Defining the willingness to pay for avoiding flood risk in The Netherlands

Master Thesis: Spatial Transport and Environmental Economics
Date: June 2022
Author: Tobias Lasschuijt - 2032309
Supervisor: Prof. dr. Hans Koster

Table of Contents

Introduction

As global warming continues and world leaders struggle to agree on effective measure to limit our influence on the world's ecosystem, climate adaptation become more and more important. One could interpret the recent acceptation of talking about climate adaptation as a way to cope with climate change, as an indication that climate change prevention might not be successful and adaptation strategies will play a more and more important role.

This opens up a need to assess the potential damages due to climate change and the investment needed to adopt society to this new reality. In this light, the concern of flooding is especially relevant within The Netherlands since 70% if the current real estate stock is located either below sea or river level (Kok et al. 2003). This thesis will focus specifically on the topic of flood risk, which is naturally intertwined with climate change and rising water levels. Future economic impact of flood risk is hard to assess, partly due to increasing unpredictability and severity of flooding events, but also due to the fact that future economic impact is also dependent on one's preferred discount rate. However, using current flood risk levels and changes in those levels, an estimation can be made of the economic impact of the current risk. Thereby providing a solid base for the estimation of future impact of changes in flood risk. Which leads to the definition of the research question which will lead this paper, what is the average willingness to pay of housing buyers for reducing or eliminating flood risk.

The valuation of flood risk can be performed using stated or revealed preference techniques. Much has been written on these topics, for instance Freeman (2003) gives a great overview of non-market valuation techniques. This paper will apply revealed preference methods, the benefits include avoiding potential bias resulting from stated preferences as opposed to actual behaviour. Additionally, this technique permits the analysis of consumer behaviour also where the risk of flooding is accounted for in the case of subjective perceptions. While stated preference techniques tend to present flood risk as a set of scenarios and climate conditions, including a risk of dike failure or overtopping.

Under the label 'Ruimte voor de rivieren' dike diversions, high water channels, floodplain lowering, side channels, temporary water storages have been created in The Netherland between 2009 and 2017. Leading up to these projects, evaluations of mortality risks in dike ring areas along the rivers IJssel, Rijn and Waal have been published in 2012. Furthermore, in 2015 mortality risks in the same locations have been published representing the situation after finalization of these projects. The maps representing these risks will be used as a basis of evaluation of flood risk valuation in this paper.

Using hedonic price analysis methods, the variation in housing transaction prices is exploited to reveal the willingness to pay to avoid or reduce flood risk. Furthermore, discontinuity of observed prices along dike ring borders will be explored in an intent to arrive at a more accurate estimate of the valuation of flood risk.

Chapter 1 will give a short background to flood risk management in The Netherlands. Furthermore, Chapter 2 will give a summary of research previously done on the topic of flood risk valuation within the Netherlands and abroad. Chapter 3 will explain the methods applied to arrive at the estimations of flood risk valuation provided by this paper. Finally, chapter 4 will discuss the findings resulting from the analysis performed and chapter 5 will draw conclusions on these findings.

1. Flood risk The Netherlands background

As one of the most flood prone countries in the world, The Netherlands is also on the forefront of technological advancement in water management. Already in roman times areas with clay soil were made habitable through the use of small dykes and artificial mounds(´terpen´). Large scale dike building started around the 14th century when sea level and ground water in The Netherlands converged to the same height.

From the 16th to the 19th century various water bodies have been reclaimed as land. These areas, polders, are not a way of protection from the water but rather expansion of land for agriculture or habitation. It does however add another dimension to water management. Since these areas are low lying, the groundwater level is lower than the surrounding water levels. Water enters the area through pressure of ground water, rainfall and rivers. Water is then pumped out of the polder through operated devices such as windmills or steam pumps. A failing of this system will inevitably lead to flooding. Moreover, the polders can be seen as a 'bathtub' with a relatively flat base and failing of one of the dykes surrounding the polder lead to inundation of the entire polder, although polders can be subdivided in several smaller 'bath tubs'. Heijer et al. (1998) estimated the economic risk of lakes flooding polders to be almost half as large as the flooding risk coming from the North Sea.

Although there have been countless floods from the sea and rivers ever since The Netherlands have been inhabited, the 1953 storm surge brought on a drastic change in thinking about water management in The Netherlands. The event is now calculated as a once in 300 year flood event, a water level which had not been expected at the time. 1,800km2 flooded and 1836 people died. The delta committee was formed to advise on further action. The new delta plan was to device new safety standards, based on economics. A national dike improvement plan and drastic shortening of the coastline. Shortening of the coastline is a process that had been applied before and was extended after this event. By shortening the coastline flood risk and the management thereof is greatly reduced. Less stretch of dyke would have to be maintained and would be at risk of failing.

The strengthening of water defenses showed a slow progress and in 1993 and 1995 two river floods occurred. This led to the 1996 Flood Protection Act which set the safety standards of the primary water defenses in The Netherlands.

Recent advances in flood risk management

Also in the 21st century water management remains a relevant topic of government policy. The Dutch government spends over EUR1.1 billion per year on the maintenance of dikes, dunes and waterworks that protect the country. Risk analysis performed by the Ministry of Infrastructure and Water Management, under de label Veiligheid Nederland in Kaart (VNK2), shows large differences in flood risk within The Netherlands. The safety standards are not the same for all locations in The Netherlands, protection systems on the coast are characterized by higher safety norms than systems along the rivers. However, these differences only partly explain variance in flood risks. Also within dike ring areas risk varies greatly, while the safety standard is the same for the entire dike ring. The source of this variation in risk has several causes. The effects of flooding differ greatly between and within dike rings. Moreover, cascade effects can appear when a flood in one dike ring causes floods in other nearby dike rings. Furthermore, the mechanism of piping, which is mostly observed in river areas, has shown to be much more important than previously thought.

It is often perceived that the risk along the coast is the largest, mainly due to the 1953 flooding and the unpredictability of storm surges. However, the VNK project has shown that flood risk along the dunes is usually small. Every year millions of cubic meters of sand are added to maintain a very safe coastline. Risks along the rivers are relatively high, the safety norms are also lower in these areas. As the VNK project shows, dike failures in these areas could cause flooding of vast amounts of land with huge damages as a result. In the, sometimes small, dike ring areas and downstream areas of the longer and more sloping dikes, the water could reach meters high. Moreover, the relatively narrow dikes, structure of the underground and relatively long duration of high water makes that the chances of flooding in the river areas is relatively high. Since higher risks with more disastrous consequences are more likely to be on the forefront of people's mind when making decisions, the river areas in The Netherlands make for a more interesting case for the determination of the effect of flood risk on housing valuation than coastal areas (VNK end report, 2015).

Flood risk does not only vary greatly between dike ring areas but also within dike ring areas. Due to differences in land heights and the creation of compartments within dike rings, the maximum water height, flow rate and speed of water ascent can vary greatly. The local individual risk (LIR) used for the estimations in this study considers the cascade effect and evacuation probability and possibility. On the other hand, concentration of economic value and population are kept out of the equation. However, when talking about flood risk more generally, the concentrations of people in cities and the variable presence of economic value in different locations clearly is another source of variation in flood risk.

Leading up to the VNK project, in 2012 reports have been published depicting the flood risk at the time. This information formed the basis of the definition of projects to be executed under de Ruimte voor de Rivier project. The maps published in 2015 in the final VNK2 report depict the situation once Ruimte voor de Rivier projects are finished, even though some projects were not finished in 2015. These 2 sourced of risk information will be used for the estimations performed in this study and will be further discussed under chapter 3.

2. Literature review

River flooding is per definition a spatial phenomenon and specifically suitable for exploratory and confirmatory spatial data analysis since flood risk tends to be defined for quite large areas and potentially causes large human and economic damages. There is a vast amount of literature available using non-market valuation techniques, specifically revealed preference methods. Environmental risks are among the most studied fields. Boyle and Kiel (2001) provide an extensive review of valuation studies regarding environmental risks using hedonic pricing methods.

Moreover, the valuation of flood risk is studied extensively, especially in the United States. Daniel et al. (2009) provides a comprehensive review of studies applying hedonic price methods to valuate flood risk specifically. They conclude that an increase of the probability in flood risk with 0,01 per year relates to a negative price effect of 0.6%, although the studies included in this meta-analysis show a range of effects from -52% to +58%. It is noted that the variation in estimates may merely represent sampling or estimation variance, but it might also be due to systematic variation in the unobserved population value of the willingness to pay.

Looking at individual studies, Speyrer & Ragas (1991) examine the impact of flood risk and mandatory flood insurance on property values in New Orleans from 1971 through 1986, the analysis confirms the finding of earlier studies that location in a floodplain does reduce property values. They a statistically significant and negative effect of flood risk of 4.2% in suburban areas, or USD6.100,00. However, they reveal that much of this reduction can be attributed to mandatory flood insurance coverage. A similar effect is found by Samarasinghe and Sharp (2010), they note that the discount for a location in a 100 year flood plain is 6.2%.

More recently, Bin et al. (2008) argue that coastal amenities and risk a so closely correlated that it would be near impossible to identify them separately within a more traditional hedonic framework.

They therefore offer an interesting solution to disentangling risk factors from spatial amenities. Using GIS, a 3-dimensional measure of ocean view is developed, and successful isolation of risk and amenity is reached. They conclude that location in a flood risk zone reduces a home's value by 11%, which represents a mean of USD36,082 in the studied area in North Carolina, USA.

Within The Netherlands, important contributions in the field of flood risk valuation analysis were made by Daniel et al. (2006). Exploiting the designation of emergency inundation areas, the flood risk announcement effect is studied on housing prices. It is concluded that prices in the studied areas are about 17% lower than the control areas. Furthermore, once these inundation area plans were finally abolished, prices moved back up.

Probably the most recent contribution to flood risk valuation using hedonic price methods in The Netherlands is provided by Bosker et al. (2019). Using an extensive set of housing transaction data and maximum water levels during floods as a proxy for flood risk, a negative price effect on housing transactions of approximately 1% is found for being located in a flood prone area versus a flood safe area. Although the usage of maximum flood water levels as a proxy of flood risk is not completely uninformative, it does seem to be flawed. Maximum flood water levels in the case of all possible floods happening, are not the ideal measure of flood risk in case of the most likely flood. Mainly because these maximum flood water levels are calculated for situations in which all flood defences fail, while in more realistic scenarios one or a few flood defences will fail. Therefore, one of the main contributions of this thesis regarding the identification of the willingness to pay to avoid flood risk in The Netherlands, is the usage of spatially extremely detailed information on so called local individual risk (LIR). The chances of mortality of a person living in a specific location, considering evacuation possibilities and likelihood. One could consider this risk measure as ideal since it is not affected by concentrations of wealth or economic activity. The measure simply tells someone the likelihood of decease if one is to move to a certain location. The value of belongings present in someone's house or in the form of someone's house is also not included. This could be identified as a possible caveat of this measure. On the other hand, these belongings become arguably invaluable with the loss of life.

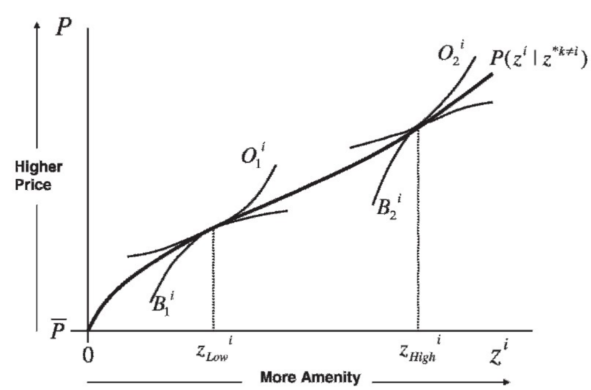
3 Methods

3.1 Hedonic price method and the willingness to pay

Assessing the willingness to pay for avoiding natural disasters is known to be a complicated task. One should not expect credible results when simply asking a focus group their stated preferences regarding risks that involve a very large damage but a very small chance of happening. However, revealed preference in the form of housing transactions could be the solution.

On the other hand, the usage of revealed preference also provides its own issues. First, housing transactions are observed, while generally there is no information on the choice set which was available for a particular transaction. Moreover, imperfect information might be an issue, not all buyers are fully informed of the risk of disaster. Specifically for the present thesis, flood risk maps are publicly available for the studied areas and the publication of these maps was given national media attention. However, some people might still not be aware of the presence of these maps. Finally, collinearity could be an issue. If homeowners in risk areas are less willing to invest in their houses, more expensive houses might not be located in these risk areas.

The observed housing transactions and published flood risk maps can be combined to construct a hedonic price function. Standard hedonic theory (Rosen, 1974) defines a house by its neighborhood characteristics, environmental amenities, and structural attributes. Given an existing stock of houses, the market equilibrium can be defined by a hedonic price function, which relates the market price of a house to these attributes. The hedonic price function is determined by the equilibrium interactions between consumers and sellers in a competitive housing market. An example of a hedonic price function $P(z^i)$ is shown below, in this case representing the relation between housing prices and a positively valued amenity of housing. Keeping all other housing attributes equal, the diagram shows the relation between the price P of property and the amount of amenity Z^i present. Moreover, the bid curves of consumers B_1 and B_2 and offer curves of sellers O_1 and O_2 are shown.



Differentiation of the hedonic price function with respect to a particular attribute, such as flood risk, then yields the implicit price of that attribute, or the willingness to pay. Since housing is not mobile, the value of location amenities should be capitalized in the transaction value of a house. Although limited to only capturing those amenities which accrue to a nearby resident, this method still could provide a partial estimate of the implicit price. (Irwin, 2002)

Endogeneity

Endogeneity occurs within the hedonic price function when a housing (dis)amenity is correlated with the error term in the price function. A variety of issues could cause that to happen but considering the relation between flood risk and housing prices especially omitted variable bias is key. Omitted variable bias occurs when important variables that affect the housing price and are correlated with the included variables are omitted from the price function. This could include neighborhood characteristics, price trends which differ over time and space and of course houses might have certain unobserved qualities. Moreover, flood risk and water related recreation could both affect housing prices in exactly the opposite direction, thereby causing underestimation of the effect of flood risk. In extend of this, areas with higher flood risk might be more likely to be close to water recreation facilities. The same line of thinking can be applied to properties having nice views of water.

3.3 Empirical design, data selection

A detailed dataset on house transactions is used, it contains house prices as recorded by the NVM real estate brokers association. This association manages approximately 80% of all housing transactions in The Netherlands, therefore the dataset is considered as a rather reliable sample of all housing transactions. Besides the transaction price, comprehensive information on the housing characteristics and specifics of the sale is present. The regressions presented in this paper use the variables house size, plot size, isolation level, interior/exterior maintenance level and construction period.

House sales that occurred between 2012 and 2019 are selected based on their location within and near specific dike ring areas adjacent to the main rivers in the Netherlands. These years are chosen on the basis of flood risk information availability in 2012 and 2015. In 2020 new flood risk information was spread in The Netherlands in a different format, therefore the years after 2019 are excluded. Specifically locations along the IJssel, Nederrijn, Waal, Merwede, Lek, Bergsche Maas, Oude Maas and Zoommeer are selected due to their presence within the Ruimte voor de Rivier projects and related documentation. The selected dike ring containing the housing transactions chosen for analysis are shown in appendix I.

The data on house transactions is supplemented by other spatial data generated on the basis of analysis using QGIS. In an intent to address the omitted variable bias in relation to the presence of nearby water bodies, the percentage of water coverage within a 1000m radius of a property is calculated for every housing unit. Furthermore, for the purpose of border discontinuity design, housing transactions are supplemented with a distance to their relative dike ring border.

Previous research within The Netherlands for the identification of the effect of flood risk on housing prices, relied on the maximum height of water levels in case of floods. Availability is great since this data is published often and since many years. However, the maximum flood level is not a great measure of risk since it assumes all flood defenses in the country fail at the same time and it does not consider the possibility of loss of life as evacuation plans are not considered.

Maps as published by Rijkswaterstaat, part of the Dutch Ministry of Infrastructure and Water Management in 2012 and 2015 for the purpose of documenting risks prior and after the Ruimte voor de Rivier projects, provide for every location of a selected house sale the local indidual risk (LIR). This risk is the annual probability that a person at a particular location will decease because of flooding in the area. These calculations include evacuation possibilities, expected time available for evacuation and expected duration of evacuation, into the equation. This gives an idea of the current level of flood protection in the selected area, but most importantly it might be of most interest to consumers looking to buy a house.

The same documentation also provides indicators of the probability of economic losses. Clearly, besides the loss of life this risk is also of significant importance to people. However, these maps show especially high risks in areas where great economic value exists and low risk where few buildings are located. This is of great interest for policy makers but not so much for an individual buying a house. This individual in the end is interested in his own economic loss and losses of his neighbors might not be important for his investment decision. Therefore, for the current paper the LIR is chosen as the preferred measure of risk.

The observed LIR varies greatly, also within dike ring areas. This is due to the varying land heights within the dike ring areas, additional flood risk protection within a dike ring area and the varying risks of flood protection failures applied to the different dikes making up the ring.

Image 1 – LIR 2012

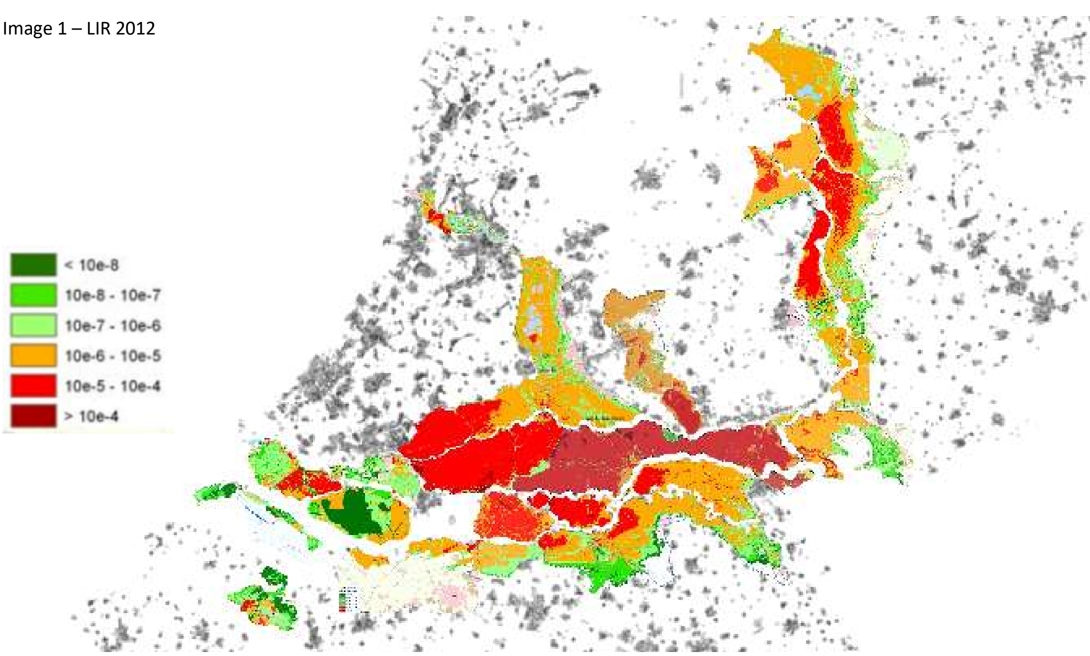
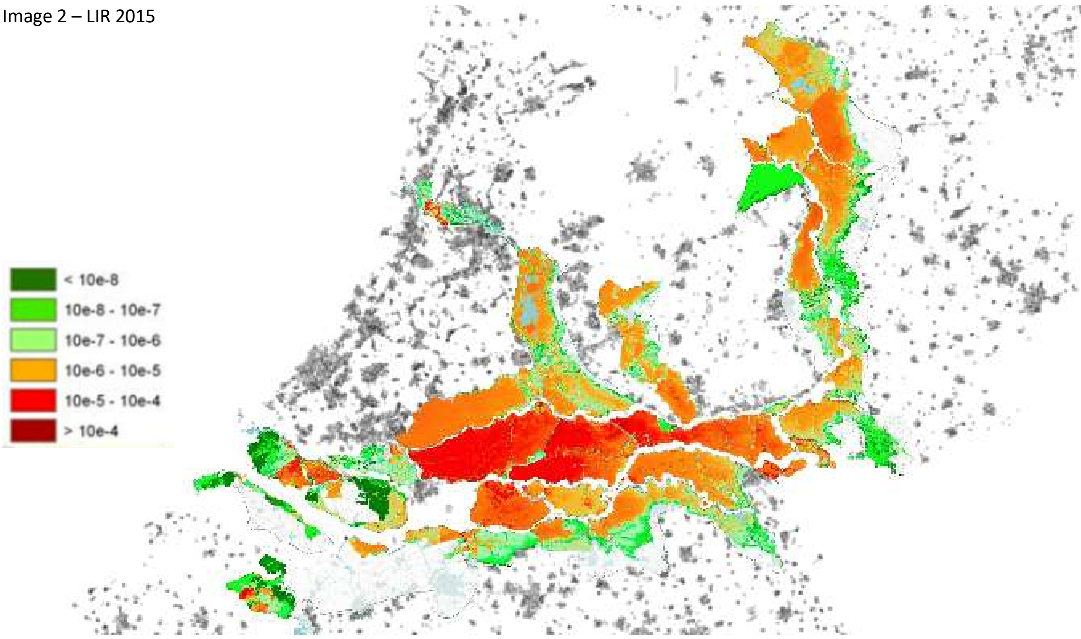


Image 2 – LIR 2015



The images 1 and 2 depict the maps indicating the LIR as published before and after the Ruimte voor de Rivier projects in 2012 and 2015. The difference between the two maps is interpreted as the effect of the Ruimte voor de Rivier project on the LIR once completed.

These maps are published per separate dike ring area during 2012 leading up to the Ruimte voor de Rivier Projects and in 2015 containing the newly calculated risks after the Ruimte voor de Rivier projects. Using QGIS the risk maps are merged and every housing transaction in the NVM dataset is given a LIR risk value based on the transaction year. The years 2012-2014 are considered to be affected by the maps as published in 2012. The maps published in 2015 also include risk as calculated for projects to be finished between 2016-2019. However, as housing buyers are considered to be forward looking, these values are applied on the years 2015-2019. In 2020 new risk maps are published using a different terminology, therefore the years after 2019 are excluded. The chances observed within the selected locations for this study lie between 0,001 and 0,00000001, while locations with 0 risk are also observed (Dijkringrapporten, 2012) (VNK end report, 2015).

The maps as published in 2015 are slightly more detailed and include an additional very high-risk category. To create an equal range of risk values for both sets of data the higher risk value is reduced to the highest risk value as present in 2012/2014 maps. Both data sets, one including a higher risk category and another not including the risk category will be used for analysis further on.

Treating the dataset for outliers and odd values, housing transactions under EUR100,000 per unit, over EUR5,000,000 unit and over EUR7,000/m2 are taken out. I consider that transactions over EUR5,000,000 and over EUR7,000/m2 are extreme outliers for the studied areas, and above all, the price is so high that the amount someone is willing to pay for a reduction of flood risk becomes nihil in relation to the transaction price. Also, it is unlikely that transactions under EUR100,000 represent real housing transactions, this price is very low for The Netherlands. These transactions could also be more technical transactions such as a property split or a transaction with family members for a reduced price. Furthermore, properties with lot size over 20,000m^2 have also been eliminated, it is likely that the land value overshadows the value of the housing unit in these cases. Finally, for properties located in apartment complexes the floor on which these properties are located is unknown. The effect of LIR on especially the apartments located on the higher floors is unclear and therefore all apartment properties are not included in the analysis.

The final dataset includes 130,569 transactions. Each observation represents a specific sale, geographically identified at the address level. As such, it was possible to identify 13,389 sales that took place more than once on the same property within the dataset.

3.4 Empirical design, Identification strategy

Using beforementioned dataset, we identify people's AWTP to avoid flood risk, using the following hedonic specification:

$$ln(P_{it}) = \alpha ln(F_{it}) + \beta X_{it} + \eta_i + \zeta_t + \mu_{it}$$

Where P is the observed price for housing transactions, i denotes the 4PPC/6PPC in which a transaction took place and t the year of the transaction. X_{it} represents the other determinants of house prices at the 4PPC level that could vary over the years, and β is a vector capturing the effect of these controls on housing transaction prices. Furthermore, the error term is composed of a time invariant or spatial error η_i, a time variant error represented by ζ_t and the idiosyncratic error term μ_{it}. F is the main variable of interest, the yearly local individual risk of decease due to flood defense failures (LIR).

A consistent estimate of α is only feasible if the allocation of housing and the related (unobserved) neighborhood characteristics, to areas with higher or lower LIR is random conditional on the applied control variables. I note that any omitted variables determining house prices which are correlated with the LIR would blur the identification of α. Moreover, a misspecification in the linear way we specify X could also lead to a misestimation of α. To address these issues, the spatial detail of the present dataset will be used and the following empirical design will be applied.

First, in the OLS regression analysis, spatial fixed effects at a 4 digit and 6 digit zip code (4PCC, 6PCC) and municipality level are included. Additionally, time fixed effects are applied at a year basis. This model allows to control for time varying and time invariant omitted variables present in the population error within the 4PCC/6PCC area. These time variant variables might include the ups and downs of the housing market over time. While the time invariant variables could be neighborhood characteristics such as the type of neighbors, community taxes and the quality of public services. A combination of municipality*year fixed effects will also be tested.

One of the time invariant omitted variables specifically relevant when determining the willingness to pay for reducing the LIR are water related amenities. It is possible that risks are higher near waterbodies, yet water related amenities also increase getting closer to waterbodies. Therefore, the computed water coverage within a 1,000m range of a housing transaction is used as one of the variables in the analyzed regressions. This covariate is used to investigate whether the amenities that come with the presence of water bodies might affect the observed effect of LIR on housing prices, in an intent to mitigate a possible underestimation of α due to increasing water amenities with increased LIR. Of course, since spatial fixed effects are included, this will resolve only the variance in water coverage present within the 4PCC/6PCC areas.

Furthermore, an attempt is made to regress the flood risk variable only on the available repeat sales of the same property in the dataset. This analysis property ID and year fixed effects is considered the most spatially detailed, exploiting the variance in flood risk due to Ruimte voor de Rivier projects versus exactly the same properties being sold through time. The benefit here is that this type of analysis almost completely removes the issue of omitted variable bias. While maintenance levels and isolation values are included in the regression analysis, some unobserved changes in the same house might of course still have taken place due to a home renovation or significant changes in neighborhood attractiveness. But it is reasonable to assume these unobserved variables are minimal compared to the changes made in flood risk levels.

Border discontinuity design (BDD)

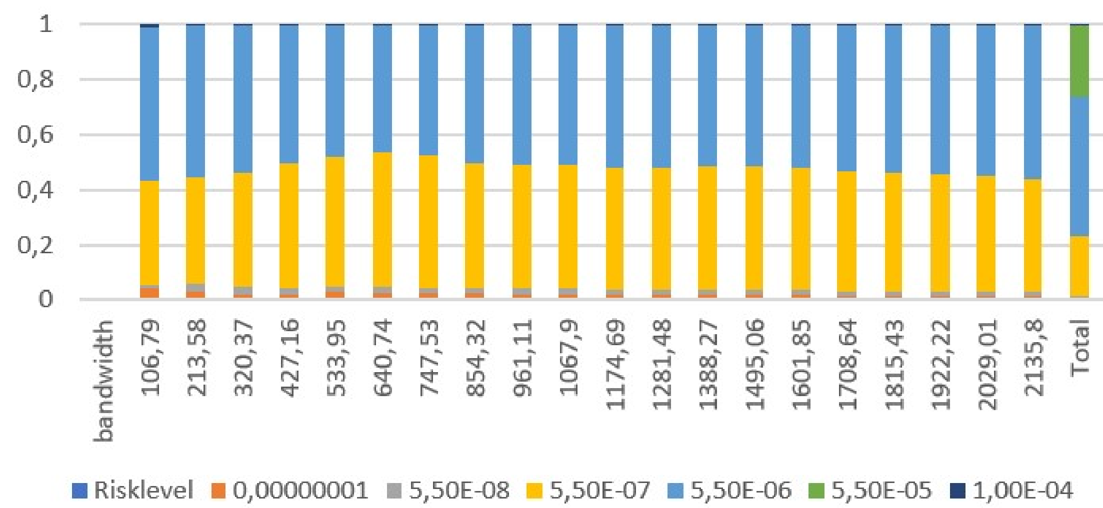
In a final attempt to approximate the effect of flood risk on housing values as precise as possible, a regression-discontinuity type design (RDD) is explored. Exploiting difference in flood risk on two sides of dike rings, a discontinuity in the effect of flood risk on prices could be observed. 3 areas have been selected as depicted in Appendix II. These areas have been selected along dike rings because of the presence of housing transactions not involving flood risk on one side, and housing transactions characterized by flood risk on the other. Moreover, in the selection of these area care was taken to select areas where these transactions of interest also are located very close to the border. Specifically, 24,382 housing transactions are observed within the selected areas, of which 10.260 are part of the control group which are not located in a flood risk area.

Effectively, the willingness to pay is identified, comparing areas where no risk of decease due to flood risk is present, versus areas where a relevant flood risk is present. In the context of LIR as published by The Ministry of Infrastructure and Water Management, relevant Is defined as bigger than $1E^{-3}$. However, in the chosen sample a risk of $1E^{-5}$ is the most prevalent. This strategy of looking at properties along a border further increases the likelihood of comparing homes that differ in their assigned LIR and not by other variables. The causal inferences derived from RDD is potentially more credible than more typical natural experiment strategies such as difference-in-difference or the instrumental variable approach. As Lee et al. (2008) showed, it is not necessary to assume the RDD design isolates the treatment variation that is randomized because this randomized variation is a consequence of an agent's inability to precisely control the assignment variable near the studied border.

In the regressions which are performed to analyse any possible border discontinuity, fixed effects are applied for year multiplied by the dike ring border specific to each of the 3 areas. This way the effect over space can be identified while keeping constant any effect caused by price appreciation over time or difference in unobserved variables between the selected areas.

In BDD the estimated effect can be very sensitive to the choice of bandwidth. A smaller bandwidth area increases the chances of comparing houses which are more similar to each other, while a larger bandwidth allows for including properties located more deeply in the dike ring areas. These properties deeper into the dike ring areas could be characterized by different levels of risk. While cross validation methods and ad-hoc determination of convenient bandwidths are not uncommon, Imbens and Kalyanaraman (2009) developed a method to choose an optimal bandwidth for estimating τ, the observed discontinuity.

An issue with this research method is the possibility that the risk levels observed in the restricted sample located in the selected area are not representative of the entire available sample of housing transactions. The graph I below shows the risk levels present in the samples for different bandwidths applied and for the total sample. The risk levels present in the sample is quite stable for different bandwidth. Clearly the share of 5,5E-6 takes presence deeper into the dike ring after 2135,8m distance from the threshold. It is important to note that this higher risk value being present especially further away from the border might lead to underestimation of the effect of flood risk when choosing a smaller bandwidth. By looking at the results for different bandwidth results, it might be possible possible to shed some light on this possible underestimation.



Graph I – Flood risk level present within a certain bandwidth from the dike ring border, in meters.

Furthermore, the selected transactions run the risk of being exactly those transactions where flood risk is most visible since the chosen borders are dike ring borders directly adjacent to water bodies. Therefore, the effect could be an overestimation. Again, extending the bandwidth used in the analysis might help to estimate the size of this issue, although this effect would be exactly opposite of the previously mentioned possibility of underestimation.

4 Results

4.1.1 Descriptive statistics

I start by showing descriptive statistics on the variables used in the regression analysis. OLS(1) On the left shows the full sample of all 130,569 housing transactions that took place during the 8 years the research covers. RDD(2) Shows the restricted sample used in the RDD design only containing those transactions in the areas selected for this purpose, a total of 24,382 observations. The mean values of the RDD sample are relatively close to the mean of the entire sample. The lot size is somewhat worrisome, it is 1.6 time the standard deviation higher than the mean lot size in the entire sample. Moreover, the mean price is higher but only 0.4 time the standard error of the mean of the total population.

OLS(1)					RDD(2)			
Observations:	130.569				24382			
Variable	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
price	269945	126576	100000	2450000	322943	169957	100000	3040000
average price per m2	2026	571	422	6806	2,328	757	674	6810
size	132	42	32	535	137	47	38	527
lot size	447	119	1	19975	633	158	1	19765
maintanance inside	0.76	0.13	0	1	0.76	0.76	0	1
maintanance outside	0.73	0.13	0	1	0.75	0.75	0	1
isolation value	2.67	1.82	0	5	2.48	1.73	0	5
constructed before 1905	0.04	0.20	0	1	0.04	0.18	0	1
constructed 1906-1930	0.07	0.25	0	1	0.09	0.28	0	1
constructed 1931-1944	0.04	0.20	0	1	0.08	0.26	0	1
constructed 1945-1959	0.06	0.24	0	1	0.09	0.28	0	1
constructed 1960-1970	0.13	0.34	0	1	0.17	0.37	0	1
constructed 1971-1980	0.18	0.38	0	1	0.15	0.36	0	1
constructed 1981-1990	0.14	0.35	0	1	0.11	0.31	0	1
constructed 1991-2000	0.14	0.35	0	1	0.09	0.28	0	1
constructed after 2000	0.18	0.39	0	1	0.18	0.39	0	1
year	2016	2.42	2011	2021	2017	2.91	2011	2021
floodrisk (cat. adjusted)	0.0000145	0.0000219	0	0.0001	0.0000164	0.0000222	0	0.0001
watercoverage (%)	4.79	5.75	0.00	55.34	2.81	4.18	0.00	33.00
absolute distance to border					1643	1500	0	9347

Table 1 – Descriptive statistics

4.1.2 Balanced variables

The following table shows the average of different observable housing attributes for different LIR levels. This should shed light on the ability to balance the observed price determinants across the different LIR levels. Unbalanced price determinants could indicate that the unobserved determinants of house prices are unbalanced as well. Moreover, if the determinants are balanced consistent estimation of the AWTP does not depend on functional form assumptions made (Imbens et al., 2015).

The mean values for all observations for the variables used for regression analysis are indicated in the first column. The subsequent columns show the mean value for these variables grouped per LIR risk category. All risk categories show mean values which are within half a standard error of the mean of the whole population, which means the price determinants used in the following regressions are well balanced over the different risk categories. The water coverage percentage within 1000m of a sold house is an exception, the highest risk category shows a mean which 0.8 standard error in distance to the mean of the whole population. It does indicate that higher risk is especially prevalent in areas with a high presence of water bodies in the vicinity, this would does support the strategy to include this variable in the regressions which will follow, in an intent to disentangle flood risk from water amenities.

LIR	All observations		0	1E-08	5.50E-08	5.50E-07	5.50E-06	5.50E-05	1.00E-04
Obs	130569		16089	3498	2514	26982	54172	27891	287
Variable	Mean	Std. Dev.	Mean	Mean	Mean	Mean	Mean	Mean	Mean
price	269945	126576	289496	265514	256148	283684	261114	264247	250547
avprice sqm	2026	571	2099	2070	1988	2104	1981	1995	1984
size	132	42	137	128	128	134	131	131	126
lotsize	447	1191	527	424	438	414	419	489	437
maintoutside	0.73	0.13	0.70	0.74	0.65	0.71	0.74	0.74	0.76
maintinside	0.73	0.13	0.70	0.72	0.64	0.72	0.74	0.73	0.67
isol	2.67	1.82	2.30	2.34	2.14	2.61	2.76	2.85	2.34
constructed before 1905	0.04	0.20	0.08	0.05	0.06	0.05	0.04	0.03	0.03
constructed 1906-1930	0.07	0.25	0.13	0.10	0.12	0.07	0.05	0.05	0.07
constructed 1931-1944	0.04	0.20	0.08	0.09	0.09	0.06	0.03	0.02	0.02
constructed 1945-1959	0.06	0.24	0.09	0.07	0.10	0.08	0.05	0.05	0.05
constructed 1960-1970	0.13	0.34	0.12	0.17	0.16	0.16	0.13	0.11	0.42
constructed 1971-1980	0.18	0.38	0.17	0.14	0.15	0.15	0.20	0.18	0.15
constructed 1981-1990	0.14	0.35	0.08	0.08	0.11	0.12	0.17	0.14	0.07
constructed 1991-2000	0.14	0.35	0.09	0.12	0.11	0.13	0.16	0.15	0.03
constructed after 2000	0.18	0.39	0.14	0.19	0.11	0.17	0.17	0.24	0.13
year	2016	2.4	2016	2015	2014	2016	2018	2015	2014
water coverage(%)	4.79	5.75	2.90	5.13	6.11	4.77	5.36	4.60	9.52

Table 2

A similar exercise is shown in table 3, this time differentiating the data for different ranges of the running variable used in the regression discontinuity analysis as presented in paragraph 4.3. Since the properties compared in this analysis are located near each other, in theory they should mainly differ in terms of flood risk but not otherwise. Indeed, observing the values in the various bandwidths, similar mean values are observed. Although properties on both sides have a lot size which is significantly bigger than the mean of the entire sample, especially the control group has mean lot sizes which differ more than one standard deviation from the entire sample. Also mean maintenance levels differ more than one standard error from the mean for several bandwidths. These variances might be attributed to the relatively small sample sizes.

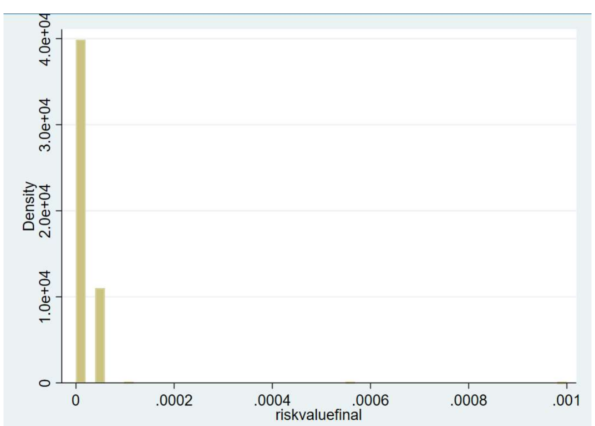
However, it is reassuring that most of the observed variables are well balanced over the different bandwidths that will be used in the regression analysis that will follow. The observed variances are not necessarily problematic values but can be kept in mind when analyzing the results from the regression analysis.

distance to border (m)	All observations		0<>50	-50<>0	0<>100	-100<>0	0<>200	-200<>0	0><400	-400><0	0><1600	-1600><0	0><3200	3200><0
Obs	130569		255	165	566	376	1193	783	4004	2625	10107	5596	16222	10072
Variable	Mean	Std. Dev.	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean
price	269945	126576	321818	365929	312840	358143	311821	346450	297098	332423	295205	352760	280655	359264
average price per m2	2026	571	2323	2491	2272	2491	2263	2460	2212	2381	2187	2486	2104	2510
size	132	42	138	138	138	139	138	137	135	137	135	140	133	142
lotsize	447	119	469	803	554	752	551	636	458	543	395	578	398	680
maintoutside	0.76	0.13	0.61	0.88	0.72	0.89	0.83	0.83	0.90	0.74	0.86	0.79	0.82	0.82
maintinside	0.73	0.13	0.59	0.92	0.71	0.90	0.81	0.89	0.90	0.77	0.82	0.82	0.81	0.87
isol	2.67	1.82	1.47	2.37	1.85	2.24	2.30	2.31	2.41	2.55	2.56	2.39	2.64	2.33
constructed before 1905	0.04	0.20	0.09	0.09	0.09	0.10	0.07	0.08	0.07	0.07	0.05	0.08	0.04	0.05
constructed 1906-1930	0.07	0.25	0.11	0.12	0.14	0.17	0.10	0.18	0.09	0.14	0.07	0.18	0.06	0.15
constructed 1931-1944	0.04	0.20	0.13	0.19	0.12	0.14	0.12	0.15	0.08	0.13	0.05	0.15	0.05	0.14
constructed 1945-1959	0.06	0.24	0.17	0.21	0.16	0.18	0.14	0.14	0.12	0.10	0.07	0.14	0.06	0.13
constructed 1960-1970	0.13	0.34	0.02	0.07	0.07	0.11	0.10	0.11	0.14	0.09	0.15	0.10	0.16	0.16
constructed 1971-1980	0.18	0.38	0.07	0.09	0.11	0.08	0.12	0.09	0.17	0.10	0.15	0.08	0.18	0.11
constructed 1981-1990	0.14	0.35	0.03	0.07	0.05	0.06	0.04	0.08	0.06	0.08	0.06	0.08	0.09	0.08
constructed 1991-2000	0.14	0.35	0.02	0.09	0.02	0.10	0.03	0.10	0.04	0.10	0.06	0.08	0.12	0.07
constructed after 2000	0.18	0.39	0.37	0.08	0.24	0.05	0.26	0.07	0.20	0.18	0.29	0.13	0.23	0.11
year	2016	2.4	2017	2016	2017	2016	2017	2016	2017	2016	2017	2016	2017	2017
watercoverage(%)	4.79	5.75	11.92	2.24	7.30	2.60	5.99	2.08	6.60	3.92	5.33	2.32	4.70	1.35

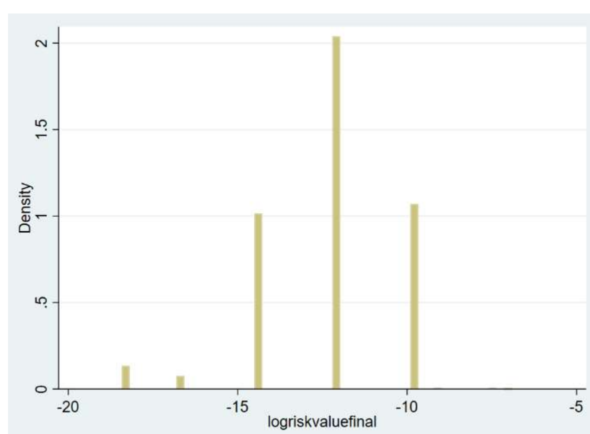
Table 3

4.1.3 Adjusted risk categories and log variables

Analyzing the distribution of the variable used in the regressions to represent flood risk, LIR, it is noted that the variable does not follow a normal distribution. While the mean is 1,3E-5, the median value of this variable is 5,5E-6. Graph 3 shows strong skewness and although hardly visible also shows kurtosis. While it is not essential for approximation of the regression coefficients, it could distort the observed relation in a regression analysis. Moreover, the result of significance tests can be affected. Therefore, a log variable is created for this LIR variable. To do so, the observations with 0 risk have to be removed, these are only 15820 of the total of 130569 observations. In the next paragraph the regular LIR variable and the log variable will be explored. The Histogram of the newly created variable is depicted in graph 4, which does show more similarity to a bell curve, although left skewed.



Graph 2 – Distribution of the density of risk values



Graph 3 Distribution of the density of the log of risk values

An added benefit of this transformation is that the regression of a log risk value on the log price variable greatly improves the interpretation of the results. The results can now be interpreted directly as a percentage on percentage effect.

Furthermore, LIR is represented in publications from 2015 with two additional higher risk categories, which are not present in the publications from 2012, namely 5,5E-4 and 1,00E-03. Therefore, a variable is created merging the highest risk categories with the next available category from 2012. This conversion only affects 133 observations, or 0.1% of all observation, and both variables will be explored in the next paragraph.

Risk category	frequency	percent
0	16089	12.24%
1.00E-08	3498	2.66%
5.50E-08	2514	1.91%
5.50E-07	26982	20.53%
5.50E-06	54172	41.22%
5.50E-05	27891	21.22%
1.00E-04	151	0.11%
5.50E-04	119	0.09%
1.00E-03	14	0.01%
Total	131430	

Table 4 – distribution of risk levels before re-categorization.

Finally, for ease of interpretation of the regular LIR variable a demeaned and standardized variable is created. The LIR varies between 0,001 and 0. Therefore, a regression coefficient indicating the price change for a change in risk between 0 and 1, or better said, a change in risk indicating 100% chance of decease in one year to 0% risk in the same year, become hard to interpret. By subtracting the mean LIR and dividing by the standard deviation, we can obtain a regression coefficient that indicates a price change for change in LIR risk of one standard deviation, or 0.000029.

4.2 OLS

Leading up to a proper estimation of the AWTP for flood risk reduction, table 5 shows the various results obtained from OLS regressions performed on the various variables present in the dataset. The tables shown in this chapter only show the main variables of interest and the entire table is present in Appendix III.

dependent variable: log(Price)

	1	2	3	4	5	6
floodrisk	-0.00454***		-0.00313***		-0.00472***	
-demeaned/stand. dev.	(0.000653)		(0.000661)		(0.000898)	
floodrisk (cat. adjusted)		-0.00900***		-0.00650***		-0.00867***
-demeaned/stand. dev.		(0.000859)		(0.000900)		(0.00118)
4PPC FE	Y	Y	Y	Y	N	N
6PPC FE	N	N	N	N	Y	Y
House ID FE	N	N	N	N	N	N
Year FE	Y	Y	Y	Y	Y	Y
Municipality ID* year FE	N	N	Y	Y	N	N
Obs dropped if risk = 0	N	N	N	N	N	N
Observations	130558	130558	130556	130556	121208	121208
R^2	0.810	0.810	0.813	0.813	0.907	0.908

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 5 – OLS regression with spatial and time fixed effects

First of all, regressions using the LIR levels as presented by Rijkswaterstaat versus log house prices lead to a coefficient α between -0.00472 and 0.00313(column 1, 3 and 5). These results are based on different applications of fixed effects in the model, (1) 4PPC with year fixed effects as the most general specification. However, this specification rests on the assumption that trends in flood risk are uncorrelated with local time varying trends in unobserved determinants of house prices. Therefore, in column 2, 4PPC with year and municipality*year fixed effects is included, in an intent to resolve for this issue.

As an alternative to adding municipality*year fixed effects, 6PCC fixed effects are applied in column 5. This is a significantly smaller area than the 4PCC area of approximately 60m*60m which would greatly reduce the chance of comparing housing with varying unobserved characteristics. However, the coefficient of -0.00472 is not much different from the 4PCC fixed effects result, which would indicate the contrary. The combination of 6PCC fixed effects with municipality*year fixed effects is also tested but results in insignificant regression coefficients and is therefore omitted.

Subsequently, another dataset is used for regression where the LIR levels of post 2015 risk maps are adjusted to the risk levels present in the 2012 maps. This is an adjustment of only 133 observations on the total of 128.463 observations present in the dataset. Therefore, only a small effect on the regression outcomes is to be expected here. Surprisingly, the resulting values nearly doubled to a range between -0.00900 and -0.00650, as can be seen in column 2, 4 and 6 of table 4.

The results of previous regressions using two different datasets with or without recategorized risk levels show great differences. This supports the idea that the estimation of regression coefficients is heavily affected by the non-normality of the flood risk data and is a good reason reason to analyze a regression using the log of flood risk and the log of house prices. Not only do the results become easier to interpret, this also normalizes the distribution of flood risk levels and could lead to a more reliable estimation of the effect of interest.

For the creation of a log variable for flood risk, it is necessary to drop all observations where the LIR is 0. This eliminates 16,089 observations but the remaining 114,684 is still very large. To assure this adjustment does not affect the outcomes of regression very strongly, the regressions of column 3 and 4 are repeated after dropping these observations. The results are presented in columns 7 and 8 in table 5 and the complete table is present in Appendix III. The resulting coefficients are very similar which is taken as an indication that the next step of a log-log regression analysis should lead to reliable results.

dependent variable: log(Price)

	7	8	9	10	11	12
floodrisk	-0.00282^{***}					
-demeaned/stand. dev.	(0.000648)					
floodrisk (cat. adjusted)		-0.00572^{***}				
-demeaned/stand. dev.		(0.000889)				
log(floodrisk)			-0.00432^{***}		-0.00350^{**}	
			(0.000461)		(0.00157)	
log(floodrisk (cat. adjusted))				-0.00432^{***}		-0.00357^{**}
				(0.000462)		(0.00158)
4PPC FE	Y	Y	Y	Y	N	N
6PPC FE	N	N	N	N	N	N
House ID FE	N	N	N	N	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Municipality ID* year FE	Y	Y	Y	Y	N	N
Obs dropped if risk = 0	Y	Y	Y	Y	Y	Y
Observations	114684	114684	114684	114684	11959	11959
R^2	0.814	0.814	0.814	0.814	0.966	0.966

* p < 0.1, ** p < 0.05, *** p < 0* p < 0.1, ** p < 0.05, *** p < 0.01

Table 5 – OLS using log-log specification

Column 9 and 10 show a regression analysis applying 4PPC, year and Municipality*year fixed effects. Moreover, flood risk is now represented by log values of the previously used risk levels. This leads to an effect of -000432, which is the same for both the dataset containing risk levels as presented by Rijkswaterstaat and the dataset with adjusted risk levels. This effect now being the same for both datasets is a strong argument for this specification being robust.

In an attempt to get as close as possible to the effect of interest resulting from exogeneous variation in flood risk only, a repeat sales specification is explored. A similar result to previous regressions is found applying the house ID fixed effects, looking only at repeat sales of the same property. In this repeat sales specification, a log variable for flood risk is also applied and an effect between -0.00350 and -0.00357 is found, which is located in between the two previous estimations. The estimate is significant only at the 5% level, but this is most likely the result of the greatly reduced number of observations. As noted, this type of specification should eliminate nearly all omitted variable bias which makes this estimation perhaps the most precise so far.

Arguably, the results from the 4PCC*year fixed effects are the most reliable, since it is considered to resolve for time varying municipality effects such as changes in taxation or policy changes. Also, the results from repeat sales analysis are very strong, omitted variables are almost completely removed. This comes with the caveat of a smaller number of observation and thus a lower statistical significance. Considering these two results from the regression, between -0.00432 and -0.00350, an interpretation can be made using the standard deviation from the mean. Considering a mean LIR of 0.0000145 and a standard deviation of 0.0000219, a standard deviation change in flood risk from the mean represents a 151.03% change in flood risk and a change in housing prices between 0.52% and 0.63%

4.3 BDD

In the subsequent attempt to estimate the effect of flood risk on housing prices, looking further than simple OLS, a border discontinuity design is explored. Looking only at properties near a dike border, dividing an area with and without flood risk, is another way of estimating an effect with the intent of greatly reducing the chances of unobserved variables affecting the observed effects. It is noted that there could also be unobserved variables present, unobserved reasons why people would prefer one side or the other. However, this method can be a good addition to the previously executed regressions. Three areas along the IJssel, Nederrijn and Waal have been selected for the specific characteristic of flood risk being present on one side of a dike ring but not on the other. These areas are depicted in appendix II. Looking at small bands around the threshold, the expectation would be that properties should be rather similar in many ways. Therefore, if flood risk has no effect on housing prices, only smooth changes in prices over space would be expected and any jump in prices could be interpreted as the effect flood risk has on prices. Moreover, the coefficients found by running a regression for different bandwidths can be interpreted as the effect on prices of the existence of flood risk.

The optimal bandwidth to apply in border discontinuity design can be determined using the theory as developed by Imbens and Kalyanaraman (2012), in this paper they show that the optimal bandwidth is best estimated as:

$$b^* = C_K \, x \left(\frac{\hat{\sigma}_-^2(c) + \hat{\sigma}_+^2(c)}{\hat{f}(c) \, x \, ((\widehat{m}_+^{(2)} - \widehat{m}_-^{(2)})^2 + (\widehat{r_+} + \widehat{r_-})} \right)^{\frac{1}{5}} x \, N^{\frac{-1}{5}}$$

The optimal bandwidth is represented by b^* and the constant C_K is defined as 3,4375. $\hat{\sigma}_-^2$ and $\hat{\sigma}_+^2$ are the conditional variance of P_{it}, the price of a house in dike ring location i in the year t and the assignment variable indicating whether property is located on the risk or non-risk side of the border, respectively and on both sides of the threshold border as indicated by + and -. This, given that $d_i = c$. $\hat{f}(c)$ indicates the estimated density of d_i at c. $\hat{m}_+^{(2)}$ and $\hat{m}_-^{(2)}$ are estimates of the second derivatives of a function of the dependent variable on the distance to the variable d_i. $\widehat{r_+}$ and $\widehat{r_-}$ are estimated regularization terms that correct for potential error in the estimation of the curvature of m(d) on both sides of the threshold.

Since the variance in price for different years and locations is exploited to obtain this estimate, the observed prices are first demeaned using the mean of the specific year and dike ring location. With the conditional variance on the demeaned housing prices and the mentioned formula and estimate can be made. Using distance to the dike ring as a running variable and thereby assuming all properties on one side of the border are characterized by risk and all observations on the other side are not at risk, a bandwidth, b^*, of 106,79m is obtained. However, introducing several covariates in the regressions that follow, adds additional variance. Also, I use a continuous variable for different levels of flood risk, which is not available for all observations on each side of the dike ring. These factors greatly affect the optimal bandwidth. Therefore, this number can be considered as only a guide, several bandwidths will be tested and a trade-off between increased bias due to variance I property characteristics and reduced standard errors due to larger numbers of observations, will have to be made.

Regular ordinary least square(OLS) analysis, applying year*dike ring fixed effects, for a set of bandwidths around the selected dike ring borders is shown in table 6. The entire table with all covariates can be found in Appendix IV. Significant effects at the 1% level can be found at bandwidths largen than 200m and range from -0.0327 to -0.0232. Interestingly, the 50m bandwidth also shows a statistically significant effect of -0.0244.

Dependent variable: Log(price)

Column number	1	2	3	4	5	6	7
Distance to border (two sides)	50	100	200	400	800	1600	3200
floodrisk (cat. adjusted)	-0.0244^{**}	-0.00988	-0.0232^{***}	-0.0272^{***}	-0.0327^{***}	-0.0320^{***}	-0.0267^{***}
-demeaned/stand. dev.	(0.0111)	(0.00856)	(0.00627)	(0.00450)	(0.00297)	(0.00227)	(0.00195)
Dike ring* year FE	Y	Y	Y	Y	Y	Y	Y
Observations	314	666	1372	2872	6238	11324	15653
R^2	0.797	0.737	0.721	0.715	0.730	0.748	0.752

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6 – BDD regression analysis for various bandwidth.

Considering the small number of observations within the 50m bandwidth from the dike ring border and the somewhat odd and statistically insignificant results at the 100m threshold, perhaps the 200m distance band gives a best estimate of -0.0232 or an effect of 2.32% of a standard deviation change in LIR on prices. Although the results further away from the border do not differ much. The slightly increasing effect of flood risk on price when increasing the size of the bandwidth does indicate that underestimation might be an issue when looking only at small bandwidths.

5 Conclusion

A need is established to assess the impact of flood risk on housing prices due to increased risk over time in relation to climate change. By estimating the average willingness to pay through revealed preference techniques, a valuation of flood risk can be approximated. In relation to this the following research question is defined: what is the average willingness to pay of housing buyers for reducing or eliminating flood risk?

During the first regression analysis using OLS estimation and applying various types of fixed effects, coefficients between -0,00996 and -0,00313 are found. This can be interpreted as an effect of -1,00% to -0,31% on housing prices for an increase of 1 standard deviation in LIR risk.

The regression discontinuity design results in an estimated coefficient of -0.00232 or an effect of - 2.32% on housing prices for a 1 standard deviation reduction in risk. This estimation results from a bandwidth chosen with the intention to balance precision of the estimate by limiting variation in housing characteristics on the one hand, versus the size of standard errors by increasing observations on the other.

The data obtained from the Brokers association contains also housing transactions for the most recent year of 2021. This year shows a median house price of EUR429,942.90 for the analysed regions. Considering the effect of a standard deviation change in flood risk, represented by a change in housing prices of 0.63% obtained from OLS and 2.32% by RDD analysis, people would be willing to pay an additional EUR2,708 – 9,974 for a standard deviation change in local individual risk. Assuming a mortgage interest rate of 3.5%, The average willingness to pay would come down to EUR95 – EUR350 per year.

To put this in perspective, the Dutch government currently invests for every citizen EUR400 per year in flood risk management. An amount which is previsioned to grow to nearly EUR500 with the implementation of the National Water Programme. Another way to make use of these results is by looking at the economic effect of complete elimination of flood risk by massive government intervention. Although not very realistic, it gives an idea of the potential economic gain of flood risk reduction. According to the different dike ring reports studied for this thesis, approximately 3.6 million people live in the area of interest, affected by Ruimte voor de Rivier and surrounding the main rivers of The Netherlands. Considering a mean household of 2.2 people and the median house price in the area, building structures and systems to reduce flood risk in the area to virtually zero would result in economic gains of EUR466 billion. One might apply the same line of thinking to an increase of flood risk, for example due to climate change. It might not be most obvious, but rising sea levels also cause the water levels in rivers to rise and reduces the flow speed of these rivers. Concluding, a doubling of flood risk in the river areas around the main rivers due to a changing climate could cost society EUR466 billion.

6 Discussion

Previous research shows that outcomes of estimations of the average willingness to pay vary greatly. In that sense the outcomes of the previously described results are in line with previous research. I note however that the outcomes of the OLS regressions are much closer to the outcomes of studies using larger datasets and applying more solid techniques to reduce omitted variable biases.

This relates to some doubt that exists on generalizability and preciseness of the outcomes from the border discontinuity design. Due to the limitations that I experienced from a limited amount of data outside of the areas affected by Ruimte voor de Rivier Projects, the RDD had to be limited to three relatively small areas. This inevitably leads to a smaller number of observations, which affects the accuracy of the outcomes. Furthermore, it enlarges the chance that variables which are very specific to these areas are picked up and interpreted as an effect of flood risk. Moreover, possible systematic spatial differences in the marginal willingness reduce generalizability. More specific to the observations in this dataset, it is noted that the lot sizes are relatively large versus the entire population and prices on the control side of the dike ring border are also quite high.

The results from the RDD regressions become somewhat larger with an increased sized of the running variable. At the same time, the descriptive analysis has shown that riskier properties tend to be located further away from the dike ring. This might indicate a possible underestimation of the actual average willingness to pay for avoiding flood risk, using this method, and choosing smaller bandwidths.

For the OLS regression generalizability might also be an issue since only areas affected by flood risk from rivers are analysed. People might not perceive risk the same for other types of water bodies than rivers and risk perception might not be linear for different risk levels. For example, Rijkswaterstaat notes that within The Netherlands, flood risk along the rivers is larger than along the coast due to legally established protection levels, yet people perceive the risk to be higher along the coast, perhaps due to the mere size and unpredictability of the sea.

Nevertheless, the outcomes from the OLS regression are still sizeable and significant in both statistic and economic terms. considering to the way the flood maps have been translated to risk levels, marked by categories and not into truly continuous variables, underestimation might have taken place due to the error this could have caused in the independent variable. Therefore, although the found effects are somewhat small, these values might in reality be a bit higher.

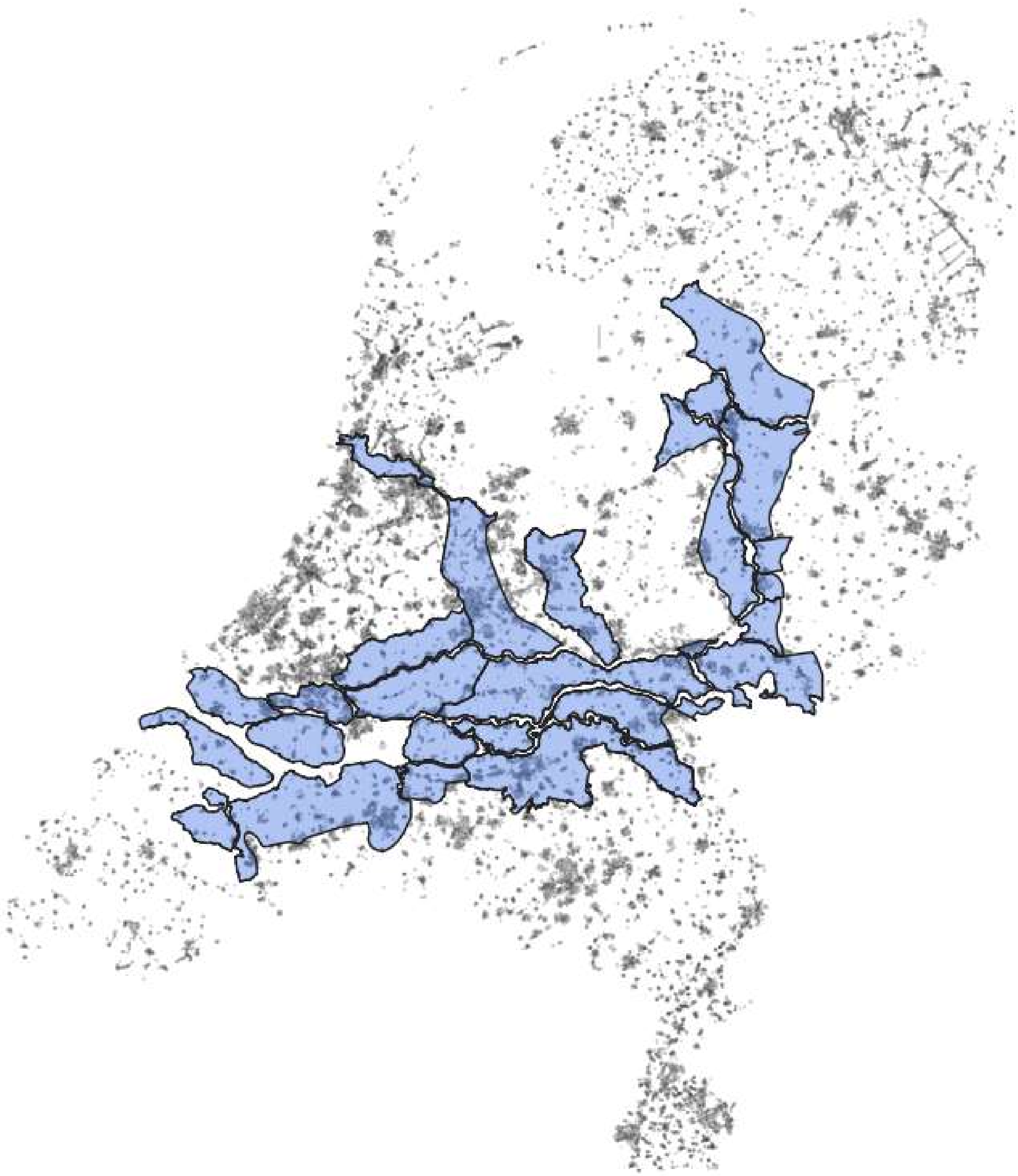
Future research

Clearly more data has shown to lead to better results. Therefore, a similar research setting using the more meaningful LIR could be applied to a dataset spanning the entire Netherlands area. But more ambitious projects are also possible. Since some years flood risk information is available on a worldwide level, using the same measure of flood risk for the entire world. Clearly, data collection on a multinational level and combining this with multinational uniform flood risk data could lead to much more generalizable outcomes for the average willingness to pay for a reduction in flood risk.

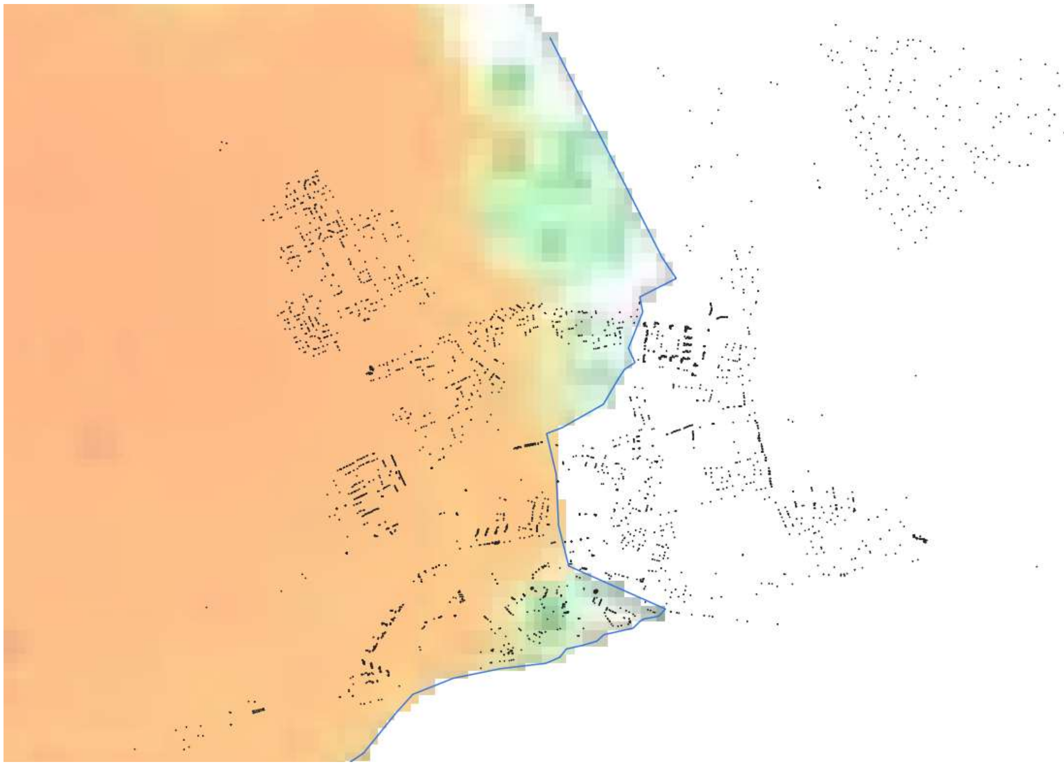
On the other hand, over the last years information on flood risk becomes more widely available and published more frequently within The Netherlands. Sadly, the form of risk indication varies greatly and is not always comparable over the years. Improvement on these measures would enable better comparative research studying the effect of improvements on risk levels regarding flooding.

Finally, I would like to note that regression analysis as performed in this paper is designed to estimate the effect of marginal changes in, for example flood risk. However, the effects of government policy or intervention is often not marginal. As the outcome of these regressions often struggle to address non marginal changes, the estimated outcomes could be severe underestimations. While Rosen's (1974) procedure is the basis of an entire line of literature, it is somewhat problematic due to endogeneity problem inherent to the approach. More recent novel methods as applied by for example Bajari and Benkard (2005) do not recover individual preferences in the more traditional sense. As Bishop and Timmins (2018) indicate the power of a buyer panel including repeat sales information from specific individuals or couples, it allows more rich descriptions of individual preferences. I also believe that especially with a large enough dataset these techniques could greatly improve the estimation of the marginal willingness to pay for reductions in flood risk, despite the caveat that comes with this method, longer time horizons to obtain the data.

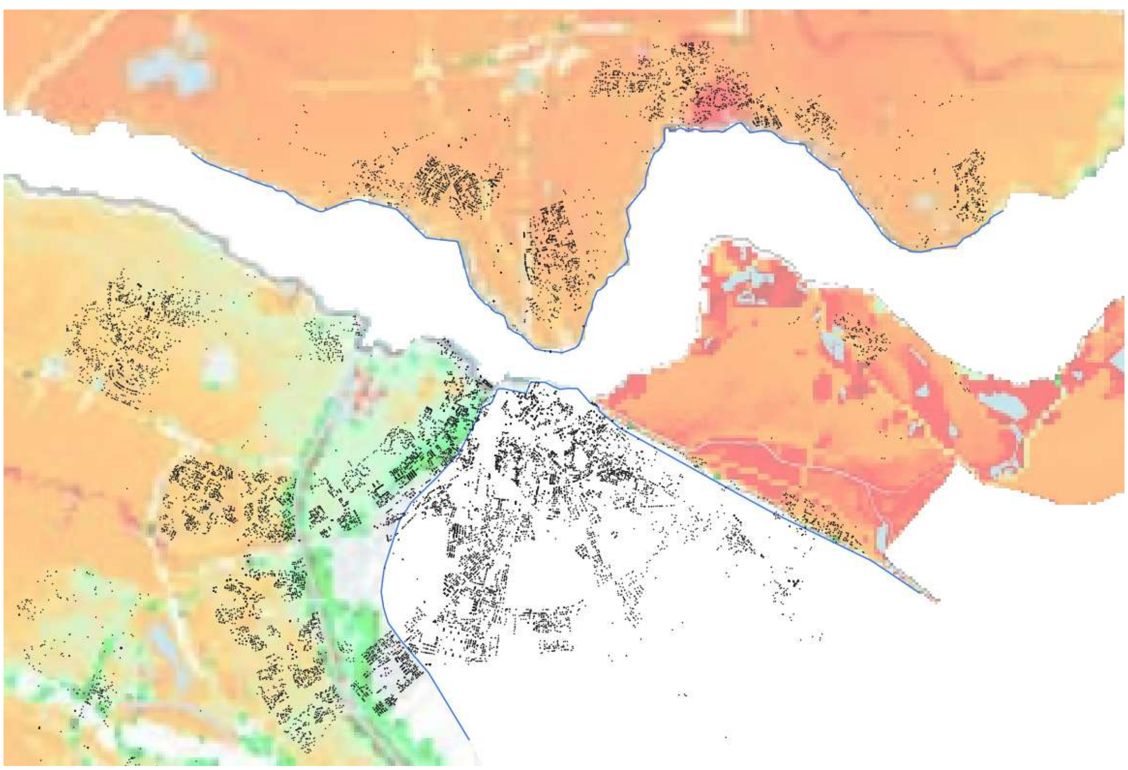
Appendix I – Selected areas for analysis



Appendix II – Selected RDD areas



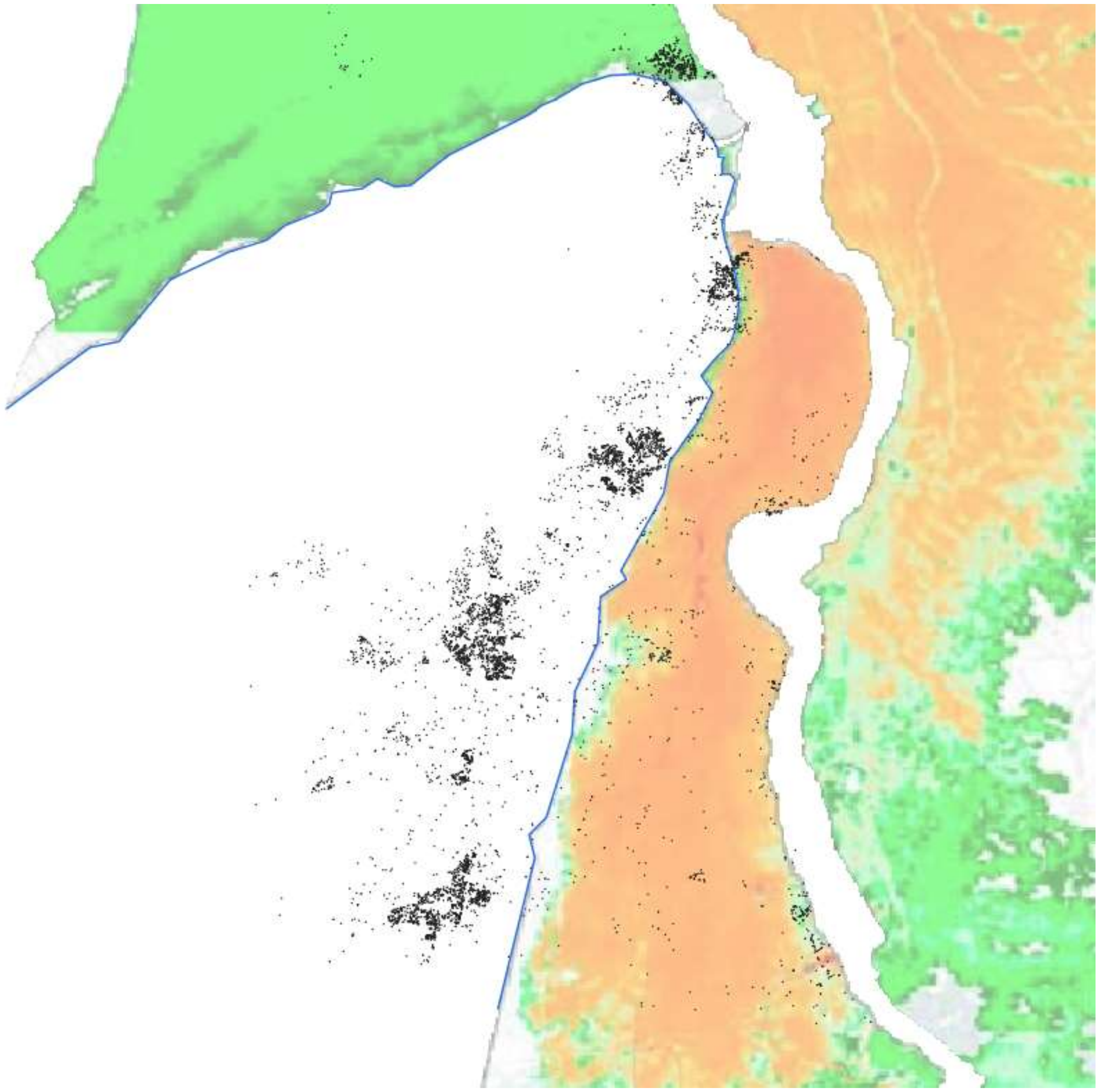
Area 1 – Gelderse vallei



Area 2 – Bommelerwaard, Land van Maas en Waal and Betuwe, Tieleren

Culemborgerwaarden

Appendix II continued – Selected RDD areas



Area 3 – Ijsseldelta and Oost-Veluwe

Appendix III – OLS regressions

dependent variable: log(Price)

	1	2	3	4	5	6	7	8	9	10	11	12
floodrisk	-0.00454***		-0.00313***		-0.00472***		-0.00282***					
-demeaned/stand. dev.	(0.000653)		(0.000661)		(0.000898)		(0.000648)					
floodrisk (cat. adjusted)		-0.00900***		-0.00650***		-0.00867***		-0.00572***				
-demeaned/stand. dev.		(0.000859)		(0.000900)		(0.00118)		(0.000889)				
log(floodrisk)									-0.00432***		-0.00350**	
									(0.000461)		(0.00157)	
log(floodrisk (cat. adjusted))										-0.00432***		-0.00357**
										(0.000462)		(0.00158)
log(floorspace)	0.656***	0.656***	0.657***	0.657***	0.523***	0.523***	0.656***	0.655***	0.655***	0.655***	0.210***	0.210***
	(0.00222)	(0.00222)	(0.00222)	(0.00222)	(0.00259)	(0.00259)	(0.00235)	(0.00235)	(0.00235)	(0.00235)	(0.0294)	(0.0294)
log(lotsize)	0.182***	0.182***	0.182***	0.182***	0.155***	0.155***	0.182***	0.182***	0.182***	0.182***	0.0120***	0.0120***
	(0.000780)	(0.000780)	(0.000777)	(0.000777)	(0.00105)	(0.00105)	(0.000821)	(0.000821)	(0.000822)	(0.000822)	(0.00407)	(0.00407)
watercoverage (%)	0.00117***	0.00108***	0.00121***	0.00114***	0.000176	0.000152	0.00168***	0.00160***	0.00161***	0.00161***	0.0114	0.0114
	(0.000124)	(0.000124)	(0.000124)	(0.000124)	(0.000312)	(0.000312)	(0.000136)	(0.000137)	(0.000136)	(0.000136)	(0.0123)	(0.0123)
isolation value	0.0173***	0.0173***	0.0174***	0.0174***	0.0152***	0.0152***	0.0165***	0.0165***	0.0165***	0.0165***	0.0114***	0.0114***
	(0.000334)	(0.000334)	(0.000334)	(0.000334)	(0.000331)	(0.000331)	(0.000348)	(0.000348)	(0.000348)	(0.000348)	(0.000930)	(0.000930)
maintanance inside	0.0000128***	0.0000128***	0.0000124***	0.0000124***	0.00000979***	0.00000980***	0.00000842**	0.00000840**	0.00000848**	0.00000848**	0.0000115	0.0000115
	(0.00000388)	(0.00000387)	(0.00000386)	(0.00000386)	(0.00000363)	(0.00000363)	(0.00000402)	(0.00000402)	(0.00000402)	(0.00000402)	(0.00000901)	(0.00000901)
maintanance outside	0.00000325	0.00000333	0.00000248	0.00000257	0.00000174	0.00000164	0.00000549	0.00000559	0.00000558	0.00000558	-0.00000518	-0.00000523
	(0.00000400)	(0.00000400)	(0.00000398)	(0.00000398)	(0.00000376)	(0.00000376)	(0.00000414)	(0.00000414)	(0.00000414)	(0.00000414)	(0.00000952)	(0.00000952)
constructed before 1905	-0.0683***	-0.0700***	-0.0694***	-0.0708***	-0.135***	-0.135***	-0.0698***	-0.0711***	-0.0724***	-0.0724***	0	0
	(0.00471)	(0.00471)	(0.00474)	(0.00475)	(0.00628)	(0.00628)	(0.00501)	(0.00501)	(0.00501)	(0.00501)	(.)	(.)
constructed 1906-1930	-0.0863***	-0.0876***	-0.0870***	-0.0880***	-0.148***	-0.148***	-0.0896***	-0.0906***	-0.0919***	-0.0919***	0	0
	(0.00448)	(0.00448)	(0.00452)	(0.00452)	(0.00607)	(0.00607)	(0.00475)	(0.00475)	(0.00476)	(0.00476)	(.)	(.)
constructed 1931-1944	-0.0381***	-0.0393***	-0.0390***	-0.0399***	-0.119***	-0.119***	-0.0375***	-0.0384***	-0.0400***	-0.0400***	0	0
	(0.00470)	(0.00471)	(0.00474)	(0.00474)	(0.00647)	(0.00647)	(0.00500)	(0.00500)	(0.00501)	(0.00501)	(.)	(.)
constructed 1945-1959	-0.110***	-0.111***	-0.111***	-0.112***	-0.159***	-0.159***	-0.110***	-0.110***	-0.111***	-0.111***	0	0
	(0.00450)	(0.00450)	(0.00455)	(0.00455)	(0.00628)	(0.00628)	(0.00476)	(0.00476)	(0.00476)	(0.00476)	(.)	(.)
constructed 1960-1970	-0.143***	-0.144***	-0.145***	-0.145***	-0.151***	-0.151***	-0.146***	-0.146***	-0.147***	-0.147***	0	0
	(0.00424)	(0.00424)	(0.00429)	(0.00429)	(0.00609)	(0.00609)	(0.00444)	(0.00444)	(0.00444)	(0.00444)	(.)	(.)
constructed 1971-1980	-0.125***	-0.126***	-0.126***	-0.126***	-0.117***	-0.117***	-0.127***	-0.127***	-0.128***	-0.128***	0	0
	(0.00416)	(0.00416)	(0.00421)	(0.00421)	(0.00607)	(0.00607)	(0.00435)	(0.00435)	(0.00435)	(0.00435)	(.)	(.)
constr 1981-1990	-0.0715***	-0.0725***	-0.0732***	-0.0739***	-0.0786***	-0.0786***	-0.0749***	-0.0756***	-0.0759***	-0.0759***	0	0
	(0.00420)	(0.00420)	(0.00425)	(0.00425)	(0.00617)	(0.00617)	(0.00439)	(0.00439)	(0.00439)	(0.00439)	(.)	(.)
constructed 1991-2000	-0.00349	-0.00406	-0.00487	-0.00529	-0.0282***	-0.0282***	-0.00442	-0.00483	-0.00504	-0.00506	0	0
	(0.00413)	(0.00413)	(0.00418)	(0.00418)	(0.00579)	(0.00579)	(0.00432)	(0.00432)	(0.00432)	(0.00432)	(.)	(.)
constructed after 2000	0.0241***	0.0238***	0.0231***	0.0229***	0.000382	0.000532	0.0242***	0.0240***	0.0238***	0.0238***	0	0
	(0.00396)	(0.00396)	(0.00401)	(0.00401)	(0.00452)	(0.00452)	(0.00413)	(0.00413)	(0.00413)	(0.00413)	(.)	(.)
4PPC FE	Y	Y	Y	Y	N	N	Y	Y	Y	Y	N	N
6PPC FE	N	N	N	N	Y	Y	N	N	N	N	N	N
House ID FE	N	N	N	N	N	N	N	N	N	N	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Municipality ID* year FE	N	N	Y	Y	N	N	Y	Y	Y	Y	N	N
Obs dropped if risk = 0	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y
Observations	130558	130558	130556	130556	121208	121208	114684	114684	114684	114684	11959	11959
R^2	0.810	0.810	0.813	0.813	0.907	0.908	0.814	0.814	0.814	0.814	0.966	0.966

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix IV – RDD regressions

Dependent variable: Log(price)

Column number	1	2	3	4	5	6	7
Distance to border (two sides)	50	100	200	400	800	1600	3200
floodrisk (cat. adjusted)	-0.0244**	-0.00988	-0.0232***	-0.0272***	-0.0327***	-0.0320***	-0.0267***
-demeaned/stand. dev.	(0.0111)	(0.00856)	(0.00627)	(0.00450)	(0.00297)	(0.00227)	(0.00195)
log(floorspace)	0.707***	0.751***	0.798***	0.773***	0.797***	0.815***	0.806***
	(0.0494)	(0.0329)	(0.0231)	(0.0162)	(0.0108)	(0.00805)	(0.00707)
log(lotsize)	0.115***	0.114***	0.106***	0.114***	0.113***	0.122***	0.128***
	(0.0132)	(0.00957)	(0.00651)	(0.00472)	(0.00320)	(0.00251)	(0.00217)
watercoverage (%)	0.00312*	0.00560***	0.00494***	0.00475***	0.00182***	-0.00229***	-0.00454***
	(0.00189)	(0.00141)	(0.00103)	(0.000740)	(0.000553)	(0.000480)	(0.000444)
isolation value	0.0282***	0.0218***	0.0241***	0.0219***	0.0159***	0.0156***	0.0176***
	(0.00922)	(0.00702)	(0.00481)	(0.00309)	(0.00202)	(0.00146)	(0.00129)
maintanance inside	-0.000184**	-0.0000222	-0.00000291	0.0000397	0.0000431**	0.0000214	0.0000111
	(0.0000824)	(0.0000631)	(0.0000431)	(0.0000290)	(0.0000192)	(0.0000136)	(0.0000111)
maintanance outside	0.000417***	0.000130*	0.0000642	-0.0000281	-0.0000320	-0.0000254*	-0.0000127
	(0.0000909)	(0.0000673)	(0.0000443)	(0.0000297)	(0.0000198)	(0.0000140)	(0.0000117)
constructed before 1905	-0.00360	0.189	0.186	0.205***	-0.0148	0.0546***	0.0268*
	(0.0569)	(0.159)	(0.129)	(0.0752)	(0.0262)	(0.0163)	(0.0157)
constructed 1906-1930	-0.0303	0.102	0.170	0.202***	-0.0205	0.0482***	0.0370***
	(0.0517)	(0.157)	(0.128)	(0.0743)	(0.0245)	(0.0141)	(0.0137)
constructed 1931-1944	0.0221	0.172	0.185	0.209***	-0.00878	0.0820***	0.0806***
	(0.0522)	(0.157)	(0.128)	(0.0741)	(0.0247)	(0.0142)	(0.0138)
constructed 1945-1959	-0.0549	0.163	0.185	0.157**	-0.104***	-0.0354**	-0.0303**
	(0.0518)	(0.157)	(0.128)	(0.0742)	(0.0246)	(0.0142)	(0.0137)
constructed 1960-1970	-0.0113	0.00961	0.0667	0.0717	-0.202***	-0.177***	-0.178***
	(0.0705)	(0.159)	(0.129)	(0.0742)	(0.0243)	(0.0136)	(0.0130)
constructed 1971-1980	-0.185***	-0.0312	0.0177	0.0514	-0.243***	-0.192***	-0.220***
	(0.0595)	(0.158)	(0.128)	(0.0739)	(0.0243)	(0.0137)	(0.0130)
constr 1981-1990	-0.0701	0.0660	0.0920	0.121	-0.137***	-0.0753***	-0.153***
	(0.0641)	(0.159)	(0.129)	(0.0747)	(0.0261)	(0.0149)	(0.0138)
constructed 1991-2000	-0.105*	0.0928	0.183	0.203***	-0.0621**	-0.00641	-0.0208
	(0.0599)	(0.157)	(0.128)	(0.0742)	(0.0259)	(0.0147)	(0.0142)
Constructed after 2000	0	0.0938	0.111	0.129*	-0.106***	-0.0193	-0.0154
	(.)	(0.157)	(0.128)	(0.0735)	(0.0247)	(0.0138)	(0.0135)
constructed after 2000	8.450***	8.070***	7.851***	7.921***	8.093***	7.908***	7.916***
	(0.225)	(0.208)	(0.162)	(0.101)	(0.0532)	(0.0372)	(0.0329)
Dike ring* year FE	Y	Y	Y	Y	Y	Y	Y
Observations	314	666	1372	2872	6238	11324	15653
R^2	0.797	0.737	0.721	0.715	0.730	0.748	0.752

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Bibliography

Aerts, Jeroen C.J.H., and W.J. Wouter Botzen. "Climate Change Impacts on Pricing Long-Term Flood Insurance: A Comprehensive Study for the Netherlands." *Global Environmental Change* 21, no. 3 (August 2011): 1045–60. https://doi.org/10.1016/j.gloenvcha.2011.04.005.

Ahlfeldt, Gabriel M., and Nancy Holman. "Distinctively Different: A New Approach to Valuing Architectural Amenities." *The Economic Journal* 128, no. 608 (February 2018): 1–33. https://doi.org/10.1111/ecoj.12429.

Anderson, Soren T., and Sarah E. West. "Open Space, Residential Property Values, and Spatial Context." *Regional Science and Urban Economics* 36, no. 6 (November 2006): 773–89. https://doi.org/10.1016/j.regsciurbeco.2006.03.007.

. "Open Space, Residential Property Values, and Spatial Context." *Regional Science and Urban Economics* 36, no. 6 (November 2006): 773–89. https://doi.org/10.1016/j.regsciurbeco.2006.03.007.

Bajari, P. and C. L. Benkard (2005). Demand estimation with heterogeneous consumers

and unobserved product characteristics: A hedonic approach. Journal of political economy 113 (6), 1239–1276.

Beltrán, Allan, David Maddison, and Robert J. R. Elliott. "Assessing the Economic Benefits of Flood Defenses: A Repeat-Sales Approach." *Risk Analysis: An International Journal* 38, no. 11 (November 2018): 2340–67. https://doi.org/10.1111/risa.13136.

Beltrán, Allan, David Maddison, and Robert J R Elliott. "Is Flood Risk Capitalised Into Property Values?" *Ecological Economics* 146 (April 2018): 668–85. https://doi.org/10.1016/j.ecolecon.2017.12.015.

Bin, Okmyung, Thomas Crawford, Jamie Kruse, and Craig Landry. "Viewscapes and Flood Hazard: Coastal Housing Market Response to Amenities and Risk." *Land Economics* 84 (February 1, 2008). https://doi.org/10.3368/le.84.3.434.

Bishop, Kelly C, and Christopher Timmins. "Using Panel Data To Easily Estimate Hedonic Demand Functions," n.d., 33.

Bitter, Christopher, Gordon F. Mulligan, and Sandy Dall'erba. "Incorporating Spatial Variation in Housing Attribute Prices: A Comparison of Geographically Weighted Regression and the Spatial Expansion Method." *Journal of Geographical Systems* 9, no. 1 (March 12, 2007): 7–27. https://doi.org/10.1007/s10109-006-0028-7.

Black, Sandra E. "Do Better Schools Matter? Parental Valuation of Elementary Education." *Quarterly Journal of Economics* 114, no. 2 (May 1999): 577–99. https://doi.org/10.1162/003355399556070.

Bosker, Maarten, Harry Garretsen, Gerard Marlet, and Clemens van Woerkens. "Nether Lands: Evidence on the Price and Perception of Rare Natural Disasters." *Journal of the European Economic Association* 17, no. 2 (April 1, 2019): 413–53. https://doi.org/10.1093/jeea/jvy002.

Boyle, Melissa A., and Katherine A. Kiel. "A Survey of House Price Hedonic Studies of the Impact of Environmental Externalities." *Journal of Real Estate Literature* 9, no. 2 (2001): 117–44.

Cavailhès, Jean, Thierry Brossard, Jean-Christophe Foltête, Mohamed Hilal, Daniel Joly, François-Pierre Tourneux, Céline Tritz, and Pierre Wavresky. "GIS-Based Hedonic Pricing of Landscape." *Environmental and Resource Economics* 44, no. 4 (December 2009): 571–90. https://doi.org/10.1007/s10640-009-9302-8.

Chiang Hsieh, Lin-Han. “Is It the Flood, or the Disclosure? An Inquiry to the Impact of Flood Risk on Residential Housing Prices.” *Land Use Policy* 106 (July 2021): 105443. https://doi.org/10.1016/j.landusepol.2021.105443.

Cho, Seong-Hoon, J M Bowker, and William M Park. “Measuring the Contribution of Water and Green Space Amenities to Housing Values: An Application and Comparison of Spatially-Weighted Hedonic Models,” n.d., 42.

Daniel, Vanessa E., Raymond J. G. M. Florax, and Piet Rietveld. “Long Term Divergence between Ex-Ante and Ex-Post Hedonic Prices of the Meuse River Flooding in The Netherlands ∗,” n.d.

Gibbons, Stephen. “Gone with the Wind: Valuing the Visual Impacts of Wind Turbines through House Prices.” *Journal of Environmental Economics and Management* 72 (July 2015): 177–96. https://doi.org/10.1016/j.jeem.2015.04.006.

Hellman, Kelly, Jeffrey Wagner, Daniel Lass, Karl Korfmacher, and Bríd Gleeson Hanna. “Estimating the Economic Impact of Stormwater Runoff in the Allen Creek Watershed.” *Ecological Economics* 145 (March 2018): 420–29. https://doi.org/10.1016/j.ecolecon.2017.11.022.

Hibiki, Akira, and Shunsuke Managi. “Does the Housing Market Respond to Information Disclosure?: Effects of Toxicity Indices in Japan.” *Journal of Environmental Management* 92, no. 1 (January 2011): 165–71. https://doi.org/10.1016/j.jenvman.2010.08.018.

———. “Does the Housing Market Respond to Information Disclosure?: Effects of Toxicity Indices in Japan.” *Journal of Environmental Management* 92, no. 1 (January 1, 2011): 165–71. https://doi.org/10.1016/j.jenvman.2010.08.018.

“History | Dutch Dikes.” Accessed February 1, 2022. http://dutchdikes.net/history/.

Husby, Trond Grytli, and Martijn I Droes. “The Great North Sea Flood of 1953, The Deltaworks and the Spatial Distribution of People,” n.d., 28.

Imbens, Guido, and Karthik Kalyanaraman. “Optimal Bandwidth Choice for the Regression Discontinuity Estimator,” n.d., 27.

Imbens, G. W., & Rubin, D. B. (2015). Causal inference in statistics, social, and biomedical sciences. Cambridge University Press.

Irwin, Elena G. “The Effects of Open Space on Residential Property Values.” *Land Economics* 78, no. 4 (November 2002): 465. https://doi.org/10.2307/3146847.

Kok, Matthijs, J. Vrijling, P.H.A.J.M. Gelder, and M. Vogelsang. “Risk of Flooding and Insurance in The Netherlands,” January 1, 2003.

Lee, David S, and David Card. “TECHNICAL WORKING PAPER SERIES,” n.d., 34.

Lee, David S, and Thomas Lemieux. “NBER WORKING PAPER SERIES,” n.d., 111.

Ministerie van Infrastructuur en Milieu, Unie van Waterschappen & Interprovinciaal Overleg. Veiligheid Nederland in Kaart 2 Overstromingsrisico dijkring 9, 10, 11, 15, 16, 17, 21, 22, 25, 27, 34, 35, 36, 38, 41, 42, 43, 47, 49, 50, 51, 53.

Ministerie van Infrastructuur en Milieu, Unie van Waterschappen & Interprovinciaal Overleg. Veiligheid Nederland in Kaart 2 eindrapport.

Newton, H Joseph, and Nicholas J Cox. “The Stata Journal,” n.d., 36. ———. “The Stata Journal,” n.d., 36.

Nicholls, Sarah. “Impacts of Environmental Disturbances on Housing Prices: A Review of the Hedonic Pricing Literature.” *Journal of Environmental Management* 246 (September 2019): 1–10. https://doi.org/10.1016/j.jenvman.2019.05.144.

Nichols, Austin. 2011. rd 2.0: Revised Stata module for regression discontinuity estimation.

Ologunorisa, T. E., and M. J. Abawua. “Flood Risk Assessment: A Review.” *Journal of Applied Sciences and Environmental Management* 9, no. 1 (n.d.): 57–63.

Pope, Jaren C. “Do Seller Disclosures Affect Property Values? Buyer Information and the Hedonic Model.” *Land Economics* 84, no. 4 (November 2008): 551–72. https://doi.org/10.3368/le.84.4.551.

Rosen, Sherwin. “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition.” *Journal of Political Economy* 82, no. 1 (1974): 34–55.

Samarasinghe, Oshadhi, and Basil Sharp. “Flood Prone Risk and Amenity Values: A Spatial Hedonic Analysis.” *Australian Journal of Agricultural and Resource Economics* 54, no. 4 (2010): 457–75. https://doi.org/10.1111/j.1467-8489.2009.00483.x.

Segura-Beltrán, F., C. Sanchis-Ibor, M. Morales-Hernández, M. González-Sanchis, G. Bussi, and E. Ortiz. “Using Post-Flood Surveys and Geomorphologic Mapping to Evaluate Hydrological and Hydraulic Models: The Flash Flood of the Girona River (Spain) in 2007.” *Journal of Hydrology* 541 (October 2016): 310–29. https://doi.org/10.1016/j.jhydrol.2016.04.039.

Speyrer, JanetFurman, and WadeR. Ragas. “Housing Prices and Flood Risk: An Examination Using Spline Regression.” *The Journal of Real Estate Finance and Economics* 4, no. 4 (December 1991). https://doi.org/10.1007/BF00219506.

Statistiek, Centraal Bureau voor de. “Bestaande koopwoningen; verkoopprijzen prijsindex 2015=100.” Webpagina. Centraal Bureau voor de Statistiek. Accessed February 6, 2022. https://www.cbs.nl/nl-nl/cijfers/detail/83906NED.