Erasmus University Rotterdam

Erasmus School of Economics
Urban, Port & Transport Economics
Master Thesis

ERASMUS UNIVERSITEIT ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

The Effect of Amateur Sport Facility Openings on Housing Prices

Abstract:

This paper uses a hedonic difference-in-difference price model analysis examining the effect of sport facilities openings on house prices to evaluate if people are willing to pay a premium for living close to a sport facility. It utilises cross-sectional data across time of house transactions and sport facility openings in the province of Gelderland from the period 2008-2017. Three complementary methods are used to find robust results. A distinct possibility exists that the effects found by previous literature are spurious if a more elaborate identification strategy is used and additional controls are added. This study finds that sport facilities openings do not seem to have a conclusive or direct effect on house prices, as the magnitude and sign of the results are inconsistent. This may indicate that potential homeowners do not consider the accessibility of sports facilities in weighing their willingness to pay for a new house. Although this might differ for the level of income.

Name: Sven Verhoeven
Student number: 546142
Supervisor: Jeroen van Haaren
Second assessor: Frank van Oort

Date final version: 17-12-2020

Table of contents

I. Introduction

House price literature is becoming more and more extensive. Early literature on house prices primarily focussed on structural or internal attributes. Most of these attributes have been implicitly estimated using hedonic price models. More recently house price literature looked into external factors that could affect house prices. The range of these external factors is broad in various directions and can be generally categorised in three groups. Some papers focus on locational attributes, like accessibility (Ibeas, Cordera, Dell'Olio, Coppola, & Dominguez, 2012; Bowes & Ihlanfeldt, 2001). Others focus on environmental attributes like taxation (Palmon & Smith, 1998) or air quality (Smith & Huang, 1995; Chay & Greenstone, 2005) while others research neighbourhood attributes, like crime (Thaler, 1978; Lynch & Rasmussen, 2001). These external factors were all found to have some effect on house prices.

A specific category of external factors are amenities. Parks (Crompton, 2001; Song & Knaap, 2001) and schools (Jud & Watts, 1981) were found to have a positive effect on nearby house prices, disamenties like hazardous waste sites negatively affect property values (Deaton & Hoehn, 2004). Some researchers have also examined the effect of sports on house prices. Sport can be categorised in professional sport and amateur sport. Research has primarily focused on the professional sport facilities on house prices (Dehring, Depken & Ward, 2007; Ahlfeldt & Maennig, 2007; Feng & Humphreys, 2012). However, amateur sports did not receive much attention to this day. Little empirical evidence exists about the effects of amateur sport facilities on house prices. This information can be valuable. Amateur sport facilities are often more widespread than professional sport facilities, that were found to positively affect house prices. Furthermore, especially in the Netherlands they are closely located near homes, often present in each neighbourhood. The relationship between sport facilities and house prices is relatively unknown, however. It is about time to disclose this relationship, as amateur sport in becoming increasingly important in our contemporary consumer society. This study critically looked at the existing literature on sport facilities and house prises, as a relationship between both is suggested. However, the results found in the literature are often unclear, contradictory or specific for a certain type of sport facility or certain geographical areas. This study will therefore critically asses if the results found in the literature are not spurious.

The importance of sports and sport facilities in our contemporary society is not to be underestimated. Western societies are becoming more individualistic (Veenhoven, 1999). This leads to more loneliness, depressions and suicides. As society is increasingly focused on the individual, social contact is becoming more and more valuable. Sport plays an important role in social contact, especially team sports. It brings people together and improves social cohesion between individuals. Amateur sports and amateur sport clubs are very important in this process. Sport clubs are the node in local sport activities because of their potency to bring people of different background together to participate in sport activities with each other (Van der Roest, 2018).

All over the world diseases as diabetes, obesity and heart disease are the most prominent causes of death. The age of death is for the first time in history declining instead of rising, indicating that health of an average individual is increasingly becoming worse (Bailey, Cope & Parnell, 2015). Accompanied by human costs are also massive financial costs driven by more expensive healthcare and loss of income due to illness related to unhealthy lifestyles. Sports can offer a solution for these problems as it improves health of individuals. The most effective and cost-efficient way to tackle obesity and other related non-communicable diseases are physical activities and sport (Bailey et al., 2015).

Good health trough sport and activity is a primary subject in local and supralocal government policies. Health policies of the largest cities in the Netherlands recognise the importance of exercise and sport in improving health. These cities actively and passively stimulate citizens to participate in some form of sport activity. The city of Rotterdam had set a goal for the period 2016-2020 to improve the renovation, modernization and flexibility of the supply of sport facilities and sport clubs (Municipality of Rotterdam, 2016). The policy of the municipality of Amsterdam aims to reduce health problems by encouraging citizens to have a healthy lifestyle with enough physical activity. It also wants to plan the city in such a spatial way that it unknowingly promotes exercise (Municipality of Amsterdam, 2017). The cities of Arnhem and Utrecht explicitly state that the emphasis of their health policies lies with promoting exercise (Municipality of Arnhem, 2011; Municipality of Utrecht, 2016). The city of Arnhem wants to enlarge the number of amateur sport participants with more than 10%. Likewise, the province of Gelderland has similar policies (Province of Gelderland, 2020). The question then arises if these policies are measurable in house prices.

Sport facilities as the name says facilitate sport. Field hockey is played on a special field with artificial grass, sometimes with sprinklers, goals and a club house with dressing rooms and a bar to get some drinks and food. A hockey facility makes it possible to play hockey. So do football facilities for football, tennis facilities for tennis and so on for each sport facility and appurtenant sport. Sport facilities are simply the places where sports are played. They are the spatial footprint of most sport activity. So, if sport is arguably becoming more and more important in our society and in government policies sport facilities are too. This leads to the question what exactly the value of sport facilities is to people. Are they willing to pay a certain premium to live close to sport facilities? House prices reflect the value of certain attributes of the house itself but also locational characteristics, in this case sport facilities. The more people value sports, the more likely they want to live close to sport facilities, the higher the demand for houses near sport facilities. Which in turn leads to higher prices for houses near sport facilities. Hence, the main research question of this paper:

What is the effect of amateur sport facility openings on house prices?

This paper tries to identify the effect of amateur sport facilities on nearby house prices using a difference-in-difference method within the hedonic price model framework. It adds to the existing literature in several ways. Firstly, this paper ads knowledge to the almost non-existing field of amateur sport facilities and house values. This field has yet to be examined extensively. Secondly, this study is the first to try to find a causal relationship between amateur sport facilities and house prices by using a difference-in-difference model, whereas previous literature only identified associations between the two. The causal relationship can result in different outcomes and implications than only associations. Thirdly, it uses a large dataset covering the whole province of Gelderland. Past papers performed analyses on city level. A larger dataset can give more conclusive answers. This paper analyses a province level data with both urban and rural environments. Therefore, an urban bias is less likely to be present in the results, as these can be different for both environments (Hoekman, Breedveld & Kraaykamp 2017).

This study could result in several policy implications. For regional governments it can be useful to have an understanding of the costs and benefits that are associated with amateur sport facilities. The existence of external benefits for homeowners may justify some of the large public subsidies for construction and operation of the sports facilities. The results can also have implications for the housing market. The research could give a transparent insight in the, until recently, unclear effect of sport facilities on house prices. Transparent insights in this effect can have implications for sellers as well as buyers of houses nearby sport facilities and can help municipalities with revaluing WOZ value, which in turn can help financing new sport facilities. Furthermore, the results of this study can help house price literature by making clear the effect of the neighbourhood characteristic of sport clubs. This can improve hedonic price model analyses leading to more accurate results and therefore conclusions.

The remainder of the paper will be structured in the following way. The literature review section will discuss previous literature related sports and house prices. Also, the hypotheses of this paper that result from the literature will be discussed. The data section discusses the different datasets used and their sources, summary statistics of the variables and exploratory data analyses. The methodology section discusses the hedonic price model, develops the models used in this study and addresses the analysis strategy. The result section summarises the key findings and results found in the regressions. The discussion section will discuss elaborately on the results and discusses limitations in this study. The last section gives concluding remarks and recommendations for further research.

II. Literature review

Amenities and house prices

Research in house prices has shown that somehow locational features influence values of homes. These locational features or amenities increase attractiveness of a location, which is reflected in house prices. Evidently, schools and quality of schools have a significant influence on house prices. Using an estimation model of housing prices Jud and Watts (1981) found that half a grade level increase in student achievement can boost house values by approximately $675. Ibeas et al. (2012) modelled the interaction of transport and real estate values in the urban area of Santander. They found that each additional transit line in proximity of a house can increases the price with 1,8%, while each additional minute travel time to the CDB decreases house prices with 1,1%. However, houses next to train stations experienced a reduction in value. Bowes and Ihlanfeld (2001) found similar results in their research on the impact of train stations on property values. Both papers prove that accessibility increases house prices. Crompton (2001) reviewed 25 studies on the proximity of parks on property values and concluded that 20 studies found support for a positive impact on house prices. In line with Crompton (2001), Song and Knaap (2001) evaluating mixed land use concluded that house prices increase when public parks are in their proximity. Burkhardt and Flyr (2019) examined the impact of marihuana dispensary openings on house prices and found that marihuana dispensaries can increase house prices in a half-mile radius with up to almost 8%.

Not only house characteristics or locational characteristics and locational amenities can influence house prices. Environmental characteristics can also affect prices. Palmer and Smith (1998) examined property tax capitalization and concluded that taxes can influence house prices. Air quality is also valued by the housing market, as examined by Smith and Huang (1995).

Housing values can also be negatively affected by location specific factors. For example, crime significantly lowers house values (Thaler, 1978; Lynch & Rasmussen, 2001). According to Thales (1978) crimes can decrease the value of a home with approximately $2000. Whereas Dekkers and van Straaten (2009) in their paper concluded that a higher level of aircraft noise result in lower house prices.

Previous examples show that a lot of different and divers drivers and factors play part in the values of house prices and the attractiveness of a location. These need to be carefully considered when examining house prices. This also shows that a lot of different factors are already researched by scholars. Regarding the effect of sport and sport facilities on house prices the literature is still relatively small-scaled. Nonetheless several papers investigate this relationship.

Sport and house prices

Locational features are often amenities or can be regarded as amenity. The importance of amenities in the urban environment is stressed in the paper of Glaeser, Kolko and Saiz (2001). They identified four critical urban amenities, of which one is an abundant variety of consumer and service goods. The authors in their paper mention restaurants and theatres as examples. Numerous other amenities qualify, but the main criterium is that they are local goods. These critical local amenities are important because they can increase the quality of life in a certain place, which consequently increases attractiveness for rich and well-educated workers (Glaeser et al., 2001).

As public goods, sport facilities as spatial stamp of amateur sports can significantly enhance the living quality of urban areas. They facilitate sport, which improves health of individuals. Sport is the most effective way to tackle obesity and other related non-communicable diseases (Bailey et al., 2015). Besides physical health, sport can also improve mental health. Exercising can clear the head and reduce mental issues like stress, depression and anxiety (Taylor, Sallis & Needle, 1985). The social factor that sport facilities offer is also important. Sport clubs are the node in local sport activities because of their potency to bring people of different background together to participate in sport activities with each other (Van der Roest, 2018). Moreover, sport facilities and sport clubs have a favourable value for the society as a whole because of positive externalities like crime prevention, youth promotion and social integration (Breuer, Feiler & Wicker, 2015).

However, sport does not only consist of amateur sport. Sport can be divided in two categories, professional sport and amateur sport. The value of professional sport to people primarily comes from the enjoyment and sociability of watching a sport (Gau, 2007). People therefore also might be willing to pay a premium for living close to professional sport facilities. Research on sport facilities and house prices primarily focusses on this question and the effect of professional sport facilities and sport stadia on house prices.

Professional sport facilities literature

Carlino and Coulsen (2004) appear to be the first to study the topic of sport stadia on property prices. They argue that professional sport franchises are public goods. People want to relocate in proximity of these stadiums, which leads to higher house prices. They used a semi-logarithmic model with monthly rent data and found that NFL franchises boosts inner city house prices with up to 8%. Tu (2005) took into account that Carlino and Coulsen (2004) were not sure whether the premium on rent was caused by the league franchise or the stadium. Analysing pre-, during- and post-development house prices near the FedEx field in Washington, Tu (2005) found that the stadium has a positive effect on nearby house prices, and the closer to the stadium the greater the effect. Ahlfeldt and Maennig (2007) explored the impact of two professional sport arenas on land values in the city of Berlin and found that arenas have a positive effect on land values in a range as far as three kilometres. However, for one stadium the closest effect, between 0 and 1 kilometer, was insignificant.

Using a hedonic model with spatial autocorrelation Feng and Humpreys (2012) estimated the effect of closeness of house values. Their results show that in a five-mile radius of each facility from the NFL, MLB, NBA and NHL property values are higher for houses in proximity of the facilities. They argue that this suggests that sport facilities have positive externalities that are captured in house prices. Ahlfeldt and Kavetsos (2014) in their paper investigate whether positive stadium effects are more related to its design or the functionality as facility. They found a positive effect of new stadia on nearby house prices and argue that this is partly due to architectural spillover effects, although hard results remain absent.

Huang and Humpreys (2014) used a difference-in-difference model with data from 56 professional sport facilities and found that new facilities lead to a boost in mortgage applications. However, most of the increase is driven by locational factors. Feng and Humpreys (2018) estimated that for each 10% decrease in distance to a professional sport stadium the value of a house increases with 1,75%. Dehring, Depken and Ward (2007) researched what the impact is of the announcement of a new Football league stadium on property prices. In the first part of their study they focussed on the announcement effect of a new stadium of the Dallas

Cowboys and found that it positively affects nearby house prices. Nonetheless they also found in a later part of their study that another new stadium in a different city had a negative impact, although this effect was not significant. Kavestsos (2012) studied the announcement effect of the London Olympics on property prices and found that properties in host neighbourhood were sold approximately 3% higher. Within a three-mile radius, house prices near the main Olympic stadium increased with 5%. Davies (2005) applies a different perspective vis-à-vis a quasi-experimental difference-in-difference model. Using an analysis with interviews and surveys, Davies (2005) found that the Millennium Stadium in Cardiff and the City of Manchester Stadium positively affect proximate house prices.

Amateur sport facilities literature

All of aforementioned papers investigate the relationship of professional sport facilities on house prices. Professional sport stadiums are much larger than amateur sport facilities and have one on one a much bigger impact. However, amateur sport facilities are present in far greater numbers and are often spread through cities. Often each neighbourhood has its own sport facilities, a swimming pool, football club, or whatsoever. These amateur sport facilities can also be regarded as amenity. They fit in the criteria of Glaeser et al. (2005) that they are a local good and can increase quality of life. They do affect the locational attractiveness.

Nonetheless amateur sport facilities on house prices has received less attention. Sirmans, Macpherson and Zietz (2005) in their literature review on hedonic house pricing only report golf courses as subject of amateur sport in house pricing. The nine reported studies all found a positive impact of golf courses on house prices. Do and Grudnitski (1995) examined the impact of golf courses on the selling price of abutting houses. They found that houses neighbouring golf courses sell for 7,6% higher prices than comparable homes. Similar results were found by Nichols and Crompon (2007). Chung and Chang (2010) use a hierarchical linear model to examine the impact of leisure and sport facilities on housing prices in Taiwan. They use one variable that measures satisfaction with both leisure and sport facilities in the neighbourhood and found that is positively affect house prices. This result gives an indication but is not very useful as the exact impact of amateur sport facilities is not known. It is tangled with leisure and the variable is a subjective measure. Chung and Chang (2010) did also not define clearly if the sport facilities they involve in their research are amateur or professional. Mok, Chan and Cho

(1995) use a hedonic model with a Box Cox analysis to investigate the impact of several structural, locational and neighbourhood attributes on house prices in Hong Kong. In their study sport facilities as swimming pools, squash courts and tennis courts turned out to positively affect property values. They however use a dummy variable, were 1 means that sport facility is in the estate and 0 if it is not. They thus do not examine externality effect from proximity to an independent sport facility, but only if a sport facility is present in the estate itself. Also, in the situation of Hong Kong, as described in the paper, the estates are flats and the houses are apartments in the flat. Tse and Love (2000) performed a similar study measuring house prices in Hong Kong but found a negative impact of sport facilities in 10-minute walking distance on property prices. They note however that this may be due to overlapping with another variable that measures if sport facilities are present in the estate, what usually is the case in Hong Kong. However, the housing market in Hong Kong and that of the Netherlands are poles apart, so those results may not be relevant in this study. This is, in lesser extent, the case for all other mostly American literature.

Hummel (2018) is the first to research the effect of amateur sport facilities on housing values in the Netherlands and found opposing results. But in generally, sport facilities are negatively associated with house prices. The paper presumes that the negative results stem from the morphology of urban areas and the nature of sport facilities. They locate in areas with enough space and relatively cheap land value. These places are often at the edges of cities near houses that are already cheap due to that same low land value.

Professional sport facilities seem to positively affect house prices, some literature shows. Other literature indicates a negative effect. The scares literature on amateur sport facilities does not come to conclusive answers on their effect on house prices. The question rises if an effect really exists. Also, the effects can have a different outcome for different proximities to a sport facility. More in-depth analysis of the subject is required. This paper tries to take on this task.

Hypotheses

Some effect of amateur sport facilities on house prices is likely. Sport facilities as spatial stamp of amateur sports can significantly enhance the living quality of urban areas. They facilitate sport, which improves health of individuals. Sport is the most effective way to tackle obesity and other related non-communicable diseases (Bailey et al., 2015). Besides physical health, sport can also improve mental health. Exercising can clear the head and reduce mental issues like stress, depression and anxiety (Taylor, Sallis & Needle, 1985). The social factor that sport facilities offer is also important. Sport clubs are the node in local sport activities because of their potency to bring people of different background together to participate in sport activities with each other (Van der Roest, 2018). Moreover, sport facilities and sport clubs have a favourable value for the society as a whole because of positive externalities like crime prevention, youth promotion and social integration (Breuer, Feiler & Wicker, 2015). Sport facilities create positive neighbourhood effects, therefore homeowners value living near a sport facility higher than living far away from a sport facility. The closer people are living to a sport facility the less time and effort has to be put into getting to the sport facility. So, living in proximity to a sport facility is convenient. Living very far away is inconvenient. If amateur sport is valued as positive, homebuyers are willing to live close to a facility for convenience.

Property values are important as they reflect the attractiveness of certain place. In property prices it is possible to, controlled for several variables, identify the attractiveness of a location. If two identical homes in two different cities are valued with different prices a difference in locational attractiveness can be revealed. Therefore, property prices can function as a sort of proxy for locational attractiveness, which is increasingly crucial for especially the future of cities, according to Glaeser et al. (2001). If people do positively value living close to sport facilities to capitalize on the positive effects of participatory amateur sports this value can be measured spatially in house prices. House prices reflect the value of certain attributes of the house itself and also its locational characteristics. The more people value sports and the higher they value it, the more likely they want to live close to sport facilities. The more this occurs the higher the demand for houses near sport facilities. This, through scarcity of houses near sport facilities and market mechanisms, in turn leads to higher prices for houses near sport facilities. People are willing to pay premium prices for living in high-quality of living environments, in this case enhancing the quality through sport. The premium prices are reflected in the house prices of those areas. This theory explains the mechanism of participatory amateur sport on the value of house prices and leads to the first hypothesis:

> *H1: The introduction of a sport facility within a range of 800 meters has a positive impact on nearby house prices.*

However, sport facilities can also bring negative externalities with them. A negative effect can be noise nuisance. Especially outdoor sports can be accompanied with noise coming from people yelling, cheering or the sound of rackets and sticks hitting balls. Regarding outdoor sport facilities, light pollution can be disturbing, mainly in the evening. Hockey and football field are illuminated with huge light poles. Furthermore, for both indoor and outdoor facilities people grouping together and meeting close to the facility can be disturbing. Sometime at peak hours, especially in the weekends, lots of traffic flows from and to the facilities, leading to traffic jams, air pollution and noise pollution. This can make it unattractive for people to live close to a sport facility. All in all, the negative effects are stronger the closer to the sport facility than they are further away. So, the effect of a new sport facility on house prices is expected to be negative for house that are in very close proximity. This leads to the following hypothesis:

> *H2: The introduction of a sport facility in a range of 350 meters has a negative impact on nearby house prices.*

Negative effects of sport facilities like noise and light pollution are more perceivable in the open air than between four wall and a roof. For example, a football match on a field in the open air brings some noise from kicking the ball and shooting and cheering players and spectators. This can be annoying for nearby dwellers. This is much less the case when the football match is played inside a hall. The externalities of the sport are captured inside the hall. This brings a lot less annoyance for nearby dwellers. In the same way, the positive effects can also be captured between four walls. This leads to the fourth hypothesis:

> *H3: Outdoor sport facilities have a stronger effect on house prices than indoor sport facilities.*

Sport facilities can offer certain values to people as health bonusses and pleasure. But the social aspect of sport is also an important factor. In teams sports the social aspect is greater than in individual sports. The value of team sport and therefore the value for team sport facilities can be valued higher than for individual sport facilities. This leads to the last hypothesis:

> *H4: Team sport facilities have a stronger effect on house prices than individual sport facilities.*

III. Data

House data

The data that this research uses consist of two datasets. The first one is data consisting of house transactions. This dataset originally contains data of 154,735 houses sold in the province of Gelderland in the period from 2008 until 2017. The dataset comes from the NVM, an association of around half of all estate agents in the Netherlands. The primary data is the date of the transaction, the transaction price of the house sold and its location. It also contains multiple house characteristics, like size, number of certain rooms, the type of home, state of garden, closeness to the city centre and attractiveness of the location.

Data preparing & cleaning

Each variable in the dataset was checked for correctness, plausibility and credibility. Some houses had for example a transaction value of a few euros while others had values nearing one billion euros. Some houses had 40 rooms while only having 100 m^2 floor area, while others had a floor area of over one million m^2. Houses containing variables with weird or incorrect values were correspondingly dropped from the dataset. The variable ceiling height was created by dividing m^3 by m^2. To correct for erroneous ratios between volume and square meters, all house transactions with ceiling heights below 2 meters and above 5 meters were dropped from the dataset. Also, to correct for outliers, the 0,5 percentage highest and 0,5 percentage lowest transaction values were dropped from the dataset.

Several ratios were created to test the coherency of the data. A ratio was created to test the number of square meters per room. Datapoints that had an average above 100 square meters per room were dropped from the dataset. The ones larger than 100 were all family homes with only one room, which seems unsound. The houses below 8 square meters per room all had 14 or 15 rooms, which seems highly unlikely. They were also dropped from the dataset. A ratio was created to check for the number of rooms per floor. Datapoints that had the number of rooms divided by the number of floors plus attic below 1 were dropped. Another ratio was created to check for the number of toilets in a house. If the ratio rooms plus bathrooms plus kitchen divided by toilets was smaller than 0,5 the datapoints were dropped. These datapoints all had 7 or more toilets which seems not credible.

To perform analyses on the effect of sport facilities and house prices it is necessary to know the locations of both houses and sport facilities and their reciprocal spatial relationship. It is important to know which transactions are close to which sport facilities and are possibly affected by the sport facility, as hypothesised in this paper. Therefore, for further analyses the exact coordinates of the house transactions are necessary. The dataset was delivered with postal codes and addresses but without coordinates, however. To retrieve the coordinates of each house transaction an opensource excel formula was used. The formula derives the correct geocoordinates from a selected postal code and house number by using the georegister database from the base register of addresses and buildings of the Dutch government (BAG, 2020). For each house transaction the corresponding x and y coordinates were retrieved and then added to the dataset as two new variables. For a small number of house transactions the right coordinates were not found. These were dropped from the dataset. After cleaning and preparing the dataset consists of 143,715 house transactions.

The period from 2008 until 2017 was subject to large fluctuations in house price. To correct for these fluctuations, a new variable was created that multiplies the transaction value with the trend of the house price in the certain period of the transaction. The trend is retrieved from the

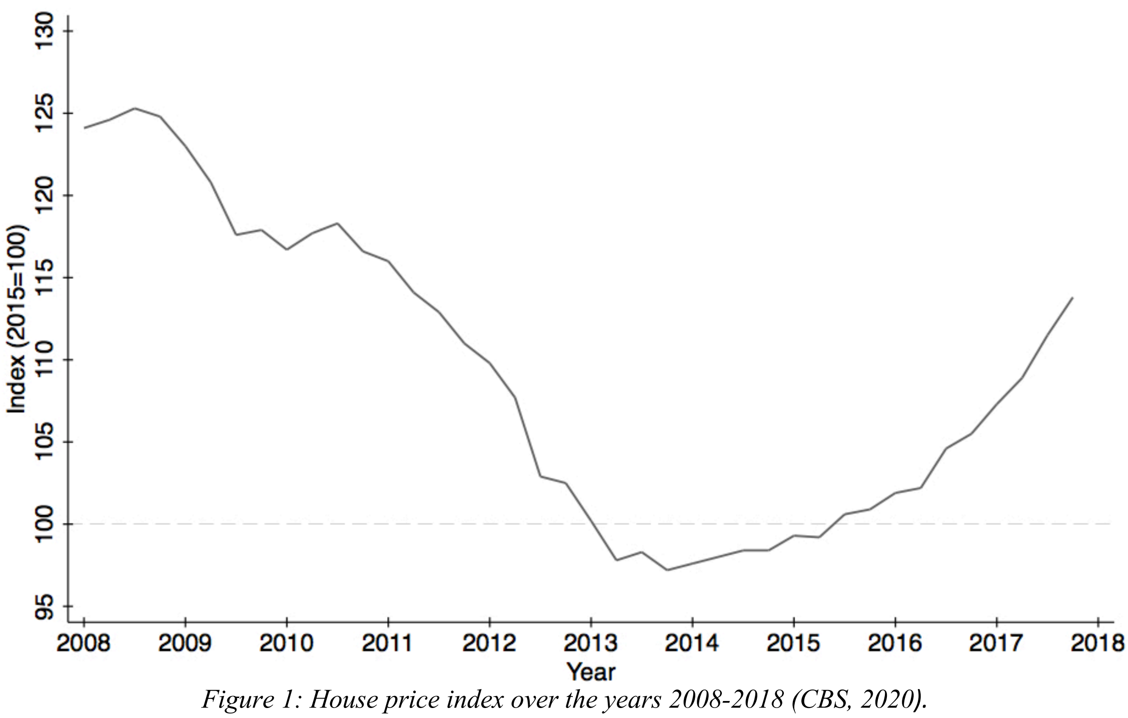


Figure 1: House price index over the years 2008-2018 (CBS, 2020).

CBS. It calculated the average house price per quarter of each year relative to 2015 as the base year. As figure 1 shows, the overall house price for the province of Gelderland was in decline until around 2014 and began to grow again from that time on.

After all processes previously discussed the dataset was added as a shapefile layer into QGIS and as a data file into Stata for further analysis and descriptive statistics.

Dependent variable

The dependent value in this study is the transaction value of houses sold. To perform regression analysis under correct assumptions, the transaction value has to be normally distributed and continuous. A histogram with transaction value on the x axis and frequency on the y axis shows that the variable transaction price is a bit skewed to the left (figure 2). To correct for this skewness the log of price is taken and plotted in a histogram (figure 3). By taking the natural logarithm of price the variable comes closer to normal distribution. The same is applied to the quarterly trend corrected price. This data has a slight skewness to the left but becomes closer to normal distribution when taken in logarithm (figure 4 & figure 5). As the logarithms of both the price and the trend corrected price approach normal distribution the closest, these variables will be used in the analyses. Descriptive statistics furthermore show that the dataset contains more transactions in recent relative older years (figure 6).

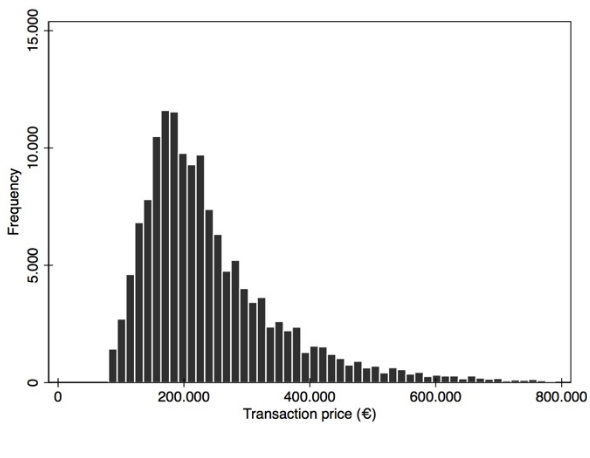


Figure 2: Distribution of transaction price.

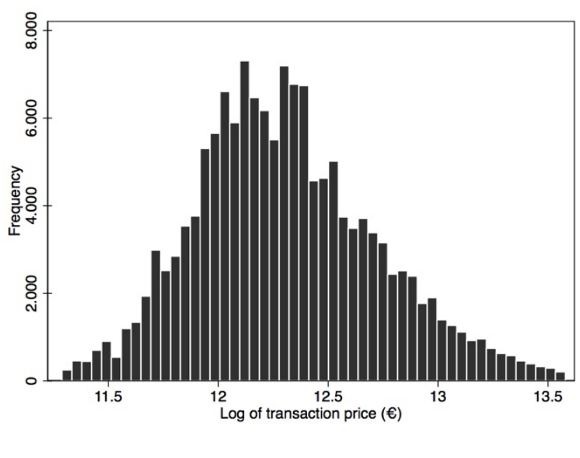
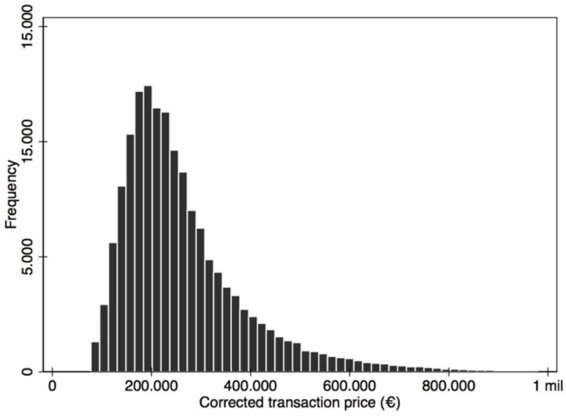


Figure 3: Distribution of the logarithm of transaction price.



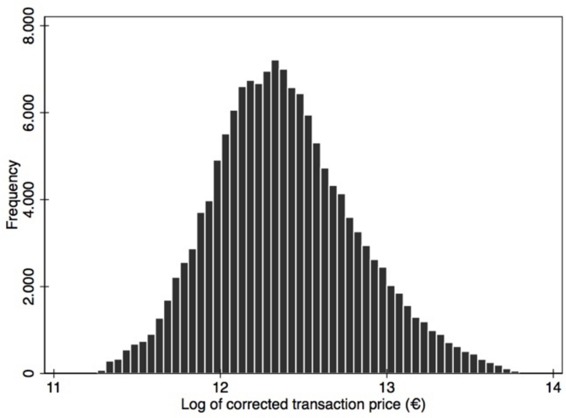


Figure 4: Distribution of transaction price corrected by the trend index.

Figure 5: Distribution of the logarithm of transaction price corrected by the trend index.

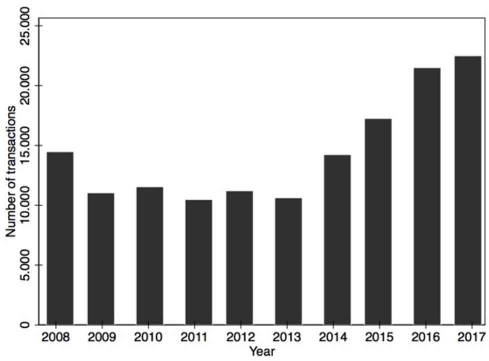


Figure 6: Distribution of transactions by year.

Categorical variables

The categorical variables of the dataset are displayed in table 1. The variables practice rooms, outdoor balcony, roof terrace, parking, garden, heating, scullery and fine location are created by transforming similar variables into binary variables, as the numbers in the additional categories were very small or the original categorisation was useless. In those cases, binary categories would make more sense. Scullery for example was continuous with most of the house having none or one scullery and only 44 having 2 sculleries. Fine location was originally grouped into 5 categories, houses not in proximity of a fine environment, houses close to a forest, houses close to water, houses close to a park and houses with a clear view. As their impact is arguably similar, beautiful locations have a positive impact on house values, all fine locations were grouped into one category. The output in table 1 shows furthermore that most houses were built in the periods 1961-1970 and 1971-1980. Pre Second World War period houses are far less common. Also, most houses have good maintenance, which true for the interior as well as the exterior. Houses on quiet and medium quiet roads are roughly evenly divided, while only a small portion of houses is built on a busy road.

Category	Number
House	115724
Apartment	27991
With lift	9378
Without lift	18613

Dwelling type	Number
Other house	1
Simple house	4935
Boat house	19
Recreational house	862
Single family house	90293
Canal house	14
Mansion	6108
Living farm	3449
Bungalow	5249
Villa	4039
Country house	737
Other apartment	9
Ground floor apartment	3440
Upstairs apartment	4204
Maisonette	1986
Porch apartment	11093
Gallery flat	7147
Welfare flat	37
Ground floor with stairs	93

Building period	Number
1500-1905	5694
1906-1930	11392
1931-1944	8007
1945-1960	13179
1961-1970	24060
1971-1980	25159
1981-1990	18502
1991-2000	18873
2001≤	18849

Living room shape	Number
Other living room	120337
L-shape	10799
T-shape	232
Z- or U-shape	2155
Open room	7765
Room en suite	2427

Attic	Number
Yes	39567
No	104148

Garret	Number
Yes	12428
No	131287

Practice room	Number
Yes	1204
No	142511

Scullery	Number
Yes	36224
No	107491

Dormer	Number
Yes	22276
No	121439

Roof terrace	Number
Yes	8217
No	135498

Indoor balcony	Number
Yes	6836
No	136879

Outdoor balcony	Number
Yes	30044
No	113671

Maintenance interior	Number
Terrible	401
Very bad	140
Bad	2066
So-so	683
Mediocre	10677
Reasonable	4706
Good	104719
Very good	4241
Excellent	16082
Total houses	**143715**

Maintenance exterior	Number
Terrible	338
Very bad	124
Bad	1426
So-so	432
Mediocre	7938
Reasonable	4317
Good	110654
Very good	3751
Excellent	14735

Garden	Number
Yes	95873
No	47842

Insulation	Number
None	16922
Terrible	35985
Bad	23523
Mediocre	18461
Good	18712
Excellent	30112

Heating	Number
Yes	139013
No	4702

Busy road	Number
Quiet	75237
Medium	64784
Busy	3694

Fine location	Number
Yes	44345
No	99370

Central location	Number
Outside built area	5180
Unknown	32753
In residential area	94023
In centre	11759

Parking	Number
Yes	72044
No	71671

Table 1: Overview of information on the categorical variables.

Continuous variables

The continuous variables of the house transaction dataset are displayed in table 2. The m^2 and m^3 variables seem to fall between reasonable values of smallest 26 and 70 and largest 419 and 1089 respectively. Plot size varies massively with a large standard deviation. This is mainly because apartments have a plot size value of 0. Floors, rooms, kitchen, toilet and bathroom seem to have reasonable mean, standard deviation and minimum. The maxima however seem quite high. This is kept this way as the values originally were much higher and are already corrected for incorrect values. Also, although unlikely, it is possible that houses with high values for said variables exist. Ceiling height has a credible mean of 3.1 meters. The minimum and maximum fall between 2 and 5 meters, as was described earlier. The mean of kitchen falls below 1, indicating that some houses do not have a separate room for a kitchen.

Variables	Number	Mean	Standard deviation	Min	Max
M2	143,715	122.4	40.43	26	419
M3	143,715	386.5	145.0	70	1089
Plot size	143,715	530.1	2,226	0	318190
Floors	143,715	2.453	0.820	1	8
Rooms	143,715	4.670	1.344	1	15
Kitchen	143,715	0.701	0.467	0	4
Toilet	143,715	3.977	1.511	1	9
Bathroom	143,715	0.912	0.450	0	4
Ceiling height	143,715	3.133	0.390	2	5

Table 2: Summary statistics of the continuous variables.

Sport facilities data

The second dataset contains data of sport facilities in the Province of Gelderland. This dataset is acquired from the Department of Sport from the Province of Gelderland. It consists of all sport facilities in Gelderland and was filtered for all accommodations opened between 2009 and 2016. This time period is chosen since for the analyses house transactions from one year before and one year after sport facility openings are needed, and the house transaction data spans from 2008 until 2017. Adjusted for errors in the data, this leaves a dataset of 115 sport facilities. Furthermore, the dataset contains the name, location in coordinates and postal codes, year built, the type of sport facility and if the facility is within the city limits or not. This dataset was also added to QGIS and Stata. The distribution of sport facility openings is displayed in figure 7. The figure shows that the sport facility openings are evenly distributed over the province, with more sport facility openings in denser urban areas and less openings in rural less populated areas of Gelderland. Figure 8 displays the spatial distribution of both sport facilities and house transactions over the period from 2008 until 2017. The figure makes clear that all sport facility openings are in proximity of at least some house transactions. More detailed larger figures of the distribution of sport facilities and house transactions are presented in the appendix.

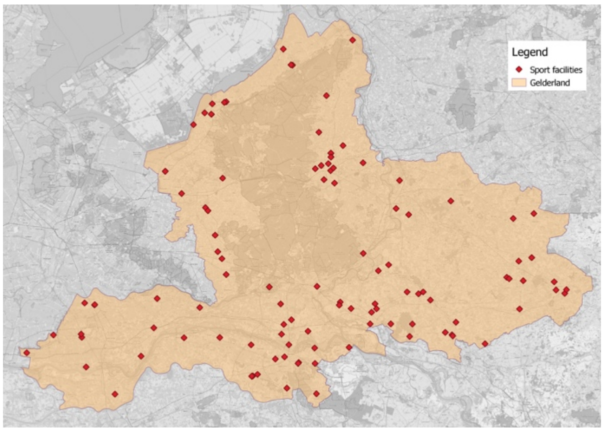


Figure 7: Distribution of sport facility openings.

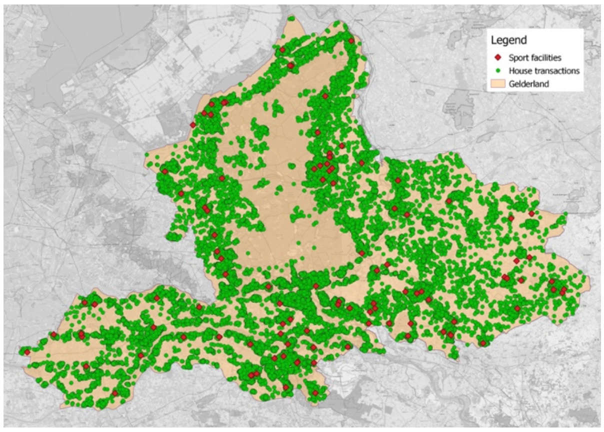


Figure 8: Distribution of sport facility openings and house transactions.

Hoekman et al. (2017) in their study about sport participation in the Netherlands distinguish four types of sport facilities: sport fields, sport halls, fitness centres and swimming pools. This paper uses the same categorisation. However, as some of the data does not fit in one of the categories, a fifth category is added. This fifth category consist mostly of motocross, lawn bowling and equestrian sport facilities. In the dataset, the different kind of sport facilities are grouped in one of the five aforesaid corresponding categories. Figure 9 shows that the sport facilities that were opened in Gelderland in the years 2009-2016 mostly fall in the categories sport fields and sport halls. Figure 10 shows that the distribution of openings is not evenly spread over the years, more sport facilities were opened in the first years of the selected data and less in the last years. 2008 is with a quantity of 30 the year with the most sport facility openings, while in 2014 only six sport facilities were opened in Gelderland of the total of 115 openings in the period 2008-2016.

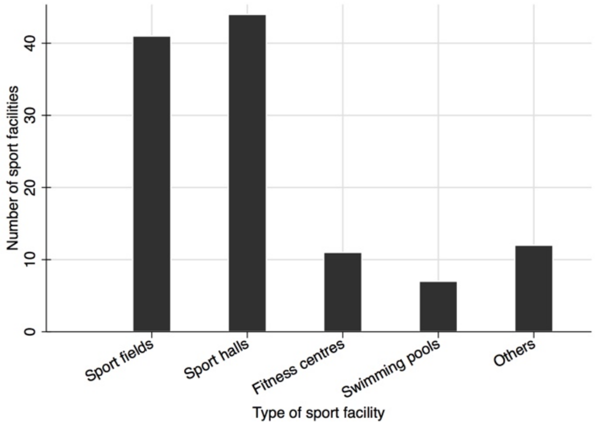


Figure 9: Distribution of sport facility openings by category.

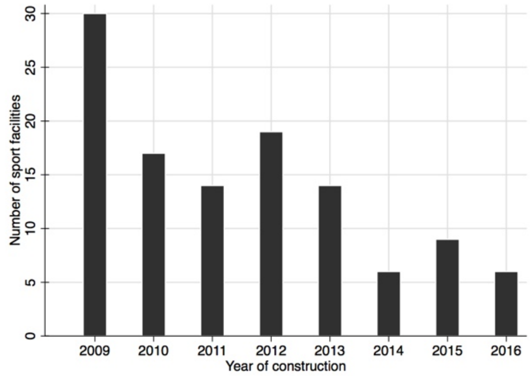


Figure 10: Distribution of sport facility openings by year.

Exploratory data analyses

The house transactions were paired with the sport facility data in one dataset in QGIS. Subsequently, a buffer of 800-meters was drawn around the sport facility openings and for each house transaction was counted if it was located in the range of the sport facility, using a python script (appendix). A buffer dummy was created, all house transactions in the 800-meter buffer of a sport facility opening were assigned with a value of 1, others a value of 0. A pairwise correlation was carried out between transaction price and the buffer dummy. The results are displayed in table 3. This result is significant with a p-value of 0,000 and a coefficient with a value of -0,0688. The same correlation is run with a 350-meter buffer and displayed in table 4. Here, the result is significant with a p-value of 0,000 and a coefficient with a value of -0,0348. Both correlations are small but significant, indicating that a negative correlation between sport facility opening and house prices exists.

	Price	800m buffer
Price	1.000	
800m buffer	-0.0688* 0.000	1.000

Table 3: Correlation matrix of price and the 800-meter buffer dummy.

	Price	350m buffer
Price	1.000	
350m buffer	-0.0348* 0.000	1.000

Table 4: Correlation matrix of price and the 350-meter buffer dummy.

To further explore the data, a simple OLS regression of price against both buffer dummies with several control variables[1] is run. The results are shown in table 5. The results of the exploratory OLS regression show that sport facility opening have a small and significant negative effect on house transactions.

Variables	(800m buffer) price	(350m buffer) price
Dummy 800	-3,247***	
	(395.2)	
Dummy 350		-3,491***
		(741.0)
Control variables	Yes	Yes
Observations	143,715	143,715
R-squared	0.769	0.769

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Exploratory OLS regression of both the 800m dummy and the 350m dummy.

Both the correlation results and the OLS results display a significant relationship between sport facilities and house transactions. These results are in line with the results found in the paper of Hummel (2018). The results indicate that there exists some negative spatial relationship between both subjects. Presumably, sport facilities locate often at the edges of cities in spacious and inexpensive areas near houses that are already cheap due to that same low land value. This can explain the negative relationship between the sport facility opening and house prices. Therefore, a new variable distance to nearest city centre will be included in the regression analyses as robustness to cope with this spatial problem. This variable is created by calculating the distance of a house transactions to the nearest one of the ten biggest cities in Gelderland and six big cities adjacent to the border of Gelderland. Further and more in-depth analyses are however necessary to disclose the exact nature of sport facilities and house prices and the extent of their relationship. The methodology section will elaborate on the methods used to tackle this issue.

[1] The control variables include structural controls, locational controls, municipal controls and year controls.

IV. Methodology

Hedonic difference-in-difference

In general, two approaches can be used to determine the value of objects in the absence of a market; revealed preference methods and stated preference methods. In the latter, data comes from questionnaires about what people would do in a certain situation or what they would value a certain attribute. In the revealed preference methods behaviour of people who are subject to utility maximizing constraints is observed. The most important difference between both methods is that while stated preference method data comes from hypothetical situations, revealed preference method data comes from actual real-life observed choices (Freeman, Herriges & Kling, 2014).

For determining amenity values the hedonic price model is one of the most popular revealed preference methods. The general idea of a hedonic price model is that utility from different characteristics or components of a product can be derived and valued. This was first proposed by Court (1939) in the automotive industry, and later extended by Lancaster (1966) who provided microeconomic foundations on how to estimate the value of characteristic that create utility, and Rosen (1974) who developed non-linear hedonic pricing. Hedonic pricing is more useful the more differentiated the products in markets are. Real estate is a very differentiated market, as houses differ enormously in various characteristics liken size, number of bedrooms and location (Sirmans et al., 2005). So, hedonic pricing is a very useful method to determine amenity and, in this case specifically, sport facility values.

Witte, Sumka and Erekson (1979) were among the first to introduce the hedonic price model in the housing market. The idea explained in their paper basically means that the hedonic model functions as a basked, encompassing the different components of a house. This idea is perfectly explained by Malpezzi (2002). The house as a whole has a certain value. Another house with different components has a different value, and so on for multiple houses with different components and different values. From the difference in components and values it is possible to derive the values of single component. So, for example, one can derive the value of swimming pool included in a house. In this way it is possible to see what difference a swimming pool makes to the value of a house, relative to a house that does not have a swimming pool. A swimming pool is one component, but this idea of capturing a value of a component can be applied to numerous product components. A swimming pool can be associated with a high positive value, while poor maintenance for example can be associated with a negative house value. Housing components can be for example specific house characteristics as bedrooms, kitchens or square meter of liveable surface. Together, all the components of the house in sum reflect the house value.

Sirmans et al. (2005) reviewed the literature on hedonic price models in real estate. In their paper they identified that different hedonic house characteristics used in approximately 125 papers can be categorised in eight different groups. These groups are *structural features, internal features, external features, natural environment features, neighbourhood & location, public services, marketing, occupancy & selling factors and financial issues*. Sirmans et al. (2005) found that house characteristics are most often researched. These specific house characteristics can be captured in the hedonic price model as explained earlier. But this can also be applied for components that are not attached to the house itself but closely related to it in its spatial nature (Rosen, 1974; Sirmans et al, 2005). This are for example neighbourhood characteristics like quality of schools, air quality or accessibility to transportation. These are characteristics that Sirmans et al. (2005) identify as locational features. Linneman (1980) found that in some cases neighbourhood effects or locational features can explain up to almost 48% of the variation in house prices. For identical houses this can be as much as 100%. The components that in total reflect the price of a house do thus not only exist of house specific characteristic, but also locational characteristics. Both house characteristics and locational components fill the basket of components that together sum the price of a house. Hedonic models therefore need to incorporate both components. This paper tries to clarify the locational component of house prices.

Within the hedonic price model framework, this paper uses a difference-in-difference method. The paramount advantage of this method is that is removes the bias occurring from omitted time-invariant factors. This quasi-experimental difference-in-difference model method is not uncommon. Lang, Opaluch and Sfinarolakis (2014) use this method to evaluate the impact of wind turbines is on nearby residential property values. Dehring, Depken and Ward (2007) use the method to investigate the impact of the announcement of a new professional sport stadium on house values. Burkhardt and Flyr (2019) used the in their paper so called 'traditional doughnut difference-in-difference method' to assess the effect of marihuana dispensary openings on house prices.

Inspired by aforementioned articles, this paper uses cross-sectional data and a difference-in-difference hedonic price model analyses to examine the effect of sport facilities openings on house prices. In this way it is possible to measure a potential effect of sport facilities on house prices while controlling for several interfering variables.

Models

After cleaning of the datasets, both the house transaction data and sport facility data were imported into QGIS. In QGIS the spatial relationship between the house transactions and the sport facility opening can be investigated. To test the hypotheses, house transactions that lie within a 350-meter buffer, within an 800-meter buffer and houses outside those buffers of a sport facility opening have to be identified. Also, it is necessary to know if a transaction is pre or post opening of a sport facility. Then the house transactions are divided into six groups. One group of houses outside the 800-meter buffer but not further away than 5km sold in the year before the opening of the nearest sport facility and one group sold a year after the opening, a group of houses inside the 800-meter buffer sold in the year before the opening of the sport facility and one group sold a year after the opening and groups within the 350-meter buffer one year before and one year after opening. With these groups difference-in-difference analyses are possible to perform. All datapoints outside the 5km buffer are not needed and therefore not used. Visualisation of the six groups, two house transaction periods, pre and post opening of a sport facility, times three spatial groups, within 350-meter buffer, within 800-meter buffer and outside both buffers but not further away than 3km, is illustrated in figure 11. The figure is only one example, this process of identifying groups is repeated for each sport facility. Table 6 on page 28 provides an exemplified overview of the possible outcomes or groups in table format.

This paper hypothesizes that house prices within 0- to 350-meters decrease due to sport facility openings while within 350- to 800-meters house prices will increase. These two buffers were chosen deliberately. The 350-meters buffer captures the relatively close effects of sport facility openings. Sport facilities are often bigger than only a point on the map, so taking the scale of the sport facility into account this buffer contains the houses that are next to or almost next to a sport facility opening. These first and second row houses can absorb most of the negative, or

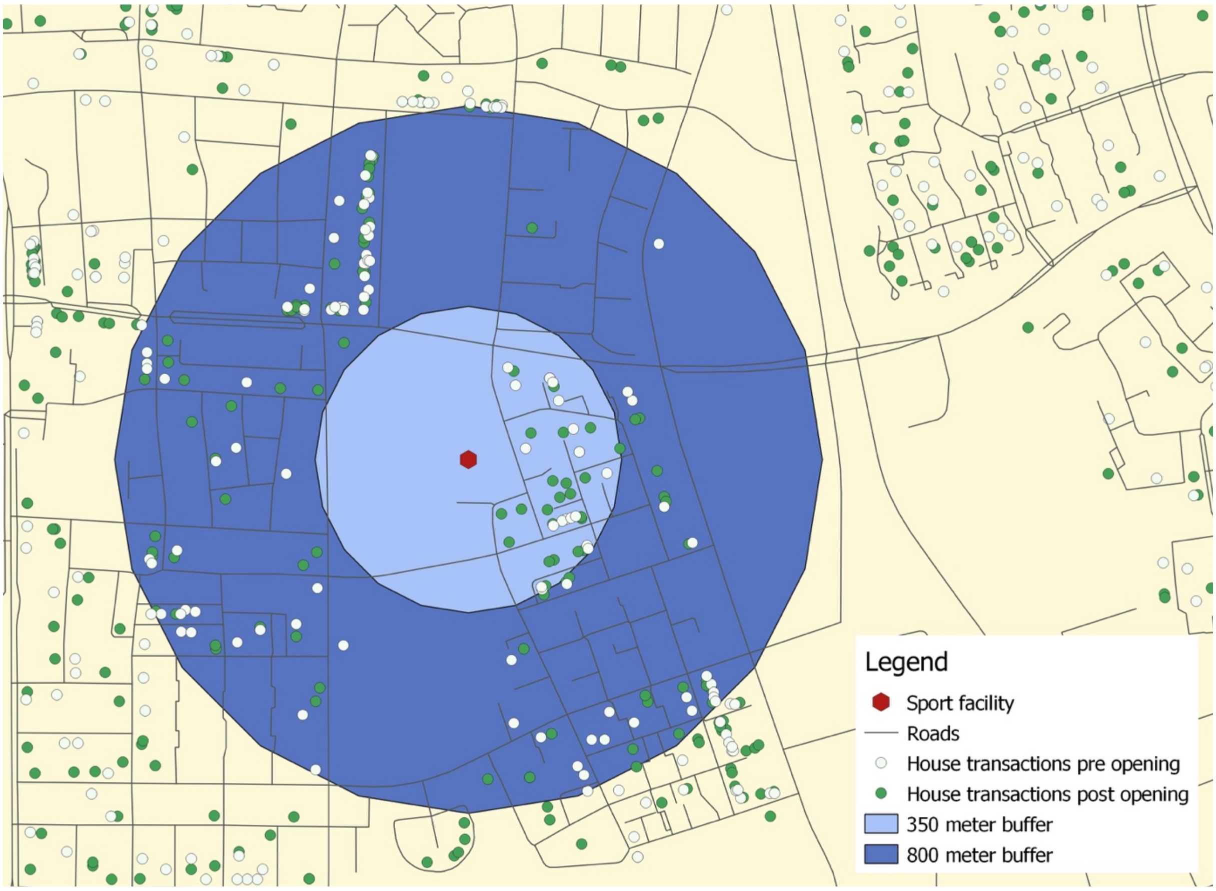


Figure 11: Map of representative sport facility opening and house transactions.

positive, effects of an amenity (Portnov, Genkin & Barzilay, 2009). Therefore, the 350-meter buffer is chosen to account for these strong local effects.

With most of the effects absorbed by the first and second row houses, in this case hypothesized negative effect, the remaining effects of proximity and accessibility can be identified. The effects are captured by the house transactions in the 350- to 800-meter buffer. With the speed of 5km/h one can walk 800 meters in a straight line. Houses in this buffer are a bit further away from a sport facility but still quite close. The sport facility is on the other side of the neighbourhood or in the adjacent neighbourhood, but not much further. The 350- to 800-meter buffer therefore accounts for the accessibility and proximity effects of a sport facility in the same or adjacent neighbourhood.

Effects outside this buffer are hypothesized to be neglectable. Houses outside this buffer therefore acts as the control group in the difference-in-difference analyses. A 5,000-meter buffer is however set to disregard the house transactions that took place on the other side of the province. These transactions have nothing to do with a sport facility half a province away. So, only house transactions in the same or neighbouring settlements within 5,000 meters are taken into the analyses.

This paper uses three difference-in-difference methods to analyse the data and to test the hypotheses stated previously, a difference-in-difference analyses for each year of sport facility openings, a house difference-in-difference analysis including all years with price corrected by the quarterly trend as discussed in the data section and a sport facility based difference-in-difference analysis including all years with price corrected by the quarterly trend. The three methods built upon each other. The first method is a regular difference-in-difference model for one year that a sport facility is opened. This model is then be repeated for each year the dataset has. The second model is an extension of the first model. It is almost the same difference-in-difference model but not just for one year but for all years combined. A side effect specific for this model is that each datapoint is weighted by its frequency in the dataset and that the data has to be corrected by the quarterly trend. The third model also uses a difference-in-difference method over multiple years but with a different approach. In this method for each sport facility that is opened the four groups[2] necessary for the difference-in-difference are created as if there were a house. They are created by taking the average of all house variables that fall in each group. So, for each of the four groups per sport facility one "house" is left that represents all houses in that group. An advantage of this model compared to the previous one is that this model can include sport facility characteristic. This makes it possible to test the team sport and outdoor sport hypotheses. The various analyses will be discussed more extensively one by one.

Base model

But first, several base models are constructed. The first base model is a linear regression model with the logarithm of transaction price as dependent variable and several control variables as independent variables. This model was subsequently extended with more control variables and the model was later duplicated with the logarithm of transaction price corrected by the trend index. Eventually, the main base model is a linear regression model with the logarithm of the trend corrected transaction price ($lnCP$) as dependent variable, a vector of multiple house characteristics as house structural controls ($\boldsymbol{S}$), a vector of locational controls ($\boldsymbol{L}$), a vector of

[2] Pre, post opening and in range, not in range of a sport facility opening

year controls ($\boldsymbol{Y}$) and a vector of municipal controls ($\boldsymbol{M}$) as independent variables. The model looks as follows:

$$lnCP = a + \boldsymbol{\beta}_1\boldsymbol{S}_1 + \gamma_1\boldsymbol{L}_1 + \boldsymbol{\kappa}_1\boldsymbol{Y}_1 + \boldsymbol{\lambda}_1\boldsymbol{M}_1 + \varepsilon$$

where a is a constant, β, γ, κ and λ are the coefficients of the independent variables and ε represents the unobserved characteristics.

The control variables included in the model explain the main determinants of the transaction price of a house, but do not have to describe the transaction price completely. The structural control variables in the model are partially based on Kain Quigley (1970), Sirmans et al. (2005) and other related literature. They identify the structural characteristics of a house as essential in hedonic pricing models, with square meters of living space, the type of house and the number of rooms as most researched attributes. The locational controls are house specific characteristics that refer to the location of the house. These are important to include as the locational characteristics can have a significant impact on the transaction price. The broader locational factors are captured in the municipal controls. The urbanisation level, population density or municipal specific laws are heterogeneous throughout the province of Gelderland. They impact and affect house prices in different ways in the different municipalities. The municipal controls take the municipal heterogeneity into account. The year controls take into account the various years in which the transactions took place, as yearly fluctuations and trends or news housing laws affect house prices each year differently.

Difference-in-difference per year

This method is the basis for analysing the data with difference-in-difference in this paper and builds upon the prior basic hedonic model. In this method only sport facilities opened in one and the same year are used. This method is simple as it analyses over two time periods, as is normal for basic difference-in-difference models. One time period is the pre-opening of a sport facility, the other the post opening period. Due to limitations in the dataset and the software used for the analyses this paper chooses to use house transactions one year prior to opening as the pre group and house transactions one year after opening as post group. For example, if sport facility openings in 2010 are chosen as year of analysis the pre group are house transactions from the year 2009 and the post group are house transactions from the year 2011. Using a Python script, the house transactions are grouped by a dummy in the six categories described earlier. As mentioned before, house transactions outside the maximum buffer of 5,000 meters are not relevant and automatically dropped by the script. An example of the categorisation with sport facility opening in the year 2010 is provided in table 6.

Period		**In 350m buffer**		**In 800m buffer**	
In words	**Numeric**	**In words**	**Numeric**	**In words**	**Numeric**
2009	0	No	0	No	0
2009	0	No	0	Yes	1
2009	0	Yes	1	No	0
2011	1	No	0	No	0
2011	1	No	0	Yes	1
2011	1	Yes	1	No	0

Table 6: Example of categorisation of the data by the Python script.

Now that the house transactions are placed in one of the categories analyses in Stata is possible. The following model will be used with the logarithm of the corrected price as dependent variable ($lnCP$), a vector of house structural controls ($\boldsymbol{S}$), a vector of locational controls ($\boldsymbol{L}$), a vector of year controls ($\boldsymbol{Y}$) and a vector of municipal controls ($\boldsymbol{M}$) as independent variables:

$$lnCP = a + \boldsymbol{\beta}_1\boldsymbol{S}_1 + \gamma_1\boldsymbol{L}_1 + \boldsymbol{\kappa}_1\boldsymbol{Y}_1 + \boldsymbol{\lambda}_1\boldsymbol{M}_1 + \mu_1 dB + \tau_1 dP + \chi_1 dB * dP + \varepsilon$$

where a is a constant, β, γ, κ, λ, μ, τ, χ are the coefficients of the independent variables, dB is the dummy variable for the treatment and control group, dP is the dummy variable for the period groups and ε represents the unobserved characteristics.

In this method only sport facility openings in a single year can be analysed. It is possible to perform the analyses repeated for each year, but it is preferred to have all years together in one analysis. Therefore, a second method is proposed.

Difference-in-difference for all years, house based

This method is built upon the previous method. In this method the data used in the analyses of each year individually is put together in one dataset. So, the data contains house transactions divided in the six groups discussed previously for each year of sport facility openings. House transactions from 2008 and 2010 for 2009 as sport facility opening year, house transactions from 2009 and 2011 for 2010 as opening year, house transactions 2010 and 2012 for 2011 as opening year and so forth, all in one dataset. The advantage of this method is that now the complete data from all the years between 2008 and 2017 can be examined in a single analysis. However, using this method, some datapoints can appear two times in the dataset. For example, a house transaction in 2010 can be in the post and in the 800-5000-meter buffer group for a sport facility opened in 2009 but can also be in the pre-group and in the 350-800-meter buffer group for a sport facility opened in 2011. To resolve this problem the analyses will be weighted by the number of times a house transaction appears in the dataset. So, a house transaction that only appears once will have a weight while a transaction that appears two time will have a weight corrected by their frequency. Hence, each house transactions will have a weight of one in total. Also, as discussed in the data section the average house value of a house transaction in the province of Gelderland fluctuated over the period 2008-2017. Therefore, the house transactions prices are multiplied by the average quarterly trend of the province. This corrected price is subsequently used as dependent variable in the analysis. The fluctuations in the house transactions over time are now accounted for. Altogether the following model will be used with the logarithm of the corrected price as dependent variable ($lnCP$), a vector of house structural controls ($\boldsymbol{S}$), a vector of locational controls ($\boldsymbol{L}$), a vector of year controls ($\boldsymbol{Y}$) and a vector of municipal controls ($\boldsymbol{M}$) as independent variables and weighted by frequency:

$$lnCP = a + \boldsymbol{\beta}_1 \boldsymbol{S}_1 + \gamma_1 \boldsymbol{L}_1 + \boldsymbol{\kappa}_1 \boldsymbol{Y}_1 + \boldsymbol{\lambda}_1 \boldsymbol{M}_1 + \mu_1 dB + \tau_1 dP + \chi_1 dB * dP + \varepsilon$$

(weighted by frequency)

where a is a constant, β γ, κ, λ, μ, τ and χ are the coefficients of the independent variables, dB is the dummy variable for the treatment and control group, dP is the dummy variable for the period groups and ε represents the unobserved characteristics.

This method is house transaction based. It is from the perspective of the house transactions. Each transaction is categorised in one of the six groups, and sometimes in two periods. To check the robustness of the results and the third and fourth hypotheses a third analysis with a slightly different method is used.

Difference-in-difference for all years, sport facility based

This method builds upon the previous methods but is sport facility based. It is from the perspective of sport facilities. For each sport facility opening four groups are created, a year before and in buffer of 0-350 meter of opening, a year before and in the 350-800-meter buffer, a year after opening and in the in buffer of 0-350 meter of opening and a year after and in the 350-800-meter buffer. The groups are created by taking the average value of house transactions and control variables for each group, using a script. So, for example, one year before opening and in the 0-350-meter buffer of sport facility X are 100 house transactions. All their variable values are summed and divided by 100. This leaves one single representative average house transactions for this group for this sport facility opening. This is also done for the other 3 groups that belong to this sport facility and subsequently repeated for all sport facility opening. Consequently, the same difference-in-difference method as aforementioned can be applied but with the additional advantage that the control variables of the sport facilities can also be included in the analysis. Therefore, it is possible to distinguish the effects for the type of sport facility and the type of sport, team sport or individual sport. Just as in the previous method the trend corrected price is used as independent variable. Altogether the following model will be used with the logarithm of the average corrected price as dependent variable ($lnACP$), a vector of average house structural controls ($\overline{\boldsymbol{S}}$), a vector of average house locational controls ($\overline{\boldsymbol{L}}$), a vector of sport facility year controls ($\boldsymbol{J}$), a vector of sport facility municipal controls ($\boldsymbol{N}$), team sport (T) city limits (C) and outdoor sport (O) as independent variables:

$$ln\overline{CP} = a + \boldsymbol{\beta}_1\overline{\boldsymbol{S}}_1 + \gamma_1\overline{\boldsymbol{L}}_1 + \boldsymbol{\nu}_1\boldsymbol{J}_1 + \boldsymbol{\rho}_1\boldsymbol{N}_1 + \zeta_1 T_1 + \xi_1 C_1 + \vartheta_1 O_1 + \mu_1 dB + \tau_1 dP + \chi_1 dB * dP + \varepsilon$$

where a is a constant, β, γ, ν, ρ, ζ, ξ, ϑ, μ, τ, and χ are the coefficients of the independent variables, dB is the dummy variable for the treatment and control group, dP is the dummy variable for the period groups and ε represents the unobserved characteristics. The results of three models and the base model are discussed in the next section.

V. Results

Base model

A base model is composed and tested for its reliability and explanatory power. The base model is constructed based on the main variables used by the Sirmans et al. (2005), Chau, & Chin (2003), Malpezzi (2002) and other related literature. These are modelled to the Dutch housing market situation by among others Kagie and van Wezel (2006) and Debrezion, Pels and Rietveld (2011) and are useful in their current form in this research. A basic base model was first constructed with a selection of the most important variables that explain house prices. This base model was run with the logarithm of price and again with the logarithm of the corrected price as dependent variable. The results are displayed in a summary[3] output in table 7.

The basic base model was extended with the pragmatic use of available variables received with the dataset and the other variables of less importance mentioned in the literature. The r-squared of the models ranges from 0,751 to 0,817 with the extended model with the logarithm of the trend corrected price yielding the highest value. In this model 81,6% of the variance of the logarithm of the corrected price is explained by the independent variables. The model therefore has a reasonably high explanatory power. The main variables m^2, ceiling height and plot size all have a positive significant effect on house prices, mirroring the result in the literature. The model therefore seems to be reliable. The housing market in the province of Gelderland seems to reflect the housing market illustrated earlier in the literature.

The control variables yield mostly expected outcomes. The construction period controls show a typical vintage effect specific to the Dutch housing market. Pre Second World War period houses are more favoured than post Second World War period houses. The year controls show negative significant results, although the house prices are corrected with the quarterly trend. This means that the impact of the yearly trend is not correctly captured by the quarterly trend.

[3] A full regression output of the basic base model and extended base model is displayed in the appendix

Variables	(Basic) Log price	(Basic) Log corr. price	(Extended) Log price	(Extended) Log corr. price
M2	0.00612***	0.00613***	0.0115***	0.0115***
	(1.66e-05)	(1.67e-05)	(6.27e-05)	(6.32e-05)
Ceiling height	0.191***	0.191***	0.323***	0.319***
	(0.00163)	(0.00164)	(0.0147)	(0.0149)
Plot size			2.29e-05***	2.29e-05***
			(3.40e-07)	(3.42e-07)
Main variables squared	Yes	Yes	Yes	Yes
Dwelling type	No	No	Yes	Yes
Structural controls	No	No	Yes	Yes
Maintenance controls	Yes	Yes	Yes	Yes
Locational controls	No	No	Yes	Yes
Year controls	Yes	Yes	Yes	Yes
Municipal controls	Yes	Yes	Yes	Yes
Constant	10.64***	10.86***	10.23***	10.46***
	(0.0142)	(0.0143)	(0.0278)	(0.0280)
Observations	143,715	143,715	143,715	143,715
R-squared	0.751	0.766	0.806	0.817

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Summary regression output of the basic model and extended basic model.

Per year

Table 8 shows the regression output of the difference-in-difference analyses for each year sport faculties were opened. The in 800m buffer and in 350m buffer variables show the effect of the transactions being in the range of a sport facility regardless of if it is opened yet or not. They yield no significant results other than in 2011 and 2016. The period variable shows whether house prices are different for the year before opening or the year after opening regardless of the house being in the range of a sport facility or not.

The period variables are all significant. They are negative for the first years until 2013. From 2014 and on the values are positive. This means that in the first years the houses after opening were less valuable and in the last years the houses were more valuable in the year after a sport facility opening. This mirrors the fluctuations in house index trend over the year 2008-2017. In the beginning of this period the house prices declined due to the financial crisis. The house transactions are corrected by the quarterly house index trend, which is an average of all the house transactions in Gelderland in a certain quarter. However, a yearly trend is still visible in the results. The dataset may not be representative for the province as a whole. Possibly the trend in the dataset is stronger than the average over the whole province. The dataset contains mostly urban areas and less rural areas. The effects of the economic crisis could have been stronger in the cities than in rural areas. In the last years house prices increased again due to the revival of the economy after the crisis.

The DD variables show the difference-in-difference outcomes. The results are quite heterogeneous through the years. For the 800-meter buffer the results vary over the years, some years positive, other years negative. The impact of all coefficients is low. Also, all coefficients are not significant. So, it is not possible to say with enough certainty that the outcomes differ from zero. For the 350-metere buffer the results are even more heterogeneous. In the first year the result is negative and significant. The years thereafter range from insignificant and negative to insignificant positive. The result for the last year is positive and significant, making that both significant results are contrary. Thus, the results do not give a conclusive answer.

The r-squared of the model ranges from 0,811 to 0,840. This means that for each model more than 80% of the variance of the logarithm of the corrected price is explained by the independent variables. Other regression without the DD variables were run and in these regressions the r-squared yields similar results. This means that the explained variance which is added by the difference-in-difference variables is very limited. Therefore, those variables do not add much in explaining house prices in this model.

Variables	(2009) Log corr. price	(2010) Log corr. price	(2011) Log corr. price	(2012) Log corr. price	(2013) Log corr. price	(2014) Log corr. price	(2015) Log corr. price	(2016) Log corr. price
In 800m buffer	-0.0102	-0.0100	0.0186*	0.0134	0.00357	-0.00692	-0.0205	-0.0363***
	(0.00718)	(0.00922)	(0.0111)	(0.0140)	(0.0112)	(0.0190)	(0.0131)	(0.0118)
In 350m buffer	-0.0208	-0.00413	0.0179	-0.0228	0.00333	-0.0452	-0.00182	-0.0812***
	(0.0152)	(0.0136)	(0.0200)	(0.0229)	(0.0232)	(0.0481)	(0.0213)	(0.0193)
Period	-0.109***	-0.0962***	-0.213***	-0.257***	-0.0870***	0.0407***	0.101***	0.203***
	(0.00281)	(0.00473)	(0.00575)	(0.00627)	(0.00432)	(0.00700)	(0.00475)	(0.00474)
DD 800m	-0.00452	0.0104	-0.00360	-0.0180	-0.000460	0.0148	0.0114	0.0127
	(0.0112)	(0.0128)	(0.0152)	(0.0184)	(0.0145)	(0.0233)	(0.0159)	(0.0151)
DD 350m	-0.0351*	-0.0194	0.0150	0.0124	0.0133	0.0512	0.00725	0.0563**
	(0.0195)	(0.0233)	(0.0286)	(0.0323)	(0.0280)	(0.0615)	(0.0269)	(0.0280)
Constant	10.75***	11.12***	10.79***	11.32***	10.77***	10.43***	10.48***	9.762***
	(0.123)	(0.171)	(0.200)	(0.139)	(0.148)	(0.218)	(0.190)	(0.145)
Structural controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipal FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Robust	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	15,840	6,106	4,635	6,450	8,182	2,466	5,837	6,250
R-squared	0.836	0.818	0.840	0.824	0.811	0.814	0.821	0.824

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

Table 8: Regression output of the difference-in-difference analyses for each year.

House based

The results of the house based difference-in-difference analysis with all years combined in one dataset is displayed in table 9. The analysis with all controls included yields the highest r-squared with a value of 0,825. The 800m buffer and 350m buffer variable coefficients are very low negatives and are insignificant, meaning that house prices within the buffer of a sport facility regardless of the sport facility is built yet or not do not differ from house prices outside this buffer with enough certainty. The period coefficient has a low negative value and is significant. This means that house prices regardless of if the houses are near a sport facility or not are lower in the year after sport facility opening than the year before opening. As in the previous results section on each year individually in this output also the adverse effect of the economic recession is visible. Although its impact over the data as a whole is mitigated by the revival of the economy in later years. Both the 800m and 350m DD results are positive but insignificant. This indicates that sport facility openings do not have an effect on house prices within a 0-350 meter nor 350-800 meter range that with enough certainty differs from zero.

Variables	(1) Log corr. price	(2) Log corr. price	(3) Log corr. price	(4) Log corr. price
In 800m buffer	-0.0140***	-0.0115**	-0.00378	-0.00188
	(0.00529)	(0.00527)	(0.00471)	(0.00382)
In 350m buffer	-0.0184*	-0.0130	0.00155	-0.00434
	(0.00999)	(0.0101)	(0.00874)	(0.00696)
Period	-0.0708***	-0.0705***	-0.0704***	-0.0114***
	(0.00213)	(0.00211)	(0.00191)	(0.00203)
DD 800m	0.00429	0.00503	0.00599	0.00311
	(0.00724)	(0.00723)	(0.00642)	(0.00535)
DD 350m	0.00719	0.00945	-0.00462	0.000102
	(0.0138)	(0.0138)	(0.0120)	(0.00980)
Constant	11.78***	11.76***	11.52***	11.38***
	(0.0873)	(0.0912)	(0.0740)	(0.0705)
Structural controls	Yes	Yes	Yes	Yes
Location controls	No	Yes	Yes	Yes
Municipal FE	No	No	Yes	Yes
Year FE	No	No	No	Yes
Robust	Yes	Yes	Yes	Yes
Observations	55,766	55,766	55,766	55,766
R-squared	0.684	0.688	0.748	0.825

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

Table 9: Regression output of the house-based difference-in-difference analysis.

Sport facility based

The regression output of the sport facility based difference-in-difference analyses for both the 800m buffer and 350m buffer is displayed in table 10. The results for the 800m buffer analysis with all controls variables show that the in-buffer variable is negative and not significant, and the period variable is negative and significant, the same results as in the previous regressions. The difference-in-difference coefficient has a very low negative value and is insignificant. The model explains more than 83% of the variance of the logarithm of the corrected price with a value of 0,829 for the r-squared.

Results for the 350m buffer show that both the period variable and in buffer variable are negative and significant, meaning that houses post opening and houses in a 350m range are less valued than the pre-opening and outside range houses. The DD variable is however insignificant, so houses that are both post opening and in the buffer do not differ with enough certainty from the control group. The model explains a lot of variance of the dependent variable with an r-squared of 0,869. For both the 800m buffer and 350m buffer regressions the DD results are insignificant. These results suggest that there is a non-trivial possibility that sports facilities and house prices are unrelated in this model.

The team sport coefficient is significant in the regressions without year fixed effect. However, with year fixed effects included in the regression the results become very insignificant with a p-values nearing 1. These results do not differ much in a regression without difference-in-difference variables. This implies that team sport facilities do not have a different effect on house prices than non-team sport facilities. This is not unexpected knowing that the results suggest that sport facilities do not have an effect on house prices at all.

The outdoor sport facility variable for both the 800m buffer and 350m buffer regressions with all control variables is positive and insignificant. This implies that outdoor sport facilities do not have a different effect on house prices than indoor sport facilities. These results are also not surprising due to the insignificance of an effect of sport facilities on house prices in general.

Similar regressions without difference-in-difference variables yield comparable results, meaning that the explained variance added by the thoses variables is very limited. Therefore, those variables do not add much in explaining house prices, as in previous analyses.

	800m buffer				350m buffer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	Log corr. price	Log corr. price	Log corr. price	Log corr. price	Log corr. price	Log corr. price	Log corr. price	Log corr. price
Team sport	-0.0420***	-0.0477***	-0.0828***	-0.00127	-0.115***	-0.115***	-0.154***	-0.0171
	(0.0147)	(0.0162)	(0.0215)	(0.0198)	(0.0254)	(0.0267)	(0.0346)	(0.0368)
Outdoor	-0.000978	0.0944***	0.181***	0.0610	-0.0232	0.0111	0.143	0.0457
	(0.0151)	(0.0302)	(0.0463)	(0.0542)	(0.0214)	(0.0646)	(0.0936)	(0.0875)
Period	-0.0453***	-0.0465***	-0.0544***	-0.0907***	-0.0732***	-0.0719***	-0.0826***	-0.107***
	(0.0110)	(0.0113)	(0.0142)	(0.0138)	(0.0186)	(0.0189)	(0.0219)	(0.0181)
In buffer	-0.0375	-0.0354	-0.0272	-0.0246	-0.0400	-0.0400	-0.0524	-0.0618**
	(0.0245)	(0.0244)	(0.0212)	(0.0196)	(0.0378)	(0.0376)	(0.0338)	(0.0263)
DD	-0.00891	-0.00882	-0.00668	-0.00319	0.0406	0.0405	0.0401	0.0362
	(0.0284)	(0.0282)	(0.0257)	(0.0224)	(0.0395)	(0.0398)	(0.0369)	(0.0299)
Constant	10.60***	10.32***	8.291***	3.931*	9.234***	9.247***	8.958***	8.949***
	(2.894)	(2.846)	(2.625)	(2.366)	(1.351)	(1.393)	(1.397)	(1.024)
Structural controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Municipal FE	No	No	Yes	Yes	No	No	Yes	Yes
Year FE	No	No	No	Yes	No	No	No	Yes
Robust	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	400	400	400	400	268	268	268	268
R-squared	0.699	0.703	0.778	0.829	0.728	0.729	0.800	0.869

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

Table 10: Regression output of the sport facility based difference-in-difference analyses for both the 800m buffer and 350m buffer.

VI. Discussion

Discussion of the regression results

As discussed in the literature section previous studies are suggestive of a positive relationship between sport facilities and house prices (Nichols & Crompon, 2007; Sirmans et al., 2005; Feng & Humpreys, 2012). This resulted in the first hypothesis:

> *H1: The introduction of a sport facility within a range of 800 meters has a positive impact on nearby house prices.*

The output of the per year analyses shows insignificant results. The sign of the results ranges from positive to negative, without a distinct pattern. The house based result shows a small positive effect, suggesting that house prices are affected positively by sport facilities. However, the coefficient is insignificant. The sport facility based analysis shows a small negative insignificant effect of sport facilities on house prices. Although most coefficients are not significant, the results suggest that the coefficients that were found are the most probable value. But the values around the coefficient, including zero, are also possible true values. These results do not allow us to reject that sports facilities and house prices are unrelated. The estimates change in sign and magnitude between specification, suggesting at the very least heterogeneity in results, but more likely this is suggestive of spurious relationships between house prices and sports facility openings within an 800-meter buffer. The results do not support hypothesis 1, therefore it is rejected. Sport facilities facilitate health improvements, social cohesion, leisure and pleasure. They might create positive neighbourhood effects, prevent crime and promote social integration, as the literature suggests. However, these effects are not visible in house prices in this study. The discrepancy between these results and the literature can be explained by the sample of data analysed. Feng and Humpreys (2012) and other related literature investigated mostly the effect of professional sport facilities on house prices, while this study focusses on amateur sport facilities. Sirmans et al. (2005), Nichols and Crompon (2007) and related studies did research on amateur sport facilities, but narrowed down to only golf courses, while this study includes varies types of sport facilities.

Based on the Tse and Love (2000) and Hummel (2018) as discussed in the literature section, a negative impact of sport facilities on house prices was hypothesised:

H2: The introduction of a sport facility in a range of 350 meters has a negative impact on nearby house prices.

The results of the per year analyses show mostly insignificant results. The result that do show significant effects are contradictory. The coefficients range between small negative and small positive values. The house based result shows a small positive but insignificant sign, indicating that house prices are affected positively by sport facilities. The sport facility based analysis shows a small positive insignificant effect of sport facilities on house prices. These results suggest that there is a non-trivial possibility that sports facilities and house prices are unrelated. The heterogeneity in the results is likely suggestive of spurious relationships between house prices and sports facility openings within a 350-meter buffer. The results do not support hypothesis 2, therefore it is rejected. Both negative spill over effects like noise, light pollution, traffic increase and positive spill over effects like the previously mentioned, proximity and accessibility are not accounted for in house prices. So, the housing market does not take those effects into account in house transactions. Another possibility is that both negative and positive effects cancel each other out. Nonetheless, dwellers do apparently not value the presence of sport facilities in their consideration when buying a house. A reason why both the house based and sport facility based regressions find different outcomes might be that the sport facility based regression does not account for morphology of cities, while the house based regression includes municipal and distance to city centre controls. This might also explain the difference in some of the results between this study and Hummel (2018) and Tse and Love (2000).

As discussed earlier in the literature section previous literature suggested that outdoor sport facilities have a stronger effect on house prices than indoor sport facilities (Hummel, 2018). Therefore, hypothesis 3 was proposed:

H3: Outdoor sport facilities have a stronger effect on house prices than indoor sport facilities.

The results for hypothesis 3 show that both for an 800-meter and 350-meter buffer the outdoor sport facilities coefficient is positive. These indicate that outdoor sport facilities have a more positive effect on house prices than indoor sport facilities. This is in line with the results found by Hummel (2018), who concluded that indoor sport facilities are negatively associated with house prices while outdoor sport facilities were positively associated with house prices. The results are however not significant, so it is not possible to say the true value of the result is not zero. Therefore, hypothesis 3 was rejected.

Hummel (2008) suggested that team sport facilities have a stronger effect on house prices than individual sport facilities, as was discussed in the literature section. Therefore, hypothesis 4 was proposed:

> *H4: Team sport facilities have a stronger effect on house prices than individual sport facilities.*

The results show for both the 800-meter and 350-meter buffer that team sport facilities have a negative impact on house prices, indicating that team sport facilities do have a larger impact on house prices and that this impact is negative. The type of sport facility does not have any significant influence on the effect of a sport facility opening on house prices, however. Although team sport facilities had a negative effect in the regressions without all controls, team sport facilities or individual sport facilities do not differ significantly in their effect in the analysis with all controls and fixed effect. These results suggest that there is a non-trivial possibility that team sport facilities have a stronger effect on house prices than individual sport facilities. But the uncertainty if the results truly differ from zero is too high. Hypothesis 4 is therefore rejected. Similar insignificant results were also found by Hummel (2018), although that study used associations.

The results seem plausible. The models and their regression results mirror those in the literature. The housing market in the province of Gelderland as in the dataset seems to reflect the housing market illustrated the literature. Also, the r-squared of all models is over 0,80. The models therefore seems to be reliable and with enough explanatory power.

Encountered problems

A possible reason why this study finds different results than other papers could be that several limitations specific for this study are encountered.

Firstly, the urban morphology could still be a problem in this study. Although it includes a municipal dummy that partially controls for the level of urbanisation and for more robustness also included a variable measuring the distance to the nearest (big) city centre to control for centrality, the real centrality might still not be correctly controlled for. The negative, but insignificant, results that appear in the regressions might possibly be the consequence of the urban morphology. Sport facilities have a tendency to locate at the edges of cities. There is the most space for sport facilities and the land prices there are low. Houses near sport facilities are in the same low land price area at the edge of the city. Houses prices here are therefore mostly lower than in the city centre. So, in general houses that are further away, in the city centre, are more valuable than houses nearby sport facilities. In this way, the relationship between sport facilities and house prices becomes negative trough space. This makes it difficult to capture a, as this paper expected, positive effect of sport facilities on house prices. The same problem is encountered by Hummel (2018). He also identified negative results due to the spatial morphology of sport facilities locating near city edges where house prices are relative cheap.

Secondly, this study encounters a negative trend in the period coefficents in all regression results. The house transactions are corrected by the quarterly house index trend. Nonetheless a yearly trend is still visible in the results. This may indicate that the dataset is not representative for the whole province of Gelderland. The dataset contains mostly urban areas and less rural areas. The effects of the economic crisis could have been stronger in the cities than in rural areas, making the trend in the dataset possible stronger than the average over the whole province.

Thirdly, the effect of marginality is not included in the analyses. This paper does not account for if houses are already near a lot of sport facilities. An additional one would impact more when it is the first in the proximity or the tenth. The Netherlands is already highly concentrated with sport facilities. In other countries where this is not the case a sport facility opening can have a different effect on house prices.

Lastly, this study does not include multiple years and does not investigate a post public announcement phase or post construction phase. A reason why no significant results were found in this study could be that analyses with multiple years before and after opening could lead to different results. This study only includes house transactions one year before and one year after the opening of a sport facility in its analyses. It can be the case that the effects of a new sport facility are only measurable after a few years, not the first year after opening. House prices fluctuate more the longer the time period, so when more years before and after opening are included the longer the effects have had time to influence prices. Then the difference-in-difference outcomes are arguably greater. In this way conclusions can be made with a lot more certainty. The post public announcement phase or post construction phase were also not investigated in this study. The opening of a sport facility can already be anticipated on in the transaction value well before the actual opening. The announcement that a sport facility will be opened can already have an effect. When a sport facility actually opened this effect can be less strong, as it is already accounted for. In line with this, the construction period can also already spill over the anticipated effects of a new sport facility.

Further analyses

The results this paper finds are not homogenous. They range in several directions. The house based analyses found mostly insignificant positive difference-in-difference results, while the sport based analysis finds insignificant negatives. The results are spurious and likely to be related to other non-observed variables. The exact nature between all variables is not clear. Several problems and limitations specific to this study are encountered that might explain the heterogeneity in the results. The urban morphology, negative trend, marginality and multiple years analyses, including announcement and construction phase, remain troublesome. Therefore, this paper carries out further analyses to unravel the relationship between sport facilities, house prices and unobserved factors and overcome most of these limitations. The first step this study takes is to analyse the residual of one of the regression models.

Residuals

The residuals of the house based regression were predicted and saved in the dataset. Thereafter, the residuals were plotted in QGIS. The distribution of the residuals is displayed in figure 12. The residuals show the variance between the real value of the data and the predicted values of the regression model. Negative values, white and light red coloured in figure 12, are residuals that are overestimated. This means that the real effect of sport facilities on these house prices was less than the model predicted. The positive, darker red values are underestimated by the regression model. This means that the real effect of sport facilities is greater than the model predicted. As visible in figure 12 the positive values in general seem to be more clustered in urban areas while in rural areas the residuals seem to be in general mostly negative. This may indicate that the effect of sport facilities is greater in urban areas than in rural areas. A further analysis will take a look at this observation later.

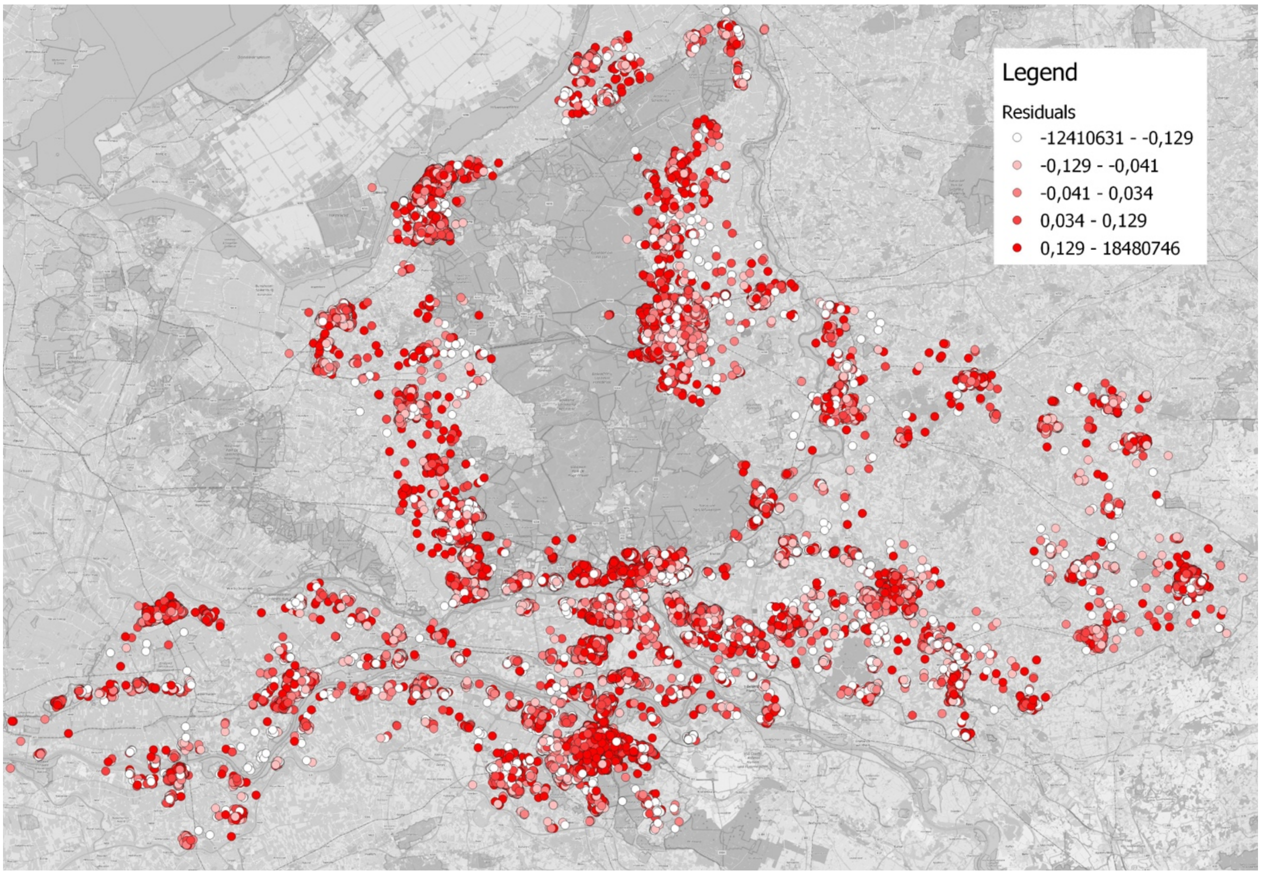


Figure 12: Distribution of the residuals of the house based regression.

The three biggest cities of Gelderland, Apeldoorn, Arnhem and Nijmegen have the highest amount of house transactions clustered together. Therefore, the residual distribution plot is zoomed in on each of these cities separately. The distribution of the residuals of these cities are displayed in figures 3, 5 and 7 in the appendix. The residual plot of Apeldoorn shows a somewhat equal distribution of the residuals through the city. However, in the north-west edge of the city the residuals are mostly red and therefore positive. The houses in this area border the forests of the Veluwe, the Apenheul Primate Park and the Palace Park Het Loo. Houses in this neighbourhood area are expensive and this neighbourhood is known as somewhat of a rich neighbourhood. This may indicate that the effect of sport facilities is dependent on the values of the house and therefore indirectly the income level of the residents. To further investigate this presumption, the income per neighbourhood was plotted with the residuals (figure 4 in the appendix). This shows indeed that in high income neighbourhoods house prices are in general underestimated. The same procedure was repeated for the city of Arnhem and the city of Nijmegen and here the same pattern as in Apeldoorn is found with income plotted together with the residuals (figure 5, 6, 7 and 8 in the appendix). In other smaller cities, this pattern is less evident.

In general, the pattern shows that for houses with a low value, sport facility openings are overestimated while for houses with a high value, the effect of a new sport facility is underestimated. An expensive house requires a high income. So, in other words, people with a high income value a new sport facility more than people with a low income. People with a high income do in general have a higher education than people with a low income, so it may also be possible that the effect of a new sport facility differs for people with a high education relative to people with a low education. The result thus may indicate that in general people that are highly educated with a high income value sport facilities more than people with a lower education and lower income. This seems plausible, as people with a high income and high education participate more in sports than lower educated people (Dishman, Sallis & Orenstein, 1985; Downward & Rasciute, 2010). The introduction a new sport facility may affect high educated people more as they are more interested, while les educated people are less interested in a new sport facility as they do less participate in sports overall.

Neighbourhood fixed effect analyses.

To account for the income effect that differs per neighbourhood and to tackle the morphology problem a neighbourhood fixed effect is added into the house based regression. This neighbourhood fixed effect can explain parts of the morphology problem as control on neighbourhood level is the smallest spatial scale possible. The results are displayed in table 11. The 800m difference-in-difference coefficient lowers a bit in the neighbourhood FE model relative to the model that does not control for it. However, the coefficient is still insignificant. The value of the 350m difference-in-difference coefficient is higher in the with neighbourhood FE model but also still insignificant. However, the model seems to better predict the effects of a sport facility opening on house prices. The neighbourhood fixed effects improve the explanatory power of the model considerably, as the r-squared rises to 0,865.

Variables	(Without neighbourhood FE) Log corr. price	(With neighbourhood FE) Log corr. price
In 800m buffer	-0.00189	-0.00447
	(0.00382)	(0.00371)
In 350m buffer	-0.00435	-0.00656
	(0.00696)	(0.00670)
Period	-0.0114***	-0.00947***
	(0.00203)	(0.00184)
DD 800m	0.00311	0.00216
	(0.00535)	(0.00484)
DD 350m	9.30e-05	0.00491
	(0.00980)	(0.00890)
Constant	11.15***	11.34***
	(0.0706)	(0.0806)
House structure controls	Yes	Yes
Location controls	Yes	Yes
Municipal FE	Yes	Yes
Year FE	Yes	Yes
Neighbourhood FE	No	Yes
Robust	Yes	Yes
Observations	55,751	55,751
R-squared	0.825	0.865

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

Table 11: Regression output of the analyses with and without neighbourhood fixed effects.

Subset analyses

The morphology is arguably well controlled for in this model with municipal, neighbourhood and distance to city centre fixed effects. The sport based regression model finds negative results but is does not account for the spatial morphology. The morphology is probably the explaining factor for the difference in results. This study find is overall positive a positive effect of sport faculties on house prices, although insignificant. Hummel (2018) found negative results. The different outcomes both studies may be the consequence of different datasets. This study examines a whole province, with both urban and rural areas, while Hummel (2018) examines only cities, Amsterdam and Rotterdam. It is possible that the results may be different in each geographical area. Therefore, a subset analyses was run to test this presumption. Also, this analysis can clarify the problem of the negative trend in the regression results. The dataset was divided between the rural and urban areas. The urban area subset consists of all house transaction in the biggest 15 cities of Gelderland. The rural subset contains the remaining datapoint. The output of the subset analysis in reported in table 12. The output shows that the difference-in-difference coefficents are slightly higher in urban areas than in rural areas. House prices in urban areas are more affected by sport facility opening than house prices in rural areas. Although, all effects are insignificant. The period coefficient is both analyses negative, indicating no bias in the economics crises effects as presumed earlier.

Another subset analysis was executed on the effect within the cities. The morphology can still be a decisive factor in the relationship between sport facilities and house prices. This may differ for houses in city centres relative to houses in city edges. The subset of cities was again divided between city centre, mostly the historic centre, and remaining city houses, in broad terms the city edge. The regressions results are displayed in table 13. The results show that for houses that are not in the inner city the difference-in-difference coefficients are positive while for inner city houses the coefficients are negative. The results show that sport facility opening have a negative impact on house prices in the inner city and a positive impact on house prices that lay outside the city centres. Note that all difference-in-difference coefficients are insignificant.

The results suggest that somehow houses in the city centre are differently affected by sport facilities than house prices outside the inner city. The morphology can probably still be a decisive factor in the relationship between sport facilities and house prices, although this study accounts for municipal, neighbourhood and distance to city centre fixed effects. It can be the case that more densely populated or larger the city more the effects tend to be negative instead of positive. The demographics of a city can play a pivotal role in the relationship. In the end, people are the decisive driver behind house prices. They consider certain factors when buying a house. So, arguably the demographics matter. This study finds in general positive results while Hummel (2018) found negative results. The demographics of the city of Amsterdam and the city of Rotterdam might explain the difference in results. Key figures from the CBS (2020) in 2015 show that the average income in Gelderland per household is 38.800, for Amsterdam 34.900 and for Rotterdam 32.100. The number of elderly people expressed in the demographic grey pressure is 32,5, 17,5 and 23,9 respectively. The ethnicity background is 85,3% Dutch, 7,7% wester and 7,0% non-western for Gelderland, 48,9% Dutch, 16,4% western and 34,7% non-western for Amsterdam and 50,7% Dutch, 11,9% western and 37,4% non-western for Rotterdam. These demographics thus differ a lot in both studies. Therefore, demographics may be an explanation why both studies found opposing results. The problem of morphology therefore may also be more a problem of demographics.

Variables	(Rural areas) Log corr. price	(Urban areas) Log corr. price
In 800m buffer	-0.00501	-0.00596
	(0.00964)	(0.00396)
In 350m buffer	-0.0300**	0.00315
	(0.0138)	(0.00759)
Period	-0.00938*	-0.00877***
	(0.00486)	(0.00191)
DD 800m	0.00184	0.00398
	(0.0114)	(0.00515)
DD 350m	0.00549	0.00710
	(0.0160)	(0.0104)
Constant	11.07***	10.28***
	(0.156)	(0.104)
House structure controls	Yes	Yes
Location controls	Yes	Yes
Municipal FE	Yes	Yes
Year FE	Yes	Yes
Neighbourhood FE	Yes	Yes
Robust	Yes	Yes
Observations	11,736	44,002
R-squared	0.846	0.876

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

Table 12: Regression output of the analyses of rural and urban areas.

Variables	(City edges) Log corr. price	(City centres) Log corr. price
In 800m buffer	-0.00796**	-0.00219
	(0.00405)	(0.0179)
In 350m buffer	0.00526	-0.0814*
	(0.00773)	(0.0465)
Period	-0.00878***	-0.00432
	(0.00194)	(0.00849)
DD 800m	0.00485	-0.0124
	(0.00524)	(0.0240)
DD 350m	0.00890	-0.0294
	(0.0107)	(0.0628)
Constant	10.40***	10.38***
	(0.1000)	(0.199)
House structure controls	Yes	Yes
Location controls	Yes	Yes
Municipal FE	Yes	Yes
Year FE	Yes	Yes
Neighbourhood FE	Yes	Yes
Robust	Yes	Yes
Observations	40,935	3,067
R-squared	0.880	0.844

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

Table 13: Regression output of the analyses with houses in city edges and city centres.

Marginal effects analysis

The problem of marginality was addressed earlier in the discussion. The opening of a new sport facility presumably has more impact on houses that do not have a sport facility in their proximity relative to houses that for example have 10 sport facilities already in their proximity. The effect of marginality was not accounted for in this study. Further analyses could clarify the effect. Additional data requested from the Department of Sport from the Province of Gelderland includes a variable containing the number of similar type of sport facilities in the proximity of a sport facility. A sport based regression with the new variable added was run. The results are displayed in table 14 and show opposing insignificant results. Although for both the 800m and 350m regressions 3 or more sport facilities of the same type have a negative effect relative to zero existing sport facilities of the same type. A marginal effect seems plausible. However, this regression method does not include the morphology, it has no municipal, neighbourhood or distance to city centre fixed effects for houses. Therefore, another analysis was tested.

An analysis with subsets of the aforementioned situation with no neighbouring sport facilities and a lot of neighbouring sport facilities is preferred. However, a distribution plot of the existing sport facilities and newly opened sport facilities in the period 2009-2016 shows that in Gelderland a lot of sport facilities are already clustered together (figure 9 in the appendix). This makes the subset analysis of all sport facilities difficult. A solution to still perform such an analysis is to look per category or per sport facility type individually. Investigation of the data shows that for tennis facilities situations exist where tennis facilities already exist in a village or part of a city and a new one opens and where no tennis facility exists and a new one opens. Also, the dataset contains relatively a lot of tennis facilities. Therefore, tennis facilities are used for a marginal analysis. The results are displayed in table 15. They show that the DD coefficient is positive for houses with no tennis facility already present whereas it is negative for houses that already had one or multiple tennis facilities in their proximity. Note that both difference-in-difference outcomes are insignificant. The results indicate that tennis facility openings have a positive impact on house prices in areas where they are not yet available and a negative effect on house prices in areas where tennis facilities are already available. Therefore, it seems plausible a marginal effect of sport facilities on house prices exists.

Variables	(800m buffer) Log corr. price	(800m buffer) Log corr. price	(350m buffer) Log corr. price	(350m buffer) Log corr. price
No neighbouring facilities		Basis		Basis
1 neighbouring facility		0.0143		-0.0170
		(0.0207)		(0.0309)
2 neighbouring facilities		0.0254		0.0204
		(0.0298)		(0.0470)
3 or more neighbouring facilities		-0.0475		-0.0211
		(0.0523)		(0.0740)
Period	-0.0907***	-0.0904***	-0.107***	-0.107***
	(0.0138)	(0.0139)	(0.0181)	(0.0182)
In buffer	-0.0246	-0.0259	-0.0618**	-0.0617**
	(0.0196)	(0.0196)	(0.0263)	(0.0267)
DD	-0.00319	-0.00387	0.0362	0.0357
	(0.0224)	(0.0226)	(0.0299)	(0.0298)
Constant	3.931*	4.362*	8.949***	8.974***
	(2.366)	(2.378)	(1.024)	(1.014)
House structure controls	Yes	Yes	Yes	Yes
Location controls	Yes	Yes	Yes	Yes
Municipal FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Neighbouring facility FE	No	Yes	No	Yes
Robust	Yes	Yes	Yes	Yes
Observations	400	400	268	268
R-squared	0.829	0.830	0.869	0.870

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

Table 14: Regression output of the sport based analyses with and without neighbouring facilities.

Variables	(No other tennis facility) Log corr. price	(Other tennis facility) Log corr. price
In buffer	-0.0679**	0.0547**
	(0.0283)	(0.0226)
Period	0.000119	-0.112***
	(0.00278)	(0.00196)
DD	0.0196	-0.0275
	(0.0275)	(0.0236)
Constant	11.25***	10.87***
	(0.0942)	(0.103)
House structure controls	Yes	Yes
Location controls	Yes	Yes
Municipal FE	Yes	Yes
Year FE	Yes	Yes
Neighbouring facility FE	Yes	Yes
Robust	Yes	Yes
Observations	37,746	27,465
R-squared	0.873	0.868

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

Table 15: Regression output of the house based analyses with and without other tennis facility.

Multiple year analysis

As stated earlier in the discussion the study does analyses with data from one year pre-opening and one year post-opening and does not consider taking multiple pre and post years. This is mostly due to the limited timespan of the data this study has. Nonetheless, a multiple year analysis is still possible to examine. This analysis takes the sport facilities that are opened in the year 2013 as treatment data. The house transactions from the years 2008 until 2012 are taken as the pre-opening data, the house transactions from the years 2014 until 2017 are taken as the post-opening data. With this data a difference-in-difference regression is run. The results of the multiple year regression are displayed in table next to the single year regression, with only 2012 as pre year and 2014 as post year. The difference-in-difference coefficient for the 800-meter buffer becomes more negative, while the coefficient for the 350-meter was positive in the single year regression but becomes negative in the multiple year regression. This is probably due to the decline in house prices in general over the period 2008 until 2017, although the transaction price was corrected by this trend. Adding multiple years looks promising as both the DD 800 and DD 350 coefficents do become more significant, although still insignificant. The outcomes are however inconsistent, showing again that the results are spurious.

Variables	(Single year) Log corr. price	(Multiple years) Log corr. price
In 800m buffer	-0.0386***	-0.0275***
	(0.0129)	(0.00536)
In 350m buffer	-0.0729***	-0.0187
	(0.0254)	(0.0119)
Period	-0.0909***	-0.178***
	(0.00400)	(0.00337)
DD 800	-0.00339	-0.00870
	(0.0128)	(0.00544)
DD 350	0.0247	-0.0150
	(0.0257)	(0.0137)
Constant	10.91***	11.06***
	(0.143)	(0.0553)
House structure & location controls	Yes	Yes
Municipal & neighbourhood FE	Yes	Yes
Year FE	No	Yes
Robust	Yes	Yes
Observations	8,182	43,077
R-squared	0.854	0.858

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

Table 16: Regression output of the single year and multiple year analysis with 2013 as treatment year.

VII. Conclusion

This paper studies the impact of sport facility openings on nearby house prices in the province of Gelderland. In examining this relationship this research provides multiple important contributions to the existing literature. Firstly, this study expands the knowledge in the almost non-existing field of amateur sport facilities and house values. Secondly, this study is the first to examine a causal relationship between amateur sport facilities and house prices by using a difference-in-difference model. Previous studies only identified associations between both. This paper uses the difference-in-difference analysis in three different models. Thirdly, this study utilises a large dataset covering the whole province of Gelderland, whereas past studies performed analyses only on city level. This study set out to answer the following research question as stated in the introduction:

What is the effect of amateur sport facility openings on house prices?

This paper proposed several complementary difference-in-difference analyses to find the answer to this question. The results that were found suggest that it cannot be rejected that an effect of sport facility openings on house prices is non-existent. Previous literature on this topic suggested that sport facilities might have an effect on house prices. This study however shows that a distinct possibility exists that the effects found by other researchers are spurious when a more elaborate identification strategy is used and additional controls are added to the analyses. The result support this presumption, as the magnitude and sign of the coefficients between the specifications are inconsistent. To directly answer the research question, this study shows that sport facilities openings do not seem to have a conclusive or direct effect on house prices. Consequently, this may indicate that people who want to buy a house do not implicitly or explicitly consider the accessibility of sports facilities in weighing their willingness to pay for a new house. Although the further analyses show that people with a high income might consider sports facilities more in weighing their willingness to pay than people with a low income and the effect might differ accounted for the number of sport facilities that already exist nearby.

The results in this study combined with the previous results in related literature provide more insight in the effect of sport facilities on house prices. This study is however not without limitations, partly as is the hedonic price model is not in general.

Firstly, houses cannot be regarded as homogeneous commodities. The housing market is likely to be composed of several interconnected submarkets, with each requiring a different hedonic function. These submarkets may be defined by structural and spatial factors, such as dwelling type, neighbourhood characteristics and structural house characteristics. Although this study controls for all above mentioned factors, it does not define different hedonic functions for each submarket, nor does it define submarkets at all.

Secondly, the hedonic price model is subject to spatial autocorrelation. Houses in the same neighbourhood often have similar structural and location characteristics. Also, the valuation process between sellers and buyers can affect the price of surrounding houses. Spatial autocorrelation can be a problem as it is more likely that results will be biased due to underestimation of the real variance to the data.

Thirdly, the hedonic price model does not consider personal preference weighing in willingness to pay. The hedonic price is useful, but it cannot objectively measure the values of characteristics. Different users value characteristic in a different way. A young family with a baby coming does probably want to pay more for an additional room than a single elderly person. So, applied for sport facilities as attribute, some people value this highly positive, while others can value this neutral or even negative. No generic objective measure for valuing sport facilities exists with hedonic pricing, only an average measure of all subjective values. In congruence with the aforementioned, the results of hedonic price model are difficult to generalize, as they are location specific.

Further research, when choosing a hedonic model, should built a firm base that deals with the aforementioned limitations. This strong foundation is a predominant step to more conclusive results. Demographic seem to influence the results in this paper, therefore it is recommended to look closely into this factor. Further research should also take into account the marginal impact of an additional sport facilities. Additionally, further research should correct more carefully for the spatial morphology of cities, taking into account that sport facilities locate often near the edges of cities in places with enough space where the land is cheap. A way to account for the morphology of cities is to include an instrumental variable, a variable that affects sport facilities but does not affect house prices. This instrumental variable can better explain the endogeneity of sport facilities and house price in relationship to the morphology of a city. Also, more years in the pre and post opening phase should be added for more robust and conclusive results. The effects of a sport facility opening can differ through the years when multiple years are added in the analysis and effects could already appear in the announcement and construction phase, as argued in the discussion. When this path is taken it is recommended to include a spatiotemporal approach in the models, weighting the datapoint both spatially and temporally.

This study does not find a direct relationship between sport facilities and house prices. It seems that homeowners do not have a willingness to pay for sports through the mechanism of house prices. This does however not mean that sport is not important from a societal perspective. Sport offers a lot of benefits, as improvements in health and social cohesion. But this study suggests that other mechanisms in this relationship are more important, such as for example membership fees. The outcomes of this study do also result in policy implications. It is not possible to capture the value of sports as an amenity externality on house prices. Through this mechanism it is not feasible to finance public contributions and justify some of the large public subsidies for construction and operation of the sports facilities. This means that the planning of sport facilities remains a non-financial, political consideration. Although the literature advocates that sport faculties are valuable and contribute positive externalities to the society as a whole, it does not work via the mechanism of house prices.

VIII. References

Ahlfeldt, G. M., & Kavetsos, G. (2014). Form or function?: the effect of new sports stadia on property prices in London. *Journal of the Royal Statistical Society: series A (statistics in society), 177*(1), 169-190.

Ahlfeldt, G. M., & Maennig, W. (2010). Impact of sports arenas on land values: evidence from Berlin. *The Annals of Regional Science, 44*(2), 205-227.

Bailey, R., Cope, E., & Parnell, D. (2015). Realising the benefits of sports and physical activity: the human capital model. *RETOS. Nuevas Tendencias en Educación Física, Deporte y Recreación*, (28), 147-154.

Banzhaf, H. S., & Farooque, O. (2013). Interjurisdictional housing prices and spatial amenities: Which measures of housing prices reflect local public goods?. *Regional Science and Urban Economics, 43(*4), 635-648.

Basisregister Adressen en Gebouwen (2020). Retrieved from the BAG geodata nationaal georegister locatieserver dataset.

Benefield, J. D. (2009). Neighborhood amenity packages, property price, and marketing time. *Property Management.*

Blomquist, G., & Worley, L. (1981). Hedonic prices, demands for urban housing amenities, and benefit estimates. *Journal of Urban Economics, 9*(2), 212-221.

Bowes, D. R., & Ihlanfeldt, K. R. (2001). Identifying the impacts of rail transit stations on residential property values. *Journal of Urban Economics, 50*(1), 1-25.

Breuer, C., Feiler, S., & Wicker, P. (2015). Sport clubs in Germany. In *Sport clubs in Europe* (pp. 187-208). Springer, Cham.

Burkhardt, J., & Flyr, M. (2019). The effect of marijuana dispensary openings on housing prices. *Contemporary Economic Policy, 37*(3), 462-475.

Caceres, B. C., & Geoghegan, J. (2017). Effects of New Grocery Store Development on Inner-City Neighborhood Residential Prices. *Agricultural and Resource Economics Review*, 46(1), 87-102.

Carey, M., & Mason, D. S. (2014). Building consent: funding recreation, cultural, and sports amenities in a Canadian city. *Managing Leisure*, 19(2), 105-120.

Carlino, G., & Coulson, N. E. (2004). Compensating differentials and the social benefits of the NFL. *Journal of Urban Economics, 56*(1), 25-50.

CBS (2020). House price index 2008-2018, province of Gelderland. Retrieved from the CBS database.

CBS (2020). Regional demographic key figures of the Netherlands. Retrieved from the CBS StatLine database.

Chau, K. W., & Chin, T. L. (2003). A critical review of literature on the hedonic price model. *International Journal for Housing Science and Its Applications*, 27(2), 145-165.

Chay, K. Y., & Greenstone, M. (2005). Does air quality matter? Evidence from the housing market. *Journal of political Economy*, 113(2), 376-424.

Conklin, J., Diop, M., & Li, H. (2020). Contact high: The external effects of retail marijuana establishments on house prices. *Real Estate Economics*, 48(1), 135-173.

Court, A. T., " Hedonic Price Indexes with Automotive Examples," in *General Motors Corporation, Automobile Demand*. New York, 1939.

Crompton, J. L. (2001). The impact of parks on property values: A review of the empirical evidence. *Journal of leisure research*, 33(1), 1-31.

Davies, L. E. (2005). Not in my back yard! Sports stadia location and the property market. *Area*, 37(3), 268-276.

Deaton, B. J., & Hoehn, J. P. (2004). Hedonic analysis of hazardous waste sites in the presence of other urban disamenities. *Environmental Science & Policy*, 7(6), 499-508.

Debrezion, G., Pels, E., & Rietveld, P. (2011). The impact of rail transport on real estate prices: an empirical analysis of the Dutch housing market. *Urban Studies*, 48(5), 997-1015.

Dehring, C. A., Depken, C. A., & Ward, M. R. (2007). The impact of stadium announcements on residential property values: Evidence from a natural experiment in Dallas-Fort Worth. *Contemporary Economic Policy*, 25(4), 627-638.

Dekkers, J. E., & van der Straaten, J. W. (2009). Monetary valuation of aircraft noise: A hedonic analysis around Amsterdam airport. *Ecological Economics*, 68(11), 2850-2858.

Dishman, R. K., Sallis, J. F., & Orenstein, D. R. (1985). The determinants of physical activity and exercise. *Public health reports, 100(2*), 158.

Downward, P., & Rasciute, S. (2010). The relative demands for sports and leisure in England. *European sport management quarterly, 10*(2), 189-214.

Feng, X., & Humphreys, B. R. (2012). The impact of professional sports facilities on housing values: Evidence from census block group data. *City, Culture and Society*, 3(3), 189-200.

Feng, X., & Humphreys, B. (2018). Assessing the economic impact of sports facilities on residential property values: A spatial hedonic approach. *Journal of Sports Economics*, 19(2), 188-210.

Freeman III, A. M., Herriges, J. A., & Kling, C. L. (2014). *The measurement of environmental and resource values: theory and methods*. Routledge.

Friedman, M. T., Andrews, D. L., & Silk, M. L. (2004). Sport and the façade of redevelopment in the postindustrial city. *Sociology of Sport Journal*, 21(2), 119-139.

Gau, L. S. (2007). Examining the values associated with spectator sports and the relationships between the values and sport spectator behaviour.

Glaeser, E. L., Kolko, J., & Saiz, A. (2001). Consumer city. *Journal of economic geography*, 1(1), 27-50.

Hoekman, R., Breedveld, K., & Kraaykamp, G. (2017). Sport participation and the social and physical environment: explaining differences between urban and rural areas in the Netherlands. *Leisure studies*, 36(3), 357-370.

Huang, H., & Humphreys, B. R. (2014). New sports facilities and residential housing markets. *Journal of Regional Science*, 54(4), 629-663.

Hummel, N. (2018). The effect of amateur sport clubs on the value of properties.

Humphreys, B. R., & Nowak, A. (2017). Professional sports facilities, teams and property values: Evidence from NBA team departures. *Regional Science and Urban Economics*, 66, 39-51.

Ibeas, Á., Cordera, R., Dell'Olio, L., Coppola, P., & Dominguez, A. (2012). Modelling transport and real-estate values interactions in urban systems. *Journal of Transport Geography*, 24, 370-382.

Jud, G. D. (1985). A further note on schools and housing values. *Real estate economics*, 13(4), 452-462.

Jud, G. D., & Watts, J. M. (1981). Schools and housing values. *Land Economics*, 57(3), 459-470.

Kagie, M., & Wezel, M. V. (2007). Hedonic price models and indices based on boosting applied to the Dutch housing market. *Intelligent Systems in Accounting, Finance & Management: International Journal*, 15(3-4), 85-106.

Kain, J. F., & Quigley, J. M. (1970). Measuring the value of housing quality. *Journal of the American statistical association*, 65(330), 532-548.

Kavetsos, G. (2012). The impact of the London Olympics announcement on property prices. *Urban Studies*, 49(7), 1453-1470.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of political economy*, 74(2), 132-157.

Lang, C., Opaluch, J. J., & Sfinarolakis, G. (2014). The windy city: Property value impacts of wind turbines in an urban setting. *Energy Economics*, 44, 413-421.

Lee, R. E., Booth, K. M., Reese-Smith, J. Y., Regan, G., & Howard, H. H. (2005). The Physical Activity Resource Assessment (PARA) instrument: evaluating features, amenities and incivilities of physical activity resources in urban neighborhoods. *International Journal of Behavioral Nutrition and Physical Activity*, 2(1), 13.

Linneman, P. (1980). Some empirical results on the nature of the hedonic price function for the urban housing market. *Journal of urban economics*, 8(1), 47-68.

Lynch, A. K., & Rasmussen, D. W. (2001). Measuring the impact of crime on house prices. *Applied Economics*, 33(15), 1981-1989.

Malpezzi, S. (2002). Hedonic pricing models: a selective and applied review. *Housing economics and public policy*, 67-89.

Mok, H. M., Chan, P. P., & Cho, Y. S. (1995). A hedonic price model for private properties in Hong Kong. *The Journal of Real Estate Finance and Economics*, 10(1), 37-48.

Municipality of Amsterdam (2017). Nota Volksgezondheid Amsterdam 2017-2020. Amsterdam: Gemeente Amsterdam

Municipality of Arnhem (2017). Sportvisie – Arnhemse visie op spot & bewegen. Arnhem: Gemeente Arnhem.

Municipality of Utrecht (2016). Utrecht sportief en gezond: sportnota 2017 - 2020. Utrecht: Gemeente Utrecht.

Municipality of Rotterdam (2016). Sport beweegt Rotterdam: sportnota Rotterdam 2017 – 2020. Rotterdam: Gemeente Rotterdam.

Nicholls, S., & Crompton, J. L. (2007). The impact of a golf course on residential property values. *Journal of Sport Management*, 21(4), 555-570.

Palmon, O., & Smith, B. A. (1998). New evidence on property tax capitalization. *Journal of Political Economy*, 106(5), 1099-1111.

Pope, D. G., & Pope, J. C. (2015). When Walmart comes to town: Always low housing prices? Always?. *Journal of Urban Economics*, 87, 1-13.

Portnov, B., Genkin, B., & Barzilay, B. (2009). Investigating the effect of train proximity on apartment prices: Haifa, Israel as a case study. *Journal of Real Estate Research*, 31(4), 371-396.

Province of Gelderland (2020) Beleidsagenda Gelderland Sport! 2020-2023.

Raymond, Y. C., & Love, P. E. (2000). Measuring residential property values in Hong Kong. *Property Management.*

Rosen, S. (1974). Hedonic prices and implicit markets - product differentiation in pure competition. *The Journal of Political Economy*, 82(1), 34–34.

Sirmans, S., Macpherson, D., & Zietz, E. (2005). The composition of hedonic pricing models. *Journal of real estate literature*, 13(1), 1-44.

Smith, V. K., & Huang, J. C. (1995). Can markets value air quality? A meta-analysis of hedonic property value models. *Journal of political economy*, 103(1), 209-227.

Song, Y., & Knaap, G. J. (2004). Measuring the effects of mixed land uses on housing values. *Regional Science and Urban Economics*, 34(6), 663-680.

Taylor, C. B., Sallis, J. F., & Needle, R. (1985). The relation of physical activity and exercise to mental health. *Public health reports*, 100(2), 195.

Thaler, R. (1978). A note on the value of crime control: Evidence from the property market. *Journal of Urban Economics*, 5(1), 137-145.

Tu, C. C. (2005). How does a new sports stadium affect housing values? The case of FedEx field. *Land Economics*, 81(3), 379-395.

Van der Roest, J. (2018). *Sportverenigingen in Nederland: Veerkrachtige verbanden voor sport* (Vol. 480). Mulier Instituut.

Veenhoven, R. (1999). Quality-of-life in individualistic society. *Social indicators research*, 48(2), 159-188.

Witte, A. D., Sumka, H. J., & Erekson, H. (1979). An estimate of a structural hedonic price model of the housing market: an application of Rosen's theory of implicit markets. *Econometrica: Journal of the Econometric Society*, 1151-1173.

Xiao Y. (2017) Hedonic Housing Price Theory Review. In: *Urban Morphology and Housing Market*. Springer Geography. Springer, Singapore.

IX. Appendix

A. Figures

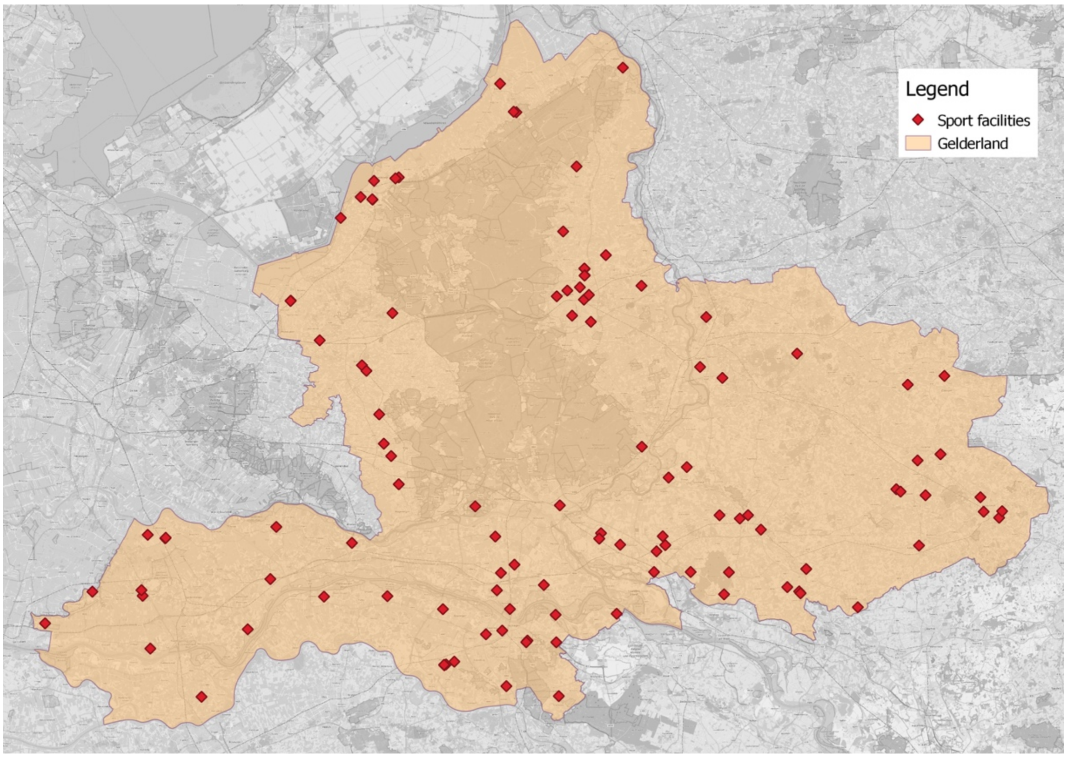


Figure 1: Distribution of sport facility openings in Gelderland.

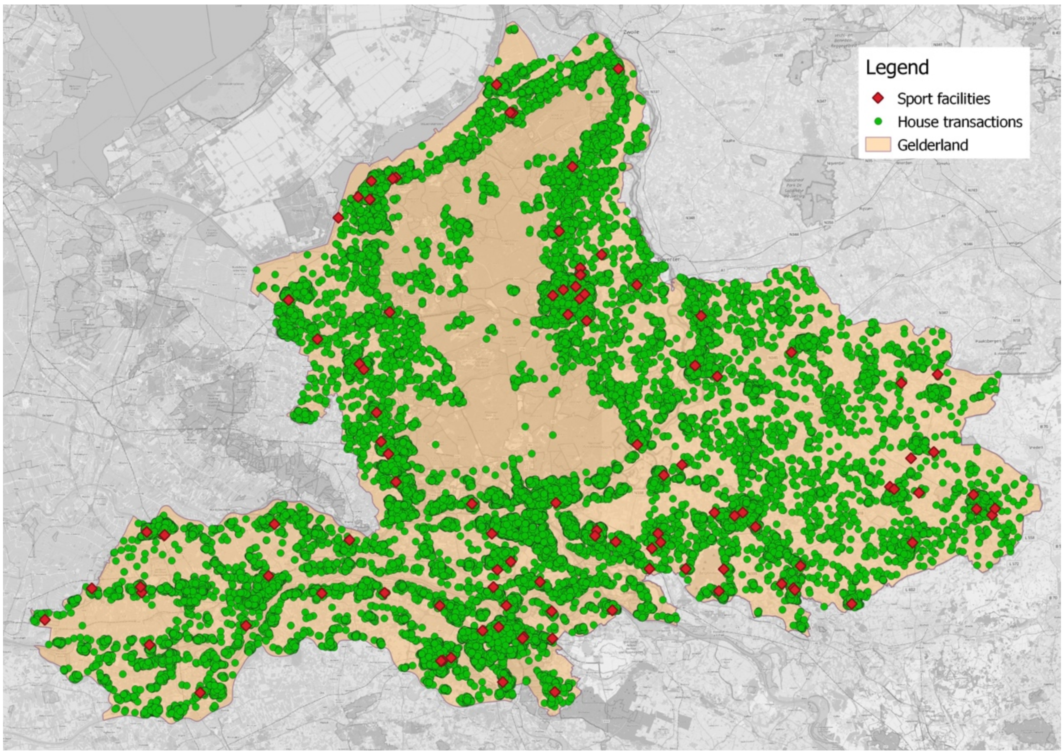


Figure 2: Distribution of sport facility openings and house transactions in Gelderland.

B. Regression output base model

Variables	(Basic) Log price	(Extended) Log price	(Basic) Log corr. price	(Extended) Log corr. price
M2	0.00612***	0.00613***	0.0115***	0.0115***
	(1.66e-05)	(1.67e-05)	(6.27e-05)	(6.32e-05)
M2 squared			-2.13e-05***	-2.13e-05***
			(1.89e-07)	(1.91e-07)
Ceiling height	0.191***	0.191***	0.323***	0.319***
	(0.00163)	(0.00164)	(0.0147)	(0.0149)
Ceiling height squared			-0.0246***	-0.0241***
			(0.00222)	(0.00224)
Plot size			2.29e-05***	2.29e-05***
			(3.40e-07)	(3.42e-07)
Plot size squared			-8.78e-11***	-8.78e-11***
			(0)	(0)
Category	-0.0940***	-0.0941***	-0.0537	-0.0561
	(0.00211)	(0.00212)	(0.0420)	(0.0424)
Dwelling type				
- Simple house			Basis	Basis
			Basis	Basis
- Other house			0.539***	0.533***
			(0.183)	(0.185)
- Boat house			0.225***	0.227***
			(0.0433)	(0.0437)
- Recreational house			-0.312***	-0.311***
			(0.00726)	(0.00732)
- Single family house			0.0157***	0.0156***
			(0.00296)	(0.00298)
- Canal house			-0.0103	-0.0131
			(0.0489)	(0.0493)
- Mansion			0.107***	0.107***
			(0.00387)	(0.00390)
- Living farm			0.0498***	0.0494***
			(0.00486)	(0.00490)
- Bungalow			0.165***	0.165***
			(0.00401)	(0.00405)
- Villa			0.241***	0.241***
			(0.00442)	(0.00445)
- Country house			0.220***	0.219***
			(0.00760)	(0.00767)
- Other apartment			-0.155**	-0.161***
			(0.0617)	(0.0622)
- Ground floor apartment			0.0905**	0.0924**
			(0.0422)	(0.0425)
- Upstairs apartment			-0.0196	-0.0173
			(0.0422)	(0.0426)
- Maisonette			-0.0899**	-0.0876**
			(0.0422)	(0.0426)
- Porch apartment			-0.0228	-0.0207
			(0.0422)	(0.0425)
- Gallery flat			-0.0729*	-0.0707*
			(0.0422)	(0.0426)

- Welfare flat			-0.0157	-0.0133
			(0.0518)	(0.0522)
- Ground floor with stairs			-0.0551	-0.0512
			(0.0449)	(0.0453)
Construction period				
- 1500-1905	Basis	Basis	Basis	Basis
	Basis	Basis	Basis	Basis
- 1906-1930	0.0559***	0.0561***	0.0487***	0.0492***
	(0.00328)	(0.00330)	(0.00326)	(0.00328)
- 1931-1944	0.0657***	0.0655***	0.0445***	0.0444***
	(0.00262)	(0.00264)	(0.00267)	(0.00269)
- 1945-1960	0.109***	0.109***	0.0853***	0.0850***
	(0.00291)	(0.00293)	(0.00290)	(0.00293)
- 1961-1970	-0.00465*	-0.00496*	0.00327	0.00306
	(0.00252)	(0.00253)	(0.00256)	(0.00258)
- 1971-1980	-0.106***	-0.106***	-0.0918***	-0.0917***
	(0.00219)	(0.00220)	(0.00228)	(0.00229)
- 1981-1990	-0.129***	-0.129***	-0.121***	-0.121***
	(0.00215)	(0.00217)	(0.00214)	(0.00216)
- 1991-2000	-0.0499***	-0.0499***	-0.0508***	-0.0508***
	(0.00229)	(0.00230)	(0.00211)	(0.00213)
- 2001 and later	0.0105***	0.0102***	-0.00368*	-0.00406**
	(0.00221)	(0.00222)	(0.00197)	(0.00199)
Lift			0.0430***	0.0428***
			(0.00256)	(0.00258)
Floors			-0.0234***	-0.0232***
			(0.00119)	(0.00120)
Rooms			0.00883***	0.00865***
			(0.000581)	(0.000586)
Living room type				
- Other living room shape			Basis	Basis
			Basis	Basis
- L-shape			0.0226***	0.0238***
			(0.00192)	(0.00193)
- T-shape			0.0181	0.0188
			(0.0120)	(0.0121)
- Z- or U-shape			-0.00665*	-0.00492
			(0.00403)	(0.00406)
- Open living room			-0.0152***	-0.0140***
			(0.00224)	(0.00226)
- Room en suite			0.0669***	0.0679***
			(0.00394)	(0.00397)
Toilet			0.00868***	0.00876***
			(0.000426)	(0.000429)
Bathroom			0.00355***	0.00447***
			(0.00130)	(0.00131)
Kitchen			-0.0197***	-0.0241***
			(0.00139)	(0.00140)
Attic			-0.00508***	-0.00524***
			(0.00130)	(0.00131)
Garret			-0.00899***	-0.00891***
			(0.00184)	(0.00185)
Practice room			0.0160***	0.0162***
			(0.00535)	(0.00539)
Scullery			0.0391***	0.0393***

			(0.00124)	(0.00126)
Dormer			0.0197***	0.0197***
			(0.00141)	(0.00142)
Roof terrace			0.0226***	0.0227***
			(0.00217)	(0.00219)
Indoor balcony			-0.0221***	-0.0220***
			(0.00240)	(0.00242)
Balcony			0.0239***	0.0241***
			(0.00151)	(0.00152)
Parking	0.149***	0.149***	0.113***	0.112***
	(0.00122)	(0.00123)	(0.00113)	(0.00114)
Garden	-0.0723***	-0.0724***	-0.0501***	-0.0502***
	(0.00145)	(0.00146)	(0.00145)	(0.00146)
Maintenance interior				
- Terrible	Basis	Basis	Basis	Basis
	Basis	Basis	Basis	Basis
- Very bad	-0.0274	-0.0279	-0.00670	-0.00725
	(0.0262)	(0.0263)	(0.0231)	(0.0233)
- Bad	0.0265*	0.0245	0.0159	0.0138
	(0.0158)	(0.0159)	(0.0140)	(0.0141)
- So-so	0.0482***	0.0473***	0.0306*	0.0295*
	(0.0178)	(0.0179)	(0.0157)	(0.0158)
- Mediocre	0.0871***	0.0856***	0.0644***	0.0628***
	(0.0156)	(0.0157)	(0.0138)	(0.0139)
- Reasonable	0.107***	0.106***	0.0805***	0.0789***
	(0.0159)	(0.0161)	(0.0141)	(0.0143)
- Good	0.184***	0.182***	0.147***	0.145***
	(0.0156)	(0.0157)	(0.0139)	(0.0140)
- Very good	0.230***	0.228***	0.190***	0.188***
	(0.0163)	(0.0164)	(0.0145)	(0.0146)
- Excellent	0.247***	0.245***	0.206***	0.204***
	(0.0159)	(0.0160)	(0.0142)	(0.0143)
Maintenance exterior				
- Terrible	Basis	Basis	Basis	Basis
	Basis	Basis	Basis	Basis
- Very bad	-0.00283	-0.000885	-0.0205	-0.0185
	(0.0280)	(0.0282)	(0.0247)	(0.0250)
- Bad	0.0173	0.0208	0.0385**	0.0420***
	(0.0174)	(0.0175)	(0.0154)	(0.0155)
- So-so	-0.0260	-0.0251	0.0176	0.0186
	(0.0201)	(0.0202)	(0.0178)	(0.0179)
- Mediocre	0.0281*	0.0308*	0.0631***	0.0659***
	(0.0169)	(0.0170)	(0.0150)	(0.0151)
- Reasonable	0.0236	0.0264	0.0650***	0.0678***
	(0.0173)	(0.0174)	(0.0153)	(0.0154)
- Good	0.0660***	0.0692***	0.101***	0.104***
	(0.0169)	(0.0170)	(0.0150)	(0.0152)
- Very good	0.0823***	0.0857***	0.110***	0.114***
	(0.0177)	(0.0178)	(0.0157)	(0.0158)
- Excellent	0.0820***	0.0851***	0.109***	0.112***
	(0.0173)	(0.0174)	(0.0153)	(0.0155)
Insulation			0.00943***	0.00949***
			(0.000399)	(0.000403)
Heating			-0.0161***	-0.0161***
			(0.00285)	(0.00288)

Fine location			0.0304***	0.0304***
			(0.00113)	(0.00114)
Busy road				
- Quiet			Basis	Basis
			Basis	Basis
- Medium			-0.0170***	-0.0171***
			(0.00104)	(0.00104)
- Busy			-0.0503***	-0.0499***
			(0.00316)	(0.00318)
Central location				
- Outside built area			Basis	Basis
			Basis	Basis
- Unknown			-0.0428***	-0.0429***
			(0.00322)	(0.00324)
- In residential area			-0.0895***	-0.0896***
			(0.00329)	(0.00332)
- In centre			-0.0509***	-0.0511***
			(0.00367)	(0.00370)
Year				
- 2008	Basis	Basis	Basis	Basis
	Basis	Basis	Basis	Basis
- 2009	-0.0476***	-0.0884***	-0.0499***	-0.0910***
	(0.00262)	(0.00264)	(0.00232)	(0.00234)
- 2010	-0.0505***	-0.111***	-0.0538***	-0.115***
	(0.00259)	(0.00260)	(0.00229)	(0.00231)
- 2011	-0.0855***	-0.180***	-0.0889***	-0.183***
	(0.00267)	(0.00268)	(0.00237)	(0.00239)
- 2012	-0.162***	-0.329***	-0.158***	-0.323***
	(0.00262)	(0.00264)	(0.00236)	(0.00238)
- 2013	-0.206***	-0.446***	-0.197***	-0.434***
	(0.00266)	(0.00268)	(0.00249)	(0.00251)
- 2014	-0.193***	-0.432***	-0.183***	-0.420***
	(0.00246)	(0.00248)	(0.00233)	(0.00235)
- 2015	-0.179***	-0.399***	-0.166***	-0.383***
	(0.00236)	(0.00238)	(0.00225)	(0.00227)
- 2016	-0.145***	-0.329***	-0.132***	-0.313***
	(0.00226)	(0.00228)	(0.00218)	(0.00220)
- 2017	-0.0934***	-0.215***	-0.0754***	-0.195***
	(0.00225)	(0.00227)	(0.00217)	(0.00219)
Municipality				
- 197 Aalten	Basis	Basis	Basis	Basis
	Basis	Basis	Basis	Basis
- 2000 Apeldoorn	0.218***	0.217***	0.168***	0.169***
	(0.00531)	(0.00535)	(0.00525)	(0.00530)
- 202 Arnhem	0.166***	0.166***	0.125***	0.125***
	(0.00532)	(0.00535)	(0.00523)	(0.00527)
- 203 Barneveld	0.336***	0.336***	0.355***	0.355***
	(0.00605)	(0.00609)	(0.00538)	(0.00542)
- 209 Beuningen	0.208***	0.209***	0.175***	0.175***
	(0.00728)	(0.00733)	(0.00663)	(0.00668)
- 213 Brummen	0.200***	0.199***	0.201***	0.200***
	(0.00745)	(0.00750)	(0.00662)	(0.00668)
- 214 Buren	0.289***	0.289***	0.248***	0.248***
	(0.00705)	(0.00709)	(0.00633)	(0.00638)
- 216 Culemborg	0.258***	0.257***	0.285***	0.284***

	(0.00668)	(0.00673)	(0.00600)	(0.00605)
- 221 Doesburg	0.0853***	0.0850***	0.0900***	0.0893***
	(0.00913)	(0.00919)	(0.00810)	(0.00816)
- 222 Doetinchem	0.0873***	0.0873***	0.0355***	0.0356***
	(0.00590)	(0.00594)	(0.00574)	(0.00579)
- 225 Druten	0.140***	0.139***	0.132***	0.131***
	(0.00768)	(0.00773)	(0.00683)	(0.00689)
- 226 Duiven	0.107***	0.107***	0.0928***	0.0924***
	(0.00700)	(0.00705)	(0.00629)	(0.00634)
- 228 Ede	0.313***	0.313***	0.268***	0.267***
	(0.00546)	(0.00549)	(0.00533)	(0.00538)
- 230 Elburg	0.269***	0.268***	0.282***	0.281***
	(0.00725)	(0.00730)	(0.00653)	(0.00659)
- 232 Epe	0.304***	0.304***	0.268***	0.268***
	(0.00646)	(0.00650)	(0.00573)	(0.00578)
- 233 Ermelo	0.385***	0.384***	0.342***	0.342***
	(0.00683)	(0.00687)	(0.00625)	(0.00630)
- 243 Harderwijk	0.296***	0.296***	0.237***	0.236***
	(0.00599)	(0.00603)	(0.00586)	(0.00591)
- 244 Hattem	0.318***	0.318***	0.280***	0.280***
	(0.00854)	(0.00859)	(0.00778)	(0.00784)
- 246 Heerde	0.305***	0.304***	0.266***	0.266***
	(0.00734)	(0.00739)	(0.00650)	(0.00655)
- 252 Heumen	0.271***	0.271***	0.239***	0.240***
	(0.00830)	(0.00835)	(0.00745)	(0.00751)
- 262 Lochem	0.178***	0.178***	0.172***	0.172***
	(0.00644)	(0.00649)	(0.00574)	(0.00578)
- 263 Maasdriel	0.224***	0.224***	0.214***	0.214***
	(0.00770)	(0.00775)	(0.00687)	(0.00692)
- 267 Nijkerk	0.358***	0.358***	0.336***	0.336***
	(0.00607)	(0.00611)	(0.00549)	(0.00554)
- 268 Nijmegen	0.265***	0.264***	0.236***	0.235***
	(0.00532)	(0.00535)	(0.00520)	(0.00525)
- 269 Oldebroek	0.274***	0.274***	0.239***	0.239***
	(0.00749)	(0.00754)	(0.00668)	(0.00674)
- 273 Putten	0.418***	0.418***	0.399***	0.399***
	(0.00714)	(0.00718)	(0.00634)	(0.00639)
- 274 Renkum	0.345***	0.346***	0.297***	0.298***
	(0.00643)	(0.00647)	(0.00592)	(0.00597)
- 275 Rheden	0.187***	0.187***	0.193***	0.193***
	(0.00620)	(0.00624)	(0.00556)	(0.00561)
- 277 Rozendaal	0.504***	0.503***	0.472***	0.471***
	(0.0188)	(0.0189)	(0.0167)	(0.0169)
- 279 Scherpenzeel	0.357***	0.356***	0.350***	0.350***
	(0.0109)	(0.0109)	(0.00964)	(0.00972)
- 281 Tiel	0.159***	0.159***	0.106***	0.107***
	(0.00633)	(0.00637)	(0.00614)	(0.00619)
- 285 Voorst	0.308***	0.307***	0.259***	0.258***
	(0.00735)	(0.00740)	(0.00674)	(0.00680)
- 289 Wageningen	0.309***	0.308***	0.257***	0.256***
	(0.00635)	(0.00639)	(0.00617)	(0.00622)
- 293 Westervoort	0.0679***	0.0683***	0.0503***	0.0511***
	(0.00814)	(0.00819)	(0.00742)	(0.00748)
- 294 Winterswijk	-0.101***	-0.101***	-0.146***	-0.146***
	(0.00691)	(0.00695)	(0.00665)	(0.00671)

- 296 Wijchen	0.196***	0.196***	0.148***	0.148***
	(0.00631)	(0.00635)	(0.00616)	(0.00621)
- 297 Zaltbommel	0.259***	0.259***	0.272***	0.271***
	(0.00729)	(0.00734)	(0.00648)	(0.00653)
- 299 Zevenaar	0.0426***	0.0429***	0.0527***	0.0531***
	(0.00639)	(0.00643)	(0.00566)	(0.00571)
- 301 Zutphen	0.0849***	0.0851***	0.118***	0.119***
	(0.00605)	(0.00609)	(0.00539)	(0.00544)
- 302 Nunspeet	0.351***	0.351***	0.347***	0.346***
	(0.00759)	(0.00764)	(0.00673)	(0.00678)
- 668 West Maas en Waal	0.141***	0.141***	0.0913***	0.0909***
	(0.00840)	(0.00845)	(0.00760)	(0.00766)
- 1509 Oude IJsselstreek	-0.0178**	-0.0181**	-0.0104*	-0.0113*
	(0.00703)	(0.00708)	(0.00628)	(0.00633)
- 1586 Oost Gelre	0.0331***	0.0331***	0.0271***	0.0270***
	(0.00900)	(0.00905)	(0.00797)	(0.00804)
- 1705 Lingewaard	0.178***	0.178***	0.138***	0.138***
	(0.00609)	(0.00613)	(0.00562)	(0.00567)
- 1734 Overbetuwe	0.193***	0.192***	0.172***	0.171***
	(0.00625)	(0.00629)	(0.00569)	(0.00574)
- 1740 Neder-Betuwe	0.258***	0.259***	0.228***	0.228***
	(0.00806)	(0.00811)	(0.00726)	(0.00732)
- 1859 Berkelland	0.0433***	0.0432***	0.0640***	0.0633***
	(0.00671)	(0.00675)	(0.00606)	(0.00611)
- 1876 Bronckhorst	0.140***	0.140***	0.103***	0.102***
	(0.00681)	(0.00685)	(0.00606)	(0.00611)
- 1945 Berg en Dal	0.198***	0.198***	0.190***	0.190***
	(0.00699)	(0.00703)	(0.00627)	(0.00633)
- 1955 Montferland	0.0153**	0.0149**	0.00946	0.00901
	(0.00668)	(0.00672)	(0.00595)	(0.00600)
- 1960 West Betuwe	0.290***	0.290***	0.299***	0.299***
	(0.00634)	(0.00638)	(0.00562)	(0.00567)
Distance closest centre			-5.76e-06***	-5.75e-06***
			(2.31e-07)	(2.33e-07)
Constant	10.64***	10.86***	10.23***	10.46***
	(0.0142)	(0.0143)	(0.0278)	(0.0280)
Observations	143,715	143,715	143,715	143,715
R-squared	0.751	0.766	0.806	0.817

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1: Regression output of the basic model and extended basic model

C. Further analyses

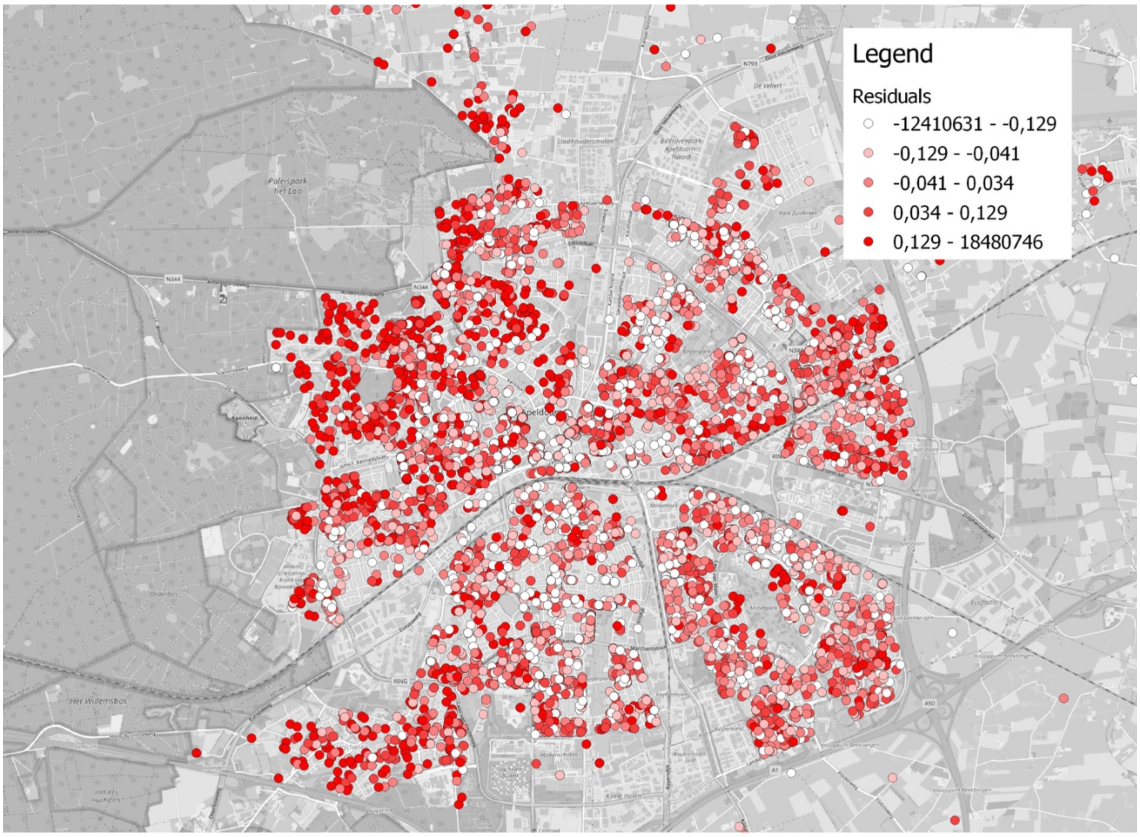


Figure 3: Distribution of the residuals of the house based regression in Apeldoorn.

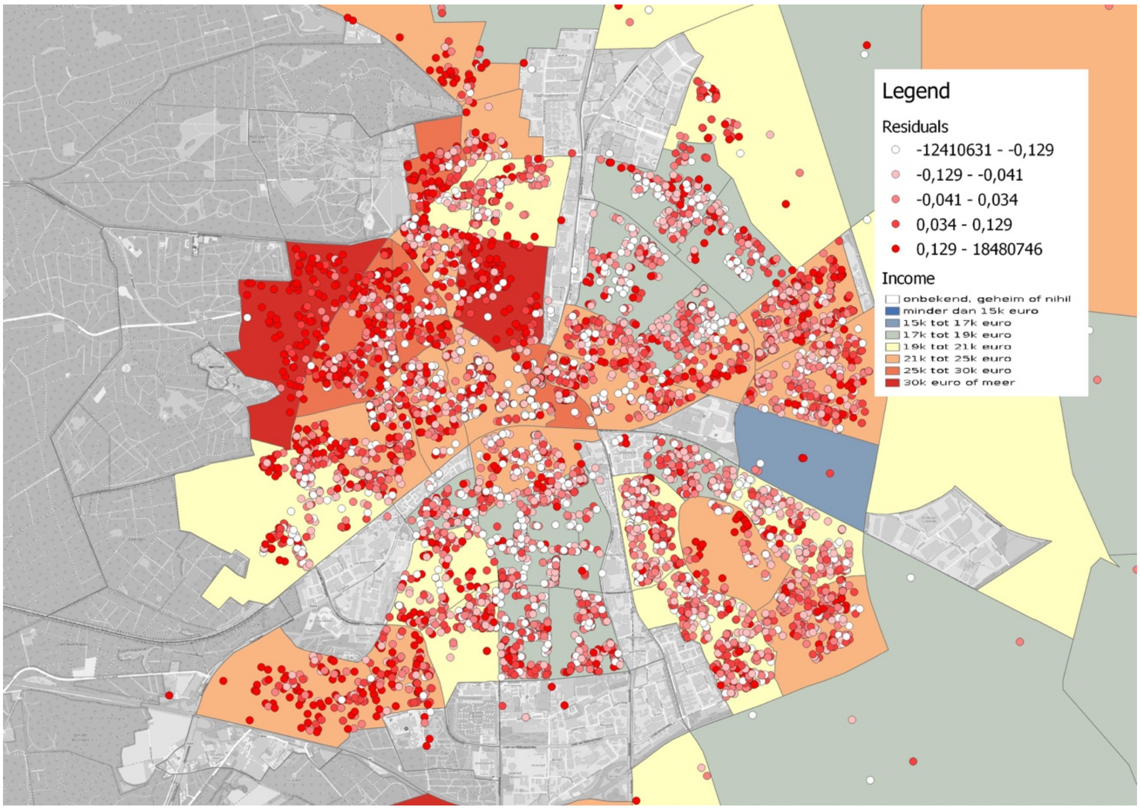


Figure 4: Distribution of the residuals of the house based regression in Apeldoorn and income per neighbourhood.

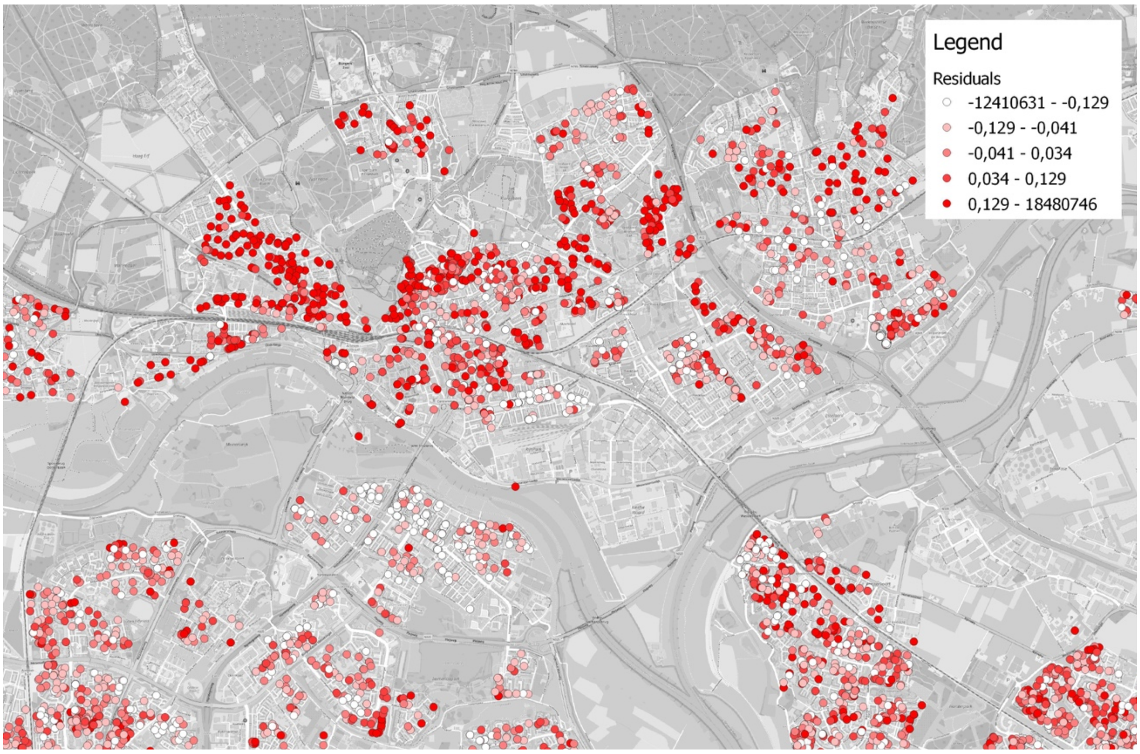


Figure 5: Distribution of the residuals of the house based regression in Arnhem.

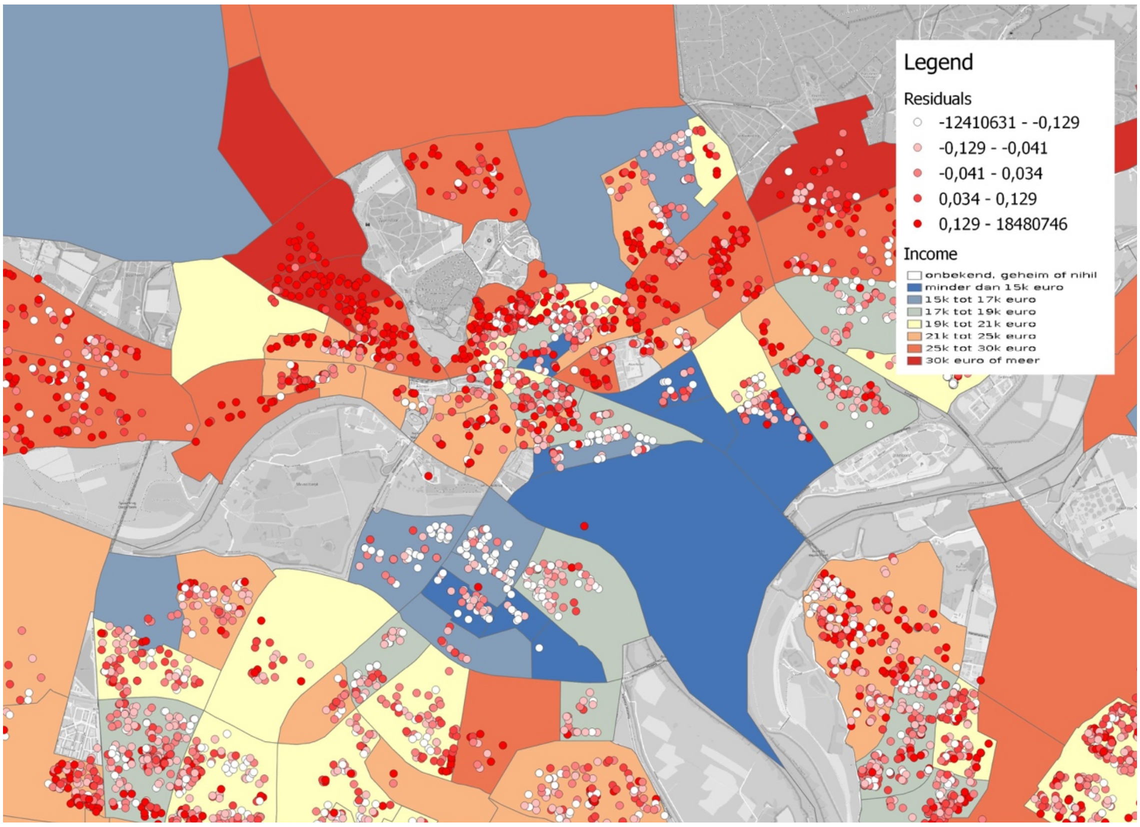


Figure 6: Distribution of the residuals of the house based regression in Arnhem and income per neighbourhood.

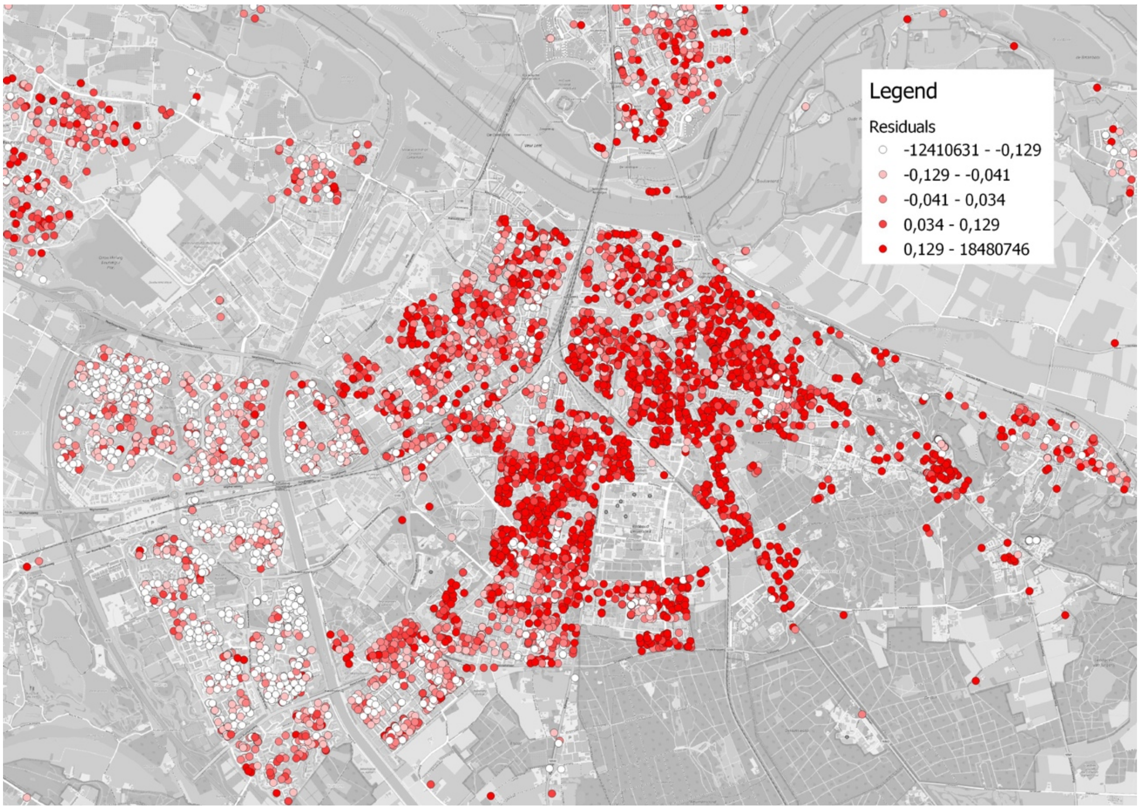


Figure 7: Distribution of the residuals of the house based regression in Nijmegen.

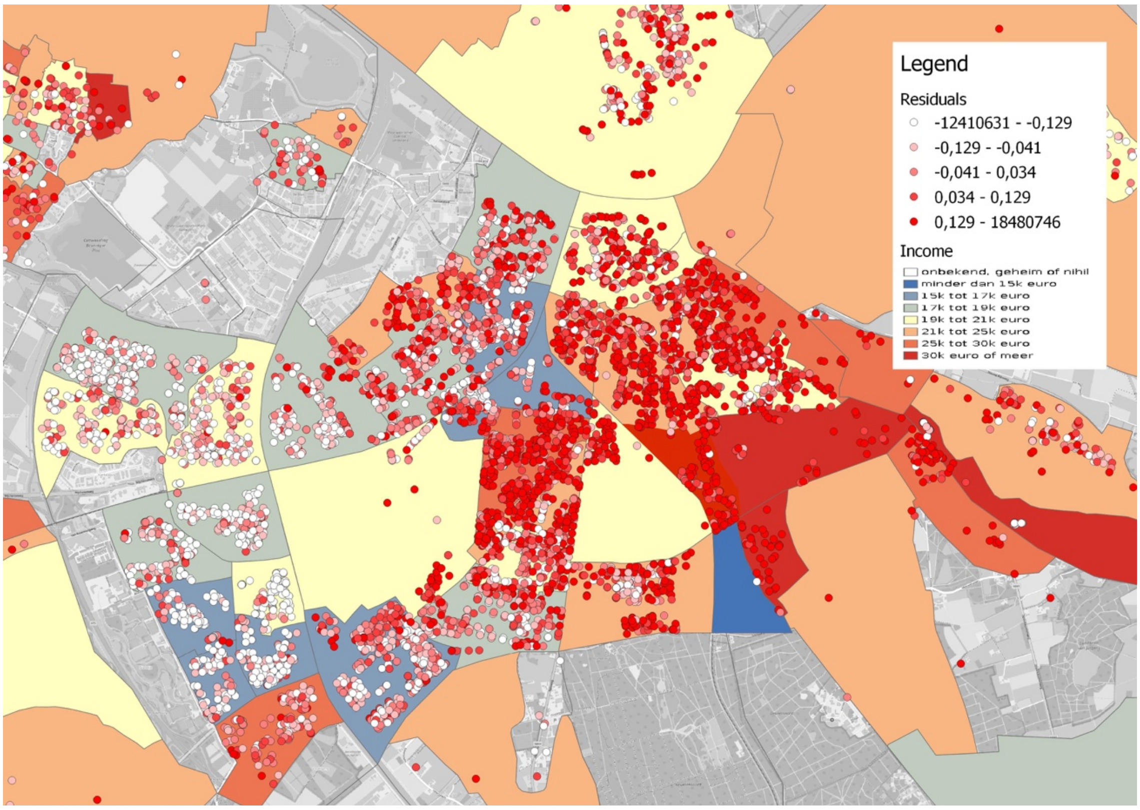


Figure 8: Distribution of the residuals of the house based regression in Nijmegen and income per neighbourhood.

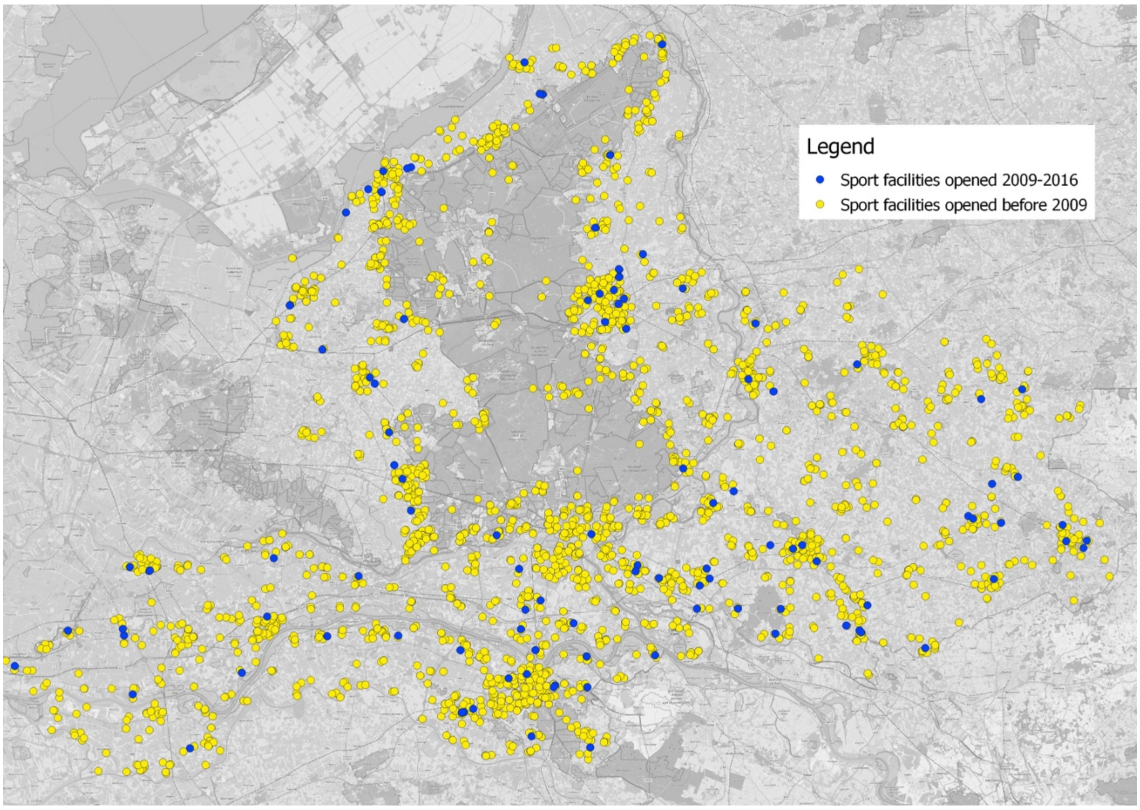


Figure 9: Distribution of existing and newly opened sport facilities.

D. Road map of the main basic analyses (Dutch)

Drie verschillende analyses zijn in deze paper uitgevoerd:

- Difference in difference met sportclubs en huizen van een bepaald jaar van opening
- Difference in difference met alle huizen
- Difference in difference met gemiddelde van huis karakteristieken per sportclubs

Datasets:
Sportfacilities.csv
Sportfacilities2009.csv
Sportfacilities2010.csv
Sportfacilities2011.csv
Sportfacilities2012.csv
Sportfacilities2013.csv
Sportfacilities2014.csv
Sportfacilities2015.csv
Sportfacilities2016.csv

Housetranscations.csv
Housetranscations2008.csv
Housetranscations2009.csv
Housetranscations2010.csv
Housetranscations2011.csv
Housetranscations2012.csv
Housetranscations2013.csv
Housetranscations2014.csv
Housetranscations2015.csv
Housetranscations2016.csv
Housetranscations2017.csv

Scripts:
Basismetmax5000.py
Diffdiff.py
GemiddeldePERSportClub.m

Programma's
Excel
QGIS
QGIS Python console
Matlab
Stata

1- Difference in difference met sportclubs en huizen van een bepaald sportclub jaar

Datasets:

Sportclubs:	Bijv: Sportfacilities2010.csv
Huizentransacties:	Bijv: Housetranscations2009.csv, Housetranscations2011.csv

Scripts:

QGIS Python console Basismetmax5000.py

QGIS:
- Zet bestanden over naar Windows computer
- Open QGIS
- Open OpenStreetMap en zet EPSG naar Amersfoort new (28992)
- Import Sportfacilities2010.csv in QGIS
- Import Housetranscations2009.csv en Housetranscations2011.csv in QGIS
- Maak nieuwe map X op bureaublad
- Exporteer Sportfacilities2010van QGIS als Shapefile naar nieuwe map X
- Exporteer Housetranscations2009 en Housetranscations2011 van QGIS als Shapefile naar map X
- Open Python Console in QGIS
- Laad Basis.py in
- Pas verwijzing naar map X aan in regel 3

Voor Housetranscations2009:
- Pas verwijzing naar Shapefile van huizen uit map X aan in regel 6
- Pas verwijzing naar Shapefile van sportclubs uit map X aan in regel 7
- Stel de range in van de eerste buffer in regel 24 (bijv 350)
- Stel de range in van de tweede buffer in regel 25 (bijv 800)
- Stel de range in van de maximale buffer voor huizen ver weg in regel 26 (bijv 5000)
- Run het script
- Wacht een paar minuten
- QGIS Python heeft een 2 nieuwe kolommen aangemaakt in het huizenbestand, Result800 als het in 350-800 meter ligt en result350 als het in 350 met van de sportclub ligt. Een -1 betekent dat het buiten dan buffer van 5km van een sportclub ligt. Een 0 als een huis niet in de range van een sportclub maar wel binnen 5km en 1 als het wel in de range van een sportclub ligt (800 of 350)

Voor Housetranscations2011:
- Pas verwijzing naar Shapefile van huizen uit map X aan in regel 6
- Pas verwijzing naar Shapefile van sportclubs uit map X aan in regel 7
- Stel de range in van de eerste buffer in regel 24 (bijv 350)
- Stel de range in van de tweede buffer in regel 25 (bijv 800)
- Stel de range in van de maximale buffer voor huizen ver weg in regel 26 (bijv 5000)
- Run het script
- Wacht een paar minuten
- Herlaad de bestanden via de browser in QGIS
- Check bij de attributentabellen of de data in de 2 kolommen bij beide huizenlagen kloppen
- Exporteer Housetranscations2009 en Housetranscations2011 als excel bestanden naar het bureaublad
- Voor elk excel file sort result800 en verwijder datapunten met -1 als waarde, maar alleen als er bij result350 een 0 staat.
- Maak van alle 2en en 3en een 1 in result800
- Sort voor result350 en maak van alle -1 bij result800 een 0 als er bij result350 een 1 staat
- Maak een nieuwe variabel aan in Housetranscations2009, Period en vul voor elk huis 0 in
- Maak dezelfde variabel aan in Housetranscations2011, Period maar vul voor elk huis 1 in
- Voeg de data van Housetranscations2011 onder die van Housetranscations2009 in hetzelfde bestand
- voeg 2 nieuwe variabelen toe, W*800 en W*350 en vul in voor alle huizen
- Importeer het excel bestand in stata
- Maak variabele van de log van corprice aan, logcp
- Run een regressie met dependent logcp en independent de control variabelen

2- Difference in difference met Huizen:

<u>Datasets:</u>

Sportclubs:	Sportfacilities2009.csv, …, Sportfacilities2016.csv
Huizentransacties:	Housetranscations2008.csv, …. , Housetranscations2017.csv

<u>Scripts:</u>

QGIS Python console	Basismetmax5000.py

<u>QGIS</u>:
- Zet bestanden over naar Windows computer
- Open QGIS
- Open OpenStreetMap en zet EPSG naar Amersfoort new (28992)
- Import Sportfacilities2010.csv in QGIS
- Import Housetranscations2009.csv en Housetranscations2011.csv in QGIS
- Maak nieuwe map X op bureaublad
- Exporteer Sportfacilities2010van QGIS als Shapefile naar nieuwe map X
- Exporteer Housetranscations2009 en Housetranscations2011 van QGIS als Shapefile naar map X
- Open Python Console in QGIS
- Laad Basis.py in
- Pas verwijzing naar map X aan in regel 3

Voor Housetranscations2009:
- Pas verwijzing naar Shapefile van huizen uit map X aan in regel 6
- Pas verwijzing naar Shapefile van sportclubs uit map X aan in regel 7
- Stel de range in van de eerste buffer in regel 24 (bijv 350)
- Stel de range in van de tweede buffer in regel 25 (bijv 800)
- Stel de range in van de maximale buffer voor huizen ver weg in regel 26 (bijv 5000)
- Run het script wacht een paar minuten
- QGIS Python heeft een 2 nieuwe kolommen aangemaakt in het huizenbestand, Result800 als het in 350-800 meter ligt en result350 als het in 350 met van de sportclub ligt. Een -1 betekent dat het buiten dan buffer van 5km van een sportclub ligt. Een 0 als een huis niet in de range van een sportclub maar wel binnen 5km en 1 als het wel in de range van een sportclub ligt (800 of 350)

Voor Housetranscations2011:
- Pas verwijzing naar Shapefile van huizen uit map X aan in regel 6
- Pas verwijzing naar Shapefile van sportclubs uit map X aan in regel 7
- Stel de range in van de eerste buffer in regel 24 (bijv 350)
- Stel de range in van de tweede buffer in regel 25 (bijv 800)
- Stel de range in van de maximale buffer voor huizen ver weg in regel 26 (bijv 5000)
- Run het script en wacht een paar minuten
- Herlaad de bestanden via de browser in QGIS
- Check bij de attributentabellen of de data in de 2 kolommen bij beide huizenlagen kloppen
- Exporteer Housetranscations2009 en Housetranscations2011 als excel bestanden naar het bureaublad
- Doe dit voor alle sportclubs jaren met bijbehorende huizentransacties
- Voor elk excel file sort result800 en verwijder datapunten met -1 als waarde, maar alleen als er bij result350 een 0 staat.
- Maak van alle 2en en 3en een 1 in result800
- Sort voor result350 en maak van alle -1 bij result800 een 0 als er bij result350 een 1 staat
- Maak een nieuwe variabel aan in Housetranscations2009, Period en vul voor elk huis 0 in
- Maak dezelfde variabel aan in Housetranscations2011, Period maar vul voor elk huis 1 in
- Voeg de data van Housetranscations2011 onder die van Housetranscations2009 in hetzelfde bestand
- voeg 2 nieuwe variabelen toe, W*800 en W*350 en vul in voor alle huizen

Doe al dit bovenstaande voor elke sportclub jaar, 2009 t/m 2016
- voeg alle bestanden samen in 1 bestand
- Importeer het excel bestand in stata
- Maak variabele van de log van corprice aan, logcp
- Run een regressie met dependent logcp en independent de control variabelen

3- Difference in difference met sportclubs:

Datasets:

Sportclubs:	Sportfacilities.csv
Huizentransacties:	Housetranscations.csv (andere file)

Scripts:

QGIS Python console	Diffdiffmetmax5000.py
Matlab	GemiddeldePERSportClub.m

Stappenplan:

QGIS:

- Zet bestanden over naar Windows computer
- Verander , in . bij variabel corprice
- Open QGIS
- Open openstreetmap en zet EPSG naar Amersfoort new (28992)
- Import BasicHousetransactions.csv in QGIS
- Import Sportfacilities.csv in QGIS
- Maak nieuwe map X op bureaublad
- Exporteer Housetransactions van QGIS als Shapefile naar nieuwe map X
- Exporteer Sportfacilities van QGIS als Shapefile naar map X
- Maak een buffer van 5000 om de sportfacilities
- Clip de huizentransacties met de buffer
- Gebruik de geclipte huizen als invoer veld in Python
- Verwijder alle huistransacties die in een range van 350 van een sportfacility vallen
- Open Python Console in QGIS
- Laad Diffdiff.py in
- Pas verwijzing naar map X aan in regel 3
- Pas verwijzing naar Shapefile van huizen uit map X aan in regel 6
- Pas verwijzing naar Shapefile van sportclubs uit map X aan in regel 7
- Stel de range in, 800
- Run het script
- Wacht ongeveer 2 uur
- QGIS Python heeft 2 bestanden gecreëerd, file1 en file2 (File1, op de rijen de huizen, in de kolommen staan sportclubs. 0 en 1 geeft aan of het huis in de range van de sportclub ligt. File2, is 0, 1 of -1. Met huizen als rijen en sportclubs als kolommen. 1 als het huis een jaar na een sportclub is gebouwd, 0 als ie een jaar daarvoor is gebouwd. -1 als ie niet mee doet)
- Check de data van file1 en file2
- Exporteer Housetransactions van QGIS als csv bestand **gescheiden met komma** op het bureaublad, als ExBasicHousetransactions.csv
- Verwijder de eerste rij met de variabel namen en met find en select en replace verwijder alle “
- Open een leeg excel bestand en bij data, get data importeer het tekstbestand van file1. Transformeer data en verwijder de lege rijen
- Save vervolgens op het bureaublad als csv bestand
- Doe hetzelfde voor file2

Matlab:

- Open Matlab
- Laad bestand GemiddeldePERSportClub.m
- Voer de naam van file1-800 in in regel 4
- Voer de naam van file2-800 in in regel 5
- Voer de naam van het huizenbestand ExBasicHousetransactions.csv
- Kopieer regel 1 van het Matlab bestand en voer in in de command window en druk op enter
- Matlab creëert 4 groepen, elk in een andere window

Excel:

- Maak een excel bestand aan waar de 4 groepen bijeenkomen
- Zet in de eerste regel de namen van de variabelen, zelfde als in bestand ExBasicHousetransactions.csv
- Zet alle groepen naast elkaar
- Zet naast de groepen de karakteristieken uit het sportfacility bestand
- Verwijder rijen zonder data NaN

- Maak een nieuwe variable aan, pre/post sportclub, P
- Maak een nieuwe variabel aan, wel/niet in range sportclub, Z
- Importeer daarna de data van groep 4: niet in cirkel, 1 jaar voor sportclub
- Voor groep 4 vul 0 in bij P en 0 in bij Z
- Importeer daarna onder groep 4 de data van groep 2: wel in cirkel, 1 jaar voor sportclub
- Voor groep 2 vul 0 in bij P en 1 in bij Z
- Importeer daarna onder groep 2 de data van groep 3: niet in cirkel, 1 jaar na sportclub
- Voor groep 3 vul 1 in bij P en 0 in bij Z
- Importeer daarna onder groep 3 de data van groep 1: wel in cirkel, 1 jaar na sportclub
- Voor groep 1 vul 1 in bij P en 1 in bij Z
- Maak nieuwe variabele aan P keer Z, met de formule =P*Z en vul in voor alle groepen

Stata:

- Importeer het excel bestand in stata
- Maak een variabele van de log van corprice aan
- Run een regressie met dependent logcorprice en als independent alle control variabelen
- Je hebt een regressie van de gemiddelde prijs en karakteristieken van alle sportclubs met een range van 800 meter
- Doe hetzelfde voor 350 meter, maar verwijder niet de huizen die in een range van 350 meter vallen

E. Scripts

Basis

```
import ogr, os, sys
os.chdir('C:/…………..')
driver = ogr.GetDriverByName('ESRI Shapefile')

fna = 'Huizen_test.shp'
fnb = 'Sport_test.shp'

dsa = driver.Open(fna, 1)
dsb = driver.Open(fnb, 0)

layera = dsa.GetLayer()
layerb = dsb.GetLayer()

layer_defn = layera.GetLayerDefn()
field_names = [layer_defn.GetFieldDefn(i).GetName() for i in range(layer_defn.GetFieldCount())]

new_field = ogr.FieldDefn('Result', ogr.OFTReal)
layera.CreateField(new_field)

featurea = layera.GetNextFeature()

dist = 800

while featurea:
  pointa = featurea.GetGeometryRef()
  buffera = pointa.Buffer(dist)
  featureb = layerb.GetNextFeature()
  counter = 0
  while featureb:
    pointb = featureb.GetGeometryRef()
    intersection = buffera.Intersection(pointb)
    if intersection.IsEmpty():
      pass
    else:
      counter = counter + 1
    featureb.Destroy()
    featureb=layerb.GetNextFeature()
  i = featurea.GetFieldIndex('Result')
  featurea.SetField(i, counter)
  layera.SetFeature(featurea)
  featureb = layerb.ResetReading()
  featurea.Destroy()
  featurea = layera.GetNextFeature()

print("ready")
```

GemiddeldePERSportClub:

```
function [groep1, groep2, groep3, groep4]= GemiddeldePERSportClub()

tabelcirkel = csvread ('file1.csv');
tabelsport = csvread('file2.csv');
huizen = csvread('huizenfile.csv');
counter = zeros(size(tabelsport, 2), 1);
counter1 = zeros(size(tabelsport, 2), 1);
counter2 = zeros(size(tabelsport, 2), 1);
counter3 = zeros(size(tabelsport, 2), 1);
groep1 = zeros(size(tabelsport, 2), size(huizen, 2));
groep2 = zeros(size(tabelsport, 2), size(huizen, 2));
groep3 = zeros(size(tabelsport, 2), size(huizen, 2));
groep4 = zeros(size(tabelsport, 2), size(huizen, 2));
for i=1:size(tabelcirkel, 1)
    disp(i);
    for k=1:size(tabelsport, 2)
        if tabelcirkel(i,k) == 1 && tabelsport(i,k) == 1

            for j=1:size(huizen, 2)
                groep1(k,j) = groep1(k,j) + huizen(i, j);
            end
            counter(k) = counter(k) + 1;
        elseif tabelcirkel(i, k)== 1 && tabelsport(i,k) == 0
            for j=1:size(huizen, 2)
                groep2(k,j) = groep2(k,j) + huizen(i, j);
            end
            counter1(k) = counter1(k) + 1;
        elseif tabelcirkel(i, k)==0 && tabelsport(i,k) == 1
            for j=1:size(huizen, 2)
                groep3(k,j) = groep3(k,j) + huizen(i, j);
            end
            counter2(k) = counter2(k) + 1;
        elseif tabelcirkel(i,k)== 0 && tabelsport(i,k) == 0
            for j=1:size(huizen, 2)
                groep4(k,j) = groep4(k,j) + huizen(i, j);
            end
            counter3(k) = counter3(k) + 1;
        end
    end
end
groep1= bsxfun(@rdivide, groep1, counter);
groep2 = bsxfun(@rdivide, groep2, counter1);
groep3 = bsxfun(@rdivide, groep3, counter2);
groep4 = bsxfun(@rdivide, groep4, counter3);
```

Diffdiffmetmax5000:

```
import ogr, os, sys

os.chdir('C:/…………….')
driver = ogr.GetDriverByName('ESRI Shapefile')

fna = 'Huis.shp'
fnb = 'Sport.shp'
file1 = open("File1.txt","a")
file2 = open("File2.txt","a")

dsa = driver.Open(fna, 0)
dsb = driver.Open(fnb, 1)

layera = dsa.GetLayer()
layerb = dsb.GetLayer()

layer_defn = layerb.GetLayerDefn()
field_names = [layer_defn.GetFieldDefn(i).GetName() for i in range(layer_defn.GetFieldCount())]


featurea = layera.GetNextFeature()

dist = 350
dist2 = 800
dist3 = 5000
iteratie = 1
while featurea:
    pointa = featurea.GetGeometryRef()
    buffera = pointa.Buffer(dist)
    bufferb = pointa.Buffer(dist2)
    bufferc = pointa.Buffer(dist3)
    featureb = layerb.GetNextFeature()
    jaara = featurea.GetField(0)

    while featureb:
        counter800 = 0
        resultjaar = 0
        pointb = featureb.GetGeometryRef()
        intersection800 = buffera.Intersection(pointb)
        intersection350 = bufferb.Intersection(pointb)
        intersection5000 = bufferc.Intersection(pointb)
        jaarb = featureb.GetField(7)
        if jaarb + 1 == jaara:
            resultjaar = 1
        elif jaarb - 1 == jaara:
            resultjaar = 0
        else:
```

```
            resultjaar = 2
        if intersection800.IsEmpty():
            pass
        else:
            counter800 = 1

        if intersection5000.IsEmpty():
            counter800 = 2

        file1.write(str(counter800))
        file1.write(",")
        file2.write(str(resultjaar))
        file2.write(",")
        featureb.Destroy()
        featureb=layerb.GetNextFeature()

    layera.SetFeature(featurea)
    featureb = layerb.ResetReading()
    featurea.Destroy()
    featurea = layera.GetNextFeature()
    file1.write("\r\n")
    file2.write("\r\n")

    iteratie = iteratie + 1

print("ready")
```

Basismetmax5000

```
import ogr, os, sys

os.chdir('C:/………………')
driver = ogr.GetDriverByName('ESRI Shapefile')

fna = 'House2017.shp'
fnb = 'Sport2016.shp'

dsa = driver.Open(fna, 1)
dsb = driver.Open(fnb, 0)

layera = dsa.GetLayer()
layerb = dsb.GetLayer()

layer_defn = layera.GetLayerDefn()
field_names = [layer_defn.GetFieldDefn(i).GetName() for i in range(layer_defn.GetFieldCount())]

new_field = ogr.FieldDefn('Result800', ogr.OFTReal)
layera.CreateField(new_field)
new_field = ogr.FieldDefn('Result350', ogr.OFTReal)
layera.CreateField(new_field)
featurea = layera.GetNextFeature()

dist = 800
dist2 = 350
dist3 = 5000
counthuis = 0
while featurea:
    counthuis = counthuis + 1
    pointa = featurea.GetGeometryRef()
    buffera = pointa.Buffer(dist)
    bufferb = pointa.Buffer(dist2)
    bufferc = pointa.Buffer(dist3)
    featureb = layerb.GetNextFeature()
    counter800 = 0
    counter350 = 0
    indicator = 0

    while featureb:
        pointb = featureb.GetGeometryRef()
        intersection800 = buffera.Intersection(pointb)
        intersection350 = bufferb.Intersection(pointb)
        intersection5000 = bufferc.Intersection(pointb)
        if intersection800.IsEmpty():
            pass
        elif intersection350.IsEmpty():
            counter800 = counter800 + 1
```

```
        if intersection350.IsEmpty():
            pass
        else:
            counter350 = counter350 + 1
        if intersection5000.IsEmpty():
            pass
        else:
            if intersection800.IsEmpty():
                indicator = 1

        featureb.Destroy()
        featureb=layerb.GetNextFeature()
    i = featurea.GetFieldIndex('Result800')
    if indicator == 0 and counter800 == 0:
        featurea.SetField(i, -1)
    else:
       featurea.SetField(i, counter800)
    i = featurea.GetFieldIndex('Result350')
    featurea.SetField(i, counter350)
    layera.SetFeature(featurea)
    featureb = layerb.ResetReading()
    featurea.Destroy()
    featurea = layera.GetNextFeature()

print("ready")
```