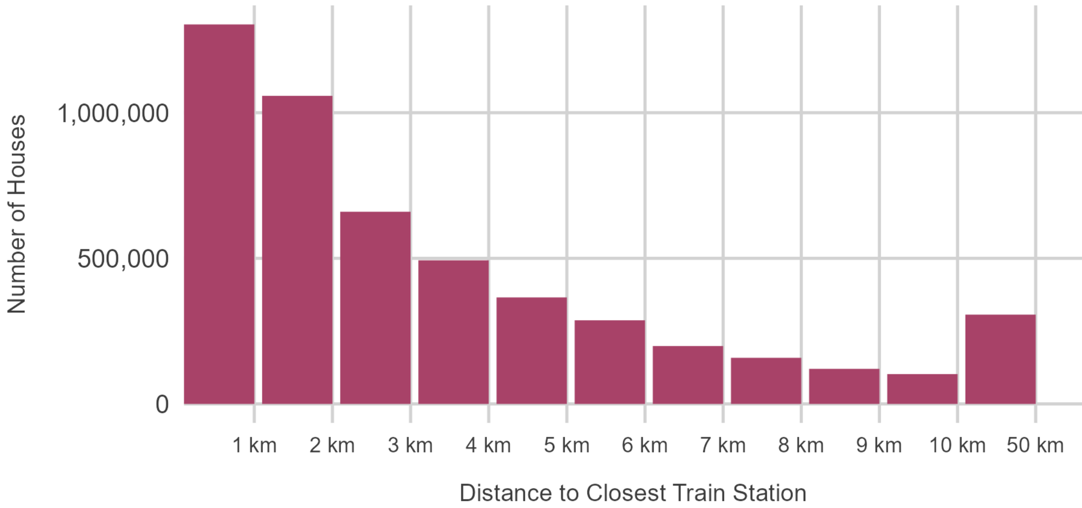

Figure 2: Number of Houses by Distance to Closest Train Station

Table 1 presents a comprehensive comparison of properties situated within a 2km radius of train stations and those beyond that threshold, confirming the notion that train stations tend to be located in more central places. Properties nearby train stations are not only in closer proximity to higher-order and other regional centers, but also located in neighborhoods characterized by substantially higher population counts (2,240 vs. 1,140) and a greater number of commercially utilized buildings (160 vs. 70). Furthermore, due to spatial constraints in central locations, properties near train stations exhibit significantly smaller plot areas compared to those situated beyond the two kilometers threshold ($644m^2$ vs. $792m^2$). This is reflected in a higher concentration of single-family houses in real estate located farther from train stations (62% compared to 55%), and a larger proportion of families living further away from the nearest train station in more rural areas (35% compared to 29%). Notably, despite the marked difference in the average plot area, the two groups demonstrate comparability in terms of living space and number of rooms ($163m^2$ and 5.6 on average, respectively), indicating that space restrictions primarily impact the size of gardens or driveways rather than overall living conditions. Furthermore, both groups exhibit similarities in various building characteristics and the purchasing power (22,300€) of the one square kilometer grid cell in which the houses are situated. However, the proportion of housing units within 500m of a railway is significantly higher for houses within 2km of a train station (46.5% compared to 8.1%). Given the general trend of higher housing costs in central areas, attributed to increased access to amenities and reduced commuting expenses, properties near train stations command a notably higher average price (€326,000 compared to €290,000) and price per square meter (€2,074 compared to €1,854) than their counterparts located farther from train stations.

For houses within 10km of any station, Figure 3 provides more detail on the negative

Table 1: House and Neighborhood Characteristics by Distance to Closest Train Station

	Train < 2km		Train > 2km		
	Mean	SD	Mean	SD	Difference
House Characteristics					
Price	326,243	191,812	290,205	164,044	36,038
Price per m^2	2,074	934	1,854	832	219
Plot Area	644.5	491.4	792.3	620.0	-147.8
Living Space	162.9	80.7	162.7	74.0	0.2
No. of Rooms	5.59	2.23	5.55	2.05	0.04
Age	33.70	37.87	32.10	37.17	1.60
Guest Toilet (%)	49.46	50.00	47.61	49.94	1.85
Basement (%)	32.87	46.97	28.46	45.12	4.41
Holiday House (%)	2.43	15.38	3.59	18.60	-1.16
Protected Building (%)	0.75	8.62	0.44	6.59	0.31
Construction (%)	8.34	27.65	8.02	27.17	0.31
Single-family House (%)	54.60	49.79	62.09	48.52	-7.49
First Occupancy (%)	21.98	41.41	20.51	40.38	1.47
Neighborhood Characteristics					
Population	2,241	1,923	1,137	1,236	1,104
Purchasing Power PC	22,313	3,922	22,262	3,628	51
No. Commercial Buildings	159.6	197.5	70.2	82.8	89.4
Unemployment Rate (%)	6.00	3.74	4.65	3.09	1.35
Share of Foreigners (%)	8.43	6.33	5.87	4.94	2.55
Share of Families (%)	29.33	17.09	35.36	20.75	-6.03
Share Above Age 60 (%)	27.96	4.75	27.51	4.85	0.46
Euclidean Distances					
Distance Highway (km)	6.64	6.75	8.52	7.74	-1.88
Distance Regional Center (km)	2.58	2.22	3.61	2.42	-1.03
Distance Higher-order Center (km)	17.41	11.96	21.24	13.52	-3.83
Railway < 500m (%)	46.45	49.87	8.10	27.29	38.35
Distance Closest Train Station (km)	0.97	0.51	5.63	3.62	-4.65

2,361,039 properties are located within 2km of a train station, and 2,692,346 are located beyond this threshold. All differences are statistically significant on the 1%-level.

relationship between prices and distance to the closest train station. The average house at 500m distance of the closest train station is valued at about €314,000, while the on average highest prices are asked for houses located at around 1.5km away from the closest train station (ca. €332,000). Beyond that, prices decrease steadily and fall below the overall mean asking price of €307,000 at around 4km. The lowest real estate values (ca. €245,000 on average) are found in properties that are located about 10km away from their closest train station.

Turning attention to those houses situated within 4km of a train station, Figure 4 shows

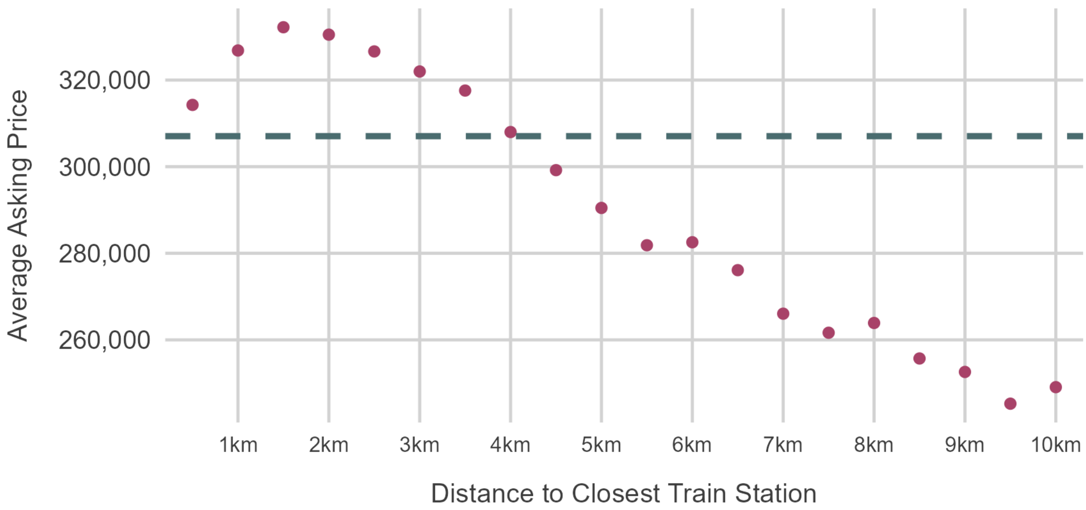

Figure 3: House Prices and Distance to Closest Train Station

The figure shows the average asking price within 500m bands of the closest train station for houses within 10km of a station, as well as the average price of all properties (dashed line).

the train stations' effect of residential house prices relative to real estate beyond 4km of the closest train station, controlling for various property and neighborhood characteristics as well as employing halfyear- and post code district fixed effects. The results demonstrate a hump-shaped pattern with lower property values in immediate vicinity to the stations, peak magnitudes at around 400m to 1.5km, and declining accessibility premiums with increasing distance. For residential real estate situated between 400m and 1.5km from a train station, premiums of more than €10,000 occur. After 1.5km the train station's positive effect declines, so that the estimate for houses between 3km and 4km is close to zero and statistically insignificant. Similarly, the coefficient for houses within 200m is small and statistically not distinguishable from zero, suggesting countervailing effects in close proximity to the stations due to - among others - noise pollution and congestion.

4 Empirical Approach

In order to estimate the causal effect of proximity to train stations on residential house prices, this study employs a Difference-in-Differences approach that utilizes train station openings to isolate variation in the supply of rail infrastructure. Following Rosen (1974) and Banzhaf (2021), the estimated changes in property values associated with the commissioning of these stations can be construed as the households' valuation of residing near a train station and consequently, their access to passenger rail services.[7] The Difference-in-Differences model is estimated using the technique proposed by Sun and Abraham

[7] While cross-sectional analyses yield coefficients that can be interpreted as households' average willingness to pay for living nearby a train station, Banzhaf (2021) demonstrates that the Difference-in-Differences approach identifies an effect that can be understood as households' average willingness to accept the foregone benefits of the realized train station opening.

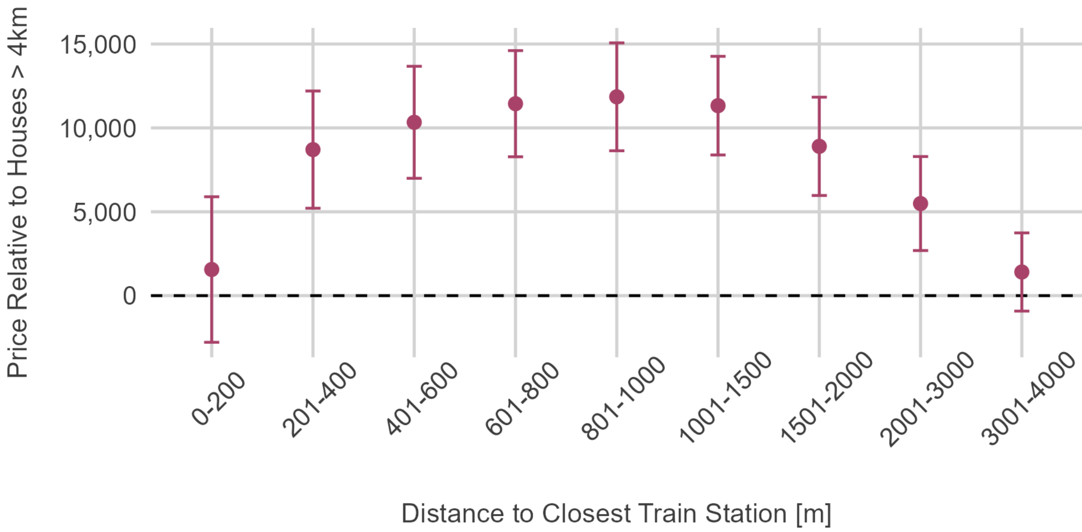

Figure 4: House Price Gradient (Pooled Fixed Effects Regression)

The figure shows the house price premium within different distance bands of the closest train station, relative to properties beyond 4km of a station. The estimates are obtained by a pooled regression of house prices on property and neighborhood characteristics as well as distances to important locations, incorporating post code district and year fixed effects.

(2021)[8] to mitigate the problem of treatment effect heterogeneity in settings of staggered treatment adoption (Goodman-Bacon 2021, Roth et al. 2023, Baker et al. 2022) and is given by

$$ln(p_{it}) = \gamma_g + \tau_{tl} + \sum_{e \notin C} \sum_{k \neq -1} \delta_{ek}(\mathbb{1}[E_i = e] D_{it}^k) + \beta X_{it} + \epsilon_{it}, \tag{1}$$

where g denotes the treatment and control group area in which house i is located, t its year of observation, and l the labor market region it belongs to. The approach developed in Sun and Abraham (2021) estimates separate treatment effects $CATT_{ek}$ for each station opening year-cohort e in the year relative to the commissioning k (adoption cohort-specific treatment effects). The proposed two-way fixed effects specification incorporates an interaction term between relative year-indicators (D_{it}^k) and adoption cohort indicators ($\mathbb{1}[E_i = e]$). The year in which the station nearby property i is taken into operation is represented by E_i. The estimated coefficients $\hat{\delta}_{ek}$ are the respective Difference-in-Differences estimators for the $CATT_{ek}$ and can be averaged over cohorts e to obtain the aggregate treatment effect in year relative to station opening k. For the aggregation, each $CATT_{ek}$ is weighted by its cohort's sample share in the respective year relative to station opening.

The dependent variable is given by the logarithm of the asking price[9] and δ_{ek} are the

[8] I implement the estimator using the R-package ”fixest” (Bergé 2018).

[9] Asking prices typically serve as a negotiation starting point, potentially leading to a downward adjustment in the eventual transaction price and hence, to inflated estimates. However, property transfer

coefficients of interest. γ_g and τ_{tl} are a station opening-treatment group fixed effect and a fixed effect on the year-labor market region level, to account for unobserved time-invariant location characteristics and region-specific shocks over time. Standard errors are clustered on the station opening-treatment group level. X_{it} is a vector of control variables for property and neighborhood characteristics as well as distances to important locations. The real estate attributes that are considered are: age of the building, plot area, living space, number of rooms, house category (e.g. mansion, single-family house, etc.), condition of the property, as well as indicator variables for whether the house is under construction, a protected building, holiday house, has a guest bathroom, or has a basement. Included neighborhood characteristics on the one square kilometer grid level are the unemployment rate, population, purchasing power per capita, the share of foreigners, families and people older than 60 years, as well as the number of commercial buildings. Additionally, I control for the distances to the closest highway entrance and railway, as well as the closest medium- and higher-order center.

In Germany, states are responsible for providing public transportation and receive federal funding to operate services as well as maintain and invest in infrastructure, which they can allocate to municipalities that contract with rail transport companies to operate passenger rail services, and to rail infrastructure companies that provide railways and stations for use by the transport companies. Proposals for where to open a train station or which railway segment to reactivate are often made by municipalities and local associations, and must pass potential and feasibility evaluations that examine various factors, such as the passenger potential, current infrastructure conditions, associated accessibility improvements and estimated costs, before detailed planning begins. The commissioning of new infrastructure is partly paid for by the infrastructure company but is largely financed by the state's public transport fund. To obtain funding, a comprehensive cost-benefit analysis evaluates the profitability of the project based on induced demand and changes in modal split, changes in travel time, CO2 emission reductions, other environmental impacts, changes in noise pollution as well as accident damages.

Due to the strategic selection of new train station locations, regions where new stations have recently commenced operation might exhibit distinctive features that differ from areas that have not gained access to passenger rail services. If those regional characteristics are unobserved and not controlled for, the estimated treatment effect is biased as it captures the effects of both rail infrastructure provision and the environment's features (Baum-Snow and Ferreira 2015). Similarly, real estate in areas with newly opened stations may systematically differ in key housing attributes from properties in other regions. However, the detailed information on property attributes available in the real estate data

taxes and notary costs that are contingent on the transaction price and thus, indirectly influenced by the presence of nearby train stations, countervail the overestimation through the usage of asking prices. Moreover, using logarithmized prices as dependent variable yields estimates expressed in percentage change, that remain unaffected by differences between asking and transaction price.

enables to control for an extensive set of housing characteristics, minimizing concerns regarding biased estimates stemming from unobserved property features.

The Difference-in-Differences approach addresses the threat of biased estimates due to unobservables at the regional level by taking first-differences over time to eliminate dissimilarities in unobserved spatial features between the treatment and control group that are invariant over time, such as the presence of the historical city center, soil quality or access to recreational areas. Additionally, the comparison of treatment and control group allows to control for time-varying unobservables that are constant across both groups, including both nationwide influences on real estate prices, such as changes in national construction regulations, as well as unobserved localized developments that are similar specifically between the treatment and control group. Consequently, the Difference-in-Differences approach identifies the causal effect as long as there are no changes in unobservable regional characteristics that differ across treatment and control group. This assumption would be violated if there are systematic, unobserved changes in the surrounding environment of the treatment group that do not occur in the control group, e.g., due to openings of new retail establishments and schools or the construction of a waste plant. In such scenarios, the property price effects of these changes would erroneously be attributed to the opening of the train station. The availability of small-scale socio-economic data, including population counts, purchasing power, unemployment rates, and number of commercial buildings facilitates the coverage of various regional developments in the analysis. However, for the identification assumption to hold, it remains imperative to select a control group that closely mirrors the treatment group, in particular with respect to time-varying unobservable regional characteristics.

In order to cover alternative versions of the formulated identification assumption, I construct three distinctive control groups based on different regions, that serve as compelling counterfactuals to areas in which a new train station is taken into operation for varying reasons. The first one is composed of houses located within a distance band of the opened train stations (green band in Figure 5), while the second variant contains properties in proximity of train stations similar in characteristics to the opened stations (dark red circle in Figure 5). The remaining control group variant consists of housing units nearby hypothetical train station locations in areas along former rail lines that have been recommended to be reactivated for passenger rail service (yellow circle in Figure 5). The treatment group is given by properties in proximity to the train station that is taken into operation (light red circle in Figure 5). For each train station opening, I construct the station-specific treatment and control groups and stack the resulting data sets separately for each control group variant, resulting in three estimation samples. The station opening-treatment group fixed effects γ_g in Equation (1) are based on the areas defining the treatment group and control group of an opened station.

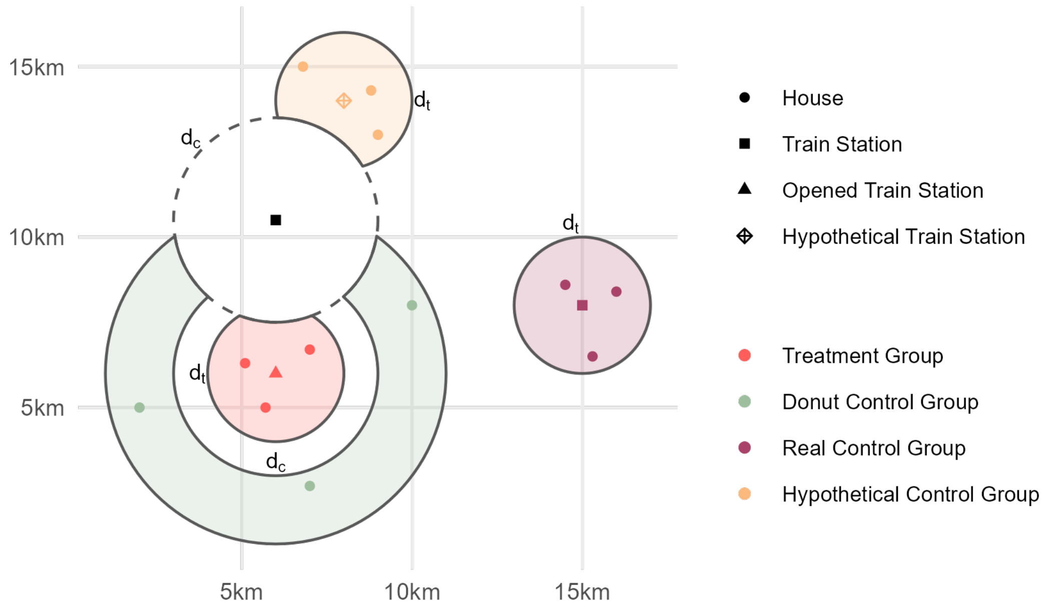

Figure 5: Illustration of Treatment and Control Groups

The treatment group consists of houses that are within treatment range d_t, e.g. 2km, of any opened station, irrespective of whether the train station was actually served by passenger transport when the house was observed. All properties that are located within distance d_c of any train station that has been served by passenger rail since before the start of the study period are discarded in order to prevent capturing the house price effects of those train stations (colorless circle in Figure 5). The distance threshold d_c exceeds d_t by 1km to ensure that no housing unit is affected by any other train station than the opened one, even in case of overly pessimistic assumptions about the spatial decay of train stations' property price effects. If a house is within range d_t of multiple opened stations and observed after they have been taken into operation, it gets discarded if the stations do not belong to the same rail line. Otherwise it is kept and allocated to the closest station. Dwellings located near multiple opened stations, that differ in whether they are serviced or not when the property is observed, are assigned to the operational train station and are considered treated by this station. Properties located in proximity of multiple stations that are scheduled for future opening, are again allocated to the closest station.

The first control group variant builds on the assumption that properties within a specific distance of an opened train station face similar changes in unobservable environment features and thus, real estate prices would develop in parallel in the absence of train station access. The control group consists of housing units far enough away from the opened station to not be affected by the opening, but close enough to be comparable in unobserved trends to houses that experience the increase in accessibility in close proximity to the station taken into operation. Effectively this means, that real estate in a distance band of 2km around the station, e.g. 3km to 5km, builds the control group. In order to ensure that no unit of the control group experiences a price increase caused by the station commissioning, I employ a buffer of 1km between the treatment ring and the control band and ignore all houses located in this buffer (the buffer's lower bound is given by d_t and the upper bound by d_c). For the same reason, I also discard any properties that are located within distance d_c of any other train station. The group design assimilates a donut and is in the following referred to as the donut control group.

The second control group variant consists of houses in proximity d_t of train stations that are similar to the station taken into operation and have been in operation throughout the entire study period. Similarly, the third control group utilizes properties within treatment range d_t of hypothetical train station locations in areas along former rail lines that have been recommended for reactivation of passenger rail service by the association of German transport companies (VDV), a major public transport industry association in Germany. The recommendations are made on the basis of the passenger potential that could be developed by, among other things, extending an existing regional rail line, creating a missing regional rail link between high-volume areas, or developing an underserved region (VDV 2020). Commonly, the neighborhoods or villages along the rail line used to have active train stations. In some cases, stations still exist and are currently served by touristic or seasonal rail, or freight transport runs on the respective railways. Conversely, in some instances the railway infrastructure has already been dismantled. The former rail lines are categorized according to their priority for reactivation and I focus only on areas near those rail segments that are classified as of highest priority, resulting in 320 hypothetical station locations.[10]

The identification assumption underlying the use of the second and third control group is that regions in which a new train station is about to open are similar in unobserved time-varying characteristics to areas that already have access to passenger rail or to regions, where train station access is considered very important to be implemented. Consequently, property prices would develop in parallel in the absence of the newly constructed train station. Since the control group variants are determined by real, operational train stations and hypothetical station locations, respectively, they are referred to as the real and hypothetical control group in the following. Notably, for the real control group, housing units that are also located within d_c of any opened station are excluded from the control group, while for the hypothetical control group, real estate within distance d_c of any train station is discarded to guarantee no capitalization of train station accessibility in the control group's property prices.

For the construction of the real and hypothetical control group, to each opened station, I assign three/one control station(s) using propensity score matching based on the stations' accessibility measures (long distance rail indicator, inversely travel time-weighted sum of

[10] Locations in which a train station used to be in operation during the study period are discarded.

accessible population, shortest travel time to the main station in a higher-order center), neighborhood characteristics (number of commercial buildings, population count and the average purchasing power per capita within 1.5km of the train station), as well as indicators for the stations' municipalities' growth and centrality, and the Euclidean distance to the closest municipality with a central function (Table A.4). Notably, since the hypothetical train station locations are not actually equipped with access to the rail network, matching based on stations' accessibility attributes is only possible for the real control group. Tables B.4 and B.5 present the balance in station characteristics between opened stations and real/hypothetical control stations, both before and after the propensity score matching. For the real control group, matching results in a more balanced sample with respect to almost all matched station characteristics. Similarly, matching opened train stations to hypothetical station locations reduces the mean difference for nearly all station attributes. However, this matching process also introduces a greater imbalance in terms of municipal growth, as the proportion of hypothetical stations in shrinking municipalities decreases, while it increases in growing municipalities.[11]

The three different control group variants are designed to allow for alternative assumptions about the determinants of train station siting and offer distinct advantages and disadvantages. The donut control group has been a popular choice in previous analyses (e.g. Ransom 2018, Diao et al. 2017, Yazdanifard et al. 2021, and Fesselmeyer and Liu 2018) as researchers argue that due to the spatial proximity, the control group area is similar in important characteristics to the treatment region and thus, provides a valid counterfactual. However, it is important to consider the possibility that prices of properties in the control group may be directly or indirectly affected by the station opening, leading to biased estimates (violation of the Stable Unit Treatment Value Assumption). While the 1km buffer I employ between the treatment ring and the control group band mitigates problems of higher price levels in the control group through improvements in accessibility, property prices in the donut control group could experience a negative impact following the commissioning of the train station because the region may become comparatively less desirable as the neighborhood around the newly opened station gains attractiveness.

In contrast, the concern of SUTVA violations due to reductions in relative attractiveness of the control group area is minimized when employing the real control group, as stations defining the control groups are on average situated about 300km away from their respective opened station (98km from the nearest station that became operational in the same year as the station opening). Conversely, using the real control group variant could introduce bias due to the presence of another station in the control group area. If the opening of a

[11] It is important to note that, due to the staggered treatment adoption, each property in the control group must be allocated to a station opening to determine the treatment timing. For the main analysis, this is accomplished through station-level matching. However, the results remain robust when matching individual properties to each other.

train station has a sustained effect on house prices that evolves over time, then house price developments near already existing stations would not represent the counterfactual trends in the treatment group, as long as the effect does not level out over longer time horizons and the train stations used to define the real control group have not been in operation for a sufficient period of time.[12] Additionally, the presence of train stations nearby real estate in the control group could introduce downward bias due to positive network effects associated with the introduction of new stations, through which the accessibility of regions proximate to interconnected train stations is augmented (Fesselmeyer and Liu 2018, Chernoff and Craig 2022). However, due to the relatively large distances between the control group train stations and their respective opened station, the probability that properties in close proximity derive advantages from enhanced accessibility diminishes.

The hypothetical control group comes closest to mimicking a true counterfactual, since it is based on the notion that both treatment and control areas would be eligible to receive a station, while only one region was actually provided with station access. Similarly, Billings (2011) and Wagner et al. (2017) utilize areas along proposed alternative rail lines to construct their control groups. Similarly to the real control group, using the hypothetical control group alleviates issues arising through reductions in the relative attractiveness of properties in the control group area as control units are not necessarily located in proximity to the opened train station. The average distance between a hypothetical station and it's respective opened station is 350km (116km for the nearest station that commenced operations in the same year as the station opening). Additionally, in comparison to the real control group, employing the hypothetical control group specification does not fall prone to potential biases occurring due to the presence of another train station.

5 Results

5.1 Main Results

Table 2 shows the estimated average effects of the train station openings on property prices within 2km ($d_t = 2km$) for the three control group variants.[13] All coefficients are positive as well as statistically significant on the 1%-level. The most substantial effect is observed when using the hypothetical control group, with an estimated increase of 7.6%. In contrast, the employment of the real control group and donut control group yields smaller coefficients, indicating an average property price increase of 5.1% and 4.9%, respectively. However, the estimated values do not differ statistically significant from each other. In absolute terms, these accessibility premiums amount to €18,700 and €18,600

[12] A one-time level increase in real estate prices after the start of operations does not pose a problem as the Difference-in-Differences approach utilizes developments over time for identification.

[13] The donut control group comprises housing units within 3km to 5km of the opened station, while the real and hypothetical control group contain dwellings within 2km of the operational and hypothetical train stations.

Table 2: The Effect of Train Station Openings on House Prices

	Donut Controls	Real Controls	Hypothetical Controls
Train Station < 2km	0.0490**	0.0511**	0.0759**
	(0.0111)	(0.0123)	(0.0147)
No. Observation	67,908	180,492	74,505
Adj. R2	0.6919	0.6383	0.7137

This table shows the average effect of train station openings on house prices based on Equation (1) with $d_t = 2km$. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level (in parentheses).
*** p < 0.01*

(€102/m^2 and €93/m^2) using the donut and real control group, and €28,400€ (€175/m^2) utilizing the hypothetical control group (Tables C.1 and C.2). The results are based on a sample of opened stations with at least 20 houses within d_t over the whole study period. For $d_t = 2km$, this leads to 90, 85 and 91 station openings considered in the analysis and treatment group sizes of 28,400 to 29,100 for the different analysis samples (Table C.3). The subset of opened stations utilized in the analysis does not differ significantly from the entire set of stations that have been put into operation (Table C.4). Table C.5 shows the estimates for all variables included in Equation (1).

The estimates remain comparable when d_t is set to 1.5km, 2.5km and 3km, indicating an accessibility premium within a radius of up to 3km from the opened train station that attenuates only modestly as the distance between the property and the station increases (Table C.6).[14] However, an analysis focusing on houses located between 2km to 3km off the opened station suggest that the commencement of station operations does not translate into higher residential house prices beyond 2km of the opened train station (Table C.7) but that the positive aggregate effect for $d_t = 3km$ is driven by dwellings in closer proximity to the station. For properties within 1km of the opened station, results indicate positive effects of 5.7% to 11.5% while residential real estate situated between 1km and 2km of newly opened train stations experiences a surge in prices following the begin of station operations of 3.1% to 4.3%.

Figure 6 displays event study estimates for $d_t = 2km$ and their 95%-confidence intervals for the years pre- and succeeding the year in which the train station was taken into operation ($relative\ year = 0$).[15] The pre-treatment estimates derived from both the

[14] The values for d_t are chosen based on the results of a cross-sectional regression presented in Figure 4, that show positive effects up to 3km that decline after 1.5km. This is in line with previous research that consistently reports positive accessibility premiums extending up to a distance of 1.5km from public transport stops that serve local rail services (Rennert 2022), and the notion that individuals might be willing to travel greater distances to access a train station serving regional rail as it facilitates more extensive travel.

[15] Due to decreasing observation counts in the years preceding and following the station opening, all dwellings that are observed in years more than 7 years before or after the opening of the station are

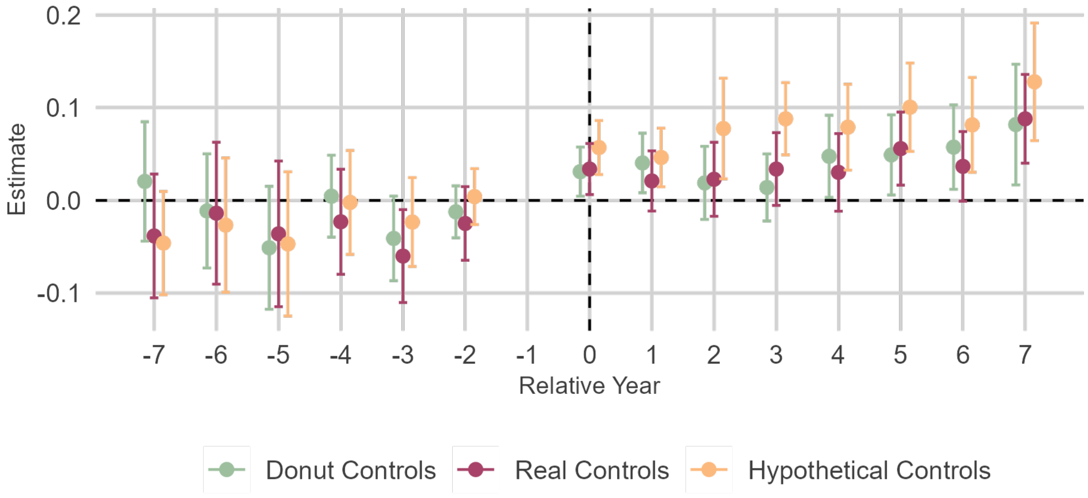

Figure 6: The Effect of Train Station Openings on House Prices Over Time

The figure shows the event study estimates and 95%-confidence intervals of Equation (1) *with* $d_t = 2km$ *for the years pre- and succeeding the year in which the train station was taken into operation (relative year* $= 0$*). Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level. All houses that are observed in years more than 7 years before or after the opening of the station are binned in relative year* -7 *and* 7*, respectively.*

donut and real control group do not indicate any discernible disparities in real estate price trends preceding the station opening. In contrast, estimates obtained using the hypothetical control group may suggest a subtle violation of the identification assumption, likely resulting in upwardly biased estimates and providing an explanation for the elevated effect sizes obtained by using the hypothetical control group. This issue prevails in a number of robustness checks that suggest a more pronounced increase in house prices in regions where stations become operational in the future, compared to properties located in areas constituting the hypothetical control group. Similar results can be found for d_t set to 1.5km and 2.5km, while for $d_t = 3km$, Figure C.2 reveals a positive development in real estate prices for the treatment group prior to the station opening relative to all control group variants. As consequence, estimates obtained by setting $d_t = 3km$ are likely to be biased upward.

Comparing the post-treatment estimates with the pre-treatment estimates, a clear upward trend in the effects becomes evident following the commencement of station operations. Moreover, the estimated effects tend to experience an increase in magnitude over time, suggesting that the opening of a train station not only has an immediate positive effect on real estate values in proximity but that this effect gradually intensifies with time in operation. The property price growth associated with the opening of an adjacent train station is particularly pronounced when using the hypothetical control group, while effect sizes increase more gradually when the donut and real control groups are used. This

binned in *relative year* -7 and 7, respectively.

pattern suggests that estimates obtained using the real control group may be biased downwards, possibly explaining the slightly lower estimates obtained by the analyses based on the real control group.

Given that the future inauguration of a former or new train station is typically announced well before the actual opening, individuals might factor in the anticipated access to regional passenger services when deciding to purchase a property, even before the station becomes operational (Billings 2011, Cohen and Brown 2017, Diao et al. 2017, Im and Hong 2018, Ke and Gkritza 2019). Unfortunately, it was not possible to collect information on station opening announcements for the stations used in the analysis. However, Figure 6 does not provide any clear evidence of anticipation effects before the year of the station commissioning. In line with this notion, omitting observations in $relative\ year = -1, -2$ to alleviate concerns of downward biased estimates due to positive anticipation effects changes the main results in Table 2 only slightly. Notably, positive effects during the year of the station opening can to a large extent be attributed to anticipation effects. In Germany, changes in train schedules, including the addition of new stops, typically occur twice a year, either at the end of June or in late December. Most of the opened train stations in the sample were inaugurated at the end of December (ca. 58%, 80% in the second half of the year), while a much smaller number opened at the end of June or on irregular dates. Consequently, many properties in vicinity of an opened station did not experience treatment in $relative\ year = 0$ but only in the subsequent year.

The results are robust to the exclusion of influential station openings, which demonstrate to alter the results significantly when omitted from the analysis, different choices of time fixed effects, control and matching variables, as well as the criteria for selecting stations based on the required number of nearby observed houses. Moreover, using a regular two-way fixed effects regression or the approach proposed by Gardner (2022), the allocation of control group units to houses in the treatment group by matching on the property level rather than on the train station level, or the estimation of a Triple-Differences model, does not significantly alter the results. For a more in-depth discussion of the robustness checks, please refer to appendix D.

5.2 Heterogeneity Analysis

To gain further insights into the spatial decay of the impact of train station openings on house prices, I utilize the Difference-in-Difference estimator proposed by Butts (2023). This method builds on the partitioning-based least squares approach developed by Cattaneo et al. (2020) to derive a treatment effect curve as a function of distance from the newly opened train station. Essentially, I estimate the relationship between residential real estate prices and proximity to opened train stations before and after the start of service, controlling for building and neighborhood characteristics, as well as distance to

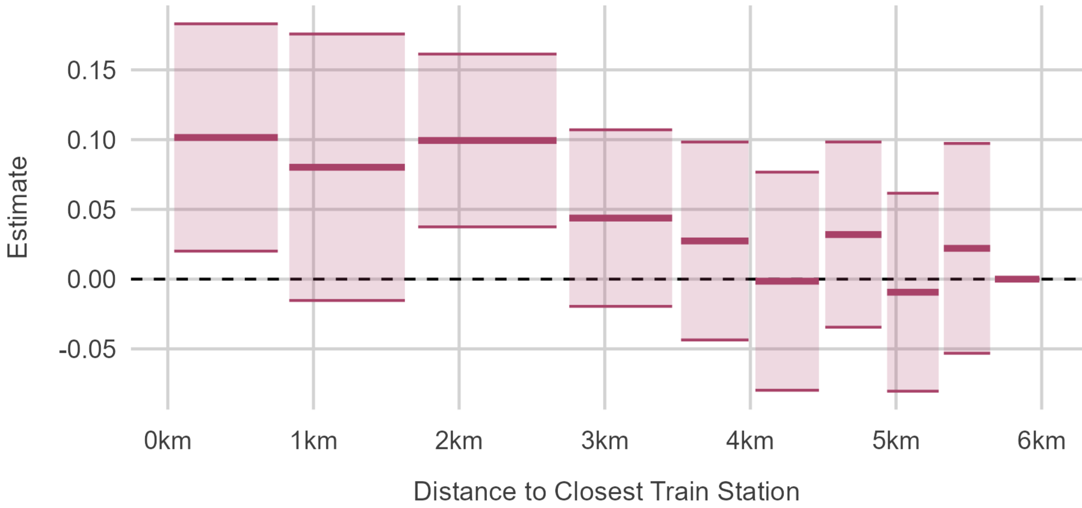

Figure 7: The Effect of Train Station Openings on House Prices Across Distance

The figure presents the treatment effect curve estimated using the approach proposed in Butts (2023) with a fixed number of 10 bins. Standard errors are clustered on post code district level.

key locations, and partitioning the distance to the train station in a data-driven manner.[16] For each distance bin, the difference in property values before and after the station opening is calculated and compared to the change in house prices in the bin furthest from the train station, which serves as the control group.

Figure 7 presents the estimated treatment effect curve derived for a a fixed number of 10 bins, Figures C.3 and C.4 show the estimates for a fully autonomous binning approach and 16 bins. The results indicate an increase in property values associated with the opening of a nearby train station, extending up to a distance of 2km to more than 3km, depending on the binning specification. The largest effect sizes are observed in close proximity to the newly operational station, suggesting that enhanced levels of air, light and noise pollution do not offset the positive accessibility effects. Beyond this immediate proximity, the effects decline steadily with increasing distance from the opened train station. The observed maximum effect ranges, extending up to more than 3km, contrast with the main analysis' findings, which suggest no increase in property values between 2km and 3km on average (Table C.7). This discrepancy may arise from differences in the econometric methods used in the two approaches. The estimator proposed by Butts (2023) employs a simple before-after comparison, adjusting house prices for property and neighborhood characteristics, but not accounting for unique unobserved regional features. In contrast, Equation (1) incorporates fixed effects on the year level and on a small regional scale that allow to control for temporal shocks as well as unobservable regional factors that could introduce omitted variable bias. Despite these differences, replicating the main analysis using the donut control group with $d_t = 3km$ and $d_c = 4km$ yields higher but comparable

[16]The binscatter regressions are estimated using the R-package ”binsreg” (`https://cran.r-project.org/package=binsreg`).

Table 3: The Effect of Train Station Openings on Prices of Houses Within 2km of Another Station

	Donut Controls	Real Controls	Hypothetical Controls
Train Station < 2km	0.0134	0.0011	-0.0022
	(0.0116)	(0.0137)	(0.0161)
No. Observation	147,892	136,104	75,329
Adj. R2	0.7123	0.6925	0.7201

This table shows the results of Equation (1) with $d_t = 2km$, where only properties within 2km of another operational train station are included in the analysis. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level (in parentheses).

results of an average accessibility premium of about 8% within 3km of a newly opened train station.

The main analysis focuses on properties gaining access to passenger rail services, that are not already situated near an operating train station. However, also properties that are already in proximity to an existing station may benefit from access to additional stations as new stations open, which could result in significant increases in property values. Table 3 demonstrates the impact of train station openings on prices of houses within 2km of the inaugurated station, that are also situated near another operational station. In contrast to the main findings, the estimates across the three control group variants are closely centered around zero and statistically insignificant. The results indicate that the commissioning of a new train station does not generate a premium for properties that already have access to passenger rail services. Thus, homeowners appear to be willing to pay a premium for access to passenger rail services in the first place, but do not value access to multiple train stations (Im and Hong 2018).

Furthermore, I utilize the distribution of station openings across Germany to examine heterogeneity in the impact of train station openings on property prices in different regional contexts. To achieve this, I categorize the station openings from the main analysis for $d_t = 2km$ into two groups and conduct separate Difference-in-Differences analyses. The categorization is based on whether the stations are situated in municipalities classified as medium-order centers and above or lower-order centers and below, and on whether the population counts and purchasing power per capita within a 1.5km radius of the train station fall below or above the sample median, respectively. The findings are presented in Table 4. Figures C.5 to C.10 show the corresponding event study graphs.[17]

Differentiating between train stations based on the surrounding population density re-

[17]Due to the minor differences between the main results obtained by the estimator developed in Sun and Abraham (2021) and a regular two-way fixed effects regression (TWFE) (Table D.8, Figures D.7 to D.9), as well as more precise estimates provided by TWFE with smaller sample sizes, particularly in the event study setting, the subsequent analysis uses a regular TWFE specification.

Table 4: The Effect of Train Station Openings on House Prices Across Regional Characteristics - Train Station < 2km

	Donut Controls	Real Controls	Hypothetical Controls
Population			
Below Median	-0.0102	0.0025	0.0343
	(0.023)	(0.0201)	(0.0214)
No. Observation	19,396	40,411	24,635
Adj. R2	0.6125	0.587	0.7195
Above Median	0.0536**	0.0372**	0.0735**
	(0.0118)	(0.0139)	(0.0114)
No. Observation	48,513	134,477	48,677
Adj. R2	0.7126	0.7014	0.7113
Municipality Type			
Lower-order Center and Below	0.0173	0.0313	0.0605
	(0.0258)	(0.0220)	(0.0410)
No. Observation	21,003	60,277	28,713
Adj. R2	0.6371	0.6344	0.6835
Medium/Higher-order Center	0.0491**	0.0392*	0.0798**
	(0.0137)	(0.0189)	(0.0161)
No. Observation	46,906	116,171	45,602
Adj. R2	0.7254	0.677	0.7498
Purchasing Power per Capita			
Below Median	0.0259	0.0113	0.0630**
	(0.0176)	(0.0163)	(0.0154)
No. Observation	32,306	81,582	39,163
Adj. R2	0.6269	0.6136	0.7069
Above Median	0.0565**	0.0834**	0.0666**
	(0.0161)	(0.0193)	(0.0198)
No. Observation	35,603	84,362	36,909
Adj. R2	0.7215	0.7133	0.7289

This table shows the results for Equation (1) with $d_t = 2km$, using different data samples based on local characteristics of the opened stations, i.e. population count, location and purchasing power per capita. The top panel shows the results for openings of stations with population counts below and above the sample median, the middle panel for station openings in medium- and higher order centers as well as lower-order centers and municipalities without a local function, and the bottom panel for stations with per capita purchasing power below and above the sample median. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level (in parentheses).

** $p < 0.05$. ** $p < 0.01$*

veals statistically insignificant effects ranging from -1% to 3.4% for station openings in less densely populated areas. In contrast, stations in regions with higher population densities show substantially larger and statistically significant effects, ranging from 3.7% to 7.4%. Densely populated areas are more likely to have a well-functioning public transit network, that facilitates easier transfers and first/last mile coverage, thereby making public transportation more attractive. Conversely, individuals in rural areas are more inclined to travel by car and may still face worse connectivity to key locations when taking the train, despite the presence of a new station. Given the substantially lower number of observations for station openings in less densely populated neighborhoods, the results suggest that the positive aggregate effects across all station openings are largely driven by those in higher population areas. An analysis of station openings in medium- and higher-order centers, as well as lower-order centers and municipalities without a central function, confirms the pattern of insignificant estimates in more rural areas and statistically significant effects in more urban municipalities, though the effects do differ less markedly in size.

A similar pattern emerges when distinguishing between stations above or below the sample median in terms of purchasing power per capita. Using the donut and real control groups yields statistically insignificant coefficients for stations below the median, and larger, significant effects for stations in wealthier areas. Conversely, utilizing the hypothetical control group yields statistically significant estimates of about 6.5% for both samples. The results are in line with previous cross-sectional evidence (Bowes and Ihlanfeldt 2001, Hess and Almeida 2007, Brandt and Maennig 2012) on higher effect sizes in wealthier neighborhoods and could suggest a higher valuation of potential time savings by high-income households. Alternatively, the findings could indicate that households residing in poorer neighborhoods are not able to pay an accessibility premium due to tighter budget constraints, resulting in a more elastic demand for housing in less wealthy regions.[18]

6 Conclusion

This study analyzes the impact of train station openings in Germany during the period of 2009 to 2020 on prices of residentially used houses located nearby the station taken into operation to estimate the value households attach to living nearby a train station, providing them with access to passenger rail services. The train station openings build quasi-experiments which are exploited for identification in a Difference-in-Differences framework. To alleviate concerns of endogenous infrastructure provision, I employ an

[18]The results should be interpreted with caution due to potential endogeneity issues in the heterogeneity analysis, as more valuable stations may be placed in more urban or higher-income areas. Additionally, the station openings may affect local population counts and income levels, potentially leading to higher values in both metrics in proximity to more valuable stations. However, categorizing stations based on their average values throughout the entire study period attenuates the latter concern.

identification strategy involving three different control groups variants that allow me to cover alternative assumptions about the comparability of regions in which a station was taken into operation and areas where no station has opened. The results underscore the critical importance of a thoughtful and well-founded control group selection in this kind of research designs. Moreover, the study's findings are based on about 80 to 100 single train station openings across the entire country of Germany over a time span of more than 10 years, leading to a greater general applicability of the findings in comparison to previous studies that predominantly focus on single rail infrastructure projects, often located in large metropolitan areas.

The results suggest a positive impact of train station commencement on residential house prices of approximately 5% (€18,000) on average within up to 2km to 3km of the opened train station, depending on the specification. Importantly, those positive effect sizes are observed exclusively for properties lacking prior access to passenger rail services. In contrast, houses situated near another train station demonstrate no increases in their values following the opening of a nearby station, indicating no appreciation for having access to a second or third train station. The results demonstrate the largest effects in immediate proximity to the train station, that steadily decline with increasing distance between house and station up to 2km to 3km, and stand in contrast to findings of previous studies about local rail services, that document increases in property values up to approximately 1.5km of an opened train station (Rennert 2022). This indicates that the presence of a train station serving regional passenger rail capitalizes into the real estate market over a greater distance, as it facilitates more extensive travel. Furthermore, separating train stations situated in different neighborhoods with respect to population density and purchasing power reveals substantial heterogeneity in the capitalization of passenger rail access in the housing market, depending on the area's socio-economic characteristics. Sizeable effect sizes are found in wealthier areas as wells as in more urban and more densely populated regions, whereas the results for stations located in more rural areas are ambiguous.

The large effects of train station infrastructure on residential house prices indicate a significant appreciation of access to rail services by homeowners. A quick back-on-the-envelope-calculation shows how these property price effects aggregate over all train station openings and demonstrates an overall valuation of access to regional passenger rail. I conservatively assume a maximum effect distance of two kilometers and use the lowest estimate for $d_t = 2km$ of an average increase in asking prices of €18,600. Furthermore, I presume a gap between offering and transaction prices of 10%, yielding an average increase of about €16,700 in property values.[19] When aggregated to the approximately

[19]The gap between asking and transaction prices varies widely over time and across regions. For the period from January 2018 to March 2021, "immobilienscout24", the internet platform from which the property data was sourced, reported that asking prices exceeded transaction prices by only 1.3% (ImmobilienScout24 2024). In 2023, however, the transaction price was 10% below the offering price on average (Handelsblatt 2023).

14,300 houses located within two kilometers of the 85 opened train stations and observed after the stations have been taken into operation, this effect translates into a cumulative rise in property values of about €238 million. These aggregated increases correspond to an average property price effect of €2.8 million per opened train station. While this signifies a substantial household valuation for access to regional passenger rail services, it is crucial to recognize that this valuation is derived from properties listed for sale on the real estate platform ”immobilienscout24”. Assuming that the houses in the analysis sample are representative of all dwellings, the findings could be extrapolated to suggest that not only households purchasing new homes benefit from improved accessibility, but also those who have perennially resided near the station that has recently commenced operations, potentially leading to a considerably larger valuation. Furthermore, as demonstrated by Fesselmeyer and Liu (2018) and Chernoff and Craig (2022), individuals residing near interconnected stations also benefit from the station commissioning, as their access to the newly connected area improves. As consequence, focusing solely on the accessibility premiums of property transactions in close proximity to newly operational infrastructure underestimates the true overall valuation.

References

Ahlfeldt, Gabriel M (2011). "The train has left the station: Do markets value intracity access to intercity rail connections?" In: *German Economic Review* 12.3, pp. 312–335.

Ahlfeldt, Gabriel M and Arne Feddersen (2018). "From periphery to core: measuring agglomeration effects using high-speed rail". In: *Journal of Economic Geography* 18.2, pp. 355–390.

Ahlfeldt, Gabriel M, Volker Nitsch, and Nicolai Wendland (2019). "Ease vs. noise: Long-run changes in the value of transport (dis)amenities". In: *Journal of Environmental Economics and Management* 98, p. 102268.

Baker, Andrew C, David F Larcker, and Charles CY Wang (2022). "How much should we trust staggered difference-in-differences estimates?" In: *Journal of Financial Economics* 144.2, pp. 370–395.

Banzhaf, H Spencer (2021). "Difference-in-differences hedonics". In: *Journal of Political Economy* 129.8, pp. 2385–2414.

Baum-Snow, Nathaniel and Fernando Ferreira (2015). "Causal Inference in Urban and Regional Economics". In: *Handbook of Regional and Urban Economics*. Ed. by Gilles Duranton, J. Vernon Henderson, and William C. Strange. Vol. 5. Elsevier, pp. 3–68.

Bergé, Laurent (2018). "Efficient estimation of maximum likelihood models with multiple fixed-effects: the R package FENmlm". DEM Discussion Paper Series 18-13. Department of Economics at the University of Luxembourg.

Billings, Stephen B (2011). "Estimating the value of a new transit option". In: *Regional Science and Urban Economics* 41.6, pp. 525–536.

Bowes, David R and Keith R Ihlanfeldt (2001). "Identifying the impacts of rail transit stations on residential property values". In: *Journal of Urban Economics* 50.1, pp. 1–25.

Brandt, Sebastian and Wolfgang Maennig (2012). "The impact of rail access on condominium prices in Hamburg". In: *Transportation* 39.5, pp. 997–1017.

Büchel, Konstantin and Stephan Kyburz (Jan. 2020). "Fast track to growth? Railway access, population growth and local displacement in 19th century Switzerland". In: *Journal of Economic Geography* 20.1, pp. 155–195.

Butts, Kyle (2021). "did2s: Two-Stage Difference-in-Differences Following Gardner (2021)". URL: `https://github.com/kylebutts/did2s/`.

– (2023). "JUE Insight: Difference-in-differences with geocoded microdata". In: *Journal of Urban Economics* 133, p. 103493.

Cao, Jason and Xiaoshu Cao (2014). "The Impacts of LRT, Neighbourhood Characteristics, and Self-selection on Auto Ownership: Evidence from Minneapolis-St. Paul". In: *Urban Studies* 51.10, pp. 2068–2087.

Cats, Oded, Yusak O Susilo, and Triin Reimal (2017). "The prospects of fare-free public transport: evidence from Tallinn". In: *Transportation* 44, pp. 1083–1104.

Cattaneo, Matias D, Max H Farrell, and Yingjie Feng (2020). “Large sample properties of partitioning-based series estimators”. In: *The Annals of Statistics* 48.3, pp. 1718–1741.

Chatman, Daniel G, Nicholas K Tulach, and Kyeongsu Kim (2012). “Evaluating the economic impacts of light rail by measuring home appreciation: A first look at New Jersey’s River Line”. In: *Urban studies* 49.3, pp. 467–487.

Chernoff, Alex and Andrea N Craig (2022). “Distributional and Housing Price Effects from Public Transit Investment: Evidence from Vancouver”. In: *International Economic Review* 63.1, pp. 475–509.

Christaller, Walter (1980). “Die zentralen Orte in Süddeutschland: eine ökonomisch-geographische Untersuchung über die Gesetzmässigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen”. Wissenschaftliche Buchgesellschaft.

Cohen, Jeffrey P and Mike Brown (2017). “Does a new rail rapid transit line announcement affect various commercial property prices differently?” In: *Regional Science and Urban Economics* 66, pp. 74–90.

DB Station&Service AG (2020). “Stationsdaten”. Dataset. URL: `https://data.deutschebahn.com/dataset/data-stationsdaten/resource/dfddd39b-b74c-40b7-9b60-d265ee2473cb.html`.

Devaux, Nicolas, Jean Dubé, and Philippe Apparicio (2017). “Anticipation and post-construction impact of a metro extension on residential values: The case of Laval (Canada), 1995–2013”. In: *Journal of Transport Geography* 62, pp. 8–19.

Diao, Mi, Delon Leonard, and Tien Foo Sing (2017). “Spatial-difference-in-differences models for impact of new mass rapid transit line on private housing values”. In: *Regional Science and Urban Economics* 67, pp. 64–77.

Dubé, Jean, Marius Thériault, and François Des Rosiers (2013). “Commuter rail accessibility and house values: The case of the Montreal South Shore, Canada, 1992–2009”. In: *Transportation Research Part A: Policy and Practice* 54, pp. 49–66.

Fesselmeyer, Eric and Haoming Liu (2018). “How much do users value a network expansion? Evidence from the public transit system in Singapore”. In: *Regional Science and Urban Economics* 71, pp. 46–61.

Gardner, John (2022). “Two-stage differences in differences”. In: *arXiv:2207.05943*.

Gibbons, Stephen, Stephan Heblich, and Edward W. Pinchbeck (2024). “The spatial impacts of a massive rail disinvestment program: The Beeching Axe”. In: *Journal of Urban Economics* 143, p. 103691.

Gibbons, Stephen and Stephen Machin (2005). “Valuing rail access using transport innovations”. In: *Journal of Urban Economics* 57.1, pp. 148–169.

Goodman-Bacon, Andrew (2021). “Difference-in-differences with variation in treatment timing”. In: *Journal of Econometrics* 225.2, pp. 254–277.

Handelsblatt (2023). “ImmoScout24-Chefin: So stark können Sie beim Hauskauf den Preis drücken”. Accessed on: April 25, 2024. Handelsblatt. URL: `https://www.handelsblatt.`

com/finanzen/immobilien/immoscout24-chefin-so-stark-koennen-sie-beim-hauskauf-den-preis-druecken/29211422.html.

Hess, Daniel Baldwin and Tangerine Maria Almeida (2007). “Impact of proximity to light rail rapid transit on station-area property values in Buffalo, New York”. In: *Urban Studies* 44.5-6, pp. 1041–1068.

Heuermann, Daniel F and Johannes F Schmieder (2019). “The effect of infrastructure on worker mobility: evidence from high-speed rail expansion in Germany”. In: *Journal of Economic Geography* 19.2, pp. 335–372.

Im, Junhong and Sung Hyo Hong (2018). “Impact of a new subway line on housing values in Daegu, Korea: Distance from existing lines”. In: *Urban Studies* 55.15, pp. 3318–3335.

ImmobilienScout24 (2024). “Verhandeln beim Immobilienkauf: So erzielen Sie den besten Preisnachlass”. Accessed on: April 25, 2024. ImmobilienScout24. URL: https://www.immobilienscout24.de/wissen/kaufen/angebots-vs-transaktionspreise.html.

Ke, Yue and Konstantina Gkritza (2019). “Light rail transit and housing markets in Charlotte-Mecklenburg County, North Carolina: Announcement and operations effects using quasi-experimental methods”. In: *Journal of Transport Geography* 76, pp. 212–220.

Liebensteiner, Mario, Jakob Losert, Sarah Necker, Florian Neumeier, Jörg Paetzold, and Sebastian Wichert (2024). “Almost Fare Free: Impact of a Cheap Public Transport Ticket on Mobility Patterns and Infrastructure Quality”. Tech. rep. CESifo.

Mayer, Thierry and Corentin Trevien (2017). “The impact of urban public transportation evidence from the Paris region”. In: *Journal of Urban Economics* 102, pp. 1–21.

Mohammad, Sara I, Daniel J Graham, and Patricia C Melo (2017). “The effect of the Dubai Metro on the value of residential and commercial properties”. In: *Journal of Transport and Land Use* 10.1, pp. 263–290.

Pilgram, Clemens A. and Sarah E. West (2018). “Fading premiums: The effect of light rail on residential property values in Minneapolis, Minnesota”. In: *Regional Science and Urban Economics* 69, pp. 1–10.

Ransom, Michael R (2018). “The effect of light rail transit service on nearby property values”. In: *Journal of Transport and Land Use* 11.1, pp. 387–404.

Rennert, Lindiwe (2022). “A meta-analysis of the impact of rail stations on property values: Applying a transit planning lens”. In: *Transportation Research Part A: Policy and Practice* 163, pp. 165–180.

Rojas, Alexandra (2024). “Train stations’ impact on housing prices: Direct and indirect effects”. In: *Transportation Research Part A: Policy and Practice* 181, p. 103979.

Rosen, Sherwin (1974). “Hedonic prices and implicit markets: product differentiation in pure competition”. In: *Journal of Political Economy* 82.1, pp. 34–55.

Roth, Jonathan, Pedro H.C. Sant’Anna, Alyssa Bilinski, and John Poe (2023). “What’s trending in difference-in-differences? A synthesis of the recent econometrics literature”. In: *Journal of Econometrics* 235.2, pp. 2218–2244.

RWI and ImmobilienScout24 (2022). “RWI-GEO-RED: RWI Real Estate Data - Houses for Sale”. RWI - Leibniz Institute for Economic Research. Dataset. DOI: `doi.org/10.7807/immo:red:hk:suf:v6.1`.

RWI and microm (2022). “RWI-GEO-GRID: Socio-economic data on grid level (wave 12).” RWI - Leibniz Institute for Economic Research. Dataset. DOI: `10.7807/microm:v12`.

Schaffner, Sandra (2020). “FDZ data description: Real-estate data for Germany (RWI-GEO-RED v3) - Advertisements on the internet platform ImmobilienScout24. 2007-06/2020”. RWI Projektberichte 242996. RWI - Leibniz-Institut für Wirtschaftsforschung.

Song, Zhe, Mengqiu Cao, Tingting Han, and Robin Hickman (2019). “Public transport accessibility and housing value uplift: Evidence from the Docklands light railway in London”. In: *Case Studies on Transport Policy* 7.3, pp. 607–616.

Sun, Liyang and Sarah Abraham (2021). “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects”. In: *Journal of Econometrics* 225.2, pp. 175–199.

Tyndall, Justin (2021). “The local labour market effects of light rail transit”. In: *Journal of Urban Economics* 124, p. 103350.

Umweltbundesamt (2024). “Klimaschutz im Verkehr”. Accessed on: September 12, 2024. URL: `https://www.umweltbundesamt.de/themen/verkehr-laerm/klimaschutz-im-verkehr#undefined`.

VDV - Die Verkehrsunternehmen (2020). “Auf der Agenda: Reaktivierung von Eisenbahnstrecken”.

Wagner, Gary A., Timothy Komarek, and Julia Martin (2017). “Is the light rail “Tide” lifting property values? Evidence from Hampton Roads, VA”. In: *Regional Science and Urban Economics* 65, pp. 25–37.

Welch, Timothy F, Steven R Gehrke, and Steven Farber (2018). “Rail station access and housing market resilience: Case studies of Atlanta, Baltimore and Portland”. In: *Urban Studies* 55.16, pp. 3615–3630.

Yazdanifard, Yekta, Masoud Talebian, and Hosein Joshaghani (2021). “Metro station inauguration, housing prices, and transportation accessibility”. In: *Journal of Transport and Land Use* 14.1, pp. 537–561.

Appendix

A Variable Descriptions

Table A.1: Variable Descriptions - House Characteristics

Variable	Description
Price	Asking price, inflation corrected with 2015 as reference year.
Plot Area	Plot area of the entire property up for sale in m^2. *As controls in: Reduced, Main*
Living Space	Area of the living space in m^2. *As controls in: Reduced, Main*
No. of Rooms	Number of rooms of the house. *As controls in: Reduced, Main*
Age	Age of the building, calculated using information on year of construction. Houses under construction are assigned a value of -0.1. Missing values are replaced by the mean age. *As controls in: Reduced, Main*
Flag Age	Dummy variable indicating whether missing age values are replaced by the average age. *As controls in: Reduced, Main*
Guest Toilet	Dummy variable indicating the presence of a second bathroom. *As controls in: Reduced, Main*
Basement	Dummy variable indicating the presence of a basement. *As controls in: Reduced, Main*
Holiday House	Dummy variable indicating use as holiday house. *As controls in: Main*
Protected Building	Dummy variable indicating the status as a protected building. *As controls in: Main*
Construction	Dummy variable indicating whether the house is still under construction. *As controls in: Reduced, Main*
House Type	Categorical variable indicating the type of house. Differentiation between: bungalow, farmhouse, mansion, semi-detached, single-family, terraced, special, other. Missing values are recoded to "No information". *As controls in: Main*
Single-family House	Dummy variable based on "House Type", indicating whether the property is a single-family house. *As controls in: Reduced*
Condition	Categorical variable indicating the house's condition. Differentiation between: by arrangement, completely renovated, first occupancy, first occupancy after reconstruction, well kempt, like new, modernised, reconstructed. Missing values are recoded to "No information". *As controls in: Main*
First occupancy	Dummy variable based on "Condition", indicating whether the property is available for its first occupancy. *As controls in: Reduced*

All information on property characteristics are given by the real estate data of the RWI. Italicized text indicates the sets of covariates used in the analysis to which the variable pertains.

Table A.2: Variable Descriptions - Neighborhood Characteristics

Variable	Description
Population	Population count in the 1 km^2 grid cell the house is located in. *As controls in: Reduced, Main*
Purchasing Power PC	Purchasing power in the 1 km^2 grid cell the house is located in, divided by its population count. *As controls in: Reduced, Main*
Unemployment Rate	Unemployment rate in the 1 km^2 grid cell the house is located in. *As controls in: Reduced, Main*
No. Commercial Buildings	Number of commercially used buildings in the 1 km^2 grid cell the house is located in. *As controls in: Reduced, Main*
Share of Foreigners	Share of households with a foreign head of household in the 1 km^2 grid cell the house is located in. *As controls in: Main*
Share of Families	Share of families with children in the 1 km^2 grid cell the house is located in. *As controls in: Main*
Share Above Age 60	Share of population that is above the age of 60 in the 1 km^2 grid cell the house is located in. *As controls in: Main*

All information on socio-economic neighborhood characteristics are on a 1 km^2 level and provided by the RWI GEO Grid data. Italicized text indicates the sets of covariates used in the analysis to which the variable pertains.

Table A.3: Variable Descriptions - Distance Measures

Variable	Description
Distance Highway	Distance to the closest highway entrance. Locations on highway entrances are taken from OpenStreetMap as of April 2022. *As controls in: Reduced, Main*
Distance Regional Center	Distance to the closest population-weighted centroid of a municipality that is considered at least a lower-order regional center. Centroids are determined using the socio-economic grid data, by calculating the population-weighted centroid of all grid cells that can be allocated to the respective municipality. *As controls in: Reduced, Main*
Distance Higher-order Center	Distance to the closest population-weighted centroid of a municipality considered a higher-order regional center. Centroids are determined using the socio-economic grid data, by calculating the population-weighted centroid of all grid cells that can be allocated to the respective municipality. *As controls in: Main*
Railway $<$ 500m	Dummy variable indicating whether the house is located within 500m of a railway. Georeferenced information on the railway network is obtained from the federal agency for cartography and geodesy as of January 2020. *As controls in: Reduced, Main*

Italicized text indicates the sets of covariates used in the analysis to which the variable pertains.

Table A.4: Variable Descriptions - Train Station Characteristics

Variable	Description
Long Distance	Dummy Variable indicating the service of long distance rail. *Source: Deutsche Bahn*
No. of Stations Accessible w/o Transfer	Number of train stations that can be accessed from the station of interest without transferring. *Source: General Transit Feed Specification*
Travel Time to Higher-order Center	Travel time to the main station of the nearest higher-order center. *Source: General Transit Feed Specification*
Population of Largest City	Population count within 1km of the train station with the highest population count that is accessible without transferring. *Source: General Transit Feed Specification, RWI Geo Grid*
Travel Time Weighted Sum of Population	Inversely travel time weighted sum of population counts within 1km of every accessible train station. *Source: General Transit Feed Specification, RWI Geo Grid*
Population	Population count within 1.5km of the train station. The population counts of grid cells that intersect a circular area with a radius of 1.5km around the station are weighted based on the overlap, and then summed up. *Source: RWI Geo Grid*
Purchasing Power PC	Average purchasing power per capita within 1.5km of the train station. The purchasing power per capita of the grid cells that intersect a circular area with a radius of 1.5km around the station is weighted by the overlap, and the averaged. *Source: RWI Geo Grid*
Unemployment Rate	Average unemployment rate within 1.5km of the train station. The unemployment rate of the grid cells that intersect a circular area with a radius of 1.5km around the station is weighted by the overlap, and then averaged. *Source: RWI Geo Grid*
No. Commercial Buildings	Number of commercially used buildings within 1.5km of the train station. The number of commercial buildings in the grid cells that intersect a circular area with a radius of 1.5km around the station are weighted based on the overlap, and then summed up. *Source: RWI Geo Grid*
Municipality Type	Categorical Variable indicating whether the train station's municipality is a higher-order center, medium-order center or lower-order center. *Source: BBSR*
Municipality Development	Categorical variable indicating whether the train station's municipality is shrinking, growing or neither of both (as of 2015). *Source: BBSR*
Distance Medium/ Higher-order Center	Distance to the closest population-weighted centroid of a municipality considered a medium or higher-order regional center. Centroids are determined using the socio-economic grid data, by calculating the population-weighted centroids of all grid cells that can be allocated to the respective municipality. *Source: BBSR, RWI Geo Grid*

B Descriptive Statistics

Table B.1: Train Station Characteristics

	N	Mean	SD	Min	Median	Max
Accessibility Measures						
Long Distance (%)	6,433	5.49	22.78	0.00	0.00	1.00
TT Higher-order Center (Minutes)	6,259	28.17	21.87	0.00	23.00	142.00
TT-weighted Access to Population	6,272	4,764	9,617	25	1,774	177,271
TT-weighted Access to Population w/o Transfer	6,272	3,830	9,033	2	1,026	174,702
No. Stations w/o Transfer	6,272	33.64	33.70	1.00	25.00	532.00
Population Largest Station w/o Transfer	6,272	24,714	12,312	31	23,912	58,114
Neighborhood Characteristics						
Population	6,433	8,676	12,644	0	4,100	118,000
Purchasing Power PC	6,431	21,650	3,188	14,254	21,289	46,883
No. Commercial Buildings	6,433	709	1,494	0	291	21,191
Unemployment Rate (%)	6,433	5.78	3.16	0.00	5.10	20.62
Location						
Medium/Higher-order Center (%)	6,433	46.71	49.90	0.00	0.00	1.00
Growing Municipality (%)	6,433	61.32	48.70	0.00	1.00	1.00
Distance Medium/Higher-order Center (km)	6,433	6.23	4.99	0.02	5.37	67.24

Table B.1 provides descriptive statistics for various attributes of German train stations. Approximately 5.5% of the stations in Germany cater to long-distance rail services. The average travel time to the nearest main station in a higher-order center is 28 minutes, with the longest travel time being two hours and twenty minutes, from "Dagebüll Mole" to Kiel main station in northern Germany. On average, travelers can reach 34 stations without the need to change trains. Frankfurt main station offers the highest connectivity, with direct access to 532 stations, while Trossingen city station is only linked to Trossingen station. German train stations are typically located in densely populated areas, with an average of 8,700 residents and 710 commercially used buildings within a 1km radius. These regions also exhibit high unemployment rates, averaging 22%, whereas the average purchasing power per capita in these areas is €21,700 and aligns with the national average. Around half of the stations are situated in municipalities classified as medium- or higher-order centers, with an average Euclidean distance of 6.2 km to the population-weighted centroid of such centers. The farthest station is Westerland, located on the island of Sylt, 67 km from its nearest medium- or higher-order center. Additionally, 60% of the stations are located in municipalities that were categorized as "growing" in 2015.

Utilizing a Linear Probability Model to assess the relationship between various train station characteristics and the initiation of train station operations during the study period indicates that newly established train stations tend to be situated to a larger extent in medium- to higher-order urban centers than stations operational throughout the entire study period. Similarly, the results highlight that newly operational train stations pri-

Table B.2: Train Station Characteristics and Station Commissioning

	Opened	Year of Opening
Accessibility Measures		
TT Higher-order Center (Minutes)	0.0005** (0.0001)	0.0082 (0.0103)
TT-weighted Access to Population (1,000)	0.0008 (0.0004)	-0.0712 (0.0521)
No. of Stations w/o Transfer	-0.0002* (0.0001)	-0.0102 (0.0136)
Population Largest Station w/o Transfer (1,000)	-0.0003 (0.0002)	-0.0413 (0.0236)
Neighborhood Characteristics		
Population (1,000)	-0.0003 (0.0004)	-0.0134 (0.0584)
Purchase Power PC (1,000€)	-0.0016 (0.0010)	0.0341 (0.1173)
No. Commercial Buildings (10,000)	0.0007 (0.0294)	7.6862 (5.5605)
Unemployment Rate (10pp)	-0.0001 (0.0106)	1.1548 (0.9410)
Location		
Medium- or Higher-order Center	0.0076 (0.0066)	0.6462 (0.6714)
Growing Municipality	0.0238** (0.0053)	0.5698 (0.5650)
Distance Medium/Higher-order Center (km)	-0.0015* (0.0006)	0.2504** (0.0719)
No. Observations	6,257	205
Adj. R2	0.0099	0.1299

Column 1 shows the results of a Linear Probability Model using a dummy indicating whether the train station was taken into operation during the study period as dependent variable. Column 2 shows the results for an linear regression with the year of station opening as dependent variable, using the subset of train stations opened during the study period.
* $p < 0.05$, ** $p < 0.01$

marily emerged in municipalities considered "growing" as of 2015. By 2020, these recently inaugurated stations tend to exhibit slightly inferior connectivity within the broader train station network, evidenced by increased travel times to the nearest higher-order center and a reduced number of accessible stations without requiring transfers. In contrast, no discernible correlation between socio-economic neighborhood characteristics and the opening of train stations is apparent (Table B.2, column 1). Additionally, the second column of Table B.2 shows no correlation between the timing of station commencement and train station attributes, except for distance to the closest medium- or higher order center.

Across all houses observed during the study period, the average asking price is €307,043 with a mean price per square meter living space of €1,957 (Table B.3).[20] The advertised properties have on average a plot area of 723 square meters and 163 square meters living space divided into 5.6 rooms. Roughly 8% of the advertised properties are still under construction at the date of the insert, 49% of the observations have a guest toilet and 31% a basement. 3% of the adverts are holiday houses, while only less than 1% is categorized as a protected building. The majority of houses in the sample are single-family houses (59%) with semi-detached (10%) and terraced houses (8%) on the runner-up spots. 479,869 observations lack information on house category and are grouped together into the category "Missing Information". 21% of the houses are advertised as "first occupancy", 18% are categorized as "well kempt" and 46% of the observations lack information on the building's condition.

On average, houses are located in one square kilometer grid cells with a population of 1,653 and a maximum population per grid cell of around 27,000. The unemployment rate has its mean at 5.3% and goes up to 40% at its maximum. The poorest neighborhood has a per capita purchasing power of 5,900€ per year, in the richest one each resident has 139,000€ yearly disposal income on average. The mean purchasing power per capita lies at 22,286€. 7% of the households residing in the properties' neighborhoods have a foreign household head, 33% are families with children. 112 commercial buildings are located in the average grid cell with the highest density being 9,700 per square kilometer. On average, houses are located 7.6km away from the closest motorway entrance, 3.1km from the closest regional center and 19.5km from the closest higher-order center. Furthermore, 26% of the houses are located within 500m of a railway.

[20] Figures B.1 and B.1 show the distribution of the (logarithmized) house price.

Table B.3: House and Neighborhood Characteristics

	Mean	SD	Median
House Characteristics			
Price	307,043	178,468	260,020
Price per m^2	1,957	888	1,788
Plot Area (m^2)	723.2	568.4	600.0
Living Space (m^2)	162.8	77.2	143.8
No. of Rooms	5.57	2.14	5.00
Age	32.85	37.51	24.00
Guest Toilet (%)	48.47	49.98	0.00
Basement (%)	30.52	46.05	0.00
Holiday House (%)	3.05	17.18	0.00
Protected Building (%)	0.58	7.61	0.00
Construction (%)	8.17	27.39	0.00
Single-family House (%)	58.59	49.26	1.00
First Occupancy (%)	21.20	40.87	0.00
Neighborhood Characteristics			
Population	1,653	1,687	1,105
Purchasing Power PC	22,286	3,768	21,803
No. Commercial Buildings	112.0	154.5	66.0
Unemployment Rate (%)	5.28	3.47	4.56
Share of Foreigners (%)	7.07	5.77	5.72
Share of Families (%)	32.54	19.36	31.36
Share Above Age 60 (%)	27.72	4.81	27.32
Euclidean Distances			
Distance Highway (km)	7.64	7.35	5.14
Distance Regional Center (km)	3.13	2.38	2.66
Distance Higher-order Center (km)	19.45	12.96	16.75
Railway < 500m (%)	26.02	43.88	0.00
Distance Closest Train Station (km)	3.45	3.53	2.22

This table shows the descriptive statistics for all 5,053,385 properties.

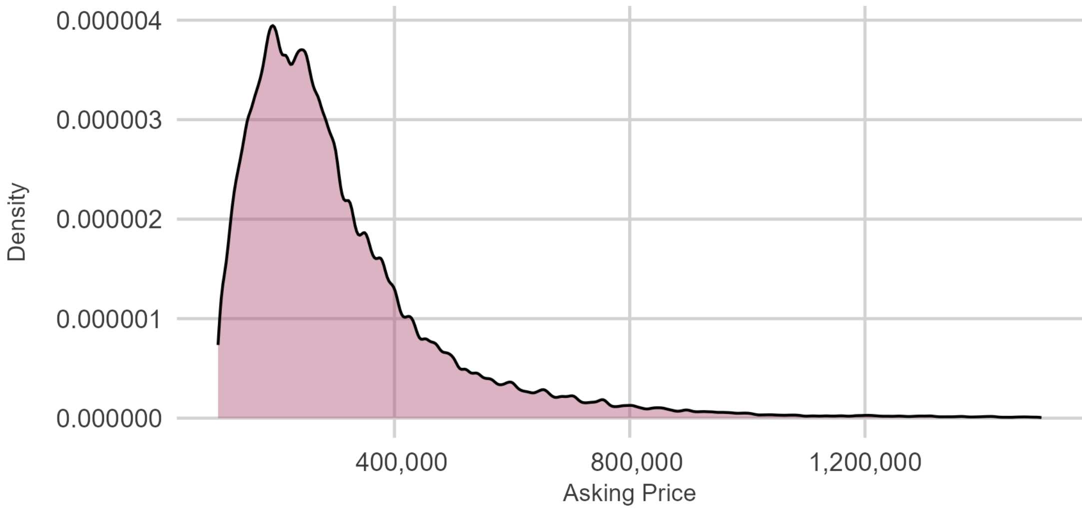

Figure B.1: House Price Density

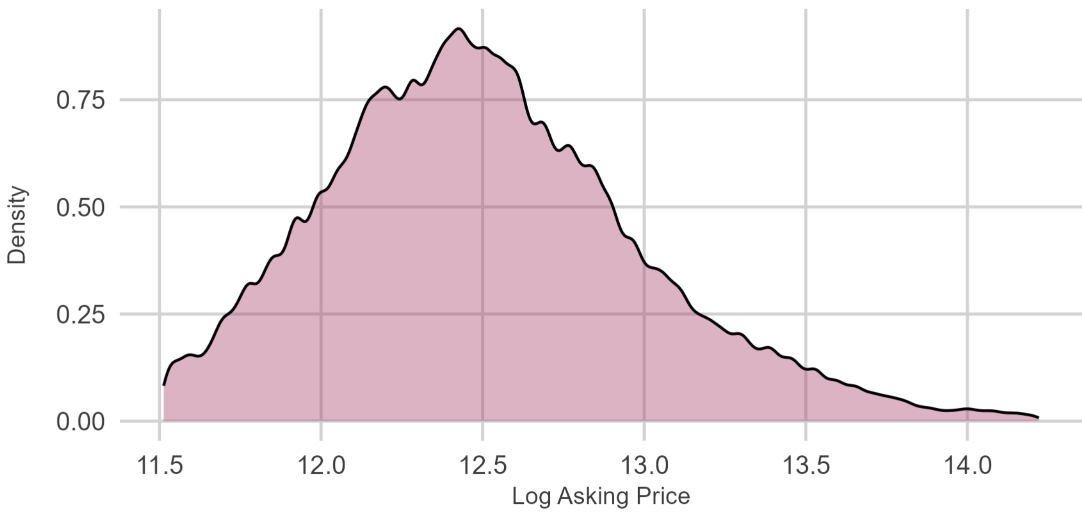

Figure B.2: Logarithmized House Price Density

Table B.4: Matching Statistics - Real Control Group

	Before Matching			After Matching		
	Treatment Group	Control Group	Difference	Treatment Group	Control Group	Difference
Long Distance (%)	1.72	6.72	-5.00*	0.94	1.56	-0.62
TT Higher-order Center (Minutes)	41.20	25.80	15.40**	41.20	38.40	2.80
TT-weighted Access to Population	2,117	5,447	-3,330**	2,117	2,102	15
Population	5,916	10,327	-4,411**	6,188	6,365	-177
Purchasing Power PC	21,255	22,097	-842**	21,192	21,465	-273
No. Commercial Buildings	396	823	-427**	409	440	-31
Distance Medium/Higher-order Center (km)	6.03	5.57	0.46	6.12	6.12	0.00
Municipality Type						
No Central Function (%)	20.70	16.60	4.10	20.60	17.10	3.50
Low-order Center (%)	26.70	30.80	-4.10	27.10	31.50	-4.40
Medium-order Center (%)	37.10	30.70	6.40	37.40	39.60	-2.20
Higher-order Center (%)	15.50	21.80	-6.30	15.00	11.80	3.20
Municipality Growth						
Municipality Shrinking(%)	16.40	18.90	-2.50	15.90	16.80	-0.90
No Clear Trend (%)	10.30	13.90	-3.60	11.20	12.10	-0.90
Municipality Growing (%)	73.30	67.10	6.20	72.90	71.00	1.90

This table shows a comparison of opened train stations and the (pool of) matched stations for the construction of the real control group, before and after the propensity score matching.
** $p < 0.05$, ** $p < 0.01$*

Table B.5: Matching Statistics - Hypothetical Control Group

	Before Matching			After Matching		
	Treatment Group	Control Group	Difference	Treatment Group	Control Group	Difference
Population	5,916	4,737	1,179	5,916	5,690	226
Purchasing Power PC	21,255	21,976	-721**	21,255	22,259	-1,004**
No. Commercial Buildings	396	329	67	396	389	7
Distance Medium/Higher-order Center (km)	6.03	7.00	-0.97	6.03	6.59	-0.56
Municipality Type						
No Central Function (%)	20.70	19.30	1.40	20.70	20.70	0.00
Low-order Center (%)	26.70	44.10	-17.40**	26.70	35.30	-8.60
Medium-order Center (%)	37.10	29.70	7.40	37.10	33.60	3.50
Higher-order Center (%)	15.50	6.93	8.57*	15.50	10.30	5.20
Municipality Growth						
Municipality Shrinking(%)	16.40	10.40	6.00	16.40	2.59	13.81**
No Clear Trend (%)	10.30	12.90	-2.60	10.30	8.62	1.68
Municipality Growing (%)	73.30	76.70	-3.40	73.30	88.80	-15.50**

This table shows a comparison of opened train stations and the (pool of) matched stations for the construction of the hypothetical control group, before and after the propensity score matching.
*** $p < 0.01$*

C Main and Heterogeneity Analysis

Table C.1: Difference-in-Difference Estimates - House Price in Levels

	Donut Controls	Real Controls	Hypothetical Controls
Train Station < 1.5km	18,109**	18,681**	36,310**
	(4,144)	(4,194)	(6,599)
No. Observation	68,881	156,293	67,131
Adj. R2	0.7012	0.6675	0.7054
Train Station < 2km	18,668**	18,575**	28,401**
	(3,921)	(4,439)	(6,082)
No. Observation	67,908	180,492	74,505
Adj. R2	0.6719	0.6248	0.6976
Train Station < 2.5km	18,685**	12,914*	27,473**
	(4,221)	(5,321)	(6,241)
No. Observation	60,213	186,267	73,406
Adj. R2	0.6489	0.6169	0.6789
Train Station < 3km	15,183**	9,459	19,607**
	(4,977)	(5,323)	(7,429)
No. Observation	56,266	212,837	76,192
Adj. R2	0.6513	0.6286	0.6546

This table shows the results of Equation (1) *with prices as the dependent variable for* $d_t = 1.5, 2, 2.5, 3km$. *The values for* d_t *are chosen based on the results of a cross-sectional regression presented in Figure 4, that show positive effects up to 3km that decline after 1.5km. This is in line with previous research that consistently reports positive accessibility premiums extending up to a distance of 1.5km from public transport stops that serve local rail services (Rennert 2022), and the notion that individuals might be willing to travel greater distances to access a train station serving regional rail as it facilitates more extensive travel. The distance threshold* d_c *exceeds* d_t *by 1km in every scenario. As a result, for* $d_t = 1.5km$, *properties within 2.5km to 4.5km to the respective opened station build the donut control group. For the remaining three distance specifications, the donut control groups' upper and lower limits are 3km and 5km, 3.5km and 5.5km as well as 4km and 6km, respectively. The real and hypothetical control groups comprise properties within 1.5km, 2km, 2.5km and 3km of the respective control and hypothetical station. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level (in parentheses).*
** p < 0.05, ** p < 0.01*

Table C.2: Difference-in-Differences Estimates - Price per Square Meter (Levels)

	Donut Controls	Real Controls	Hypothetical Controls
Train Station < 1.5km	108.78**	106.70**	219.79**
	(23.74)	(23.95)	(37.74)
No. Observation	68,881	156,293	67,131
Adj. R2	0.6663	0.6338	0.6913
Train Station < 2km	102.19**	92.96**	175.31**
	(23.62)	(21.60)	(35.85)
No. Observation	67,908	180,492	74,505
Adj. R2	0.6505	0.5944	0.6901
Train Station < 2.5km	107.89**	71.78**	156.98**
	(25.85)	(26.83)	(36.85)
No. Observation	60,213	186,267	73,406
Adj. R2	0.6146	0.5754	0.6407
Train Station < 3km	107.91**	55.51	144.23**
	(31.24)	(29.02)	(45.56)
No. Observation	56,266	212,837	76,192
Adj. R2	0.6194	0.6043	0.6523

This table shows the results of Equation (1) *with prices per square meter as the dependent variable for* $d_t = 1.5, 2, 2.5, 3km$. *The values for* d_t *are chosen based on the results of a cross-sectional regression presented in Figure 4, that show positive effects up to 3km that decline after 1.5km. This is in line with previous research that consistently reports positive accessibility premiums extending up to a distance of 1.5km from public transport stops that serve local rail services (Rennert 2022), and the notion that individuals might be willing to travel greater distances to access a train station serving regional rail as it facilitates more extensive travel. The distance threshold* d_c *exceeds* d_t *by 1km in every scenario. As a result, for* $d_t = 1.5km$, *properties within 2.5km to 4.5km to the respective opened station build the donut control group. For the remaining three distance specifications, the donut control groups' upper and lower limits are 3km and 5km, 3.5km and 5.5km as well as 4km and 6km, respectively. The real and hypothetical control groups comprise properties within 1.5km, 2km, 2.5km and 3km of the respective control and hypothetical station. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level (in parentheses).*

** $p < 0.05$, ** $p < 0.01$*

Table C.3: Difference-in-Difference Sample Sizes

		Donut Controls	Real Controls	Hypothetical Controls
Train Station < 1.5km	Opened Stations	95	90	97
	Treatment Group	27,116	26,647	27,177
	Control Group	41,765	129,646	39,954
	Treated Units	14,190	14,146	14,206
Train Station < 2km	Opened Stations	90	85	91
	Treatment Group	29,079	28,432	29,098
	Control Group	38,829	152,060	45,407
	Treated Units	14,307	14,243	14,319
Train Station < 2.5km	Opened Stations	83	77	84
	Treatment Group	29,739	29,022	29,814
	Control Group	30,474	157,245	43,592
	Treated Units	13,271	13,175	13,272
Train Station < 3km	Opened Stations	74	71	77
	Treatment Group	28,081	27,508	28,414
	Control Group	28,185	185,329	47,778
	Treated Units	12,435	12,445	12,552

For the Difference-in-Differences analysis, I restrict the sample of stations considered in the Difference-in-Differences analysis to opened train stations with at least 20 houses within d_t over the whole study period. This table shows the number of station openings, the number of observations in the treatment and control group as well as the number of treated units across the different regression variants of Equation (1)*. The decrease in the number of observations in the treatment group for $d_t = 3km$ stems from the associated increase in d_c to 4km, leading to the exclusion of more properties as they are considered in proximity to potentially confounding train stations. Similarly, the number of observations for the specification using the donut control group reduces with higher values for d_t because the control group size is negatively affected by a greater d_c. The real and hypothetical control groups increase in size with greater assumed effect ranges since in both cases the control group is determined by the choice of d_t. The specifications based on the real control group have by far the most observations since each opened station is matched to three control stations.*

Table C.4: Train Station Characteristics - Opened vs. Treatment Stations

	Opened Stations		Treatment Stations		
	Mean	SD	Mean	SD	Difference
Accessibility Measures					
Year of Opening	2014.2	3.4	2014.7	3.3	-0.5
Long Distance (%)	0.91	9.51	0.83	9.09	0.08
TT Higher-order Center (Minutes)	34.13	25.65	39.59	27.25	-5.47
TT-weighted Access to Population	4,752	8,190	2,627	4,401	2,125*
TT-weighted Access to Population w/o Transfer	3,990	7,756	2,056	4,242	1,934*
No. Stations w/o Transfer	23.85	24.01	19.44	20.07	4.41
Population Largest Station w/o Transfer	22,120	11,069	22,039	11,734	81
Neighborhood Characteristics					
Population	7,913	9,977	6,896	8,598	1,017
Purchasing Power PC	21,552	2,685	21,405	2,386	148
No. Commercial Buildings	600	953	469	650	131
Unemployment Rate (%)	5.71	3.15	5.52	3.06	0.19
Location					
Medium/Higher-order Center (%)	54.09	49.95	50.41	50.21	3.68
Growing Municipality (%)	73.18	44.40	76.03	42.87	-2.85
Distance Meidum/Higher-order Center (km)	5.64	5.64	5.92	5.89	-0.29

This table compares all stations taken into operation in the study period to the sample of opened stations used in the Difference-in-Differences analysis ("Treatment Stations"). The stations selected for analysis are situated in regions with lower population counts and tend to exhibit inferior connectivity within the rail network compared to all train stations that have been put into service (though statistically insignificant). One potential explanation is that stations taken into operation in densely populated areas and surrounded by multiple other stations, are not considered in the analysis because houses situated within distance d_t of the opened station are also located within d_c of other stations and are thus discarded from the treatment group. Consequently, the treatment groups associated with these opened stations fall short of the required threshold of 20 properties, leading to the train stations being excluded from the analysis. Regarding other neighborhood and location characteristics, as well as the year of station inauguration, the subset of opened stations utilized in the Difference-in-Differences analysis does not differ significantly from the entire set of stations that have been put into operation.

* $p < 0.05$

Table C.5: Difference-in-Differences Estimates - Full Model - Train Station < 2km

	Donut Controls	Real Controls	Hypothetical Controls
House Characteristics			
hspace0.3cmPlot Area ($100m^2$)	0.0255** (0.0024)	0.0338** (0.0022)	0.0273** (0.0023)
Plot Area ($100m^2$) sqrd	-0.0003** (0.0001)	-0.0005** (0.0001)	-0.0004** (0.0001)
Living Space ($100m^2$)	0.4230** (0.0241)	0.2584** (0.0114)	0.4118** (0.0293)
Living Space ($100m^2$) sqrd	-0.0226** (0.0035)	-0.0025** (0.0001)	-0.0228** (0.0044)
No. of Rooms	-0.0219** (0.0037)	-0.0057 (0.0035)	-0.0177** (0.0037)
Age (10 years)	-0.0196** (0.0013)	-0.0201** (0.0011)	-0.0183** (0.0012)
Guest Toilet	0.0670** (0.0085)	0.0744** (0.0046)	0.0724** (0.0067)
Basement	-0.0151* (0.0070)	-0.0114* (0.0055)	-0.0177** (0.0064)
Holiday House	-0.0595** (0.0149)	-0.0351** (0.0107)	-0.0274 (0.0211)
Protected Building	0.1332** (0.0436)	0.0726 (0.0518)	0.1667** (0.0423)
Construction	0.0159 (0.0099)	0.0236** (0.0065)	0.0197* (0.0085)
Building Type			
Bungalow	0.0867** (0.0154)	0.0527** (0.0109)	0.0732** (0.0136)
Farmhouse	-0.0886* (0.0387)	-0.1183** (0.0452)	-0.1215** (0.0418)
Mansion	0.1994** (0.0255)	0.2421** (0.0165)	0.1869** (0.0233)
Other Property	-0.1766** (0.0207)	-0.1842** (0.0117)	-0.1578** (0.0161)
Semi-detached House	-0.0295* (0.0134)	-0.0554** (0.0091)	-0.0506** (0.0114)
Single-family House	0.0425** (0.0120)	0.0186* (0.0074)	0.0281* (0.0111)
Special Property	0.0560 (0.0286)	0.0439 (0.0224)	0.0234 (0.0332)
Terraced House	-0.1104** (0.0179)	-0.1317** (0.0095)	-0.1383** (0.0141)
Condition			
Completely Renovated	0.0851** (0.0171)	0.0609** (0.0112)	0.0670** (0.0165)
First Occupancy	0.0656** (0.0103)	0.0527** (0.0062)	0.0677** (0.0092)
First Occupancy after Reconstruction	0.0681 (0.0399)	0.1547** (0.0344)	0.0579 (0.0556)
Like New	0.1969** (0.0095)	0.2007** (0.0078)	0.1947** (0.0101)
Modernised	0.0144 (0.0124)	0.0255** (0.0083)	0.0269 (0.0144)
Reconstructed	0.0593** (0.0165)	0.0862** (0.0136)	0.0537** (0.0191)
Well Kempt	0.0276** (0.0092)	0.0231** (0.0061)	0.0327** (0.0081)
Neighborhood Characteristics			
Population (100)	0.0015* (0.0006)	0.0011** (0.0003)	0.0008 (0.0006)
Purchase Power PC (1,000€)	0.0170** (0.0032)	0.0180** (0.0029)	0.0118** (0.0031)
No. Commercial Buildings (100)	-0.0180** (0.0068)	0.0021 (0.0049)	-0.0101 (0.0058)
Unemployment Rate (pp)	0.0034 (0.0021)	0.0014 (0.0024)	0.0001 (0.0025)
Share of Foreigners	0.0006 (0.0010)	0.0032** (0.0010)	0.0011 (0.0008)
Share of Families	-0.0006** (0.0002)	-0.0007** (0.0002)	-0.0002 (0.0002)
Share Above Age 60	0.0017 (0.0016)	0.0028 (0.0015)	0.0017 (0.0015)
Euclidean Distances			
Distance Highway (km)	-0.0071* (0.0036)	-0.0038* (0.0017)	-0.0083 (0.0057)
Distance Regional Center (km)	-0.0056 (0.0146)	0.0008 (0.0075)	-0.0096 (0.0201)
Distance Regional Center sqrd	-0.0030 (0.0024)	-0.0001 (0.0005)	-0.0031 (0.0041)
Distance Higher-order Center (km)	-0.0076* (0.0031)	-0.0070** (0.0013)	-0.0041 (0.0053)
Railway < 500m	-0.0054 (0.0103)	-0.0204** (0.0061)	0.0045 (0.0091)
Train Station < 2km	0.0486** (0.0112)	0.0510** (0.0123)	0.0757** (0.0150)
No. Observations	67,908	180,492	74,505
Adj. R2	0.6921	0.6384	0.7141

This table shows the results for all coefficients of Equation (1) with logarithmized prices as the dependent variable for $d_t = 2km$. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level (in parentheses).
** $p < 0.05$, ** $p < 0.01$*

Table C.6: Difference-in-Difference Estimates for All d_t

	Donut Controls	Real Controls	Hypothetical Controls
Train Station < 1.5km	0.0563**	0.0487**	0.0844**
	(0.0106)	(0.0107)	(0.0152)
No. Observation	68,881	156,293	67,131
Adj. R2	0.7205	0.6907	0.7255
Train Station < 2km	0.0490**	0.0511**	0.0759**
	(0.0111)	(0.0123)	(0.0147)
No. Observation	67,908	180,492	74,505
Adj. R2	0.6919	0.6383	0.7137
Train Station < 2.5km	0.0575**	0.0433**	0.0861**
	(0.0134)	(0.0151)	(0.0174)
No. Observation	60,213	186,267	73,406
Adj. R2	0.6635	0.6369	0.6894
Train Station < 3km	0.0551**	0.0382**	0.0681**
	(0.0151)	(0.0146)	(0.0205)
No. Observation	56,266	212,837	76,192
Adj. R2	0.6624	0.6485	0.6741

This table shows the results of Equation (1) *for* $d_t = 1.5, 2, 2.5, 3km$. *The values for* d_t *are chosen based on the results of a cross-sectional regression presented in Figure 4, that show positive effects up to 3km that decline after 1.5km. This is in line with previous research that consistently reports positive accessibility premiums extending up to a distance of 1.5km from public transport stops that serve local rail services (Rennert 2022), and the notion that individuals might be willing to travel greater distances to access a train station serving regional rail as it facilitates more extensive travel. The distance threshold* d_c *exceeds* d_t *by 1km in every scenario. As a result, for* $d_t = 1.5km$, *properties within 2.5km to 4.5km to the respective opened station build the donut control group. For the remaining three distance specifications, the donut control groups' upper and lower limits are 3km and 5km, 3.5km and 5.5km as well as 4km and 6km, respectively. The real and hypothetical control groups comprise properties within 1.5km, 2km, 2.5km and 3km of the respective control and hypothetical station. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level (in parentheses). Information on the sample of opened stations used in the analysis is presented in Table C.3.*
*** $p < 0.01$*

Table C.7: Difference-in-Differences Estimates - One Kilometer Distance Bands

	Donut Controls	Real Controls	Hypothetical Controls
Train Station < 1km	0.0803**	0.0576**	0.1154**
	(0.0120)	(0.0128)	(0.0183)
No. Observation	58,301	96,597	42,182
Adj. R2	0.7208	0.6946	0.7283
1km < Train Station < 2km	0.0429**	0.0313*	0.0341*
	(0.0120)	(0.0127)	(0.0151)
No. Observation	49,973	92,608	35,886
Adj. R2	0.7021	0.6559	0.7271
2km < Train Station < 3km	0.0340	-0.0300	-0.0488*
	(0.0178)	(0.0200)	(0.0218)
No. Observation	33,299	73,694	22,278
Adj. R2	0.6543	0.6863	0.677

The table's upper panel shows the results for estimating Equation (1) with $d_t = 1.5km$ and data that only contains properties within 1km of the opened and control stations. The negative coefficient of -0.048 obtained using the hypothetical control group is to a large extent driven by sizeable negative effects on houses observed more than five years after the train station commissioning. Omitting these observations yields more comparable results to the ones provided by the donut and real control group that are close to zero and statistically insignificant (Figure C.1). The middle panel contains the results of Equation (1) for $d_t = 2km$ with data restricted to properties between 1km and 2km of the stations, while the lower panel holds the results for estimating Equation (1) with $d_t = 3km$ and data that only contains properties beyond 2km of the treatment and control stations. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level (in parentheses).

** $p < 0.05$, ** $p < 0.01$*

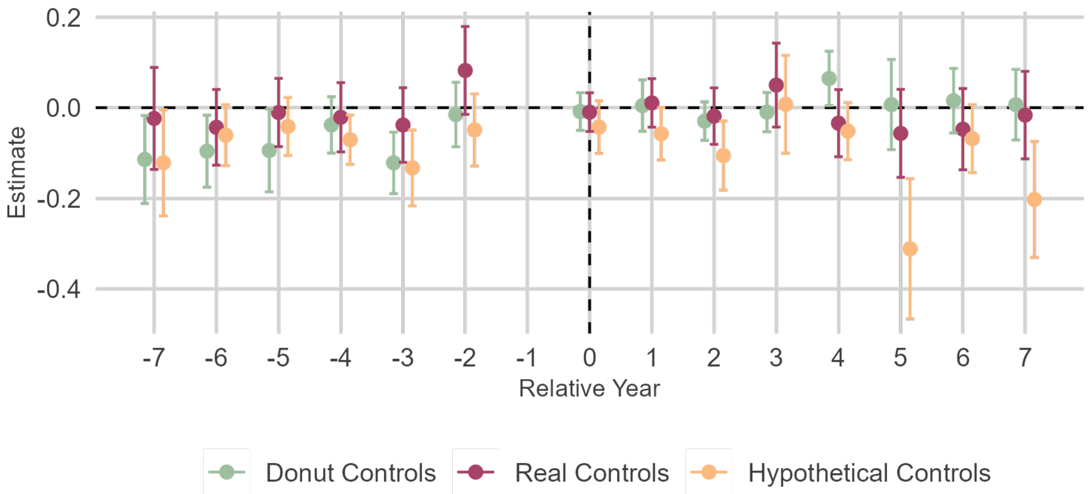

Figure C.1: DiD Event Study Estimates - 2km < Train Station < 3km

The figure shows the event study estimates and 95%-confidence intervals of Equation (1) *with* $d_t = 3km$*, estimated with data that only contains properties beyond 2km of the opened train stations. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level. Relative years exceeding* -6 *and* 6 *are binned in relative year* -7 *and* 7*, respectively.*

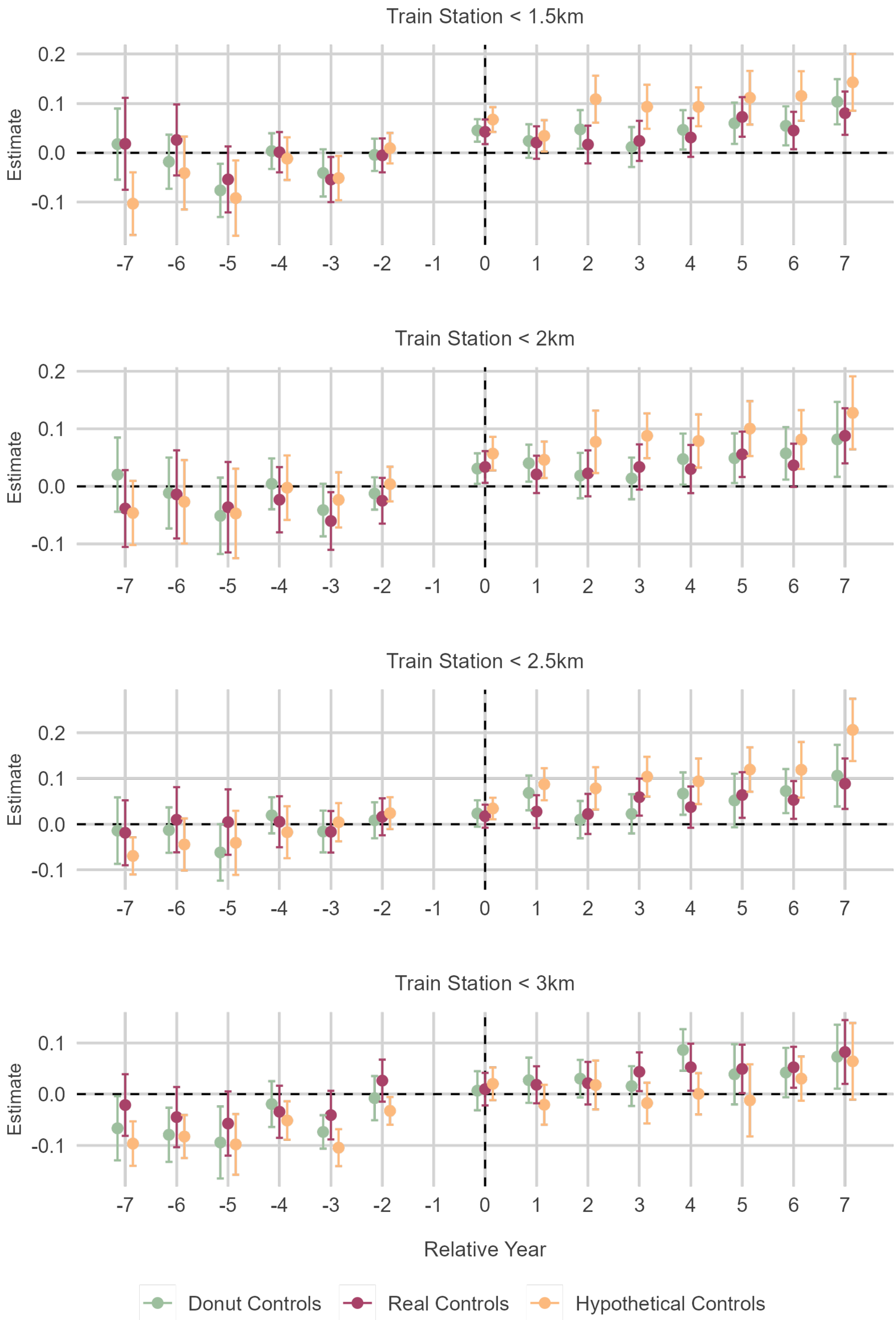

Figure C.2: DiD Event Study Estimates for All d_t

The figure shows the event study estimates and 95%-confidence intervals of Equation (1) *with* $d_t = 1.5km, 2km, 2.5km, 3km$ *for the years pre- and succeeding the year in which the train station was taken into operation (relative year* $= 0$*). Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level. Due to decreasing observation counts in the years preceding and following the station opening, all dwellings that are observed in years more than 7 years before or after the opening of the station are binned in relative year* -7 *and* 7*, respectively.*

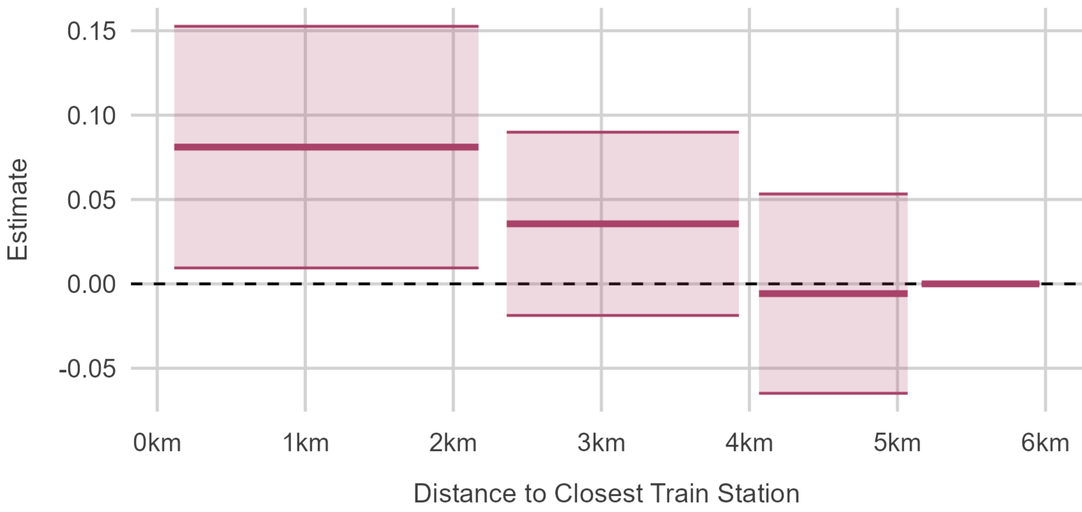

Figure C.3: The Effect of Train Station Openings on House Prices Across Distance - Autonomous Bin Selection

The figure presents the treatment effect curve estimated using the approach proposed in Butts (2023) with an autonomous binning procedure. Standard errors are clustered on post code district level.

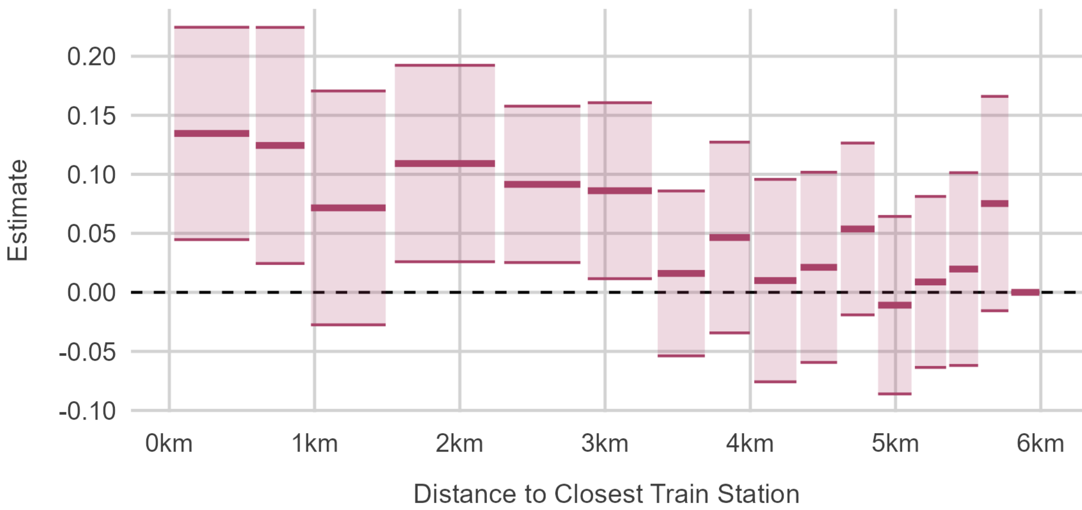

Figure C.4: The Effect of Train Station Openings on House Prices Across Distance - 16 Bins

The figure presents the treatment effect curve estimated using the approach proposed in Butts (2023) with a fixed number of 16 bins. Standard errors are clustered on post code district level.

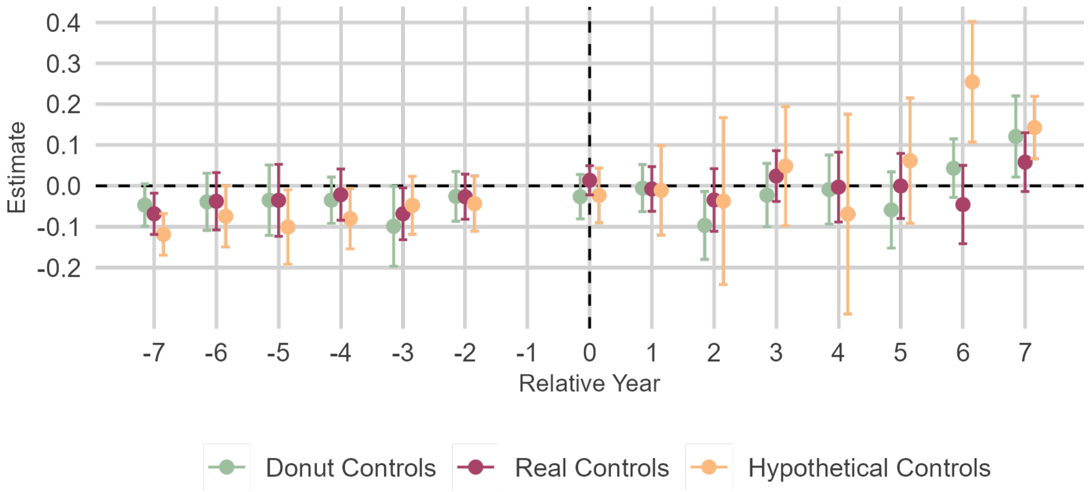

Figure C.5: DiD Event Study Estimates - Lower-Order Centers and Below - Train Station $< 2\text{km}$

The figure shows the event study estimates and 95%-confidence intervals of Equation (1) *estimated with TWFE for* $d_t = 2km$*, for train stations located in lower-order centers or municipalities without a central function. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level. Relative years exceeding* -6 *and* 6 *are binned in relative year* -7 *and* 7*, respectively.*

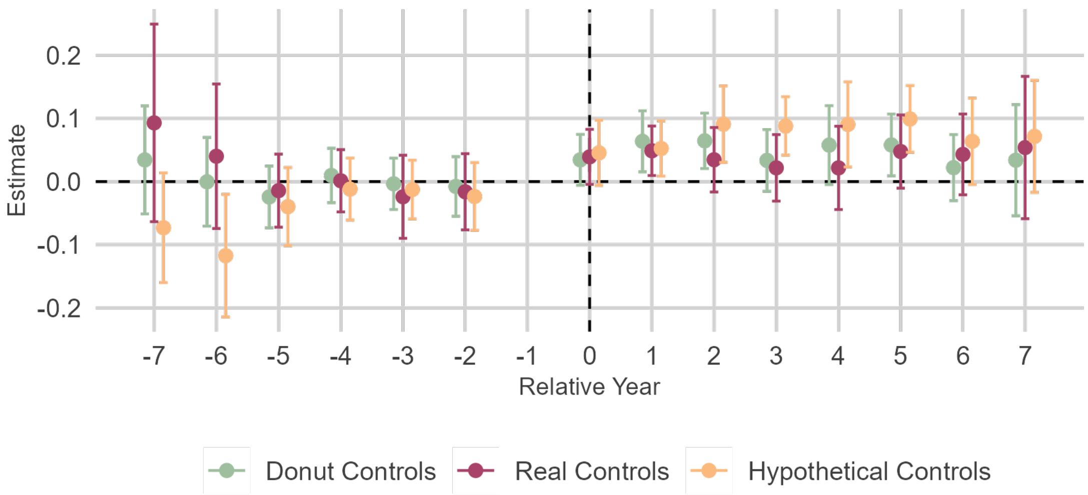

Figure C.6: DiD Event Study Estimates - Medium-Order Centers and Above - Train Station $< 2\text{km}$

The figure shows the event study estimates and 95%-confidence intervals of Equation (1) *estimated with TWFE for* $d_t = 2km$*, for train stations located in medium- and higher-order centers. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level. Relative years exceeding* -6 *and* 6 *are binned in relative year* -7 *and* 7*, respectively.*

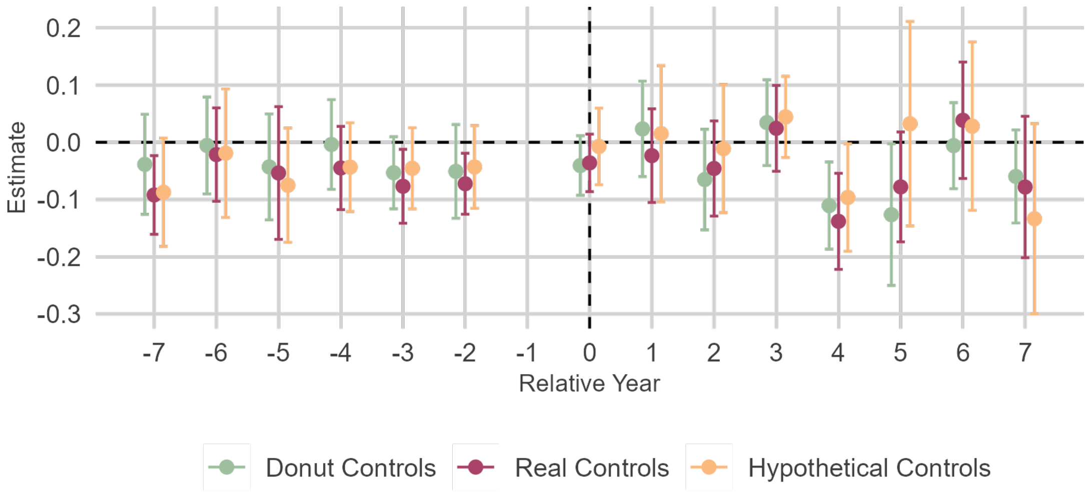

Figure C.7: DiD Event Study Estimates - Population Below Median - Train Station < 2km

The figure shows the event study estimates and 95%-confidence intervals of Equation (1) *estimated with TWFE for* $d_t = 2km$*, for train stations with population counts below the station sample's median. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level. Relative years exceeding* -6 *and* 6 *are binned in relative year* -7 *and* 7*, respectively.*

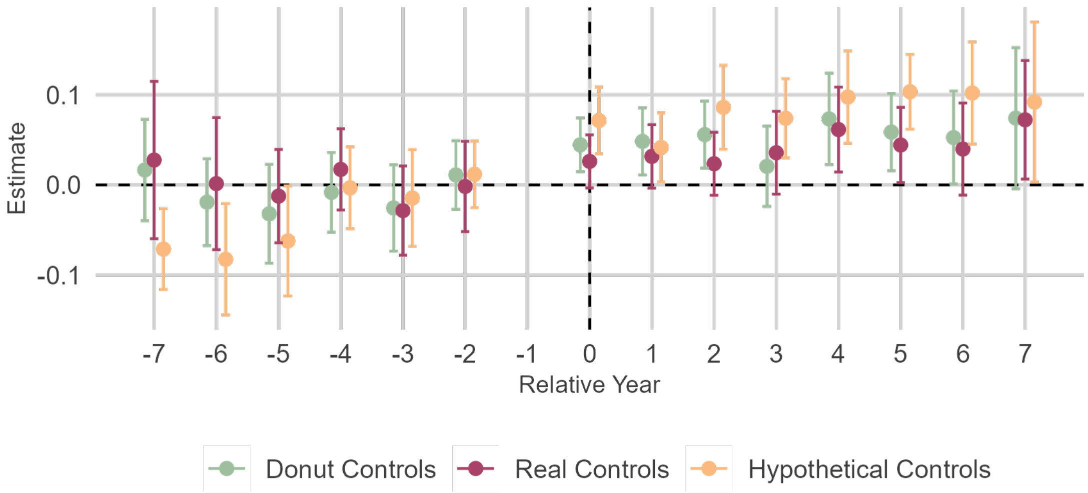

Figure C.8: DiD Event Study Estimates - Population Above Median - Train Station < 2km

The figure shows the event study estimates and 95%-confidence intervals of Equation (1) *estimated with TWFE for* $d_t = 2km$*, for train stations with population counts above the station sample's median. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level. Relative years exceeding* -6 *and* 6 *are binned in relative year* -7 *and* 7*, respectively.*

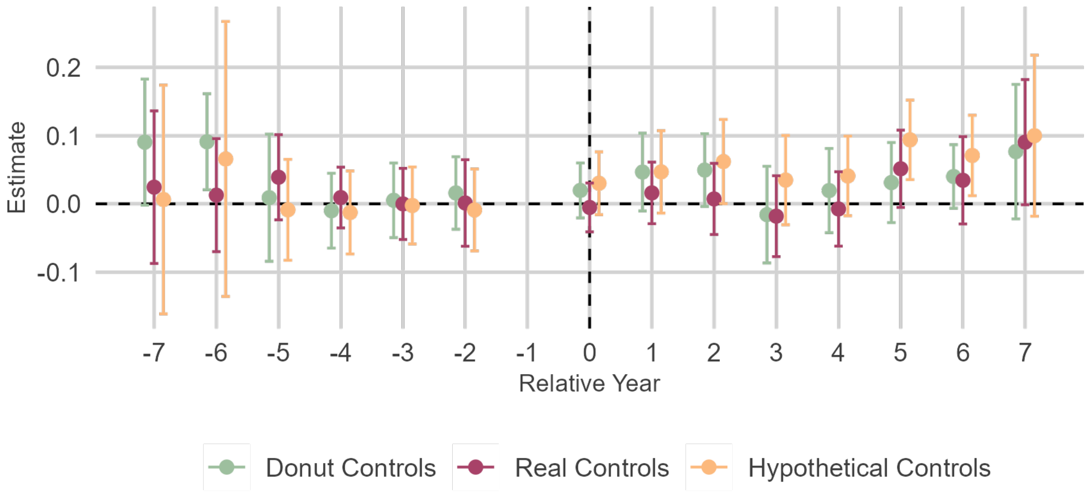

Figure C.9: DiD Event Study Estimates - Purchasing Power per Capita Below Median - Train Station < 2km

The figure shows the event study estimates and 95%-confidence intervals of Equation (1) *estimated with TWFE for* $d_t = 2km$*, for train stations with purchasing power per capita below the station sample's median. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level. Relative years exceeding* -6 *and* 6 *are binned in relative year* -7 *and* 7*, respectively.*

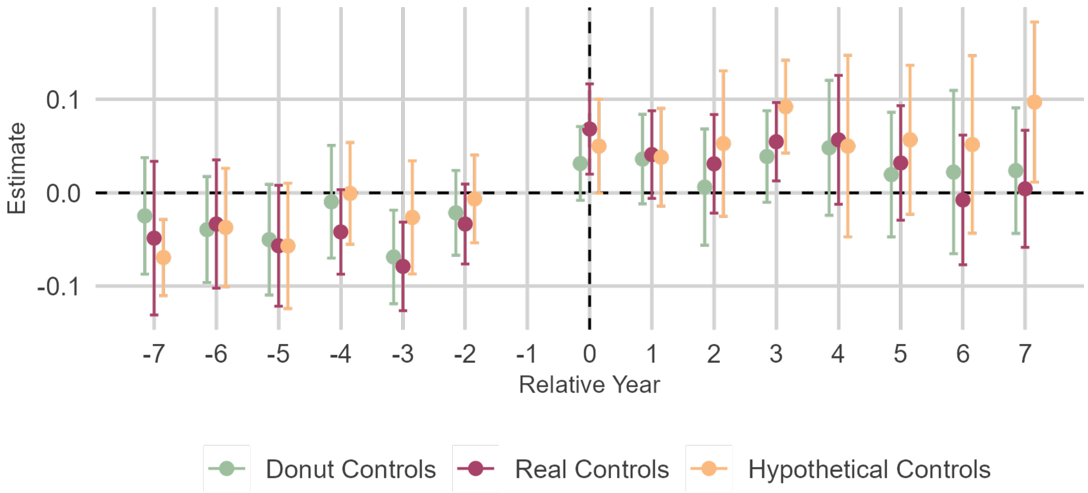

Figure C.10: DiD Event Study Estimates - Purchasing Power per Capita Above Median - Train Station < 2km

The figure shows the event study estimates and 95%-confidence intervals of Equation (1) *estimated with TWFE for* $d_t = 2km$*, for train stations with purchasing power per capita above the station sample's median. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level. Relative years exceeding* -6 *and* 6 *are binned in relative year* -7 *and* 7*, respectively.*

D Robustness Checks

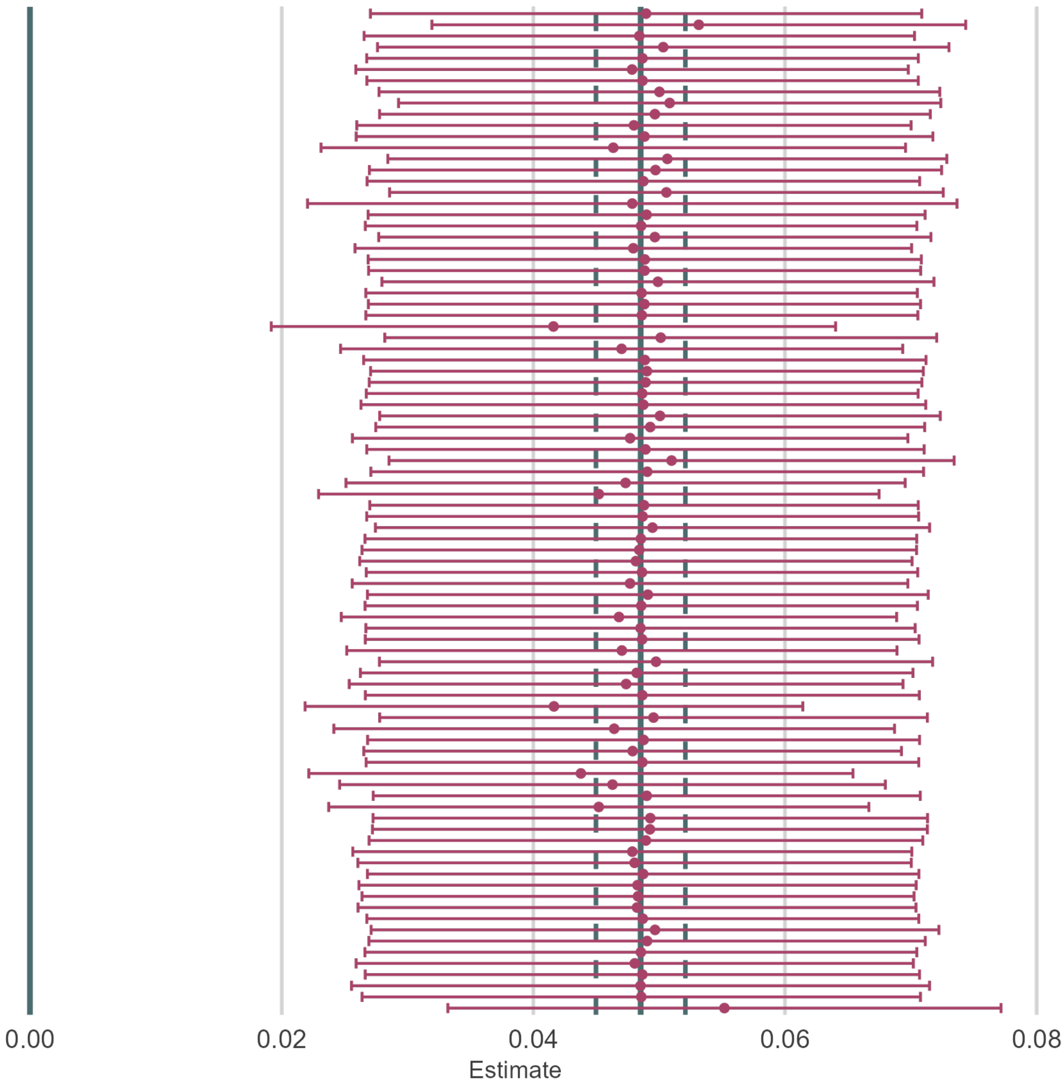

Figure D.1: DiD Estimates Omitting Station Openings - Donut Control Group - Train Station < 2km

The figure shows the aggregate treatment effects for Equation (1) *using the donut control group with* $d_t = 2km$*, systematically excluding one station opening at a time in each regression (y-axis). The two dotted lines illustrate two standard deviations from the regressions' average effect of 0.048. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level (in parentheses).*

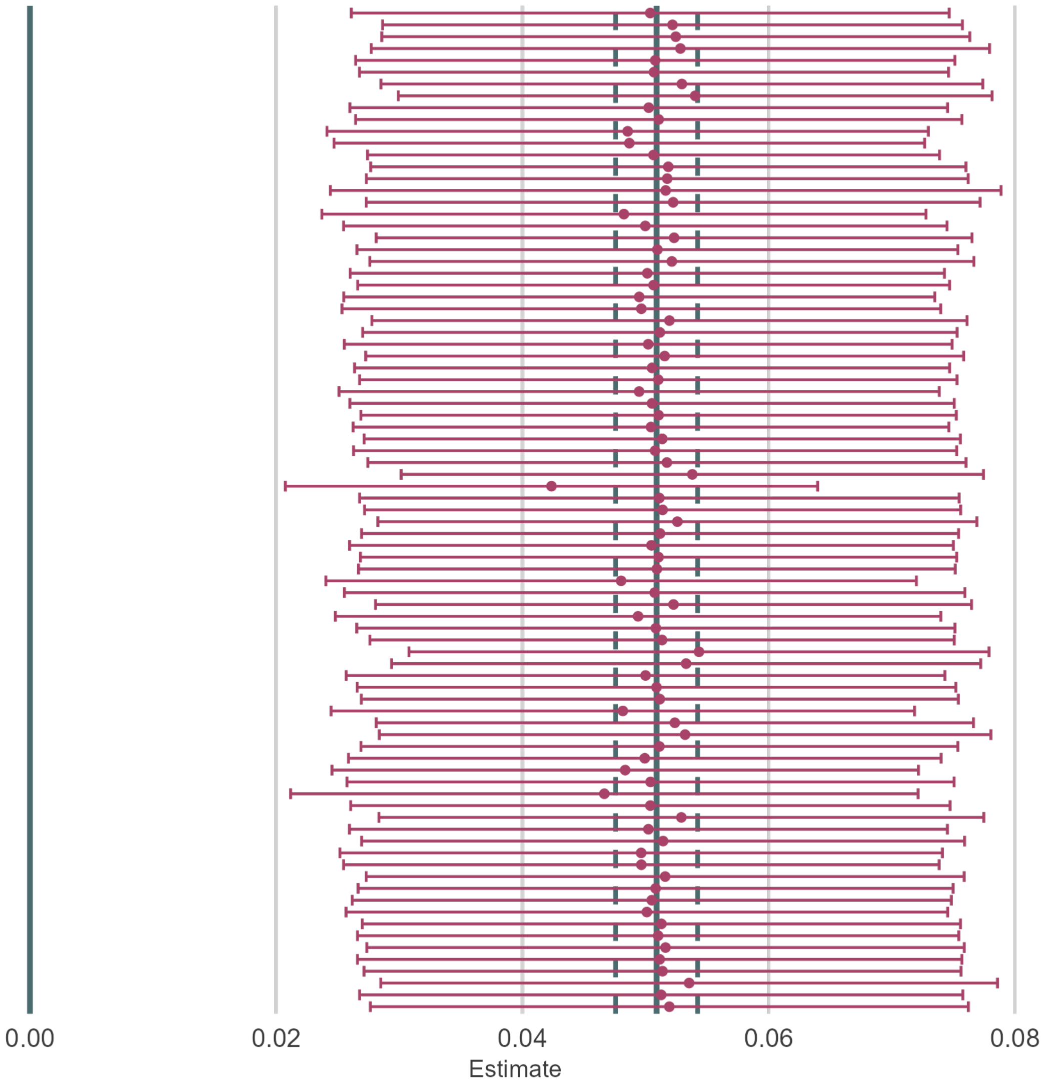

Figure D.2: DiD Estimates Omitting Station Openings - Real Control Group - Train Station < 2km

The figure shows the aggregate treatment effects for Equation (1) *using the real control group with* $d_t = 2km$*, systematically excluding one station opening at a time in each regression (y-axis). The two dotted lines illustrate two standard deviations from the regressions' average effect of 0.051. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level (in parentheses).*

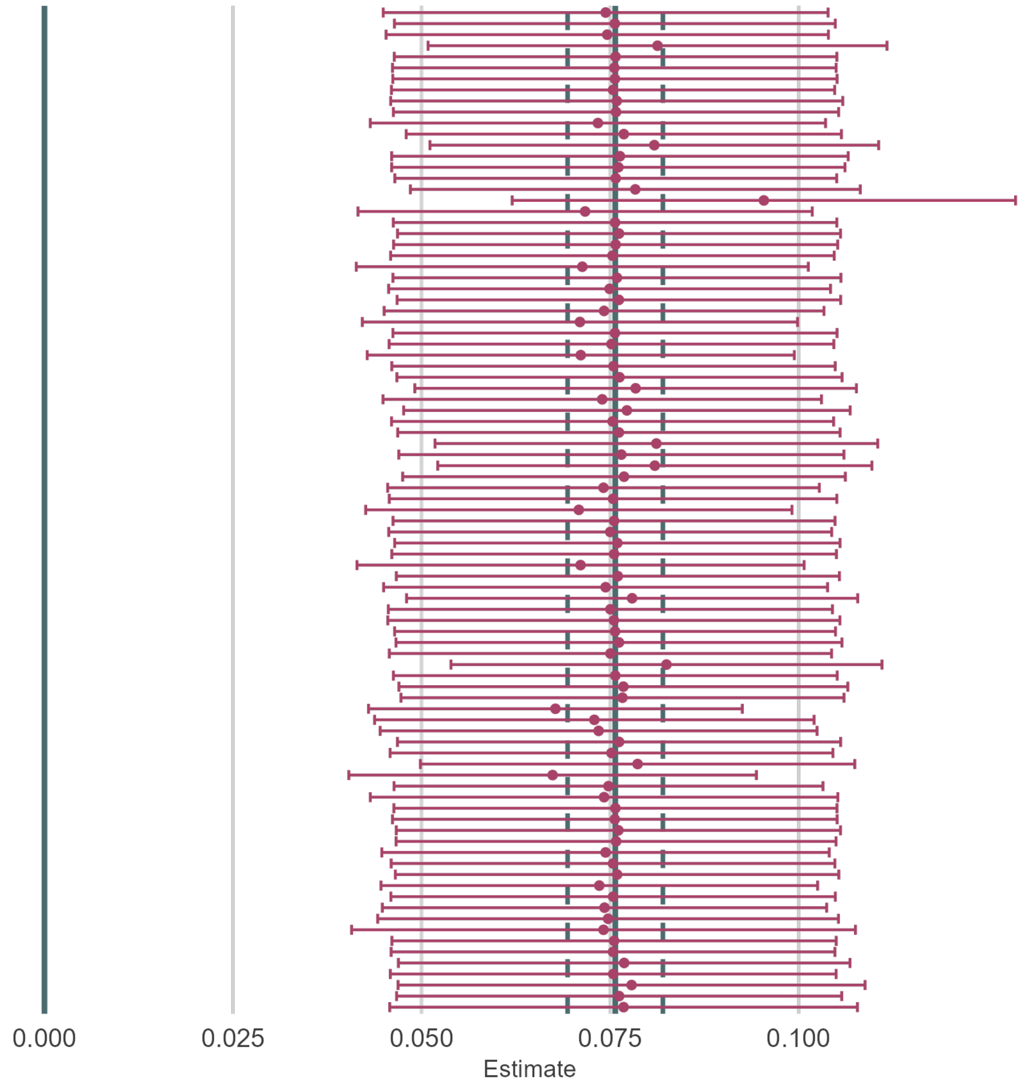

Figure D.3: DiD Estimates Omitting Station Openings - Hypothetical Control Group - Train Station < 2km

The figure shows the aggregate treatment effects for Equation (1) *using the hypothetical control group with* $d_t = 2km$, *systematically excluding one station opening at a time in each regression (y-axis). The two dotted lines illustrate two standard deviations from the regressions' average effect of 0.075. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level (in parentheses).*

To ensure that the estimated effects are not driven by a small set of train station openings, I estimate the main specification of Equation (1), systematically excluding one station at a time in each regression. The results using the donut, real and hypothetical control group are displayed in Figures D.1 to D.3. In the majority of cases, excluding individual station openings from the analysis does not result in a treatment effect that significantly deviates from the average effect observed across all regressions. However, the omission of a few specific stations leads to effects that deviate by more than two standard deviations from the mean effect. This suggests that these particular stations play a decisive role in influencing the estimated effect. To examine this further, I replicate the main results from Table 2 excluding these influential stations from the analysis (Table D.1). The results indicate a reduction in the estimated effects for the donut and real control group, and a slight increase using the hypothetical control group. However, the disparities between the replicated results and the initial estimates are not substantial. The same observation holds true for the event study estimates illustrated in Figure D.4.

Furthermore, I employ various data samples and regression specifications for estimating Equation (1) to ensure the results' robustness. Table D.2 presents the outcomes of Equation (1) with d_t set to 2km, incorporating different sets of covariates. Specifically, I estimate the model without any covariates, with a reduced set of control variables, and with an extensive set of control variables (main specification) (see Tables A.1, A.2 and A.3 for a list of variables). Furthermore, I assess the robustness of the results concerning the choice of variables utilized in the allocation of real and hypothetical control stations to the train stations that were commissioned, by conducting the propensity score matching using alternative sets of matching variables. The results can be found in Table D.3. Additionally, since I initially limit the sample of station openings to train stations with at least 20 nearby houses observed during the study period in the main analysis, I examine the robustness of the results concerning this inclusion criterion. To do so, I repeat the estimation with varying numbers of excluded opened stations. In the most inclusive

Table D.1: DiD Estimates - Without Outlier Station Openings

	Donut Controls	Real Controls	Hypothetical Controls
Train Station $<$ 2km	0.0416** (0.0093)	0.0410** (0.0109)	0.0796** (0.0190)
No. Observation	62,539	172,636	70,253
Adj. R2	0.6834	0.6431	0.7057

The table shows the estimates for the main specification of Equation (1) with $d_t = 2km$, excluding those station openings and their respective control groups that were identified as outliers in Figures D.1, D.2, D.3. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level (in parentheses).
*** p < 0.01*

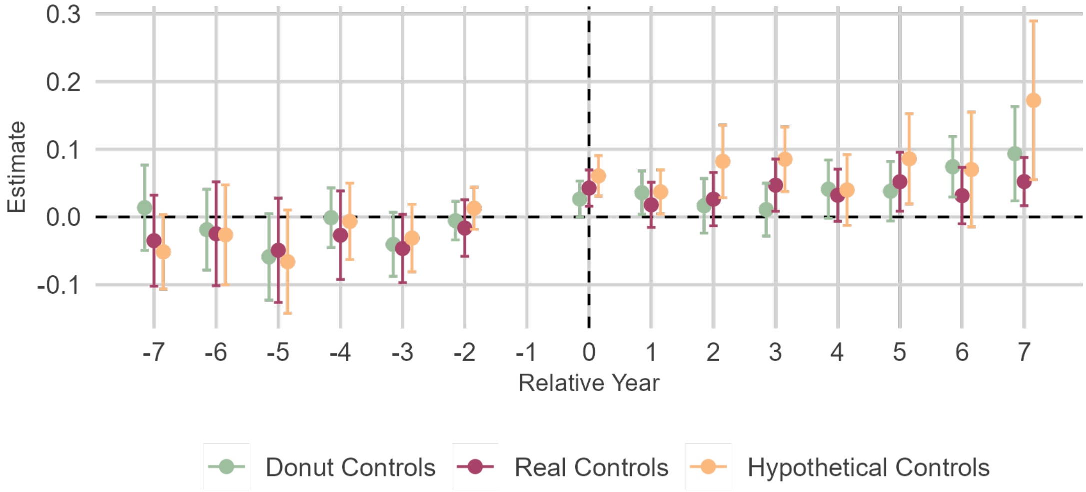

Figure D.4: DiD Event Study Estimates - Without Outlier Station Openings - Train Station < 2km

The figure shows the event study estimates and 95%-confidence intervals of Equation (1) *with* $d_t = 2km$*, excluding those station openings and their respective control groups that were identified as outliers in Figures D.1, D.2, D.3. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level. Relative years exceeding* -6 *and* 6 *are binned in relative year* -7 *and* 7*, respectively.*

variant, all stations were included, while in more restrictive variants, inclusion required the observation of 100, 200, or 400 properties in proximity to the station (Table D.4). All three robustness checks corroborate the main findings and suggest that the results are not influenced by the choice of control and matching variables or the selection of station openings.

Table D.2: DiD Estimates - Different Sets of Covariates - Train Station < 2km

	No Covariates	Reduced Set of Covariates	Main Specification
Donut Control Group	0.0388**	0.0570**	0.0486**
	(0.0143)	(0.0114)	(0.0112)
No. Observation	67,908	67,908	67,908
Adj. R2	0.3892	0.6629	0.6921
Real Control Group	0.0348*	0.0555**	0.0510**
	(0.0154)	(0.0130)	(0.0123)
No. Observation	180,492	180,492	180,492
Adj. R2	0.3103	0.6020	0.6384
Hypothetical Control Group	0.0648**	0.0763**	0.0757**
	(0.0190)	(0.0163)	(0.0150)
No. Observation	74,505	74,505	74,505
Adj. R2	0.4124	0.6884	0.7141

The table shows the estimates for the main specification of Equation (1) with $d_t = 2km$ incorporating different sets of covariates. The model is estimated without any covariates, with a reduced set of control variables and an extensive one (main specification) (see Tables A.1, A.2 and A.3). Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level (in parentheses).
** $p < 0.05$, ** $p < 0.01$*

Table D.3: DiD Estimates - Alternative Matching Variables - Train Station < 2km

	Reduced Set	Basic Set	Main Specification	Extended Set
Donut Control Group	0.0486**	0.0486**	0.0486**	0.0486**
	(0.0112)	(0.0112)	(0.0112)	(0.0112)
No. Observation	67,908	67,908	67,908	67,908
Adj. R2	0.6921	0.6921	0.6921	0.6921
Real Control Group	0.0426**	0.0432**	0.0510**	0.0499**
	(0.0120)	(0.0123)	(0.0123)	(0.0119)
No. Observation	184,528	164,845	180,492	173,766
Adj. R2	0.6845	0.6531	0.6384	0.668
Hypothetical Control Group	0.0742**	0.0695**	0.0757**	0.0762**
	(0.0149)	(0.0155)	(0.015)	(0.0169)
No. Observation	75,463	74,365	74,505	73,962
Adj. R2	0.7148	0.7140	0.7141	0.7212

The table shows the estimates for the main specification of Equation (1) with $d_t = 2km$ using alternative sets of variables that are included in the matching of a real and hypothetical control station to each opened station. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level (in parentheses).
*** $p < 0.01$*

Table D.4: DiD Estimates - Treatment Group Size Restrictions - Train < 2km

	All Stations	Treatment Group > 20	Treatment Group > 100	Treatment Group > 200	Treatment Group > 400
Donut Control Group	0.0493**	0.0486**	0.0455**	0.0616**	0.0722**
	(0.0111)	(0.0112)	(0.0121)	(0.0126)	(0.0133)
No. Observation	73,368	67,896	58,206	46,505	39,952
Adj. R2	0.7000	0.6920	0.6840	0.6919	0.6889
Real Control Group	0.0451**	0.0511**	0.0435**	0.0478**	0.0517**
	(0.0127)	(0.0123)	(0.0132)	(0.0131)	(0.0146)
No. Observation	212,855	180,321	130,617	95,879	72,392
Adj. R2	0.6376	0.6387	0.6480	0.6663	0.6548
Hypothetical Control Group	0.0763**	0.0757**	0.0670**	0.0749**	0.0808**
	(0.0142)	(0.0150)	(0.0144)	(0.0155)	(0.0164)
No. Observation	85,772	74,505	58,363	47,400	38,544
Adj. R2	0.7101	0.7141	0.7050	0.7116	0.6913

The table shows the estimates for the main specification of Equation (1) *with* $d_t = 2km$ *including different sets of station openings in the analysis sample. In the mildest variant all stations are included, in the more restrictive ones at least 20 (main specification) 100, 200 and 400 properties in proximity of the opened station need to be observed, so that the station is included in the estimation sample. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level (in parentheses).*
*** $p < 0.01$*

For the main analysis, control group units are selected based on their location in an area that is allocated as control group region to a respective station opening, e.g. the area surrounding a hypothetical or real train station that is matched to the opened train station. To ensure that the results are not driven by this definition of control groups on the train station level, I match to each real estate object in the treatment group a control group unit situated between three to five kilometers of any opened station, in proximity of any active train station and adjacent to any hypothetical stations, respectively, based on similarities in property and neighborhood characteristics. The results are presented in Table D.5 and Figure D.5 and corroborate the main findings.

Table D.5: DiD Estimates - Matching on Property Level

	Donut Controls	Real Controls	Hypothetical Controls
Train Station < 2km	0.0528**	0.0314**	0.0693**
	(0.0106)	(0.0106)	(0.0134)
No. Observation	57,313	114,551	58,199
Adj. R2	0.6847	0.6297	0.6813

The table shows the estimates for the main specification of Equation (1) *with* $d_t = 2km$*, with control group units allocated to houses in the treatment group on property level. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level (in parentheses).*
*** $p < 0.01$*

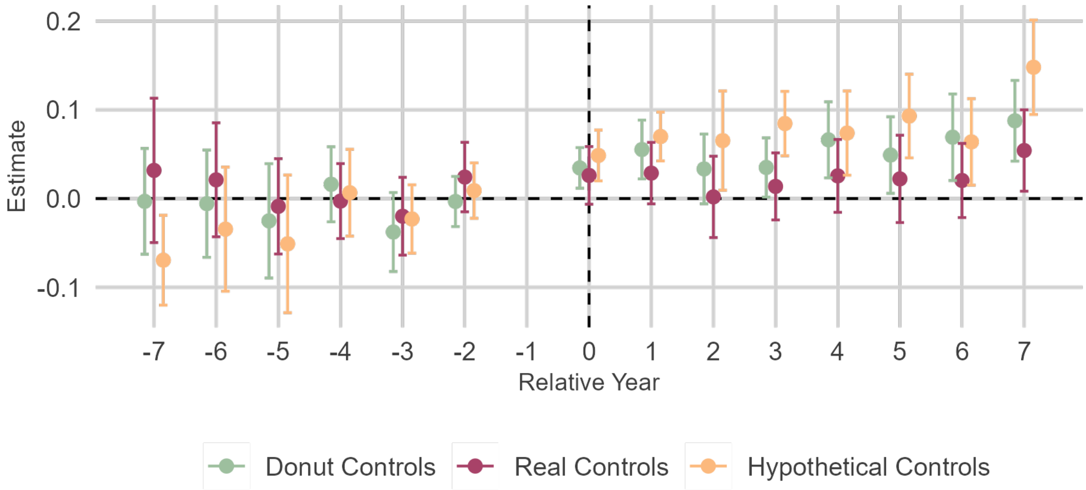

Figure D.5: DiD - Event Study Estimates - Matching on Property Level - Train Station < 2km

The figure shows the event study estimates and 95%-confidence intervals of Equation (1) with $d_t = 2km$, with control group units allocated to houses in the treatment group on property level. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level. Relative years exceeding -6 and 6 are binned in relative year -7 and 7, respectively.

Table D.6 provides the outcomes of Equation (1) with d_t set to 2km, utilizing different variants of time fixed effects. It shows that while the results obtained employing the real control group exhibit little change across the different specifications, neglecting region-specific shocks over time significantly impacts the results utilizing the donut and hypothetical control groups. Given the weaker assumption of region-specific shocks, I select fixed effects saturated in the interaction of year and region as the main specification for the analysis. However, the utilization of year-region fixed effects can potentially introduce multicollinearity issues, as a substantial portion of the data's variation is already explained by the fixed effects, particularly in regions solely encompassing properties assigned to control groups.

This issue is less prominent in the analysis sample utilizing the donut control group, as properties in both treatment and control groups are closely located to each other, often within the same region. In contrast, for the analysis samples relying on the real and hypothetical control groups, properties may not necessarily be close to their respective station openings and can be dispersed across Germany. Consequently, some control groups might fall within regions without properties near an opened train station, and therefore do not contribute to the estimation of the coefficient of interest. To address this concern, I opt for year-labor market region fixed effects in the main analysis, as the size of labor market regions exceeds the one of municipalities and districts. In order to further address this concern, I estimate Equation (1) using analysis samples constructed by matching opened stations to real and hypothetical control stations that are located within the

Table D.6: DiD Estimates - Alternative Time Fixed Effects - Train Station < 2km

	Year FE	Year-LMR FE	Year-District FE
Donut Control Group	0.0260	0.0486**	0.0457**
	(0.0234)	(0.0112)	(0.0136)
No. Observation	67,908	67,908	67,908
Adj. R2	0.6626	0.6921	0.6965
Real Control Group	0.0506**	0.0510**	0.0657**
	(0.0177)	(0.0123)	(0.0153)
No. Observation	180,492	180,492	180,492
Adj. R2	0.5683	0.6384	0.6523
Hypothetical Control Group	0.0410*	0.0757**	0.0914**
	(0.0182)	(0.0150)	(0.0168)
No. Observation	74,505	74,505	74,505
Adj. R2	0.6854	0.7141	0.7179

The table shows the estimates for the main specification of Equation (1) with $d_t = 2km$ incorporating different time fixed effects. Besides station opening-treatment group fixed effects, the models include year fixed effects, fixed effects on the year-labor market region level (main specification), and on the year-district level, respectively. Standard errors are clustered on station opening-treatment group level (in parentheses).
** $p < 0.05$, ** $p < 0.01$*

same labor market region (Table D.7). The coefficients derived using the donut control group remain identical to the main results since these data samples are unaffected by the alternative matching strategy. Conversely, for the regression variants employing the real and hypothetical control groups, notably smaller sample sizes become evident. In the former case, I chose to match only one operational control station to each station opening, while in the latter, a number of opened train stations lack a corresponding hypothetical train station location within the same labor market region. Consequently, these station openings were excluded from the analysis, leading to a reduction in the number of observations. The estimates align to a large extent with the main findings presented in Table 2, and do not indicate that the aforementioned issue significantly affects the average treatment effects. However, it is important to note, that Figure D.6 suggests a clear violation of the parallel trends assumption when using the hypothetical control group, biasing the estimates upwards.

Table D.7: DiD Estimates - Matching within Labor Market Regions

	Donut Controls	Real Controls	Hypothetical Controls
Train Station < 2km	0.0486**	0.0354**	0.0617**
	(0.0112)	(0.0130)	(0.0120)
No. Observation	67,908	72,204	38,726
Adj. R2	0.6921	0.6654	0.7108

The table shows the estimates for the main specification of Equation (1) *with* $d_t = 2km$ *restricting the pool of real and hypothetical control stations in the propensity score matching to stations that are located in the same labor market region as the opened station. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level.*
*** p < 0.01*

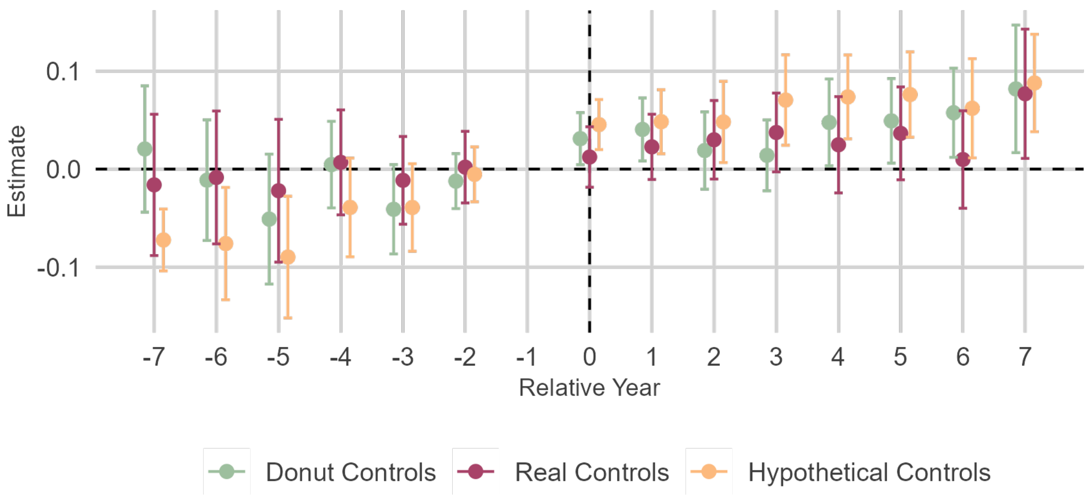

Figure D.6: DiD Event Study Estimates - Matching Within Labor Market Regions - Train Station < 2km

The figure shows the event study estimates and 95%-confidence intervals of Equation (1) *with* $d_t = 3km$, *restricting the pool of real and hypothetical control stations in the propensity score matching to stations that are located in the same labor market region as the treatment station. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level. Relative years exceeding* -6 *and* 6 *are binned in relative year* -7 *and* 7, *respectively.*

Table D.8 presents the aggregate treatment effects for $d_t = 2km$ obtained through the estimator proposed in Sun and Abraham (2021), the approach by Gardner (2022) and a conventional two-way fixed effects (TWFE) regression.[21] Across the alternative econometric approaches, the estimated effects of station openings on property values using the donut control group range from 3.8% to 5.2% and employing the real control group from 5.1% to 5.4%. Utilizing the hypothetical control group, the coefficients fall within the range of 6.1% and 8.1%. In comparison to the other estimators, the approach by Gardner (2022) yields the largest estimates (5.2% to 8.1%), whereas the estimates produced by regular TWFE exhibit the lowest variation across control group variants (3.8% to 6.1%). Figures D.7, D.8 and D.9 illustrate the event study estimates obtained by the alternative econometric approaches for the three control groups. Regardless of the control group variant considered, the event study estimates follow a similar pattern across all estimators. Notably, the approach by Gardner (2022) demonstrates precise coefficient estimates for pre-treatment years, which closely center around zero, suggesting a lack of discernible pre-treatment trends and lending support to the validity of the parallel trends assumption.

[21] I implement the estimators using the R-packages "fixest" (Bergé 2018) and "did2s" (Butts 2021).

Table D.8: DiD Estimates - Alternative Estimators - Train Station < 2km

	S & A	Gardner	TWFE
Donut Control Group	0.0486**	0.0519**	0.0380**
	(0.0112)	(0.0165)	(0.0113)
No. Observation	67,908	66,566	67,909
Real Control Group	0.0510**	0.0541**	0.0509**
	(0.0123)	(0.0164)	(0.0139)
No. Observation	180,492	179,328	180,493
Hypothetical Control Group	0.0757**	0.0805**	0.0607**
	(0.0150)	(0.0153)	(0.0137)
No. Observation	74,505	68,886	74,506

The table shows the estimates for the main specification of Equation (1) *with* $d_t = 2km$*, using the estimators proposed by Sun and Abraham (2021) and Gardner (2022) as well as regular TWFE. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level (in parentheses).*
*** $p < 0.01$*

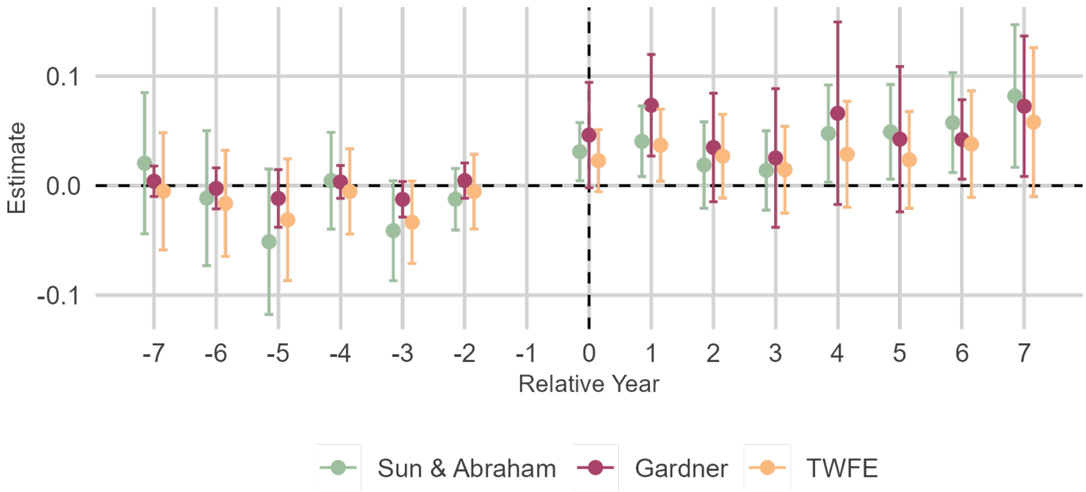

Figure D.7: DiD Event Study Estimates - Alternative Estimators - Donut Control Group - Train Station < 2km

The figure shows the event study estimates and 95%-confidence intervals obtained by the estimators proposed by Sun and Abraham (2021) and Gardner (2022) as well as regular TWFE for $d_t = 2km$*, using the donut control group. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level. Relative years exceeding* -6 *and* 6 *are binned in relative year* -7 *and* 7*, respectively.*

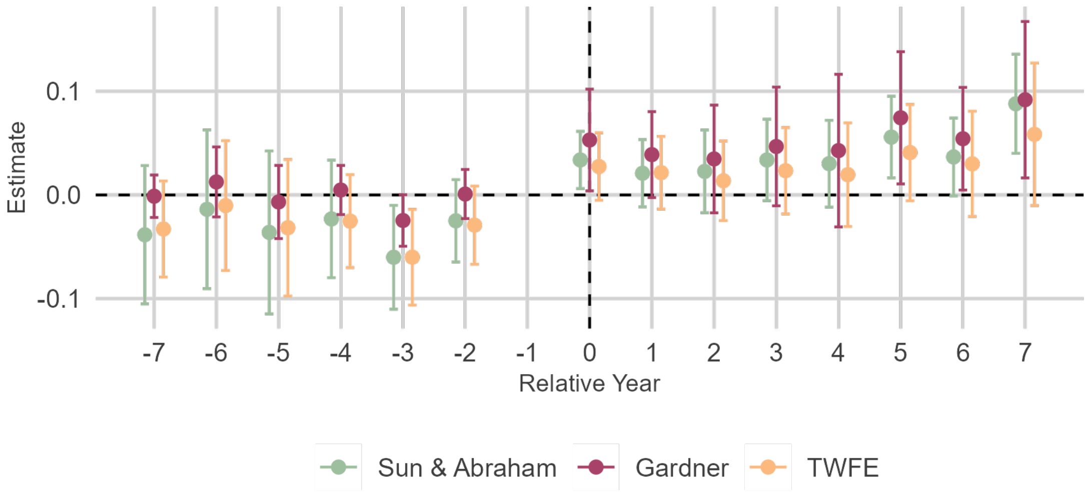

Figure D.8: DiD Event Study Estimates - Alternative Estimators - Real Control Group - Train Station < 2km

The figure shows the event study estimates and 95%-confidence intervals obtained by the estimators proposed by Sun and Abraham (2021) and Gardner (2022) as well as regular TWFE for $d_t = 2km$*, using the real control group. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level. Relative years exceeding* -6 *and* 6 *are binned in relative year* -7 *and* 7*, respectively.*

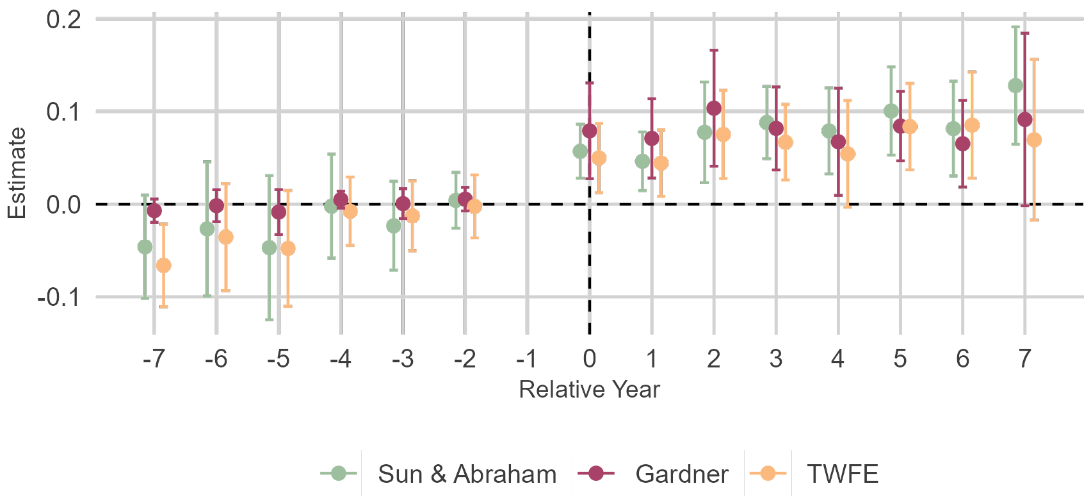

Figure D.9: DiD Event Study Estimates - Alternative Estimators - Hypothetical Control Group - Train Station < 2km

The figure shows the event study estimates and 95%-confidence intervals obtained by the estimators proposed by Sun and Abraham (2021) and Gardner (2022) as well as regular TWFE for $d_t = 2km$*, using the hypothetical control group. Fixed effects are on station opening-treatment group and year-labor market region level, standard errors clustered on station opening-treatment group level. Relative years exceeding* -6 *and* 6 *are binned in relative year* -7 *and* 7*, respectively.*

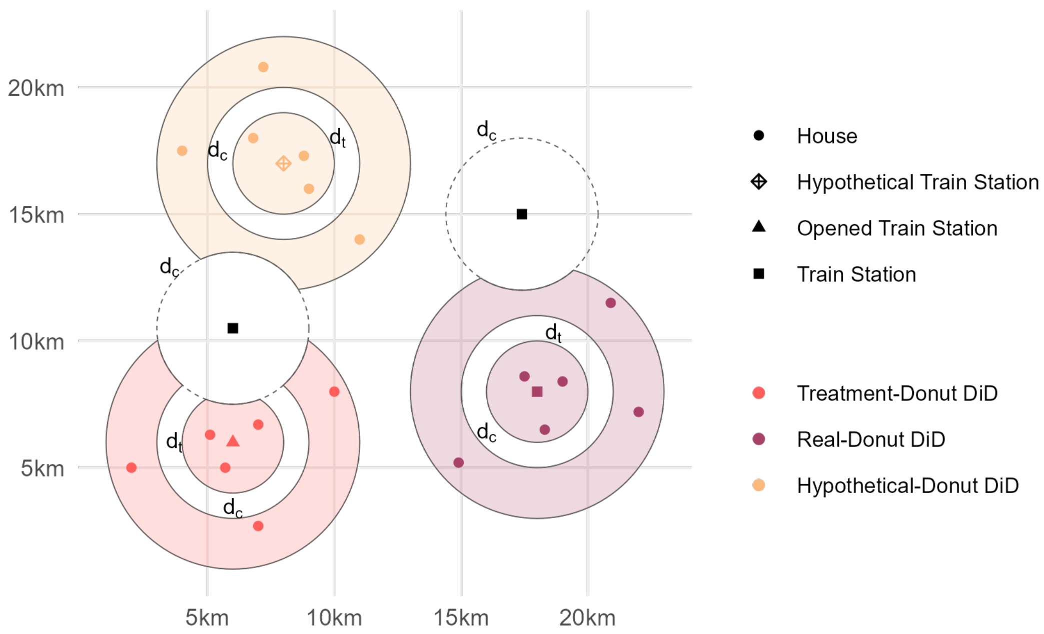

Figure D.10: Illustration of Triple-Differences Approach

Table D.9 shows the results of a Triple-Differences approach based on the comparison between the treatment group and donut control group. For each station opening's real/hypothetical control group, I construct a Difference-in-Differences setting by now considering properties in proximity of the (hypothetical) station as part of the treatment group and employing a donut control group around this area. In the Triple-Differences approach, the first two differences involve comparisons over time and between the treatment and donut control groups. The third difference compares the results of the Difference-in-Differences analysis, in which a station was actually taken into operation, to those where no station opening occurred (see Figure D.10).

The model is given by

$$ln(p_{it}) = \alpha_{gm} + \tau_{tl} + \gamma_{am} + \lambda_{gmt} + \mu_{gma} + \theta_{amt} + \delta D_{itgma} + \beta X_{it} + \epsilon_{it}, \tag{D.1}$$

in which g denotes the treatment group status of house i, m and l the municipality and labor market region its located in, and t its year of observation. a indicates whether the Difference-in-Differences, to which property i pertains, includes an actual station opening or is a placebo. The model incorporates fixed effects on the treatment group-municipality level (α_{gm}), year-labor market region level (τ_{tl}) and actual opening-municipality level (γ_{am}), as well as their interactions λ_{gmt}, μ_{gma}, and θ_{amt}. D_{itgma} is the Triple Differences interaction term and indicates whether a property is located in a treatment group area of a station that is actually taken into operation, and observed after the station commissioning. Given very comparable main results obtained from the estimator pro-

Table D.9: Triple Differences Estimates

	Real Controls	Hypothetical Controls
Train Station < 2km	0.0280 (0.0242)	0.0472* (0.0218)
No. Observation	295,565	146,371
Adj. R2	0.6913	0.7335

This table shows the results of the Triple-Differences model shown in Equation D.1. The model incorporates fixed effects on the treatment group-municipality level, year-labor market region level, and actual opening-municipality level, as well as their interactions. Standard errors are clustered on station opening-treatment group level (in parentheses).
** p < 0.05*

posed by Sun and Abraham (2021) and regular TWFE (Table D.8, Figures D.7 to D.9), the Triple-Differences model is estimated using a common fixed effects regression. However, comparable results are obtained when utilizing the method outlined by Gardner (2022). Additionally, to address concerns regarding multicollinearity, treatment group-municipality fixed effects are employed instead of station opening-treatment group fixed effects, as done in the Difference-in-Differences framework. Notably, this adjustment yields only marginal changes to the results in the Difference-in-Differences analysis.

The underlying identification assumption posits the violation of the common trend assumption for properties near inaugurated stations, compared to houses located 3km to 5km away, mirrors that between properties near a similar (hypothetical) station and those 3km to 5km distant. The results do not refute the main findings. Specifically, the Triple-Differences estimate using the real control group falls 2.1 percentage points short of the Difference-in-Differences estimate using the donut control group, while utilizing the hypothetical control group in the Triple-Differences approach yields a 0.1 percentage point lower estimate.