The effect of a new train station on housing prices in rural areas: evidence from Dronten

August 29, 2024

COLOFON

Title	The effect of a new train station on housing prices in rural areas: evidence from Dronten
Version	2
Author	Jesse Luimes
Supervisor	Dr. Frans Schilder
Assessor	Prof. dr. ir. Arno J. van der Vlist
E-mail	j.j.luimes@student.rug.nl
Date	August 29, 2024

ABSTRACT

This paper models the effects of a new train station (light rail) on the housing prices in the (rural) municipality of Dronten, Flevoland. I distinguished two different effects: within 3500m proximity to the station and homes within the Dronten Municipality further away than 3500m from the station. I found that for homes within a 3500m perimeter from the station, the effect of a new train station is a significant decrease of 8%. The effect for homes within the municipality of Dronten further away than 3500m from the station is a significant decrease of 13% in the housing prices, although the parallel trend assumption was violated, leading to a biased estimation. The most important cause of this negative coefficient are the negative externalities associated with a train station. Additionally, it was found in a literature review that people use the train for enjoyment and a sense of independence, but demand increases when the service is reliable. The key recommendation is to replicate this study with different, more homogeneous data to ensure a holding parallel trends assumption.

PREFACE

Dear reader,

In front of you lays my master thesis, which will discuss the influence of a new train station in a rural area, with the case of Dronten. While this might not directly sound interesting to you, it is highly relevant if you live in a rural area, since your wealth is directly affected by this. I hope you have a great time reading my final paper which concludes my master's degree in Real Estate Studies at the University of Groningen.

The writing process of this thesis did not go without obstacles, as the data analysis proved to be challenging with the provided data. Furthermore, it was hard to schedule working hours on this thesis from time to time, making the general process of this thesis more demanding than I anticipated beforehand, yet it was a rewarding and educational experience for which I am grateful.

Hereby, I want to take the opportunity to thank my supervisor Frans Schilder, who was of great help and support with an overdose of positivity. I also want to thank prof. dr. ir. Arno J. van der Vlist and dr. Mark van Duijn, who taught me proper econometrics in the real estate research course, through which I was able to conduct such research as well as the data provider NVM, which provided the transaction data of 7 municipalities for the years 2006 until 2018. Other people who had a significant positive influence were Lennard Kruger, for helping me out with some statistical questions, my parents for pushing me through, Berend and Janneke for possibility to stay in the Groningen area this year and Jochem Frehe for being a great mentor by providing valuable life perspectives and advice regarding personality and career. Last but not least, I want to thank my Lord and Savior Jesus Christ, who gave me the power and strength to keep on working on this paper, while being in a hectic period.

"*Come to Me, all you who labor and are heavy laden, and I will give you rest*" ~ Jesus
Matthew 11:28 – The Bible

Best,
Jesse Luimes

TABLE OF CONTENTS

1. INTRODUCTION

1.1. Motivation

About a decade ago, the town of Dronten has opened a train station in 2012. This was done as part of a new train line called the "Hanzelijn". Clearly, this would also have large implication for the town and its surroundings. Past annual monitoring of the new Hanzelijn stations provided by Vor and Buunk (2016) showed an increase over the years of 2013-2015 of 13% of inhabitants of Dronten working in the province of Overijssel, in which Zwolle is located. In total, 17.000 people make use of the Dronten Station every day and many people use the park and ride area to commute. The Hanzelijn stretches from Lelystad to Zwolle, spans about 50 kilometers and accommodates both light (*sprinter*) as well as heavy rail (*intercity*) transport. Connecting Lelystad and the Amsterdam metropolitan area to the eastern and northern part of the Netherlands, but also giving Dronten a brand new train station (RailWiki, 2013), facilitating a light rail train connection. This would make the municipality of Dronten a more attractive place to live, since it would be better connected to the rest of the country, and bigger cities like Amsterdam and Zwolle. Before 2012, Dronten was only connected to Zwolle by car, without any highways. The question remains whether people will choose the train over the car to travel to Zwolle and whether they value the amenity of a train station in a rural area like Dronten. This would then be reflected in the housing prices in Dronten, because if people value the amenity of a train station, they are willing to pay more. In this way, the study attempts to connect the appreciation of those amenities to the housing prices, to uncover potential relationships.

1.2 Academic relevance

Much research has been done on the topic of the effect of new public transport connections on housing prices, but also on land values. One example of such a research is the study of Berawi et al. (2020). This study focusses on the influences of new railway investments on commercial land values in Jakarta. Furthermore, there have been many studies on the positive and negative externalities provided by train stations (affecting the housing prices). There are studies which show the negative externalities outweighing the positive, for example a study done by Xue-Zhen, Bai and Xu (2016) in Beijing and Nelson (1992) with the case of Atlanta. However, Im and Hong (2017) as well as Rojas (2024) show a positive effect of train stations on housing prices regardless of the distance to the train station. One of the few studies on rural areas has been conducted by Bulteau, Feuillet and Boënnec (2018), who found opposing effects for urban and rural areas. A Dutch study on the negative externalities of railroads in the Netherlands is the research Ruijven and Tijm (2021), showing 5% higher price elasticity regarding distance after tunneling a railroad in Delft, exterminating nuisance.

Another Dutch example is provided by a paper written by Koster, Ommeren and Rietveld (2012). This paper sets out a study with of which the effect of the opening of 13 new railway stations in the Netherlands on the housing prices is measured by a repeated sales model. However, no statistically significant adjustment in housing prices was found and none of these 13 new stations were in a rural area. Furthermore, a paper has been written by Debrezion, Pels and Rietveld (2006) of the Tinbergen Institute which explains how the effects of proximity to a railway station and the frequency of this public transport in the whole of the Netherlands positively relate to housing prices. Beside the last-mentioned study on the effects of public transport quality on the housing prices, there have been other studies on this topic as well such as the meta-analysis which has been conducted by Rennert (2022) showing the impact of various transit service elements like frequency, fare and reliability on housing prices.

These research outputs show there is a clear research gap, since most studies focus on urban areas, rather than rural areas. The research gap this thesis tries to fill mainly consists out of this rural part: the effect of a new train station on the housing prices closer to the station as compared to housing prices further away from the station *in a rural area*.

1.3 Research problem statement

The central research aim of this study is "to estimate the effect of new train stations on housing prices in rural areas". This will be done by the following central research question (1):

How do train station openings influence housing prices in a rural area?

This central research question will be addressed using three research questions:

(1) What is the theoretical effect of a new train station on housing prices in rural areas?

This study will conduct a comprehensive literature review to examine the impact of new train stations on housing prices in other regions, as well as the connected consequences and externalities. The findings will reveal the theoretical relationship between the establishment of new train stations and fluctuations in housing prices. Consequently, this literature review will facilitate the development of a conceptual framework that describes the relationship between these two variables, providing detailed information into a structured way.

(2) What is the effect of a new train station in Dronten?

This will be researched by taking a look at the housing prices before and after the opening of the train station in a Dronten, with help of a difference-in-difference model which will be used to decide upon the effects of the shock (i.e. the new train station). This type of regression will produce an estimation, which will be the answer on the question. However, because housing prices might have risen before the train station was opened, this can lead to distortion in the analysis. To take away as much as possible of this distortion, I took multiple years of transaction data. Dronten is an example of a rural municipality as Table 1 shows Dronten having less than 1000 addresses per square kilometer.

Table 1: *Addresses per square kilometer for Dronten* (Centraal Bureau Statistiek [CBS], 2024)

	Name muni.	**Urbanity level municipality**	**Population**	**Adress-density**
Name		Description	Amount	per km2
Dronten	Dronten	Little Urban	44 337	780
Zwolle	Zwolle	Strongly Urban	133 133	2105

One particular matter which will be taken into consideration is the fact that housing prices further from the train station of Dronten, might yield a different effect. Therefore, another quantitative analysis will be done, for observations which are located further away from the train station will be taken as treated group, excluding the observations closer than 3500m.

(3) What are the reasons for this influence of the new train station on the housing prices?

This research question will be answered by a thorough literature review of existing knowledge on this topic. Specifically knowledge on the influence of transportation on housing prices and knowledge on bid rent curves. Out all of this existing knowledge, a structured overview will be given, out of which conclusions will be drawn on what the most important reasons are.

The approach of this thesis contains many new elements which have rarely been combined in the Netherlands. Among the new elements is the rural aspect of this thesis and the differentiation which is made for distance to the station, since most studies focus on urban areas. This set up led to the conceptual model visible in Figure 1 in the appendix.
The remainder of this thesis is organized as follows. Chapter 2 describes the theory, coming from an literature review. Chapter 3 describes the methodology, as well as the data cleaning and processing. Chapter 4 gives an overview of the results, which will be further discussed in chapter 5. Out of this discussion, a conclusion will be drawn.

2. THEORETICAL FRAMEWORK

During this theoretical framework, current theory and existing knowledge regarding the topic will be discussed briefly. This to give a clear overview of what has already been studied in the field. The thesis is based on the urban version of the Alonso model as already mentioned in the introduction chapter, which will be discussed in chapter 2.1. In chapter 2.2, it is discussed which factors do influence housing prices from the current literature. Chapter 2.3 sets out existing literature in the field and in chapter 2.4 it becomes clear why people value public transport.

2.1 Housing prices explained

To explain housing prices, one needs to take a look at housing consumption. Following Rosen (1974), housing products are often differentiated by their attributes (e.g. size, quality or location), which makes them heterogeneous. Rosen argues that in a competitive market, the price of such a housing product is determined by the sum of the values of those attributes, which implies that every attribute (like location) has it's own price.

Secondly, the Alonso-Muth-Mills-model, explaining the tradeoff between housing prices and transportation costs (i.e. deciding upon the price of the location attribute). The model assumes a city, with one single Central Business District (CBD), and everybody who lives in or around the city works in that particular city and commutes to that single CBD. Citing from Chau and Chin (2003), who state the following: "*In the traditional view of location, accessibility is measured in terms of access to the Central Business District (CBD)*". This would imply that there would be a tradeoff between housing prices and transportation costs. The people living in the city center have low transportation costs, since they live close to their job. Therefore, housing prices are very high, because there is high demand for those spots. Housing further located from the city center is cheaper, since the transportation costs are higher; people can't afford an expensive home anymore, because the money would be spent on commuting (Kulish, Richards and Gillitzer, 2011). Obviously, this model is a simplified version of reality and it could be problematic when applying to real life situations, especially in non-American cities, like Zwolle and surroundings. If this would be applied to our situation on Dronten, city centers like Lelystad and potentially Amsterdam are not taken into account, although they might have an influence on the housing markets of Dronten, since they get a better connection to those cities as well. One would run into questions because of which city center housing in Dronten would become more expensive, since it also gets an improved connection to Lelystad.

From this Alonso-Muth-Mills-model, a bid rent curve can be constructed. In Figure 2 and 3, a bid curve is already visible as the diagonal line. Connecting this to the previously discussed model, the lower transportation costs would mean higher housing prices, since there is created more room within the tradeoff. Although trains are generally fairly expensive in the Netherlands for most people (Hartkamp, 2024), it does in fact lower transportation costs. Since there are more transportation options to choose from, there is more choice between (cheaper) alternatives. Assuming that those transportation costs are lower, it would still be unknown whether the increase of housing prices can be assigned to better connectivity to the city center of Zwolle. Additionally, it might be that people living in rural parts of Dronten will not be affected by the train station, since they might have to use a car anyway.

2.2 Factors on housing prices

In this section, influences on housing prices will be discussed, underpinned by research. This is done to ultimately find out which variables can be relevant to function as control variables.

When looking at meta-studies, the variables which are found to be most influential are coming forth from different studies. One of these studies is a meta study by Sirmans et al. (2006), stating: "*This paper provides a meta regression analysis of the nine housing characteristics that are appear most often in hedonic pricing models for single-family housing: square footage, lot size, age, bedrooms, bathrooms, garage, swimming pool, fireplace, and air conditionin*g."

Additionally, a study on hedonic pricing models performed by Chau and Chin (2003) states that homes have locational, structural and neighborhood attributes and differentiates based on type of housing: they classify housing to underpin the heterogeneity of housing products. They show that the market price of the property is the function of those three attributes. Furthermore, it is suggested that people tend to have a higher willingness to pay for properties closer to public transportation (So et al. (1996) cited in Chau and Chin, 2003). A finding of significant importance for this thesis. Another variable which they discuss is view, which can be positive (e.g. nature) or negative (e.g. a factory). The study also mentions examples of structural attributes being: age, lot size, number of rooms (number of bedrooms and bathrooms as well), fireplaces, basements, air-conditioning, and (functional) floor area. Examples of neighborhood attributes are subdivided again into:

1. Socio-economic variables (social class of neighborhood, occupations of inhabitants)
2. Local government or municipal services (schools, hospitals, or churches e.g.)
3. Externalities (like crime, noise, shopping centers).

Other research confirm on what's being stated in this research of Chau and Chin (2003) and every study includes many common factors to account for. Another study which divides effects in locational, structural and neighborhood (or environmental) categories is the study of Ahmed, Ahmed and Kashif (2020). Both studies include factors for all categories. Several studies for example correct for the age of a home, parcel size and the size of the home (like area, but also no. of bedrooms, bathrooms, no. of floors etc.). Examples of such studies are Harding, Knight and Sirmans (2003) and Cohen and Coughlin (2008). Also the study of Muhammad, Burhan and Safian (2024), Nguyen (2020), Nepal et al. (2020) and Lieske et al. (2019) correct for the majority of the earlier mentioned factors.

However, Chau and Chin (2003) do also mention that relatively little research has been done on the structural quality of a home, since it is hard to measure. At last but not least, many studies correct for accessibility factors, such as highway or public transport proximity. This claim is underpinned by the studies already mentioned and the studies of Conway et al. (2008), MacDonald et al. (2010) and Seo, Golub and Kuby (2014).

In Table 2 is visible which variables are often mentioned in literature, underpinned with a selection of studies which use the same method.

Table 2: *Variables underpinned by earlier research*

Control variables	**Source**
Age (created from building year)	(Koster, Ommeren and Rietveld, 2012) (Debrezion, Pels and Rietveld, 2006) (Harding, Rosenthal and Sirmans, 2007) (Buonanno, Montolio and Raya, 2012)
Age^2	(Harding, Rosenthal and Sirmans, 2007) (Francke and Minne, 2016) (Knight and Sirmans, 1996)
Parcel size	(Bulteau, Feuillet and Boënnec, 2018) (Theebe, 2004)
Distance to train station	(Bulteau, Feuillet and Boënnec, 2018) (Debrezion, Pels and Rietveld, 2006)
Dwelling type fixed effect	(Koster, Ommeren and Rietveld, 2012) (Theebe, 2004)
salesyear fixed effect	(Harding, Rosenthal and Sirmans, 2007) (Buonanno, Montolio and Raya, 2012) (Theebe, 2004) (Gibbons and Machin, 2005)
Neighborhood fixed effects	(Redfearn, 2009) (Zabel, 2015) (Yinger, 2015)
Maintenance status	(Harding, Rosenthal and Sirmans, 2007) (Theebe, 2004) (Francke and Minne, 2016) (Knight and Sirmans, 1996)
Other structural variables: type of parking spot, monumental status, swimming pool, garage, basement	(Debrezion, Pels and Rietveld, 2006) (Theebe, 2004) (Pope, 2008) (Sirmans et al., 2006)

Not all control variables which were discussed in the literature section are used in the model of this thesis, like the size variables (because the dependent variable already corrects for size). It can also occur that they are not available, cause econometrical problems or contribute too little to the model (the model has been built parsimoniously). Hence, a selection is made of each group (locational, structural and neighborhood), by adding those as fixed effects. Furthermore, many of the variables mentioned can be captured in e.g. neighborhood fixed effects or time fixed effects, which will control for socio-economic differences between neighborhoods and differences developed over time. In principle, the variables in Table 2 are the starting point for the analysis of this thesis.

2.3 Existing research in the field

This chapter is dedicated to set out existing research in the field, it distinguishes itself from chapter 2.2 as that chapter was a separate literature review with a focus on how current hedonic prices models are set up and which factors could be used in such a model for this study.

One of the more known studies is the study by Koster, Ommeren and Rietveld (2012), already shortly mentioned in the scientific relevance. This study is based on a repeated sales model for 13 different cities which got a new station (yet all cities already had a station in the urban area). With this repeated sales model, in which they looked at homes which were sold at least twice before and after the opening of a station, they did not find any significant result.

When it comes to the influence of train stations in rural areas, considerably less studies have been conducted. One of the few studies is the study by Bulteau, Feuillet and Boënnec (2018), in which they found that the effects of public transportation on housing prices were different for urban and rural areas. In urban areas, they find a positive relation between the housing prices and the distance to the nearest train station, possibly because the negative externalities outweigh the provided service. In rural areas, this was the exact opposite: they found a negative relationship between the distance to the railway station and the housing prices, since accessibility gains a lot when having a station close by. That those negative externalities near stations can potentially be strong is shown by a study from Bowes and Ihlanfeldt (2001). They found that positive externalities (like better accessibility) can be countered by negative externalities. Since train stations provide neighborhood access to criminality. This would depend on the income level of neighborhood residents, distance to the station, whether the station is downtown and if it is downtown. Another finding of them was that the negative externality is magnified when the station has a parking lot. Yet, in middle- and high-income neighborhoods, the saved commuting-costs exceed the costs caused by the negative externalities.

Research by Xue-Zhen, Bai and Xu (2016) shows that within 200 meters of a transfer station, the negative externalities outweigh the positive externalities in the city of Beijing. Nelson (1992) confirms this with a study on Atlanta and shows that close proximity to heavy rail stations may have a negative impact on housing prices in high-income neighborhoods. On places where more low-income inhabitants are dependent on public transport, this is not necessarily the case. That living close to a railway does not necessarily mean higher housing prices is once again shown by Hewitt and Hewitt (2012), Zhou and Zhang (2021), and Chen, Rufolo and Dueker (1998). All these studies show that the effect of distance to station or a railway line is not linear and there could be contrasting results based on distance. However, there are also studies which find positive results regardless of the distance to the railway station, such as Im and Hong (2017) and Rojas (2024). It can be concluded that the size and direction of the effect is dependent on the specific situation.

Further statistics regarding the differences in effects on housing prices are mentioned in the study of Debrezion, Pels and Rietveld (2006). They found that homes close to a station are 25% more expensive than homes located further than 15 kilometers away from a station and that a doubling of the operating frequency leads to an approximately 2.5% increase in property values. When homes were closer located to a railway line (not station), the effects were negative, due to noise impacts. Homes closer than 250 meters to a railway line were 5% cheaper than homes which were more than 500 meters away of a railway line. The claim that homes closer to a railroad are cheaper (ceteris paribus), is underpinned by a vast amount of studies, among which: Diao, Qin and Sing (2015) found that housing prices within 400 meters of the station increased by 3.5% when the station was removed. Also Geng, Bao and Liang (2015) found that the negative externalities decreased when the distance increased, which implies housing prices in close proximity of the station experience the negative effects of those externalities.

That the negative impact on the living quality of railroads can be solved was shown by a study of the CPB (Central Planning Agency) conducted by Ruijven and Tijm (2021), studying the effects of tunneling a railroad in Delft on the housing prices, found that price elasticity was 5% higher with respect to the distance to the railroad than before the tunneling. This was due to noise and other nuisance according to Ruijven and Tijm (2021). This impact of noise on housing prices has also been studied by Rich and Nielsen (2004) who found that for every 1 dB increase in noise, a home would depreciate approximately 0.5% on average, in which houses depreciate more (0.535%) than apartments (0.473%). However, this study is based on traffic noise, and not the noise of trains. Logically the nuisance of trains also affect housing prices negatively as pointed out by Armstrong and Rodríguez (2006) and Kilpatrick et al. (2007) .

The study by Rennert (2022) has already been mentioned in the scientific relevance section and will be further discussed here, providing some valuable insights. It studied the relationship between various transit service elements and housing prices. Rennert (2022) conducted a meta-analysis on 46 studies, which produced 66 different “rail access premiums”, in the regression analysis he regresses his transit service variables on the housing price premiums, which gave an indication of how the premiums are affected by the various factors regarding the service quality of the public transport. Not all of this findings were significant; only the variables “expenditure share” and “available connections” showed significant correlation coefficients. The influence of expenditure share shows that for every 1 percentage point increase in annual expenditure share, the price premiums associated with rail access decrease by 1.4 percentage points. The effect of "available connections" is as follows: for every addition of 10 new stations, the rail access premium increases by 0.21 percentage points. These 10 new stations could be anywhere in the transit system, according to Rennert (2022), even distanced further away.

2.4 Reasons for public transport

To gain a better understanding of *why* nearby train station would influence housing prices, it is good to know why people value and use public transport. Firstly, I mention some of the users of public transport and what the general opinion is on public transport. Then several researches are mentioned to unravel the underlying reasons for using public transport. Steg (2003) describes how “*public transport is often perceived to be a poor alternative for car use*”. Car users tend to be more negative about public transport and the people who do not use a car frequently are less negative about public transport. However, even people who barely use a car, are more positive about cars than public transport. The groups of people which use cars relatively less are women, younger people, low-income groups and single people. Consequently, cars are more often used by men, older age groups, higher income groups, couples and families. The reasons why people do see public transport as an attractive option according to Steg (2003) are: independence, convenience, the fun of public transport, freedom, control, status and traffic safety. The people who do not drive only see the benefits of public transport in a few areas (safety, costs, less varied experiences). In connection to that, there is research showing by Li (2017) that people with less cars would still pay a higher premium for their homes because of railway access.

The fact that it matters how well public transport works has also been shown by Paulley et al. (2006). They find that fare increases, car ownership, quality of service and income all influence the demand for public transport. Of which income and quality of service have a positive impact. Income also has a negative impact, since people can then afford cars, which drives the demand down and fare increases have a negative impact on public transport demand. This claim is underpinned by Van Lierop, Badami and El-Geneidy (2017), who show that satisfaction and loyalty to public transport are positively influenced by various factors among which punctuality, service frequency, cleanliness, courtesy and the behavior of operators. Both Buehler and Pucher (2012) and Bresson et al. (2004) have studied upon the factors influencing demand as well. The findings of Buehler and Pucher (2012) indicate that fare prices, service quality and frequency, regional integration, car use restrictions, and land-use policies promoting mixed-use development contribute

to the success of Germany's transport system, which experiences high demand. Bresson et al. (2004) also found that an income-effect (more income) is in fact a motorization effect, showing that the more income a family has, the less the demand for pubic transport.
Out of this short set out theory, it can be concluded that people apparently value a well working train line and are willing to pay a premium, although most people perceive a car substantially more positively following Steg (2003), Li (2017), Paulley et al. (2006) and Rennert (2022).

2.5 Literature Review Conclusion

Out of the existing literature, it appears that train stations have varying effects on housing prices, depending on the distance to the train station, because this amenity adds value although that value could be outweighed by the negative externalities. One can estimate these effects when using the right controls and it becomes clear that the effect should not be expected to be large. For this study, it means that the effect is going to be estimated for various distances by using the correct controls already used in the literature.

Regarding the case of this thesis, it is theoretically predicted that housing close to the station has a negative effect (e.g. noise and nuisance), but it does not negatively affect housing further than 500m. The positive effects are expected to weigh heavier than the negative effects on the housing prices, when homes are located further away then 500m from the train station. Eventually, after 3500m the positive as well as the negative externalities fade away, hence also the effect of housing prices. Additionally, it is also predicted that housing outside of the town of Dronten in the country, is negatively affected, since they experience the disadvantages of a train line, but do not experience the benefits of a train station closeby.

3. METHODOLOGY

This section contains the data analysis part of this study in order to answer the first two sub questions, on what the specific effects of this new train station opening in Dronten is for various distances to the train station. To study this, a difference-in-difference model has been chosen. Since this type of regression can measure the effect before and after a shock-event, in this case the opening of the a new train station in Dronten. This effect will be estimated for two main models, which distinguish different radiuses of observations around the station. The process of creating those models will be discussed in this chapter, distinguishing in the data cleaning (3.1), model building (3.2) and the testing of the assumptions (3.3).

Although there are papers which talk about negative externalities in close proximity of the station such as Xue-Zhen, Bai and Xu (2016), there are many papers which do not find these negative externalities (rather positive) such as Im and Hong (2017), so I did not consider these negative effects from this moment and onwards. Furthermore, it is not feasible to do so, as there are only 32 observations within 500 meters of the train station, making it impossible to capture the potential negative externalities.

The data provided will be utilized and analyzed in accordance with the highest standards of confidentiality and anonymity. Extensive efforts were made to ensure these standards were met and to prevent any negative consequences for the source of the data, which was done by avoiding the use of any personal data.

3.1 Dataset & cleaning

The dataset is a cross-sectional dataset, with observations of multiple time periods (from 2006 until 2018) and multiple units of measurement (i.e., every housing transaction is regarded as an observation). Repeated cross-sectional data is essentially taking multiple samples, from the same population over time. When combining all these samples, one acquires a dataset of similar characteristics to the one used in this thesis. In Figure 4, a map I created with this transaction data from NVM (the national association of real estate agents) is visible. In this map, *all* observations are visible in accordance to their corresponding x,y-coordinates. The green dots are the observations of the municipalities of Dronten (i.e. the majority of those are in the treated group). The other colors represent observations in the control group.

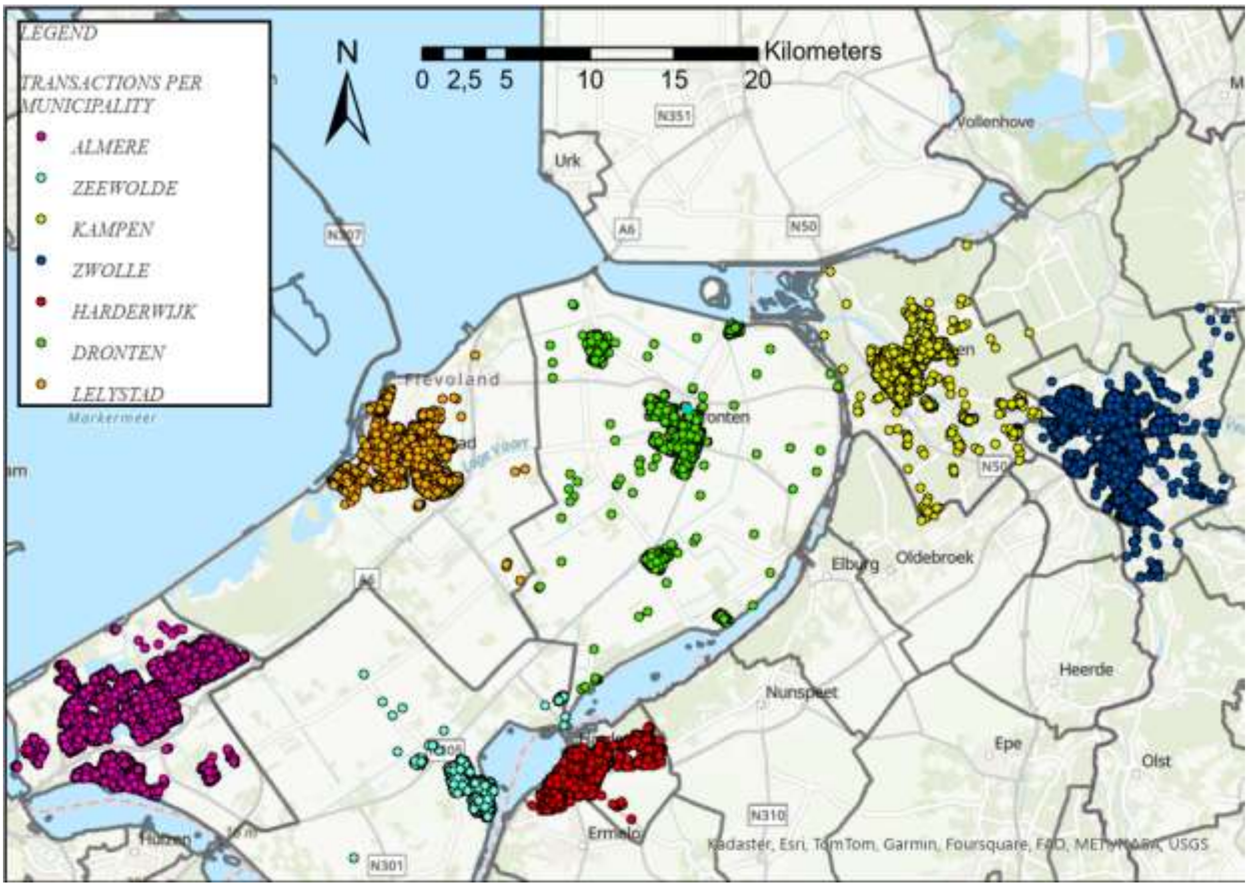

Figure 4: *Map of all transactions, sorted by municipality* (Source: NVM transaction data).

The first observations I deleted, were the transactions within the municipality of Kampen to prevent those transactions from entering the control group. Since the town of Kampen got a new station as well, those transactions are de facto a treated group, yet for the scope of this study, only the new station of Dronten is used. Now, it can almost be assumed that most of the observations within the dataset are relevant observations, only missing values and extreme values (or outliers) should be deleted. Observations which had missing values on critical variables (e.g., transaction year, price, town) were deleted. Furthermore, variables within the upper or lower 1st percentile were deleted as well, to account for extreme values. This ensures that almost all outliers were deleted. Conspicuities were the parcel variable, which was not normally distributed because apartments have no parcel (giving many values of 0). Additionally, the age variable could not be normalized, and the bathroom, kitchen, and living room surface areas had so many missing values that they were deemed unusable.

It is only after this, a first look was taken at the distribution of these variables and it was concluded that it would be better to log-transform the transaction price variable, since it was skewed. This log transformation solved the problem of non-normality in the observations (Figure 7). The distance variable was not transformed, since this made the variable extremely skewed (Figure 9); logically, the missing values were deleted.

3.2 The models

The main characteristic of a difference-in-difference model is the comparison between a treatment group and a control group (WorldBank, 2023) in the presence of a shock event. In the case of this study, the treatment group are the transactions which are located within a 3500m distance of the station which are sold between 2013 and 2018. Since the

Hanzelijn was opened in December 2012, these houses have had the treatment of adding a train station. The 3500m distance was chosen based on Keijer and Rietveld (2000) who show that until 3500m away from the station, the effect the station has on the housing prices stays relatively equal. After 3500m, the effect of a train station gradually declines. The control group is are all other transactions. One fact to take into consideration is that those groups are heterogeneous, since houses which were sold during the control group period are likely not resold during the treatment group period. Hence, the two groups will consists out of observations which are inherently different, creating potential problems such as heteroskedasticity.

Another aspect to take care of are general fluctuations in housing prices (nationwide). This is corrected for by adding time fixed effects in the form of sales years, which would capture differences in the data which were caused by for example inflation. This "salesyear" variable was newly encoded and was used to create an "age" variable and it was used as the difference-in-difference time variable. The following aspect to take into account was that the model needed to correct for unobserved variance between areas, which could not be explained by other variables which were present in the dataset. Hence, I added neighborhood fixed effects. During the regressions, it was found that there were many zip code fixed effects (10098), since many zip code had only one transaction. Therefore, I decided to replace the zip code fixed effects with neighborhood fixed effects (314 dummies), which is a more logical choice.

Furthermore, I encoded a new dependent variable: the transaction-price per square meter, which would already be corrected for the size of a home and was thought to measure a purer effect. This eventually led to the removal of the variables which controlled for housing size (i.e. no. of rooms and square meters), since the parsimonious nature of the model. Additionally, I recoded the maintenance variables (indoor and outdoor maintenance) into 0/1 dummy variables. The original scale for both variables ranged from 1 to 9, with 6 being doubly classified as "unknown" and "reasonable to good," which could lead to the incorrect incentive of rating poorly maintained homes as 6 instead of, for example, 2. This would render the maintenance variables unusable, as the true maintenance status of homes rated 6 would be unknown. Therefore, I recoded the values 7, 8, and 9 as well-maintained, and the remaining values as no maintenance.

This final model as shown in the model-equation has been the result of primarily testing the assumptions and secondarily checking whether the effects of the initial variables (Table 2) contributed enough to the model (by looking at the size of the coefficient and the t-value). When it occurred that a particular variable was responsible for violating an assumption, or did not contribute enough to the model, it was left out (to ensure parsimony). The testing of the assumptions is set out in chapter 3.3.

$$\ln Y_{i,t} = ß_0 + ß_1 * X_{1i,t} + ß_2 * X_{2i,t} + ß_3 * X_{3i} + ß_4 * X_{4i,t} + ß_5 * X_{5i,t} + ß_6 * X_{6i,t} + ß_7 * X_{7i} + Home\ Type\ FE_i + Salesyear\ FE_{i,t} + Neighborhood\ FE_i + V_i + \mu_i$$

The main goal of the model was to estimate $ß_6$, for various distances and to create a graph in which the housing prices are clearly visualized over the periods 2006-2018, in which one would expect housing prices to go up after 2012.
The following variables are used in the model: X_1 and X_2 are respectively the age and squared age the model corrects for. Age^2 was added to account for the non-linear effect of age as done earlier by Francke and Minne (2016). X_3 is the

distance to the Dronten train station in meters for every observation. X_4 is a dummy indicating whether an observations falls within the 3500m perimeter of the station. If so, it gets the value of 1 and is added to the treated group. X_5 is a dummy which gets the value of 1 when the transaction took place in 2013 or later (after 2012). X_6 is the interaction effect of X4 and X5, and the difference-in-difference-coefficient. X_7 is a maintenance dummy of the indoor and the outdoor maintenance status to control for differences in well and less well maintained home. The maintenance variable has been recoded to well maintained and poorly maintained levels. I also corrected the model with various fixed effects being: Type of Home fixed effects which control for the unobserved variance between various home types. I also corrected for the unobserved variance caused by time trends during the years, like buyer's preferences, inflation etc. by adding time fixed effects in the form of sales years. Furthermore, I added area fixed effects in the form of neighborhood fixed effects, which control for unobserved variance between neighborhoods (e.g. income differences between neighborhoods). The symbol V denotes a vector with structural property characteristics fixed effects of parking type, monumental status and the type of basement a home has. μ_i denotes the error term of the model.

In Table 3, the descriptive statistics of the final model are depicted, right before regressions were run.

Table 3: *Descriptive Statistics of Continuous Variables in Initial Model.*

Variable	Obs	Mean	Std. Dev.	Min	Max
lnpricem2	39271	7.464	0.222	6.85	8.067
lnage	39271	2.968	0.686	0.693	4.522
age2	39271	843.939	1268.99	4	8464
lnperceel	39271	5.255	0.553	0	7.112
Distance	39271	26968.6	10855.9	170.409	44389.2

3.3 Assumptions

The initial assumption tests were conducted on a model that included all variables listed in Table 2; this model, however, did not represent the final model, which is depicted in the model equation (assumption violations caused necessary changes to the initial model). The seven assumptions are tested on a regular OLS-model, because Stata does not allow those assumptions to be tested when using a "didregress"-command. The first assumption is that the error terms should have a mean of zero, which is automatically solved by STATA by adding a constant. The second assumption requires the error term to have a constant variance, a matter solved by STATA when using the "didregress"-command. The third assumption is not relevant for this study, as timeseries data is not involved. The fourth assumption required the model to be exogeneous, a characteristic the model is very likely to have, as it is unlikely to be endogenous (see appendix for detailed explanation). The fifth assumption states the error terms should be normally distributed, which is the case in Figure 6. The sixth assumption is that there is no multicollinearity present in the model. This is visible in Table 5. To meet the seventh and last assumption, extensive effort was made to ensure a correctly specified relationship, as explained in detail in the appendix.

Furthermore, the model assumes a normally distribution of observations (Figures 7, 8, 9 and, 10). This was already ensured by transforming the data and removing outliers where necessary of the variables which went into the final model (the variables which are also visible in the regression model equation). Naturally, this was not done for the factor variables (such as type of home), because they were added as fixed effects. After the assumptions were tested and none were violated, a start was made with the regression model, followed by the test of the difference-in-difference assumption. This so-called equal trends assumption is of great importance stating that: *no time-varying differences exist between the treatment and control groups* (WorldBank, 2023).

Difference in Difference Assumption

The assumption that becomes relevant specifically when conducting a difference-in-differences analysis is the equal trends assumption: *the assumption that no time-varying differences exist between the treatment and control groups* (WorldBank, 2023). This assumption was verified by creating a visual plot, with the left graph being the means of the outcome (log transformed transaction prices) of both groups put next to each other and the right graph being the predicted values of these means after the treatment. This is visible in Figure 11. Coincidentally, the ptrends-test[1] shows a different result. The H0-hypothesis was rejected, stating that there are no parallel trends. Generally, issues like these are uncommon in economic research such as these. Additionally, from a visual check, there might not be a violation of this assumption, as those tests are subject to many discussions (Rambachan and Roth, 2023) To solve this issue, I tried combinations with various sets of variables of the model for which I tested the assumption. I ended up excluding the swimming pool as fixed effect. Furthermore, I changed the garage fixed effect to a parking type fixed effect as control, the closest variable in terms of definitions, which performed better and calculated the discussed result, yet the effect is still not there as visible in Figure 11. Another noteworthy point is the fact that there is no indication of an anticipation effect, hence, the effect of a train station is not visible before the treatment, according to a conducted granger's test[2].

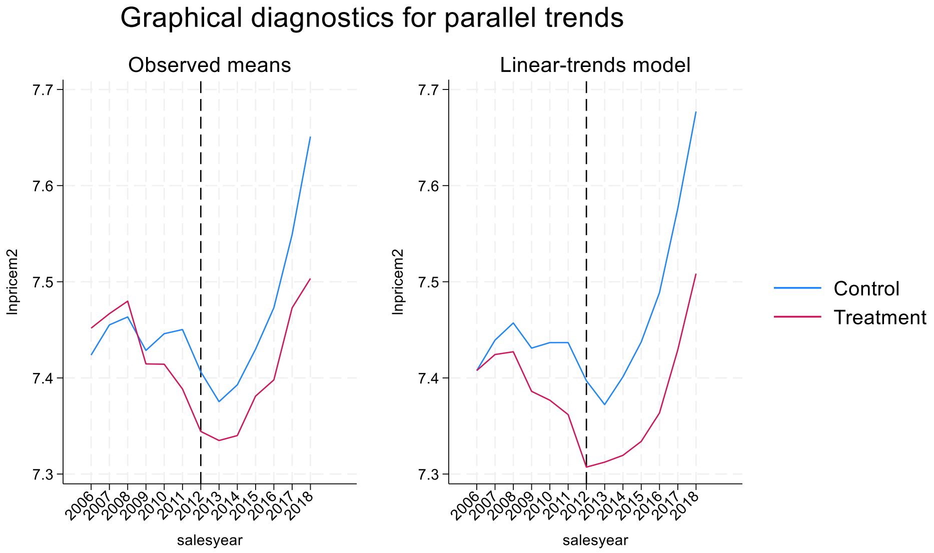

Figure 11: *Visual representation of the first model*.

The second base model was the opposite of the first base model: all observations *closer* than 3500m to the train station were deleted. This would then give only the effect on the housing prices which are situated further away from the station and are mostly out of town in Dronten. Moreover, Keijer and Rietveld (2000) state that until 3500m, the effect

[1] P-value = 0.0055
[2] P-value = 0.0754

remains relatively stable. Hence, the goal of the second model is to account for the effect after this 3500m mark and see what the effect of the station is on the housing prices further away from this station. The parallel trend assumption was violated for this model[3] and I did not find a way to solve this. When performing a granger's causality test on this second model, the result was a 0.0282, meaning there was an anticipation effect in the pre-treatment period.
The violation of the parallel trends assumption eventually leads to a biased estimator, meaning that the estimated coefficient might not be the true value (Columbia University Mailman School of Public Health, 2013). Yet the trends of this estimation look relatively similar (Figure 12). Based on visual checks, the results of these models are used.

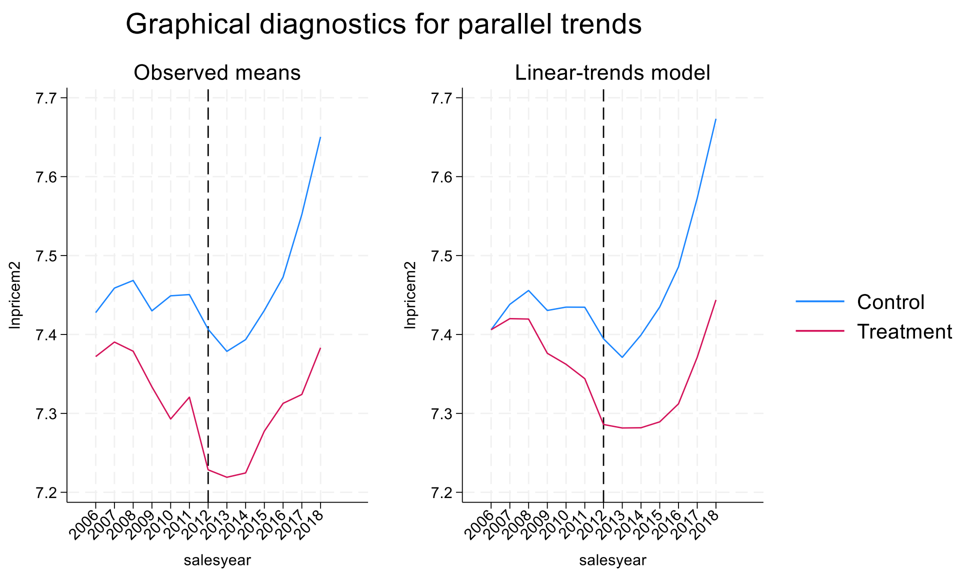

Figure 12: *Violated parallel trends assumption of the second base model visualized*.

[3] P-value = 0.0102

4. RESULTS

4.1 Regression Results

In this section, the two main results will be set out. The first regression was run with a treated group with observations which were all within a 3500m perimeter of the station. The second regression was run with a treated group in which all observations were further away than 3500m from the station, but still in the municipality of Dronten. The first part of the table, the treatment and time information, is the same for every analysis and robustness check, it will therefore be shown only once at the first analysis.

At the top, it states that "salesyear" is the time variable and "drontenafter2012" is the treated group. In the second part of the table, the Average Treatment Effect on the Treated (ATET) is found with its corresponding treated group, coefficients, standard error and significance parameters.

Table 6: *Base Difference-in-Difference Model for transactions within 3500 m proximity of the Dronten station.*

Treatment and time information
Time variable: salesyear
Control: drontenaft~2012 = 0
Treatment: drontenaft~2012 = 1

	Control	Treatment
Group		
DUMMYDISTANCE	1	1
Time		
Minimum	2006	2013
Maximum	2006	2013

Difference-in-differences regression

Data type: Repeated cross-sectional

(Std. err. adjusted for 2 clusters in DUMMYDISTANCE) Robust Number of obs = 39,271

lnpricem2	Coefficient	std. err.	t	P>t	[95% conf. interval]	
ATET drontenafter2012						
(1 vs 0)	-0.087	5.35E-06	-1.60E+04	0	-0.086857	-0.086721

Note: ATET estimate adjusted for covariates, group effects, and time effects.

The estimated effect is $e^{-0.087} \approx 0.92$, which is significant with a t-value of -1.6e4 (Table 6). This means that housing prices within a ring of 3500m around the station decreased by approximately 8% after 2012 when the station was opened. To put this percentage into context: given the current average transaction price of €415,136 in the town of Dronten (WalterLiving, 2024a), an 8% decrease would result in a reduction of €33,210.88 in housing prices.

Table 7: *Second Base Difference-in-Difference model for transactions outside of the 3500 m perimeter of the Dronten station.*

(Std. err. adjusted for 2 clusters in DUMMYDISTANCE)		Robust			Number of obs = 39,271	
lnpricem2	Coefficient	std. err.	t	P>t	[95% conf. interval]	
ATET after2012after3500						
(1 vs 0)	-0.135	0.000651	-207.26	0.003	-0.143095	-0.126563

Note: ATET estimate adjusted for covariates, group effects, and time effects.

The estimated effect is $e^{-0.135} \approx 0.87$, and is significant with a t-value of -207.26 (Table 7). This means that housing prices further away than 3500m (outside the town of Dronten) from the station went down by approximately 13% after the station was opened in 2012.

The current average transaction price for the *municipality* of Dronten is €43937.96 (WalterLiving, 2024b). In this case, the decrease in housing prices would be €399,436*0.13= €51926.68, an even larger decrease, although this could be distorted, since homes in the countryside are usually more expensive, due to bigger parcels.

4.2 Robustness check

A robustness check is executed when one wants to see how coefficients of the core regression (i.e. the first base model) behave when the model is modified (Lu and White, 2014). During this robustness check, I modify the models in three different ways by changing the observations in the treated group. Particularly the last model is interesting, since it provides insight into the anticipation effect, which was present according to the various granger's causality tests.

The first robustness check involves creating a smaller radius around the station for observations in the control group. Instead of 3500m, a 2500m model is used, and it becomes visible that the coefficient changed by 0.001, which is negligible. The untransformed coefficient is $e^{-0.086} \approx 0.92$ as comes out of Table 8, which is the same if one rounds off. Hence, this model is robust to changes in distances. It indeed appears the effect is relatively stable until 3500 m as shown earlier by Keijer and Rietveld (2000), hence, this robustness check is in line with the expectations based on the literature.

Table 8: *First Robustness Check*

(Std. err. adjusted for 2 clusters in DUMMYDISTANCE)		Robust			Number of obs = 39,271	
lnpricem2	Coefficient	std. err.	t	P>t	[95% conf. interval]	
ATET after2012and2500						
(1 vs 0)	-0.086	0.000314	-219.30	0.003	-0.09072	-0.080784

Note: ATET estimate adjusted for covariates, group effects, and time effects.

The second robustness check involved a model which used an even smaller radius, namely 1500m as treated group. The untransformed coefficient is $e^{-0.0857519} \approx 0.92$ (Table 9). This makes the effect of the train station on the housing prices -8% as well.

Table 9: *Second Robustness Check*

(Std. err. adjusted for 2 clusters in DUMMYDISTANCE)		Robust			Number of obs = 39,271	
lnpricem2	Coefficient	std. err.	t	P>t	[95% conf. interval]	
ATET after2012and1500						
(1 vs 0)	-0.089	0.0001106	-803.87	0.001	-0.0903504	-0.0875386

Note: ATET estimate adjusted for covariates, group effects, and time effects.

The third check was based on a change in the year of the shock event. I put the shock event on the year 2010 instead of the usual 2012, with the results being in Table 10.

Table 10: *Third Robustness Check.*

Treatment and time information

Time variable: salesyear

Control: robustdronten~2010 = 0

Treatment: robustdronten~2010 = 1

	Control	Treatment
Group		
DUMMYDISTANCE	1	1
Time		
Minimum	2006	2011
Maximum	2006	2011

Difference-in-differences regression

Data type: Repeated cross-sectional

(Std. err. adjusted for 2 clusters in DUMMYDISTANCE)		Robust			Number of obs = 39,271	
lnpricem2	Coefficient	std. err.	t	P>t	[95% conf. interval]	
ATET robustdronten2010						
(1 vs 0)	-0.093	0.000034ss	-2743.31	0.003	-0.0937947	-0.0929299

Note: ATET estimate adjusted for covariates, group effects, and time effects.

The observed effect size is -0.093, translating to an untransformed coefficient of approximately e^-.093 ≈ 0.91. This indicates a decrease of 9% in housing prices. However, it remains contentious whether this decline is attributable to the volatile years following the great financial crisis or to the presence of the train station. Figure 13 illustrates an upward trend in the graph post-shock event, suggesting a positive influence of the shock event despite the negative coefficient. Additionally, the p-value of Granger's causality test is now 0.0754, indicating no anticipation effect where the impact was observable earlier.

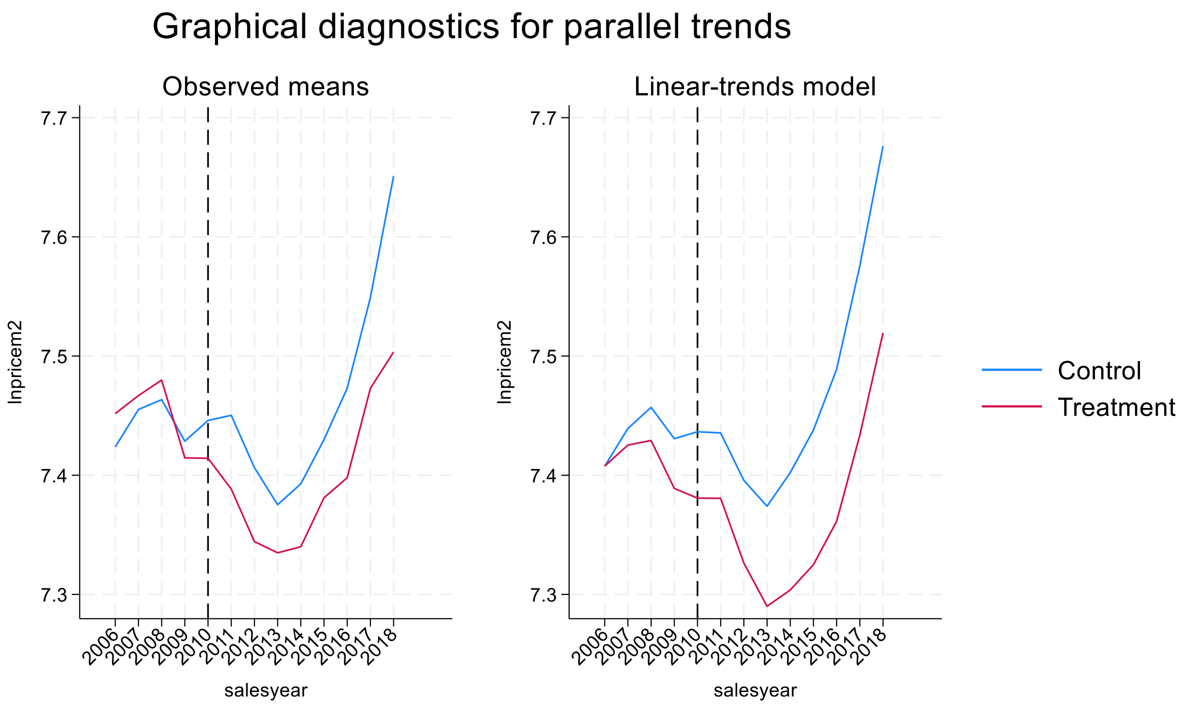

Figure 13: *Parallel Trends for moved shock event to 2010*.

After conducting several robustness checks, it can be concluded that the model is robust, with minimal changes in the coefficients and the impact on housing prices, even when the model is significantly modified. The fact that the coefficients exhibited only slight changes in the first robustness check supports the assertion made by Keijer and Rietveld (2000), who indicated that the effect would remain relatively stable up to a radius of 3,500.

4.3 Discussion of results

In this study, I estimate the effects of the Hanzelijn on the housing prices in Dronten and check for potential significant differences in distance to the Dronten station. Within a radius of 3500m from the station, there is a significant negative effect of 8% (answer first subquestion). This contradicts the theoretical prediction as I predicted a positive effect of a train station opening on the housing prices. Whether it contradicts literature remains the question, as there are studies pointing in both directions.

Although this thesis did not distinguish between distances to the station within this 3500m perimeter selection; due to the low amount of transactions within 500m of the train station (32 transactions), it is hard to believe the negative externalities outweigh the positive externalities for the whole group of transactions within a 3500m proximity, as the research of Bulteau, Feuillet and Boënnec (2018) and Bowes and Ihlanfeldt (2001) suggest (i.e. negative externalities can outweight the positive externalities). A noteworthy point of discussion is illustrated in Figure 11, where housing prices for the treated group exhibit an upward trend beginning in 2012, falling together with the construction year of the station. This trend may suggest that these are indeed the effects I intended to measure. This claim is strengthened by the performed granger's tests, which did not indicate an anticipation effect, implying that the impact of the new train station on the housing prices was not observable prior to the construction of the new train station.

The second model I estimate is a model containing transactions which are located further away than 3500 meters, which yielded a negative effect of 13%. The direction of the effect is not directly in line with the current literature, as this study contradicts the earlier begotten results of Bulteau, Feuillet and Boënnec (2018), who state the effect of train stations on urban and rural housing prices are opposite. In this case, the effects are both negative, albeit the effect of the rural model significantly more negative. From the read literature, it was predicted that observations further away get a less positive or more negative effect ("worse" effect for the housing prices as compared to the housing prices within the town), which turned out to be true. This can be linked to a scenario in which people do not care about a station in Dronten, especially not further away from the station in the rural area since those people already need to take the car, this links up with Steg (2003), who mentions the already negative opinion of car users on public transport. Following Keijer and Rietveld (2000), who state that the propensity to take the train decreases when one lives away further then 3500m from the train station. Public transport would then only lead to a detour, which might be more expensive and take longer, so it does not offer a great leap in accessibility in this regard. When looking at Figure 4, many of the people living at the places of the rural transactions of the municipality of Dronten (coloured green) would indeed need to make a detour to the bigger cities if one goes by train. Furthermore, transactions further distanced from the train station might live close to the train line but not to the train station, experiencing mainly the negative externalities, producing a negative effect on their housing prices. Hence, they have the negative externalities (landscape and noise), but they do not experience the positive externalities (better accessibility). These positive externalities are exactly why people still value public transport, and why it would positively influence the housing prices (if those positive externalities are actually there). Steg (2003) found this is the case because it is fun and can be cheaper than the car and one is more independent. Yet, the public transport connection needs to work well, fares, connections and safety/cleanliness having significant impact on the demand of public transport (Paulley et al., 2006; Van Lierop, Badami and El-Geneidy, 2017) and also on the housing prices (Rennert, 2022), which answers the third sub question.

It could be that these -8% and -13% are not the true values because the parallel trends assumptions were violated, causing biased results (Columbia University Mailman School of Public Health, 2013). Potential causes for this are that observations further away than 3500 m in the municipality of Dronten are mainly rural, while the control groups are mainly urban, making them heterogenetic. During the roaring times around the financial crisis, which happened to be the pre-treatment period, this might have caused further differences between the urban control group and rural treated group in their housing prices. Moreover, these (urban/rural) observations inherently differ from each other (the prices are build up differently, rural homes have a bigger parcel size on average, distorting the average transaction price/m^2 as well). In general, it is assumed that the control and treatment group are too heterogeneous, since the majority of transactions in the control group are urban and most transactions in the treated group are in a rural area.

5. CONCLUSION

Concludingly, it can be stated that the effect of the new train station on the housing prices in a rural area like the municipality of Dronten is negative. The size of the effect is influenced by the distance to the station. This could be due to the negative externalities close by, and the fact that one does experience the negative effects of a railway, but not the positive externalities of the train station when living further away, since those people are forced to use a car. Therefore, the effect further away from the train station is more negative than the effect closer by the station (within town).

Clearly, the results of this study went against some of the predictions, which also shines through in the policy implications following this study. Policy makers should think more on whether a train station adds value in a particular area. It appears that a train station does not add value in all areas, and certainly not the areas which were not designed for public transport (the trainline was much later than the actual polder). Policy makers should make better arguments why a train station should be built on certain spots, when the effects tend to be negative instead of positive, as well as the fact that everybody is used to using a car, which might be caused by the fact that the car is perceived more positively than the train, which also does not help building the habit of using public transport in general.

This research contains several limitations. The biggest limitation is the violated parallel trends assumption, providing biased estimates. This also made it hard to create variations on the model, which held the assumption. Furthermore, the chosen time period, which now happened to include the global financial crise of 2008 might not be suitable, as this may have caused further differences between the control and treated group. This is also my main recommendation, to continue studying the effect of train lines on rural housing prices, ideally on a different case than Dronten with a different dataset to see whether there is a clearer difference between rural homes and homes within the town itself. One of the keys in successfully replicating will be data which is homogeneous enough to do so.

REFERENCES

Ahmed, S., Ahmed, F. and Kashif, M. (2020) 'Price Determinants of Residential Properties by using Hedonic Price Model,' *ResearchGate* [Preprint]. https://www.researchgate.net/publication/342751116_Price_Determinants_of_Residential_Properties_by_using_Hedonic_Price_Model.

Armstrong, R.J. and Rodríguez, D.A. (2006) 'An Evaluation of the Accessibility Benefits of Commuter Rail in Eastern Massachusetts using Spatial Hedonic Price Functions,' Transportation, 33(1), pp. 21–43. https://doi.org/10.1007/s11116-005-0949-x.

Berawi, M. A., Miraj, P., Saroji, G., & Sari, M. (2020). Impact of rail transit station proximity to commercial property prices: utilizing big data in urban real estate. Journal of Big Data, 7(1). https://doi.org/10.1186/s40537-020-00348-z

Bowes, D.R. and Ihlanfeldt, K.R. (2001) 'Identifying the impacts of rail transit stations on residential property values,' Journal of Urban Economics, 50(1), pp. 1–25. https://doi.org/10.1006/juec.2001.2214.

Buonanno, P., Montolio, D. and Raya, J.M. (2012) 'Housing prices and crime perception,' Empirical Economics, 45(1), pp. 305–321. https://doi.org/10.1007/s00181-012-0624-y.

Bulteau, J., Feuillet, T. and Boennec, R.L. (2018) 'Spatial heterogeneity of sustainable transportation offer values: A Comparative Analysis of Nantes Urban and Periurban/Rural Areas (France),' Urban Science, 2(1), p. 14. https://doi.org/10.3390/urbansci2010014.

Buehler, R. and Pucher, J. (2012) 'Demand for public transport in Germany and the USA: An analysis of rider characteristics,' Transport Reviews, 32(5), pp. 541–567. https://doi.org/10.1080/01441647.2012.707695.

Bresson, G., Dargay, J., Madre, J. and Pirotte, A. (2004) 'Economic and structural determinants of the demand for public transport: an analysis on a panel of French urban areas using shrinkage estimators,' Transportation Research. Part a, Policy and Practice, 38(4), pp. 269–285. https://doi.org/10.1016/j.tra.2003.11.002.

Centraal Bureau Statistiek [CBS] (2010) Rural areas. https://www.cbs.nl/en-gb/news/2010/23/narrowing-gap-between-number-of-facilities-available-in-urban-and-rural-areas/rural-areas (Accessed: May 1, 2024).

Centraal Bureau Statistiek [CBS] (2024) StatLine - Gebieden in Nederland 2024. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/85755NED/table?ts=1715057490851.

Chau, K.W. and Chin, T.L. (2002) A critical review of literature on the Hedonic price model. https://ssrn.com/abstract=2073594.

Chen, H., Rufolo, A.M. and Dueker, K.J. (1998) Measuring the Impact of Light Rail Systems on Single Family Home Values: A Hedonic Approach with GIS https://www.researchgate.net/publication/253389026_Measuring_the_Impact_of_Light_Rail_Systems_on_Single_Family_Home_Values_A_Hedonic_Approach_with_GIS_Application Accessed: May 4, 2024).

Cohen, J.P. and Coughlin, C.C. (2008) 'SPATIAL HEDONIC MODELS OF AIRPORT NOISE, PROXIMITY, AND HOUSING PRICES*,' Journal of Regional Science, 48(5), pp. 859–878. https://doi.org/10.1111/j.1467-9787.2008.00569.x.

Columbia University Mailman School of Public Health (2013) Difference-in-Difference estimation | Columbia Public Health. https://www.publichealth.columbia.edu/research/population-health-methods/difference-difference-estimation (Accessed: June 27, 2024).

Conway, D., Li, C.Q., Wolch, J., Kahle, C. and Jerrett, M. (2008) 'A spatial autocorrelation approach for examining the effects of urban greenspace on residential property values,' Journal of Real Estate Finance and Economics, 41(2), pp. 150–169. https://doi.org/10.1007/s11146-008-9159-6.

Debrezion, G., Pels, E. and Rietveld, P. (2006) The impact of rail transport on real estate prices: An Empirical analysis of the Dutch housing market, Social Science Research Network [Preprint]. https://doi.org/10.2139/ssrn.895270.

Diao, M., Qin, Y. and Sing, T.F. (2015) 'Negative Externalities of Rail Noise and Housing Values: Evidence from the Cessation of Railway Operations in Singapore,' Real Estate Economics, 44(4), pp. 878–917. https://doi.org/10.1111/1540-6229.12123.

Duijn, M. van (2024) Real Estate Research Week 2 Lecture 3 [Brightspace].

Fingleton, B. (no date) Endogeneity, Exogeneity and instrumental variables. https://www.cantab.net/users/bf100/pdf/iv_slides_fingleton.pdf.

Francke, M.K. and Minne, A.M. van de (2016) 'Land, structure and depreciation,' Real Estate Economics, 45(2), pp. 415–451. https://doi.org/10.1111/1540-6229.12146.

Geng, B., Bao, H. and Liang, Y. (2015) 'A study of the effect of a high-speed rail station on spatial variations in housing price based on the hedonic model,' Habitat International, 49, pp. 333–339. https://doi.org/10.1016/j.habitatint.2015.06.005.

Gibbons, S. and Machin, S. (2005) 'Valuing rail access using transport innovations,' Journal of Urban Economics, 57(1), pp. 148–169. https://doi.org/10.1016/j.jue.2004.10.002.

Harding, J.P., Knight, J.R. and Sirmans, C.F. (2003) 'Estimating Bargaining Effects in Hedonic Models: Evidence from the Housing Market,' Real Estate Economics, 31(4), pp. 601–622. https://doi.org/10.1046/j.1080-8620.2003.00078.x.

Harding, J.P., Rosenthal, S.S. and Sirmans, C.F. (2007) 'Depreciation of housing capital, maintenance, and house price inflation: Estimates from a repeat sales model,' Journal of Urban Economics, 61(2), pp. 193–217. https://doi.org/10.1016/j.jue.2006.07.007.

Hartkamp, R. (2024) Treinkaartje dreigt nóg duurder te worden: ‘Maak komend jaar 120 miljoen vrij’ https://wnl.tv/2024/03/07/treinkaartje-dreigt-nog-duurder-te-worden-maak-komend-jaar-120-miljoen-vrij/.

Hewitt, C. and Hewitt, W. (2012) 'The effect of proximity to urban rail on housing prices in Ottawa,' Journal of Public Transportation, 15(4), pp. 43–65. https://doi.org/10.5038/2375-0901.15.4.3.

Im, J. and Hong, S.H. (2017) 'Impact of a new subway line on housing values in Daegu, Korea: Distance from existing lines,' Urban Studies, 55(15), pp. 3318–3335. https://doi.org/10.1177/0042098017741412.

Keijer, M.J.N. and Rietveld, P. (2000) 'How do people get to the railway station? The Dutch experience,' Transportation Planning and Technology/Transportation, Planning and Technology, 23(3), pp. 215–235. https://doi.org/10.1080/03081060008717650.

Kilpatrick, J., Throupe, R., Carruthers, J. and Krause, A. (2007) 'The impact of transit corridors on residential property values,' Journal of Real Estate Research, 29(3), pp. 303–320. https://doi.org/10.1080/10835547.2007.12091201.

Knight, J.R. and Sirmans, C.F. (1996) 'Depreciation, maintenance, and housing prices,' Journal of Housing Economics, 5(4), pp. 369–389. https://doi.org/10.1006/jhec.1996.0019.

Koster, H., Ommeren, J. van and Rietveld, P. (2012) 'The Gains of Trains: The effect of station openings on house prices,' Tinbergen Institute Discussion Papers 10-094/3 https://papers.tinbergen.nl/10094.pdf (Accessed: February 2, 2024).

Kulish, M., Richards, A. and Gillitzer, C'. (2011) 2. The Alonso-Muth-Mills Model. https://www.rba.gov.au/publications/rdp/2011/2011-03/alonso-muth-mills-model.html.

Li, T. (2017) 'The Value of Access to Rail Transit in a Congested City: Evidence from Housing Prices in Beijing,' Real Estate Economics, 48(2), pp. 556–598. https://doi.org/10.1111/1540-6229.12222.

Van Lierop, D., Badami, M.G. and El-Geneidy, A. (2017) 'What influences satisfaction and loyalty in public transport? A review of the literature,' Transport Reviews, 38(1), pp. 52–72. https://doi.org/10.1080/01441647.2017.1298683.

Lieske, S.N. et al. (2019) 'A novel hedonic price modelling approach for estimating the impact of transportation infrastructure on property prices,' Urban Studies, 58(1), pp. 182–202. https://doi.org/10.1177/0042098019879382.

Lu, X. and White, H. (2014) 'Robustness checks and robustness tests in applied economics,' Journal of Econometrics, 178, pp. 194–206. https://doi.org/10.1016/j.jeconom.2013.08.016.

MacDonald, D.H.,. Crossman, N.D., Mahmoudi, P., Taylor, L.O., Summers, D.M. and Boxall, P.C. (2010) 'The value of public and private green spaces under water restrictions,' Landscape and Urban Planning, 95(4), pp. 192–200. https://doi.org/10.1016/j.landurbplan.2010.01.003.

Muhammad, A.A., Burhan, B. and Safian, E.E.B.M. (2024) 'Price model for Transit-Oriented developments in Kuala Lumpur, Malaysia,' Real Estate Management and Valuation, 32(0). https://doi.org/10.2478/remav-2024-0022.

Nelson, A.C. (1992) 'EFFECTS OF ELEVATED HEAVY-RAIL TRANSIT STATIONS ON HOUSE PRICES WITH RESPECT TO NEIGHBORHOOD INCOME,' Transportation Research Board, 1359, pp. 127–132. https://trid.trb.org/View/371633 (Accessed: May 4, 2024).

Nepal, M., Rai, R.K., Khadayat, M.S. and Somanathan, E. (2020) 'Value of cleaner neighborhoods: Application of hedonic price model in low income context,' World Development, 131, p. 104965. https://doi.org/10.1016/j.worlddev.2020.104965.

Nguyen, M.-L.T. (2020) 'The hedonic pricing model applied to the housing market,' International Journal of Economics & Business Administration, VIII(Issue 3), pp. 416–428. https://doi.org/10.35808/ijeba/526.

Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., Shires, J. and White, P. (2006) 'The demand for public transport: The effects of fares, quality of service, income and car ownership,' Transport Policy, 13(4), pp. 295–306. https://doi.org/10.1016/j.tranpol.2005.12.004.

Pope, J.C. (2008) 'Do seller disclosures affect property values? Buyer information and the Hedonic model,' Land Economics, 84(4), pp. 551–572. https://doi.org/10.3368/le.84.4.551.

RailWiki. (2013). *Hanzelijn - Somda RailWiki*. https://www.railwiki.nl. Retrieved February 2, 2024, from https://www.railwiki.nl/index.php/Hanzelijn

Rambachan, A. and Roth, J. (2023) 'A More Credible Approach to Parallel Trends,' *The Review of Economic Studies*, 90(5), pp. 2555–2591. https://doi.org/10.1093/restud/rdad018.

Redfearn, C.L. (2009) 'How informative are average effects? Hedonic regression and amenity capitalization in complex urban housing markets,' Regional Science and Urban Economics, 39(3), pp. 297–306. https://doi.org/10.1016/j.regsciurbeco.2008.11.001.

Rehal, V. (2023) 'Simultaneous Equation Bias,' SPUR ECONOMICS - Learn and Excel, 11 January. https://spureconomics.com/simultaneous-equation-bias/.

Rennert, L. (2022). A meta-analysis of the impact of rail stations on property values: Applying a transit planning lens. *Transportation Research Part A: Policy and Practice*, *163*, 165–180. https://doi.org/10.1016/j.tra.2022.06.013

Rich, J. and Nielsen, O.A. (2004) Assessment of traffic noise impacts,' International Journal of Environmental Studies, 61(1), pp. 19–29. https://doi.org/10.1080/0020723032000113790.

Rojas, A. (2024) 'Train stations' impact on housing prices: Direct and indirect effects,' Transportation Research. Part A, Policy and Practice, 181, p. 103979. https://doi.org/10.1016/j.tra.2024.103979.

Rosen, S., 1974. Hedonic prices and implicit markets: Product differentiation in pure competition. Journal of Political Economy, 82(1), pp.34-55. https://www-jstor-org.proxy-ub.rug.nl/stable/1830899?seq=1.

Ruijven, K. van and Tijm, J. (2021) Housing Market Effects of a Railroad Tunneling: Evidence from a quasi-experiment. https://www.cpb.nl/en/housing-market-effects-of-a-railroad-tunneling-evidence-from-a-quasi-experiment.

Seo, K., Golub, A. and Kuby, M. (2014) 'Combined impacts of highways and light rail transit on residential property values: a spatial hedonic price model for Phoenix, Arizona,' Journal of Transport Geography, 41, pp. 53–62. https://doi.org/10.1016/j.jtrangeo.2014.08.003.

Sirmans, G.S., MacDonald, L., Macpherson, D.A. and Zietz, E.M. (2006) 'The Value of Housing Characteristics: A Meta Analysis,' Journal of Real Estate Finance and Economics, 33(3), pp. 215–240. https://doi.org/10.1007/s11146-006-9983-5.

Steg, L. (2003) 'Can public transport compete with the private car?,' *IATSS Research*, 27(2), pp. 27–35. https://doi.org/10.1016/s0386-1112(14)60141-2.

Svetunkov, I. (2023) 15.1 Model is correctly specified. https://openforecast.org/sba/assumptionsCorrectModel.html.

Theebe, M.A.J. (2004) 'Planes, trains, and automobiles: The impact of traffic noise on house prices,' Journal of Real Estate Finance and Economics, 28(2/3), pp. 209–234. https://doi.org/10.1023/b:real.0000011154.92682.4b.

Vor, F. de and Buunk, W. (2016) *Hanzelijn Monitor 2015: De derde staat van de Hanzelijn*. ISBN: 978-90-77901-77-9. HBO Kennisbank. https://hbo-kennisbank.nl/details/sharekit_windesheim:oai:surfsharekit.nl:cbf092f9-5c24-4513-8a22-c062ecfdd3df (Accessed: August 25, 2024).

WalterLiving (2024a) Housing market in Dronten (June 2024). https://walterliving.com/city/dronten (Accessed: June 27, 2024).

WalterLiving (2024b) Housing market in municipality Dronten (June 2024). https://walterliving.com/municipality/dronten/flevoland (Accessed: June 27, 2024).

WorldBank (2023) Difference-in-Differences - Dimewiki. https://dimewiki.worldbank.org/Difference-in-Differences.

Xue-Zhen, D., Bai, X. and Xu, M. (2016) 'The influence of Beijing rail transfer stations on surrounding housing prices,' Habitat International, 55, pp. 79–88. https://doi.org/10.1016/j.habitatint.2016.02.008.

Yinger, J. (2015) 'Hedonic markets and sorting equilibria: Bid-function envelopes for public services and neighborhood amenities,' Journal of Urban Economics, 86, pp. 9–25. https://doi.org/10.1016/j.jue.2014.12.001.

Zabel, J. (2015) 'The hedonic model and the housing cycle,' Regional Science and Urban Economics, 54, pp. 74–86. https://doi.org/10.1016/j.regsciurbeco.2015.07.005.

Zhou, Z. and Zhang, A. (2021) 'High-speed rail and industrial developments: Evidence from house prices and city-level GDP in China,' Transportation Research. Part A, Policy and Practice, 149, pp. 98–113. https://doi.org/10.1016/j.tra.2021.05.001.

APPENDIX

Figures

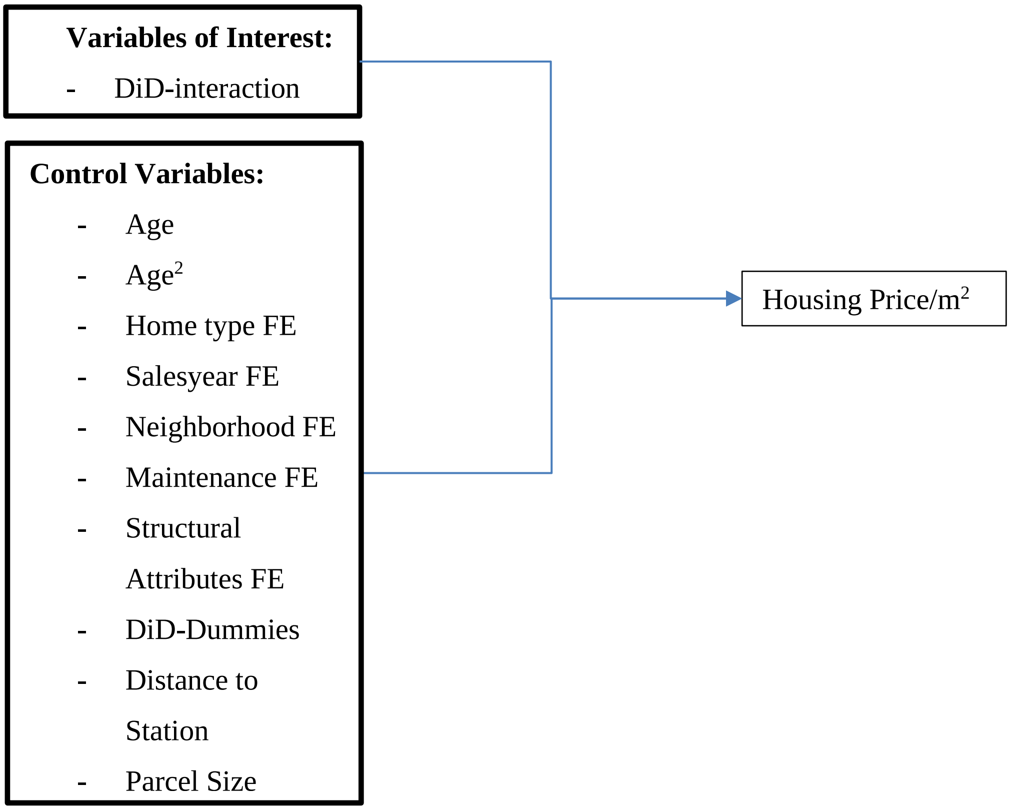

Figure 1: *Conceptual Model.*

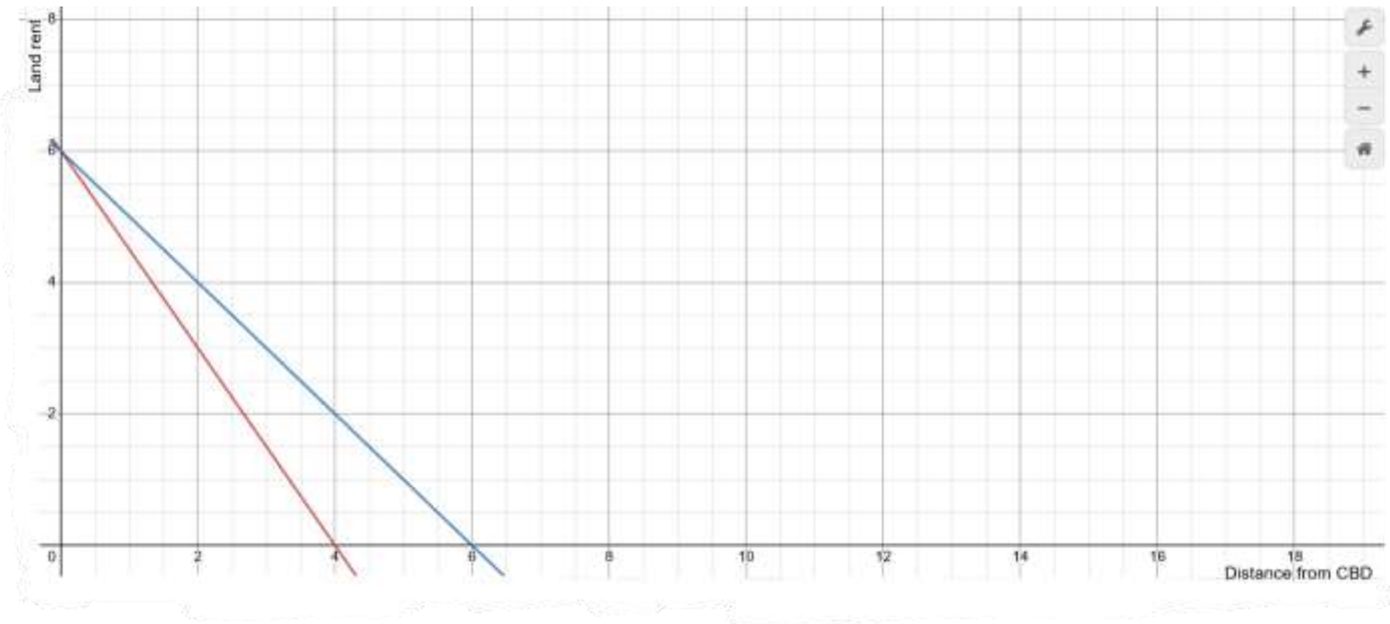

Figure 2: *Outward shift in bid rent curve.*

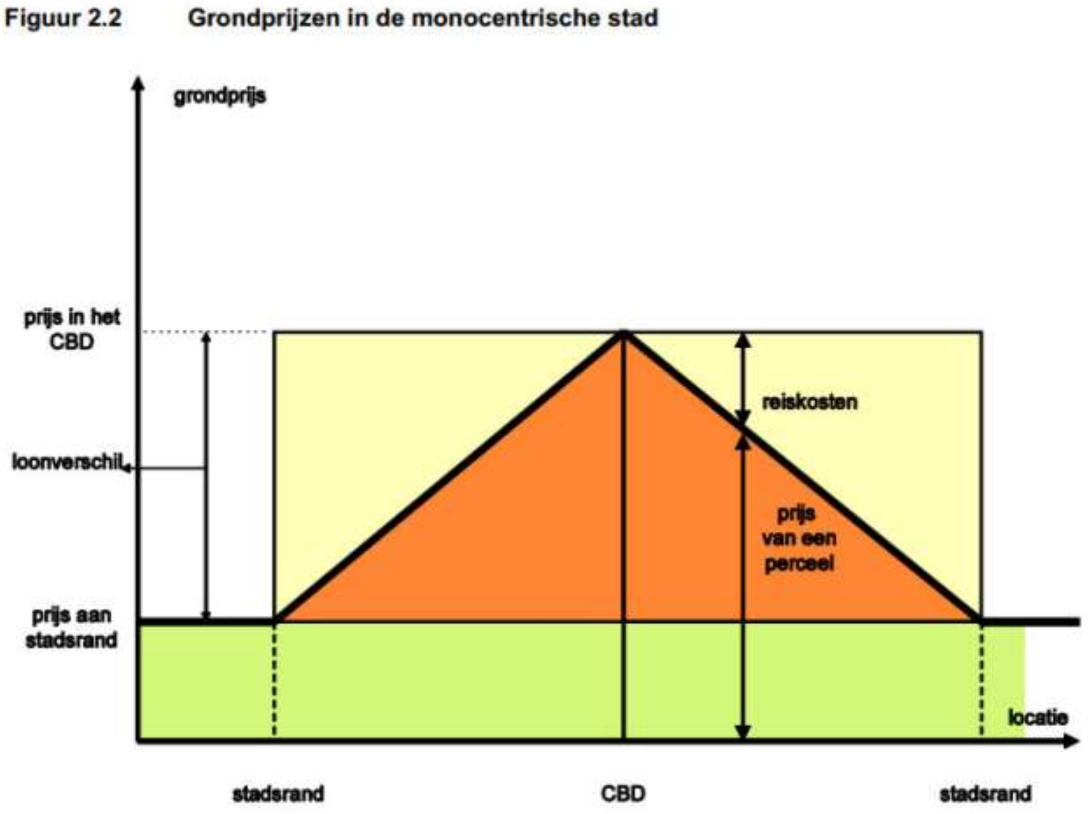

Figure 3: *Alonso model visualization*. Bron: (Duijn, 2024)

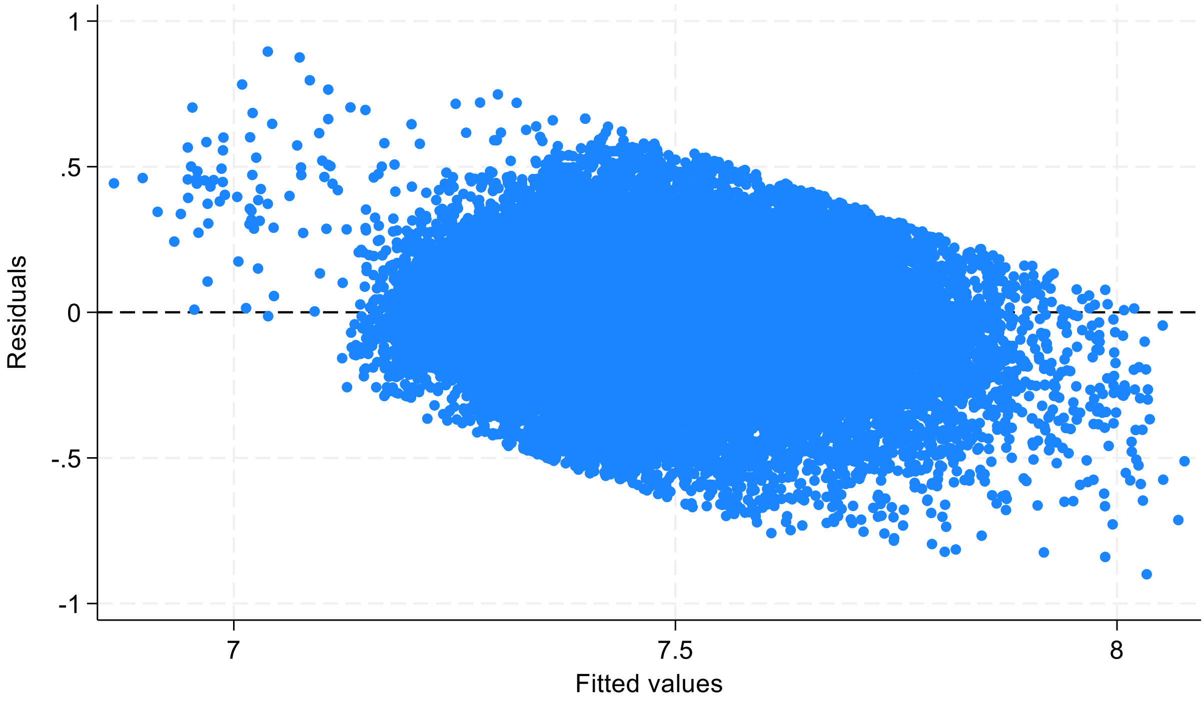

Figure 5: Residuals plotted against fitted values, showing heteroskedasticity.

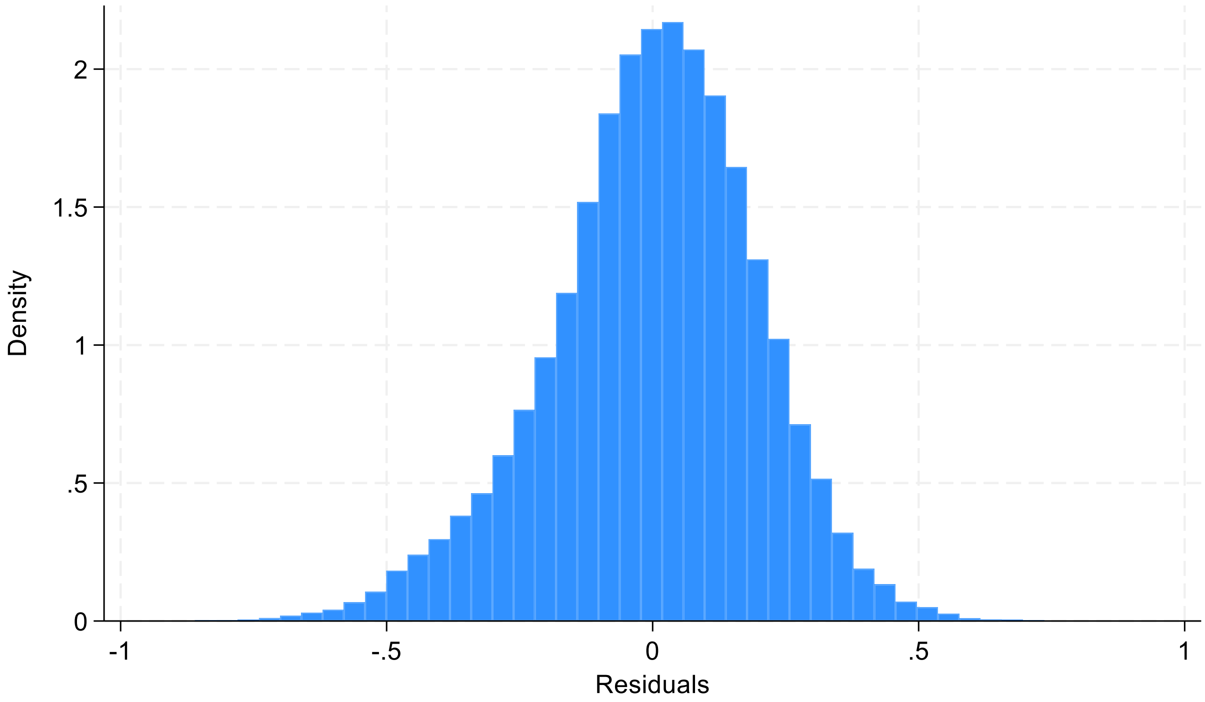

Figure 6: *Distribution of the model residuals*.

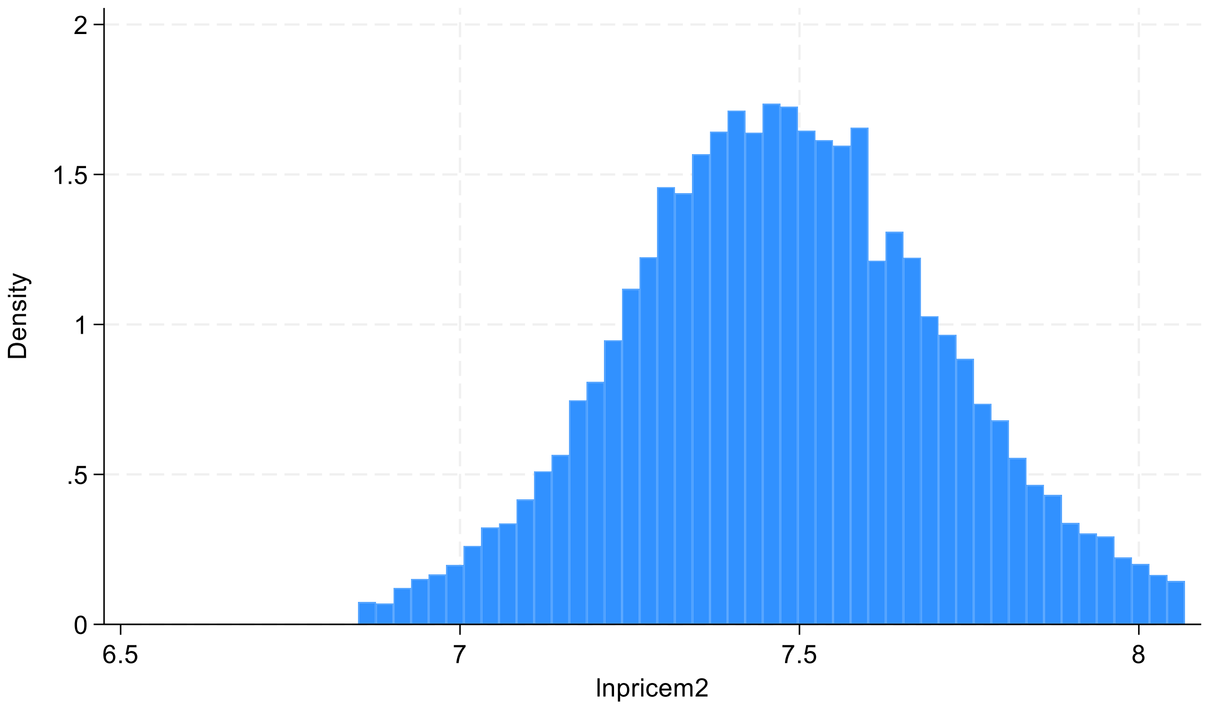

Figure 7: *Normally distributed lnpricem2 variable*.

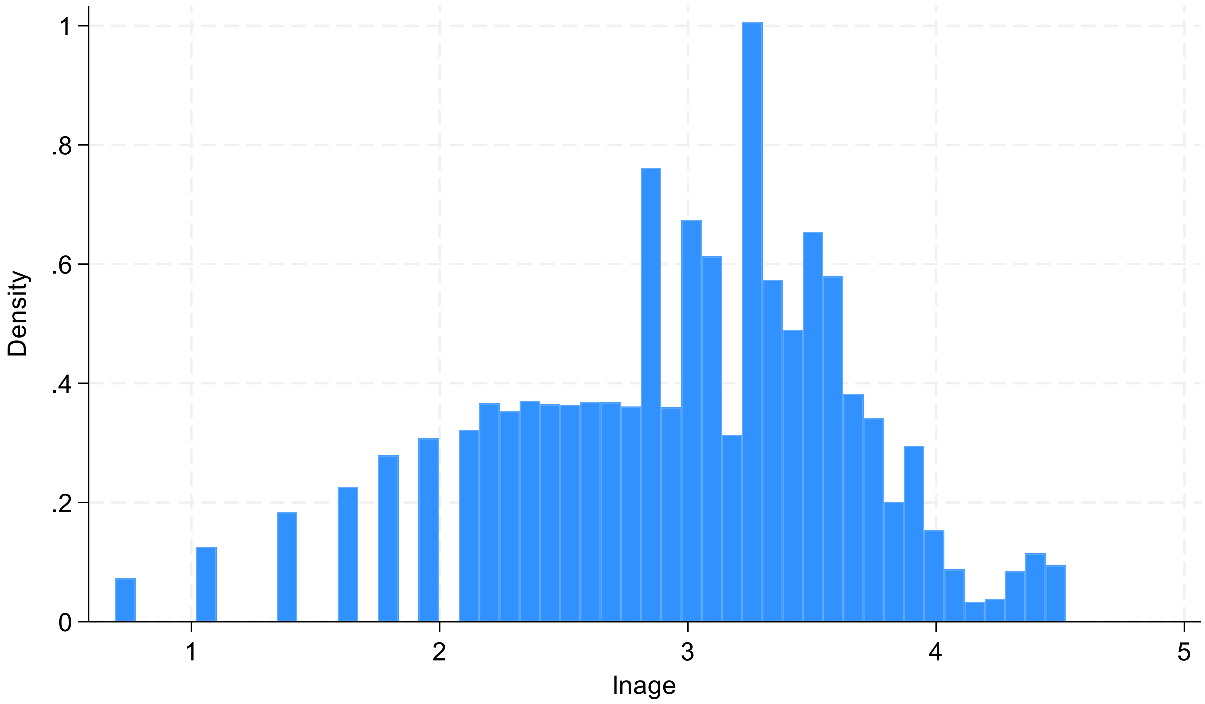

Figure 8: *Distribution of the lnage variable.*

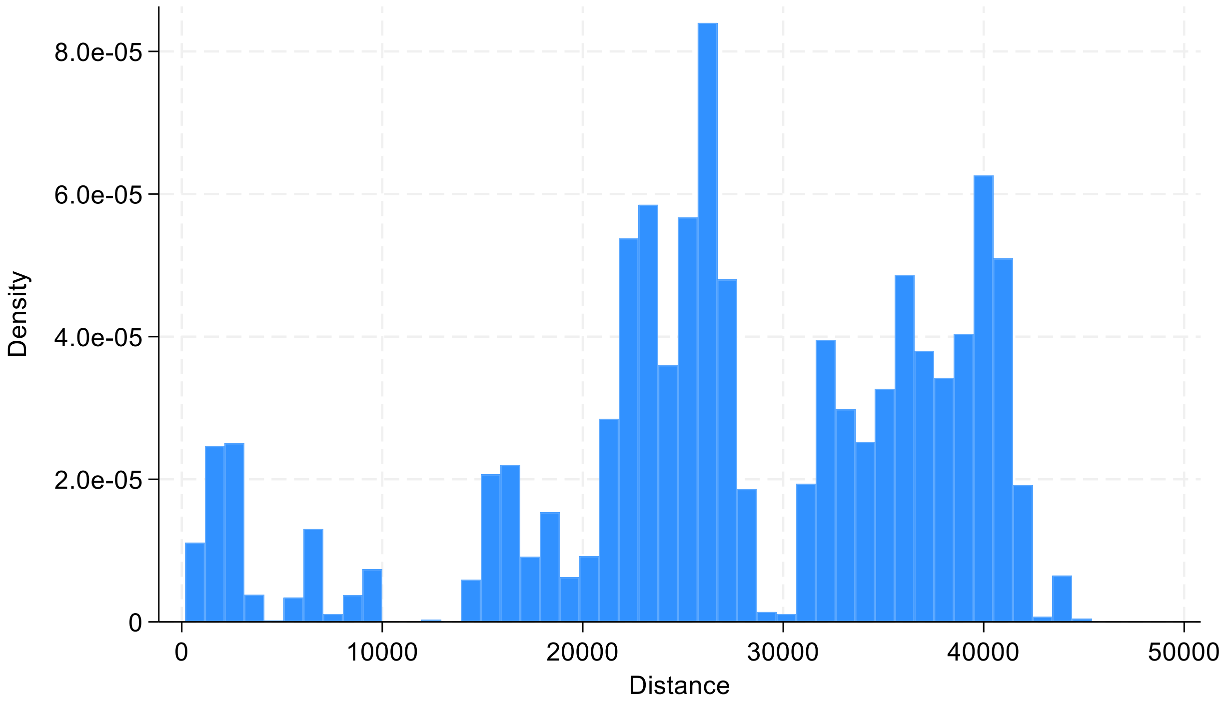

Figure 9: *Distribution of the distance variable.*

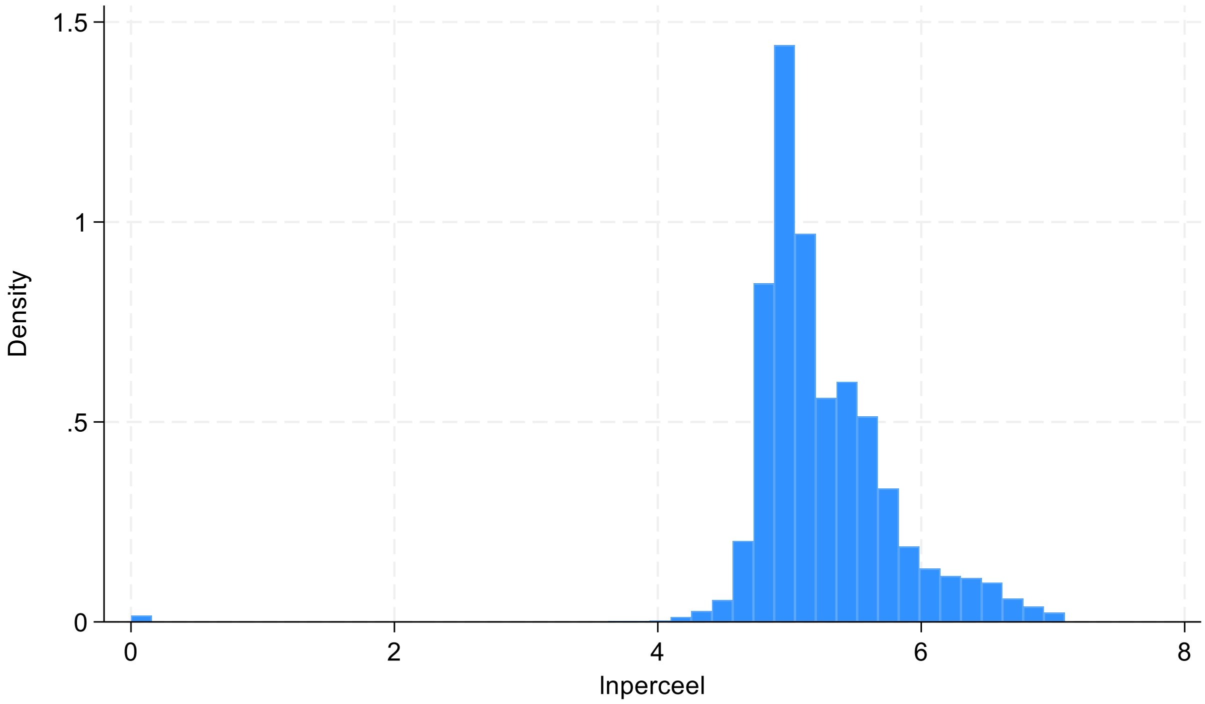

Figure 10: *Distribution of the log transformed parcel size variable*.

Tables

Table 4: *Breusch-Pagan/Cook-Weisberg Test for Heteroskedasticity*.

Breusch-Pagan/Cook-Weisberg Test for Heteroskedasticity

Assumption: Normal error terms

Variable: Fitted values of lnpricem2

H0: Constant variance

chi2(1) = 4760.23

Prob > chi2 = 0.0000

Table 5: *VIF-table*.

Variance inflation factor

VARIABLE	VIF
lnage	2.559
age2	2.251
lnperceel	1.845
Distance	1.149
Hometype: Houseboat	1.041
Hometype: Recreational	1.052
Hometype: Single-Family	6.853
Hometype: Canal Home	1.019
Hometype: Mansion	4.151
Hometype: Farmhouse	1.014
Hometype: Bungalow	2.351
Hometype: Villa	3.403
Hometype: Manor	1.039
Hometype: Ground Fl Aprtmnt	1.029
Hometype: Upstairs Aprtmnt	1.039
Hometype: Maisonette	1.034
Hometype: Porch Apartment	1.069
Hometype: Gallery Apartment	1.03
Hometype: 2 Story Apartment	1.003
Salesyear 2007	1.741
Salesyear 2008	1.629
Salesyear 2009	1.497
Salesyear 2010	1.458
Salesyear 2011	1.425
Salesyear 2012	1.468
Salesyear 2013	1.439
Salesyear 2014	1.598
Salesyear 2015	1.735
Salesyear 2016	1.875
Salesyear 2017	1.92
Salesyear 2018	1.847
Well Maintained Indoor	2.231
Well Maintained Outdoor	2.131
Parking: Parking Spot	1.117
Parking: Carport, No Garage	1.066
Parking: Garage, No Carport	1.503
Parking: Garage + Carport	1.063
Parking: Multi-car Garage	1.093
Monumental Status	1.004
Mean VIF	1.712

Assumptions detailed explanations

First Assumption

The first assumption is automatically solved by Stata, since Stata includes a constant ß0, which ensures that the error terms do have a mean of zero.

Second Assumption

It turned out there was heteroskedasticity in the first place, which was checked by doing a visual test (see Figure 5), showing heteroskedasticity, and a Breusch-Pagan/Cook-Weisberg test. The Breusch-Pagan/Cook-Weisberg test had a p-value of 0.0000, indicating that heteroskedasticity is present as visible in Table 4 (H0: Constant Variance rejected). This was automatically solved when using the didregress-command as was done in the final analysis. After this, heteroskedasticity was not an issue anymore, due to the automatic solution.

Third Assumption

The third assumption is not relevant for this analysis, since no time series data is involved.

Fourth Assumption

The fourth assumption requires the model to be exogenous; in other words, the independent variables should not be influenced by the error term. According to Fingleton (no date), there are four reasons for endogeneity: 1) simultaneous equations bias, 2) omitted variable bias, 3) inclusion of a lagged dependent variable, and 4) uncertainty due to an unmeasurable true independent variable.

1. This condition is only applicable when our dependent and independent variables are jointly dependent (Rehal, 2023), which is not the case here, as price does not significantly influence any other independent variable utilized.
2. This scenario is highly improbable, as I employed the same regressors as other researchers, as visible in Table 2.
3. This criterion applies only to time series analysis.
4. This situation does not apply, as the values exactly express what they should express.

Fifth Assumption

The fifth assumption was tested by plotting the residuals (i.e. error terms) over time. This gave the result which is visualized in Figure 6. It could be concluded that those residuals were normally distributed. This assumption was checked before the second assumption in the model, because the Breusch-Pagan and White tests assume the error terms to be normally distributed.

Sixth Assumption

The sixth assumption (no multicollinearity) was checked by creating a VIF-table, which showed no big multicollinear independent variables in the model. The VIF-values are visualized in Table 5, with absorbed neighborhood fixed effects, as the table would become too large otherwise. I also renamed the fixed effects, to clarify their actual meaning.

Seventh Assumption

The seventh assumption was ensured by checking the literature, and making sure that all relationships were of a linear order. However, the effect of the age of a home on the price is not linear, rather negatively exponential (the effect flattens out after a certain amount of years). Therefore, the variable age^2 was added, which is a quadrate of the age variable, to account for this exponentiality (Harding, Rosenthal and Sirmans, 2007; Francke and Minne, 2016; Knight and Sirmans, 1996). The reason behind this is that after a certain amount of years, there is no further depreciating effect on the housing price, since it is fully depreciated. Furthermore to ensure correct model specification, all necessary transformation have been applied, missing values and outliers are deleted, and there are no signs of the model over- or underfitting the data (Svetunkov, 2023). This was checked by making sure the variables used were normally distributed, the distributions of the used (continuous) variables are visible in the Figures 6, 7, 8, and, 9. There are only a few outliers in the residuals, which was deemed to be acceptable.

Template Research Data Management Plan

Instructions: this is the template for a data management plan. Please fill this in and discuss it with your supervisor during the design phase of the thesis. If your thesis is nearly complete, please add this as an appendix to the thesis. The purpose of making a dmp to think ahead. How will you manage the data gathered for your project? It is not about providing the 'right' answers, but making your research transparent. Some items just require ticking, some require further explanation.

1. General	
1.1 Name & title of thesis	Jesse Luimes - The effect of a new train station on housing prices in rural areas: evidence from Dronten
1.2 *(if applicable)* Organisation. Provide details on the organisation where the research takes place if this applies (in case of an internship).	

2 Data collection – the creation of data	
2.1. Which data formats or which sources are used in the project? For example: - theoretical research, using literature and publicly available resources - Survey Data - Field Data - Interviews	Provide a short description of the sources/data that you are going to use. **I used transaction data from the NVM, in the form of a repeated cross-sectional dataset they provided.**
2.2 Methods of data collection What method(s) do you use for the collection of data. (Tick all boxes that apply)	☐ Structured individual interviews ☐ Semi-structured individual interviews ☐ Structured group interviews ☐ Semi-structured group interviews ☐ Observations ☐ Survey(s) ☐ Experiment(s) in real life (interventions) ☒ Secondary analyses on existing data sets (if so: please also fill in 2.3) ☐ Public sources (e.g. University Library) ☐ Other (explain):

2.3. (If applicable): if you have selected 'Secondary analyses on existing datasets': who provides the data set?	☐ Data is supplied by the University of Groningen. ☒ Data have been supplied by an external party. (Please mention the party here). **NVM**

3 Storage, Sharing and Archiving	
3.1 Where will the (raw) data be stored *during* research? If you want to store research data, it is good practice to ask yourself some questions: • How big is my dataset at the end of my research? • Do I want to collaborate on the data? • How confidential is my data? • How do I make sure I do not lose my data? Need more information? Take a look at the site of the Digital Competence Centre (DCC)) Feel free to contact the DCC for questions: dcc@rug.nl	☐ X-drive of UG network ☐ Y-drive of UG network ☐ (Shared) UG Google Drive ☐ Unishare ☒ Personal laptop or computer ☐ External devices (USB, harddisk, NAS) ☐ Other (explain):
3.2 Where are you planning to store / archive the data after you have finished your research? Please explain where and for how long. Also explain who has access to these data NB do not use a personal UG network or google drive for archiving data!	☐ X-drive of UG network ☐ Y-drive of UG network ☐ (Shared) UG Google Drive ☐ Unishare ☐ In a repository (i.e. DataverseNL) ☒ Other (explain): The code of conduct of the NVM and the UG told me to delete it after my thesis has been finished. The retention period will be 0 years.
3.3 Sharing of data With whom will you be sharing data during your research?	☐ University of Groningen ☐ Universities or other parties in Europe ☐ Universities or other parties outside Europe ☒ I will not be sharing data

4. Personal data	
4.1 Collecting personal data	~~Yes~~/no

Will you be collecting personal data? If you are conducting research with personal data you have to comply to the General Data Privacy Regulation (GDPR). Please fill in the questions found in the appendix 3 on personal data.	
If the answer to 4.1 is ‘no’, please skip the section below and proceed to section 5	
4.2 What kinds of categories of people are involved? Have you determined whether these people are vulnerable in any way (see FAQ)? If so, your supervisor will need to agree.	My research project involves: ☐ Adults (not vulnerable) ≥ 18 years ☐ Minors < 16 years ☐ Minors < 18 years ☐ Patients ☐ (other) vulnerable persons, namely (please provide an explanation what makes these persons vulnerable) (Please give a short description of the categories of research participants that you are going to involve in your research.)
4.3 Will participants be enlisted in the project without their knowledge and/or consent? (E.g., via covert observation of people in public places, or by using social media data.)	Yes/no If yes, please explain if, when and how you will inform the participants about the study.
4.4 Categories of personal data that are processed. Mention all types of data that you systematically collect and store. If you use particular kinds of software, then check what the software is doing as well. Of course, always ask yourself if you need all categories of data for your project.	☐ Name and address details ☐ Telephone number ☐ Email address ☐ Nationality ☐ IP-addresses and/or device type ☐ Job information ☐ Location data ☐ Race or ethnicity ☐ Political opinions ☐ Physical or mental health ☐ Information about a person's sex life or sexual orientation

	☐ Religious or philosophical beliefs ☐ Membership of a trade union ☐ Biometric information ☐ Genetic information ☐ Other (please explain below):
4.5 Technical/organisational measures Select which of the following security measures are used to protect personal data.	☐ Pseudonymisation ☐ Anonymisation ☐ File encryption ☐ Encryption of storage ☐ Encryption of transport device ☐ Restricted access rights ☐ VPN ☐ Regularly scheduled backups ☐ Physical locks (rooms, drawers/file cabinets) ☐ None of the above ☐ Other (describe below):
4.6 Will any personal data be transferred to organisations within countries outside the European Economic Area (EU, Norway, Iceland and Liechtenstein)? If the research takes places in a country outside the EU/EEA, then please also indicate this.	Yes/no If yes, please fill in the country.

5 – Final comments	
Do you have any other information about the research data that was not addressed in this template that you think is useful to mention?	Already signed a code of conduct, which was sent to my supervisor Frans Schilder.