UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Master Degree Project in Economics

Property Market Valuation of Flood Risk When Insurance is Free

Steve Berggreen-Clausen

Supervisor: Hanna Mühlrad
Master Degree Project No. 2016:88
Graduate School

June 17, 2016

UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Master's Thesis in Economics, spring 2016:

"Property market valuation of flood risk when insurance is free"

Author: Steve Berggreen-Clausen

Supervisor: Hanna Mühlrad

Abstract

I study the effect of perceived flood risk on property prices in Sweden, before and after a major flood event. Data on property transactions for single-family homes are used in a difference-in-differences spatial hedonic model to study the pre- and post-flood risk discount. I find no significant price discount for the properties in the floodplain, neither before nor after the flood. This stands in contrast to findings from other countries, where prices have been found to drop significantly after a major flood. An explanatory mechanism driving the difference could be the home insurance system of Sweden, where insurance against flood damages is included free of charge, in contrast to countries investigated in previous studies.

1 Introduction

A recent study shows that floods accounted for 40 % of all economic losses from natural disaster for the period 1900-2015 (Daniell et al., 2016). Further, in the last 50 years, floods have caused an estimated 300 000 deaths and adversely impacted 3.6 billion people (EM-DAT, 2015). Due to climate change, projected increases in extreme precipitation events and rising sea levels, along with steady population growth, floods will likely become more frequent and damaging in the coming decades[1]. This thesis aims to contribute to the understanding of how individuals value flood risk, by examining property transaction data for the region around Lake Vänern, in Sweden, which 2000-2001 suffered from a severe flood. The results from such analyses provide an empirical basis for the design of the increasingly important policies that aim to minimize society's vulnerability to floods.

Previous studies on the topic have mostly used property market reactions to flood risk as a proxy for individuals' risk perception and preferences. Most have examined the property market in the US and found mixed evidence for property price adjustments to flood. The strongest effect has been found for areas that recently have suffered from a severe flood, implying that the flood has at least a temporary effect on flood risk perception within the local communities (Bin and . However, the external validity of the findings from a country such as the US to other countries, will among other factors, depend on differences in e.g. the design of the domestic flood insurance system. In Sweden, and a few other countries in the EU, there is no market specifically for flood insurance, as there is in the US. In Sweden, flood damages are included in the home insurance free of charge, which may lead to moral hazard and over-exploitation of construction near water as homeowners living near water do not have to bear the monetary costs of the increased flood risk.

In line with previous literature, I use property transaction data in combination with GIS techniques in order to identify and estimate the effect of flood risk on house prices. During the winter of 2000-2001, Lake Vänern flooded its surroundings, and resulted in one of the most severe floods in Swedish history. In comparison to previously studied flood events, this flood is unique in the length of its duration, which allows me to study not just the effect of the flood on house prices before and after the flood, but also during the flood. I hypothesize that the price discount for flood risk was the highest during the flood, and that it was lower after relative to before the flood. Using a theoretical framework that takes into account differences in insurance costs, I further hypothesize that the price differential for flood-

---

[1] See e.g. IPCC's 5th Assesment Report (IPCC, 2014)

vulnerable properties in Sweden is lower than that estimated for the US. In order to identify the effect of flood risk perception on property prices, I use Difference-in-Differences (DD) estimation where the flood constitutes the exogenous treatment, and the treatment group the properties that lie in the floodplain.

I find that only during the flood was there a significant price discount for floodplain properties, and only for the group consisting of the most vulnerable properties; the relative price reduction for this group of properties is estimated to 47-66 %. However, while previous studies in the US find a significant temporary effect of flood risk on prices after a flood, I find no significant effect neither before nor after the studied flood. An important explanatory mechanism could be the differences in insurance costs to cover flood damages, which are negligible in Sweden. My findings imply that individuals do not take into account neither the tangible nor the intangible costs of a flood to their property, suggesting that regulation of construction of homes with regards to flood risk remains important to prevent over-exploitation in flood-vulnerable areas.

The thesis is organized as follows. In section 2 I review some of the previous literature on the topic, as well as providing a backdrop on insurance policy in Sweden relative to countries like the US. Section 3 contains the theoretical framework used to model how an economic agent values flood risk when purchasing a property. Section 4 contains data and descriptive statistics, while section 5 describes the empirical strategy used. Section 6 contains the empirical results. In section 7 I discuss the empirical results within the provided theoretical framework, and compare my findings to other countries, while section 8 concludes.

I would like to thank my supervisor Hanna Mühlrad for all of the invaluable support I received writing this thesis. I dedicate this thesis to Sepideh and Freya, my wife and daughter, who have been very supportive and understanding during this process.

2 Background

2.1 Previous literature

Most literature on natural hazard risk preferences in the property market are based on the hedonic pricing method, first applied to the property market by Freeman (1974). In his paper, Freeman derives the value of certain characteristics of a good, in this case a property, by exploiting variation in properties’ exposure to air pollution and their respective prices. Given assumptions such as equilibrium in the property market (where all homeowners have maximized their utility), the difference in price between properties at different exposure to air pollution levels equal the marginal willingness to pay (MWTP) for clean air, if everything else is held constant. The benefits of the model are that it is based on revealed rather than stated preferences and is easily applied to empirical analysis policy evaluation. However, Freeman's analysis builds on the critical assumption that the benefits of lower air pollution levels are adequately captured by property market prices. For markets to be efficient, the buyers and sellers must be assumed to have perfect information as well as taking rational decisions, such that they maximize their utility (Debreu, 1959). While Freeman studied air pollution specifically, the hedonic pricing method can be used to model the value of any characteristic of a property, such as e.g. flood risk.

According to the Hedonic Pricing method, an increase in flood risk awareness among homeowners should lead to relative price reductions, so called flood risk discounts, for properties at a relatively higher risk of flood. However, previous studies have found mixed evidence that governmental flood risk information campaigns would significantly affect property prices in a certain direction. Samarsinghe and Sharp (2010) find that release of flood maps to homeowners in New Zeeland reduce the price discount on properties in flood-hazard zones relative to properties outside these zones, suggesting that homeowners may have overestimated flood risk prior to the release of the flood maps. Contrary to the previous, both Shilling et al (1989) and Speyrer and Ragas (1991) found a negative effect of flood risk on house prices following flood-hazard zone designations. However, the latter studied the US, where households are incentivized, and sometimes obliged, to purchase separate flood insurance. This leads to a direct, tangible cost of flood risk and may thus have a greater impact on property prices than the release of information alone.

To investigate whether homeowners take informed decisions or not, qualitative analyses, such as surveys, complement the quantitative analyses cited above. Chivers and Flores (2002) show in a survey sent to individuals in a flood-hazard zone in Boulder, Colorado, that the majority did not have a full understanding of the flood risk in their vicinity, nor the costs of insuring against this risk, at the time of their home purchase. Furthermore, Siegrist and Gutscher (2008) interviewed both individuals with previous flood experience, and individuals without previous flood experience. They found that individuals with no personal flood experience systematically and significantly underestimated the negative effects of a flood. This could explain why the property market tends to react differently to a real flood event, relative to information on flood risk alone.

One of the earliest attempts at estimating the effects of actual flood events on property prices started with Bin and Polasky (2004). In their work, they estimated the effects of 1999 Hurricane Floyd on residential homes in Pitt County, North Carolina using sales data from 1992 to 2002. Using OLS regression involving a floodplain and time period interaction term, they estimate a statistically significant flood risk discount to 3,8 % before the flood, which rises to 8,3 % in the three-year period after the flood, where the difference between the two is also significant. However, the period of study after the flood was limited to three years. Thus, it is not possible to say whether or not the flood had a lasting effect on the property market. Second, OLS does not account for spatial dependence between the observations, a common trait in econometric models of spatially distributed objects, which could lead to spatial autocorrelation. Not taking this into account may lead to inefficient and inconsistent parameter estimates and biased standard errors, making any inferences difficult (Dubin, 1998).

To account for the limitations in Bin and Polasky (2004) and work similar to theirs, Bin and Landry (2013) as well as Atreya et al. (2013) make significant contributions on the topic by extending the post-flood period in their studies, and adjusting their econometric models to account for spatial dependence effects. Bin and Landry (2013) investigate the same area as Bin and Polasky, but with an extended dataset on property sales stretching from 1992 to 2008. In addition they study both the 1999 Hurricane Floyd as well as the 1996 Hurricane Fran as natural experiments to identify variation in property prices following flood events. A theoretical framework for the behavior of an economic agent is provided, that is also used in this thesis, which explains how property prices may vary with what is defined as "subjective flood risk" rather than the actual flood event's probability. Subjective flood risk takes into account two important concepts. First, uninsurable and non-monetary damages are included in the model, such as the risk to human life, tediousness of reparation and restoration following a flood event. Second, they introduce a measure of information that varies with flood risk perception. The authors hypothesize that following a recent flood, individuals will be subject to the "availability heuristic", a behavioral trait first identified by Tversky and Kahneman (1973). The availability heuristic is a behavioral trait which influences decision-making under uncertainty by giving more weight to recent and traumatic events and thus leading to subjective bias. Applied to flood risk, this would mean that individuals might shift towards a stronger (negative) valuation of flood risk following a traumatic flood event, and that this might dissipate over time. In their empirical analysis, a Difference-in-Differences (DD) approach is used in which the effect of flood risk is identified by comparing properties within the floodplain with those outside, while controlling for all observable characteristics that vary across the dataset. Spatial dependence is controlled for using a spatial weights matrix to model the error term, and using Maximum Likelihood for parameter estimation. Similar to Bin and Polasky (2004) they find increases in the price discount for properties within the 100-year floodplain in the first three years following each flood respectively, ranging from 9 % to 13 %. Before the floods, no price discount was found for houses inside the floodplain. The post-flood discount is studied different time decay functions. Regardless of specification, the price discount decreases over time, and vanishes between 5-6 years after the floods, leading to the conclusion that while a traumatic flood does influence agents' decision-making under uncertainty, they do not have a long-lasting impact. Atreya et al. (2013) use a similar empirical strategy to analyze data for Dougherty County, Georgia following the Flint River Flood of 1994. Unlike the former, they use a SARAR model which includes a spatial lag effect, in addition to modeling the error term. In contrast to Bin and Landry, they find a (weakly) significant pre-flood price discount for houses in the floodplain of 9 %. Following the flood, they estimate a further 13-14 % price discount for houses in the floodplain, almost identical to that of Bin and Landry. Similar to Bin and Landry, they test different price decay function specifications, and find that the decay in price discount over time is relatively robust to specification, and ranges between 4 to 9 years. The external validity to countries with different insurance systems than that of the US is limited however. For countries with insurance systems like that of Sweden, there is a lack of knowledge to how the property market responds to the perception of flood risk, and whether this differs or not from other country contexts.

2.2 Flood insurance policy in Sweden

The majority of the previous literature on flood risk perception has studied property markets in the US, where homeowners are obliged to purchase flood insurance from the government in order to cover any damage from floods (FEMA, 2015a). However, in Sweden flood damages are by default included "free of charge"[2] in the typical home insurance. Sweden is climatologically different from US and not subject to the same magnitude of storms and flood events. Furthermore, house construction in Sweden has arguably been subjected to more centralized planning by municipal authorities, which may have reduced exploitation in flood-prone areas. This has allowed insurance firms to offer more equitable insurances that do not

[2] This is a simplification one can make given that the number of houses at risk of flood is very small relative to the number of houses not at risk of flood, such that the average insurance premium paid is not affected by this.

adversely price-discriminate those that are living near water, a policy that has been in place for the last decades (Thorsteinsson et al, 2007). However, with the production of detailed flood maps and projections showing increasing flood risk, there have been calls for more stringent limitation of house construction in the floodplains (Thorsteinsson et al, 2007), some authorities even going as far as to limit house construction in areas that are estimated to be flooded once every 10 000 years[3].

3 Theoretical framework

A convenient method for eliciting the value of a characteristic of a good, such as a house, is the Hedonic Pricing method (HP), first applied to the property market by Freeman (1974) and summarized below. Let the value (measured by its sales price) P_i of a house i be a function of a set of characteristics, grouped into structure- (S), neighborhood- (N) and environmental (E) characteristics:

$$\boldsymbol{P_i = f(S_i, N_i, E_i)} \qquad \textit{\textbf{for i = [1,n]}}$$

Structure characteristics S relate to properties of the structure, such as the square footage, whereas neighborhood and environmental characteristics N and E relate to properties that are contingent on the location of the house, such as e.g. crime rate (a neighborhood characteristics) and distance to nearby water bodies (an environmental characteristic).

The marginal implicit price of a specific environmental characteristic, such as flood risk E_{Flood} is then the first derivative of the price function with respect to the environmental characteristic:

$$\boldsymbol{\frac{\partial P_i}{\partial E_{Flood}} = P_{Flood}(E_{Flood})}$$

Thus, the hedonic model of property valuation implies that one more unit of the environmental characteristic E_{Flood} will lead to a price differential of the house that equals the marginal willingness to pay (MWTP) P_{Flood} for that characteristic (which may be negative for the case of flood risk), ceteris paribus. The advantage of using the hedonic model is that when applied to specific markets, such as the housing market, revealed preferences can be elicited for goods that are otherwise hard to measure in monetary terms. Furthermore, it is conveniently used and interpreted in a regression framework. When using e.g. OLS, the estimated coefficient for the flood risk variable corresponds directly to the hedonic price for this characteristic.

[3] For a policy brief (in Swedish) by a working group of county authorities on this, see Länsstyrelserna (2006).

Using the hedonic pricing framework, MacDonald et al. (1987) provides a model for the relationship between hedonic prices, insurance costs and incremental option value based on expected utility theory. Expected utility theory is used to explain decision-making under uncertainty. The theory explains how an individual's expected utility from a range of potential outcomes varies with the likelihood and risk of the outcome, as well as the individual's level of risk averseness (Schoemaker, 1982). For a risk neutral individual, the choice with the highest expected value directly corresponds to the choice with the highest expected utility. For a risk averse individual, however, risk itself is seen as a cost and thus leads to disutility, which is why the choice with the highest expected utility (using a risk-discounted utility) is preferred, rather than simply the highest expected value-choice. It is assumed that individuals decide their "property location MWTP" and insurance coverage with regards to their perception of flood risk such that it maximizes their expected utility (Macdonald et al., 1987). As noted, however, previous empirical results show that property location MWTP seems to depend heavily on recent flood events, i.e. there exists a subjective bias. Tversky and Kahneman (1973) name this behavioral trait the "availability heuristic", with which an individual tends to perceive risk based on previous personal experiences rather than on facts. Bin et al. (2013) extends the model of MacDonald et al. to include the "subjective probability" (i.e. an individuals perceived flood risk) of flooding as a function of information (e.g. a recent flood), as well as uninsurable losses. Uninsurable losses are defined as non-monetary, intangible damage such as the risk of human lives, loss of invaluable objects etc. In contrast to insurance costs, which are determined on the estimated probability of a flood, the expected utility of uninsurable losses depends on the subjective probability.

The model by Bin et that describes the relationship between flood risk and hedonic price includes both an uninsurable losses term and an insurance costs term:

$$\frac{\partial R}{\partial p} = \frac{V_1(a, \hat{y}) - V_0(a, \hat{y})}{[1 - p(i)]\partial V_0/\partial y + [p(i)]\partial V_1/\partial y} - \frac{\partial I(p)}{\partial p} < 0$$

R is hedonic property price, **p** is the actual probability of a flood, $\mathbf{V_0}$ and $\mathbf{V_1}$ are utility functions of housing attributes **a** and income **y** before and after a flood, where $\mathbf{V_0}$ indicates utility before flood and V_1 after. **P(i)** is the subjective probability of flood and depends on the information level **i**. **I(p)** is the insurance premium paid and is a function of actual flood probability **p**. $\mathbf{V_1 < V_0}$ since not being flooded is strictly preferred to being flooded, independent of insurance costs, and $\frac{\partial I(p)}{\partial p} > \mathbf{0}$ since a higher flood risk will entail higher insurance premiums. Thus, $\frac{\partial R}{\partial p} < \mathbf{0}$ and there should be a negative effect of an increase in flood risk on price. In practice, the model states that given two houses identical in all parameters except flood risk, the house with higher risk of flood should be valued strictly lower than the house with lower risk of flood. The model can be used both for short- and long term effects following a flood event. While the insurance costs depend on flood risk rather than recent flood event incidences, subjective flood probability is likely to be affected by recent floods due to their effect on the availability heuristic.

In Sweden, however, insurance costs for households do not even depend on flood risk. This implies that the second term can be omitted when adapting the model to Sweden[4]. In other aspects, the uninsurable losses remain identical if individuals in Sweden are assumed to be subject to the same form of flood risk information influence and availability bias as individuals in other countries. Thus, the model that forms the theoretical framework of this thesis is written:

$$\frac{\partial \boldsymbol{R_{SWE}}}{\partial \boldsymbol{p}} = \frac{\boldsymbol{V_1(a,\hat{y}) - V_0(a,\hat{y})}}{\boldsymbol{[1-p(i)]\partial V_0/\partial y + [p(i)]\partial V_1/\partial y}} < \boldsymbol{0}$$

Applying the model to Sweden gives a strictly smaller price differential due to variation in flood risk probability ($\frac{\partial R_{US}}{\partial p} > \frac{\partial R_{SWE}}{\partial p}$), holding all else constant. In addition to insurance costs, the magnitude of the difference will depend on other effects that may differ between the countries, such as the severity of the flood (which affects the price differential through the subjective probability p(i)) and cultural and lifestyle differences (which may affect the amount of uninsurable losses due to a flood). Furthermore, I hypothesize that a change in $\boldsymbol{p(i)}$ might occur after a dramatic flood event, which should affect $\frac{\partial R_{SWE}}{\partial p}$**.** In section 5 “Empirical strategy” I discuss how $\frac{\partial R_{SWE}}{\partial p}$ is estimated, before and after a dramatic flood event.

4 Data

4.1 The 2000-2001 flood of Lake Vänern, Sweden

The flood event I investigate in this study is the flood of Lake Vänern in southwestern Sweden, which began in October 2000 and lasted until May 2001 (henceforth “the flood”). Prior to this flood, the lake and its surroundings had been spared from a major flood since 1967. The location of Lake Vänern in Sweden is shown in the map in Figure 1. The flood was

[4] This is based on the assumption that the insurance premium is fixed so that construction of houses within floodplains is small enough to not significantly affect the home insurance premiums on the national scale.

one of the most disruptive floods in Sweden in modern history, and surpassed all previously recorded water levels, including the flood of 1967 (SOU, 2006). The flood lasted for a whole seven months and is unique in comparison to other studied flood events in the literature by its duration, and the fact that water levels raised slowly, increasing by 1 m over the course of three months (SMHI, 2015a). Figure 2 shows a diagram of the flood's progression during 2000-2001, with water levels peaking by the end of January 2001.

Totally, an estimated 280 private properties were reported for insurance claims after the flood (Blumenthal, 2010). However, due to the slow development of the flood, authorities could successfully forecast and prevent future flooding of many more houses within the floodplain (SOU, 2006). The implication of the generally successful effort is that the number of insurance claims likely heavily underestimates the number of properties actually at risk of floods. The total number of houses at risk of flood is much higher, possibly ranging up to a thousand or more. In fact, according to a recent study by Karlstad Universitet in the order of 2,000 buildings are at risk of floods if water levels reach near the levels of the 2000-2001 flood (Andersson et al., 2013).The proportion that consists of privately owned homes is not clear. In order to reduce the risk of a Type II-error, the sample size of transactions of houses in the floodplain should preferably be at least as high as in previous studies, especially since the flood risk price discount in Sweden is likely to be smaller than what previous studies have found for the e.g. the US[5].

4.2 Variables, data sources and data-generating processes

In order to estimate the effects of flood risk, and specifically the effect of the 2000-2001 flood on prices, I have collected geocoded data on property transactions for the period 1998-2003 and 2012-2013 respectively. The observations is on a monthly resolution and cover a period of 34 months before the flood, 6 months during the flood, 31 months after the flood, and a 24-month window 12 years after the flood. The transaction data is limited to private homes, and includes both permanent residency homes as well as vacation houses such as cottages, but excludes apartments[6]. I use single-family home properties rather than industrial or public properties, since it fits well with the theoretical framework, as well as being able to compare my findings with existing literature. Furthermore, the data for property transactions specifically for private homes is especially rich in comparison with e.g. data used in previous

[5] As a comparison, the datasets of Bin & Landry (2013) and Atreya et al. (2013) contain in the order of 300-400 observed floodplain properties sales, when adjusted to the number of years studied in this thesis. My final dataset consists of 60 to 399 floodplain property sales, depending on the definition used (see section 5.2).
[6] Apartments are only rarely on ground level and thus not often subject to flood damages.

research. For example, the data includes a value points indicator used to assess the structural quality of the property, as well as an assessed value ("taxation value") for the house and the lot, that is used for taxation purposes. Unlike the final transaction price, the taxation values are an exogenous source of information since they are predetermined by the tax authorities, and thus not affected by temporary deviations in the perception flood risk.

The main variables are the independent variable house price (time-variant), measured in thousands SEK (kSEK), and variables indicating flood risk (time-invariant). Similar to previous studies, I define flood risk as a binary characteristic, which indicates whether or not a property lies in the floodplain. The floodplain is defined as the area that lies below the maximum water level of a specific flood, often the 100-year flood. I use the maximum water level of the Lake Vänern flood to delineate the floodplain, which also approximately corresponds to a 100-year flood, similar to the definition used in some recent studies[7]. The fact that I use a lake flood event is an important distinction from previous work. Previous studies have examined river floods, which are characterized by a maximum water level that varies along the river. Floodplain location need to be extracted based on the location of the property in relation to an externally produced flood map that takes the water level variation into account. These maps may be a source of measurement error, as the relation between the water level and the elevation of the house or property may be unknown. In the case of lake floods, however, the water level is more or less identical across the floodplain during the flood[8]. This allows me identify floodplain location by inspecting the elevation of each property directly, which reduces measurement error. Another important disparity is the difference in duration of the flood. Normally, river floods occur in the order of hours or days and are generally contained within the levees. The flood studied lasted for several months. Due to backwards flow of water from the lake through the storm sewers, as well as "trapped" storm water behind the levees, floods would be likely to occur even behind the levees, making all properties in the vicinity of the lake below the maximum water level vulnerable to floods.

In order to identify the elevation of the properties, I use a raster containing a highly detailed Digital Elevation Model (DEM), with a cell size resolution of only 2 m[9]. In contrast to using externally produced flood maps, I have a direct sense of the accuracy of the elevation data.

---

[7] Atreya et al. (2013) and Bin & Landry (2013) both use 100- and 500-year floodplain indicator variables.
[8] In the case of Lake Vänern, waves and seiching during the 2000-2001 flood contributed to temporary differences in water levels in the order of 0,3 m between the communities around the lake.
[9] The average elevation error is stated as less than 0,2 m for open spaces [Metria, 2016], which I consider fully adequate for the task.

Using the coordinate data in the geocoded dataset on property transactions as well as the DEM, I extract the elevations for each property. With the elevation data, I generate floodplain variables indicating whether or not the house of property *i* lies in the floodplain ($D_i^{Flood\_House}$) and whether or not the property *i* but *not* the house lies in the floodplain ($D_i^{Flood\_Property}$). Lastly, I generate the variable D_i^{Flood}, which indicates whether or not a property belongs to any of the two variables defined above, to use in the baseline model.

For the variable, $D_i^{Property\_fp}$, only properties where the house lies in the 0 to 1 meter-range above the floodplain are included. The reason to not include all properties where part of the property is at risk of flood (i.e. where parts of the property lies in the floodplain but the house lies > 1 m above it) is because a visual inspection showed that a significant amount of properties are both steep and large, with the house lying several meters above the floodplain, indicating that a flood would likely not cause any adverse effects nor damage to the owners of these properties. Limiting the data to houses lying 0-1 meters selects the most vulnerable of the properties where the house itself may or may not be at a direct risk of flood, while still providing me with a dataset of almost 400 houses, in line with previous work. A graphic explanation is provided in Figure 4, showing three properties at the three different levels of flood risk.

Apart from floodplain location and sales price, a number of important covariates are included in order to increase precision and avoid omitted variable bias (as flood risk will correlate with e.g. proximity to water which will possibly have a positive effect on price). Of special importance are e.g. variables indicating whether or not the property is lakefront, the distance to nearby waterbodies and the elevation value itself, since these may be correlated with both price and flood risk and thus bias the parameter estimates. The covariates are best categorized in the same way as the characteristics of the Hedonic Pricing model: *structure*, *neighboorhood* and *environment* variables where the two latter both depend specifically on the location of the property and are henceforth categorized as *location* variables.

Structural variables describe the physical structure of a house, i.e. the house and lot area, the quality, number of rooms and the type of house (e.g. whether the house is detached or not). Neighborhood variables describe the quality of the neighborhood, and may include distance to the city center, distance to and qualities of nearby schools etc. Environmental variables describe amenities such as whether the house has a lakefront (which, depending on elevation, may or may not be associated with flood risk) and distance to e.g.

highways and industrial zones. Descriptive statistics for all included variables are provided in section 5.3.

Data on historical house prices have been ordered from Lantmäteriet's[10] database of property transactions, specifically for the municipalities along Lake Vänern. Other than coordinates, the data includes important information about the structure, such as the living and lot area of each property, the assessed tax value, the number of value points etc. The location data, in terms of the provided coordinate, is used to generate a set of location-specific covariates, mainly distance to features that may have either a positive or negative impact on price, such as the distance to the nearest waterbody (likely positive) or the distance to the nearest highway (likely negative).

The sample of property transactions from Lantmäteriet originally contained 19,300 transactions, dispersed across the municipalities around lake Vänern. After extracting elevation data from GIS specifically for the coastal region, the sample was condensed to 9,012 observations. Figure 3 shows a map of Lake Vänern, the municipalities encircling it, and the 9,012 observations of floodplain and non-floodplain properties respectively.

4.3 Descriptive statistics of included variables

In Table 1, descriptive statistics are provided for the variables used in the empirical analysis. The statistics are disaggregated into three groups according to the floodplain indicator variables: properties where the house lies in the floodplain ($D_i^{Flood\_House}=1$), properties where the house lies 0-1 m above the floodplain ($D_i^{Flood\_Property}=1$) and properties where the house lies more than 1 m above the floodplain ($D_i^{Flood}=0$).

The dataset consists of a total 9012 observed property sales, of which 399 properties, roughly 4 %, are defined as being in the floodplain. Of these, 60 properties have houses lying below the maximum occurred water level, while 339 have houses lying 0-1 m above it.

On average, floodplain properties tend to sell for higher than non-floodplain ones, despite the fact that non-floodplain houses are both larger and have more value points. A probable reason is that the location of floodplain properties is comparatively more valued by the market than structure, as seen by the higher lot taxation values. Also, the share of

---

[10] The Swedish National Land Survey, a government agency.

properties that lie waterfront is dramatically different between non-floodplain properties and properties, going up from 2 % to 40 %. Previous studies have found that waterfront location typically comes with a premium of about 8-16 % on the property price (see e.g. Bin and Landry, 2013 and Atreya et al., 2013).

Between the two floodplain groups, we see that properties lying partly in the floodplain are considerably higher priced than properties lying fully in the floodplain. We also see that these houses are larger and have more quality points, which could signal that owners might be more willing to invest in houses that both have a somewhat lower risk of flooding, while still lying close to water. Houses that lie too close to the water surface, however, might be too vulnerable to flooding to make it affordable to maintain a high standard. Another indicator of differences in house standard is the "Vacation property" binary variable. Floodplain properties consist of almost 37 % vacation properties, comparing with 20 % where the house is 0-1 m above the floodplain and only 10 % for non-floodplain properties.

5 Empirical strategy

For the baseline models, I use a Difference-in-Differences-approach. This constitutes a quasi-experimental approach in which one group receives a treatment in time period t (the treatment group) while another group (the control group), receives no treatment. Later on, in time period $t+1$, the effect of the treatment is evaluated by taking the measured difference between the treatment and control group at $t+1$ and subtracting by the measured difference in time period t. Any difference between the two groups over time may then be attributed to the treatment. An advantage of the DD approach is that it allows for level differences between the treatment and control group. However, for the DD approach to be valid the assumption of parallel trends is crucial; in the counterfactual scenario (i.e. if there was no treatment), both the treatment and control group should show identical trends in the outcome variable, so that any difference over time can be attributed to the received treatment. Although this assumption cannot be tested, the trends prior to the treatment can be analyzed. Optimally, the two groups should show similar trends in the period before the treatment. Further, there should be no spillover effects, such that only the treatment group receives treatment. This is controlled for by using detailed elevation data and recorded water levels to delineate the floodplain boundary that separates the treatment group from the control group.

In the baseline model, the treatment group consists of houses lying within the floodplain (as specified in the previous section), while houses outside of the floodplain make up the control group. The flood constitutes the treatment and is assumed exogenous to property market mechanisms. The baseline empirical model is defined as:

$$\ln(P_{it}) = \beta_0 + \beta_1 D_i^{Flood} + \beta_2 D_{it}^{post} + \beta_3 D_i^{Flood} * D_{it}^{post} + \beta_4' \boldsymbol{S_i} + \beta_5' \boldsymbol{L_i} + \delta_t + \varepsilon_{i,t}$$

The outcome variable is the natural log of price in kSek, $\ln(P_{it})$, for house i sold at time t. When the dependent variable is in the log form, the parameters are estimated as the relative, rather than the absolute, effect they have on price. For example, the price effect of an increase in flood risk for a given property should realistically depend on the price level itself: for upscale properties, the price discount due to flood risk should be higher than for less-valuable properties that tend to be less decorated.

D_i^{flood} indicates whether or not a house lies within the defined floodplain, while D_{it}^{post} indicates whether or not a house was sold post-flood. Thus, D_i^{flood} explains the price differential between the treatment and control group pre-treatment, while D_{it}^{post} explains the price trend given no treatment. The difference-in-difference estimate, is the parameter estimate β_3 of the interaction variable $D_i^{flood} * D_{it}^{post}$. Controlling for pre-treatment differences D_i^{flood} and the price trend D_{it}^{post}, this interaction variable captures the difference between the treatment and control group over time. If the parallel trends assumption holds, the parameter estimate of $D_i^{flood} * D_{it}^{post}$ will equal the average treatment effect. If found negative and statistically significant it would indicate that the flood incidence had a negative impact on prices of houses within the floodplain.

The baseline model defined above can easily be expanded to include multiple interaction terms for the floodplain variable, e.g. to disentangle the effect of the most vulnerable properties (indicated by $D_i^{Flood\_House}$) from the less-vulnerable properties in the floodplain (indicated by $D_i^{Flood\_Prop}$). Furthermore, the model can be expanded with interaction terms between the floodplain variable and specific years, to model the eventual decay of the flood risk discount over time.

Other variables included are a vector of structural variables $\mathbf{S_i}$ and a vector of locational variables $\mathbf{L_i}$ (which include neighborhood- and environmental characteristics, but excludes flood risk). The functional form of these variables (i.e. the effects might be non-linear) need not be linear, and are evaluated in the Results section. Controlling for structural and locational variables for both the treatment and control group, the only observable difference between the treatment and control group will be the difference in flood risk.

Coefficients β_1 and β_3 will therefore be equal to the pre-flood price discount in percent and the post-flood price discount in percent, respectively.

As a starting point, OLS is used to estimate the model parameters. Thus, in addition to the parallel trends assumption, the Gauss-Markov assumptions should be fulfilled in order for OLS to be the preferred estimator. For the topic of study, the assumptions on strict exogeneity and spherical errors are of special importance. First, omitted variable bias might violate the assumption of strict exogeneity. To minimize this potential bias, all relevant structural and location control variables that might both correlate with floodplain location and affect price are included (e.g. the variable that indicates whether or not the property lies waterfront). Second, the assumption on spherical errors is often violated in models of spatially distributed data, such as properties. The error terms in OLS models of e.g. property prices tend to show spatial autocorrelation, due to the fact that neighborhoods tend to share many similar characteristics that are also unobserved (Dubin, 1998). This can lead to inefficient and inconsistent estimates which may make any inference unreliable (Anselin and Bera, 1998). To test for spatial autocorrelation, I use Moran's I statistic, which is a measure of spatial correlation that ranges from -1 to 1, where the former indicates perfect negative correlation, and the latter perfect positive correlation. 0 indicates no correlation at all. The test can be done on observations directly to describe the spatial dependence processes in the dataset, as well as a regression diagnostics tool if used on residuals from an OLS model. However, the test requires that a model of the error term with an assigned spatial weights matrix (SWM) is set-up, which is used to compare with the null hypothesis of a model with normally and independently distributed error term (OLS). The spatial weights matrix contains pair-wise spatial relationships between all observations, and is thus an *n*-by-*n* matrix where *n* is the number of observations in the dataset. One commonly used measure of spatial relationship, which is used here, is inverse distance, which uses the inverse of the distance between each observed variable as weights in the SWM, based on the logical assumption that objects located near each other share similar unobserved features (Dubin, 1998). The SWM is then used to model the error term, which is written:

$$\varepsilon_{it} = \lambda W_i \varepsilon + u_{it}$$

Where ε_{it} is the modelled error term, u_{it} is a normally and independently distributed error term, W the spatial weights matrix and λ the coefficient . If λ is found significant, the error term shows spatial autocorrelation, and thus the OLS model has potential to be improved in terms of efficiency and consistency by e.g. modeling the error term based on a priori information on spatial correlation. The standard errors can also be corrected using clustering, however, this technique gives unnecessarily conservative measures of the standard errors, relative to alternative specifications that exploit information on the spatial correlation (Gibson et al., 2010). Instead, exploiting the known spatial covariance structure between the observations, the standard errors can be estimated more efficiently using a General Methods of Moments (GMM) estimator which allows for spatial dependence and is robust to heteroskedasticity, also known as the Spatial HAC estimator, where HAC stands for Heteroskedasticity- and Autocorrelation-consistent (Conley, 1999). Since OLS is a special form of GMM when the model is just-identified, point estimates for models estimated using Spatial HAC are identical to those estimated with OLS, while the standard errors differ (Conley, 1999),.

The results are analyzed with regards to sensitivity to chosen spatial weights parameters as well as time period and floodplain variable definitions.

6 Empirical results

6.1 Main results of DD-estimation

Results for four baseline models with floodplain-time period interaction terms, are presented in Table 2. The floodplain variable D_i^{Flood} indicates all properties deemed vulnerable to flooding according to the definition set out in section 4. Year, zip code and municipality fixed effects are included for all models. All models account for spatial autocorrelation, using spatial HAC, as described in section 5 “Empirical Strategy”, with a distance cutoff of 2 km. The value of this parameter is tested, motivated and discussed in section 6.3.1. Model 1 and Model 2 use the full sample, while Model 3 and Model 4 use a sample restricted to properties with distance < 300 m from nearest water body to see whether properties closer to water are more adversely affected by changes in flood risk perception. For Model 2 and 4, the structure and location variables described in Table 1 are included in the regressions, while they are omitted in Model 1 and 3, in order to see which effect they have on the flood risk estimates.

The coefficient of D_i^{Flood} is negative, indicating that there was a flood risk discount prior to the flood event in the order of 4 to 8 %. However, it is only significant for the full sample with control variables. When restricting to the properties closest to water, the estimate is both smaller and becomes insignificant, implying that there can be other factors that affect this estimate other than flood risk perception. Since I control for time effects with yearly fixed effects, the coefficient of D_{it}^{post} is included to control for the specific difference over time between the treatment and control group, thus it is best interpreted in conjunction with the coefficient of $D_i^{Flood} \times D_{it}^{post}$. None of the estimates of the latter is significant, and across all the models the estimate is positive, implying that there was no flood risk discount after the flood.

It is, however, possible that the potential effect of the flood on prices is attenuated by the less-vulnerable properties in the $D_i^{Flood\_Prop}$ treatment group, since the sample size in this group is 339 relative to the 60 observations found in $D_i^{Flood\_House}$. In order to disentangle the effect of flood risk on price by how vulnerable the properties are to flood, D_i^{Flood} is replaced by $D_i^{Flood\_House}$ and $D_i^{Flood\_Prop}$ in the baseline models, with results presented in Table 3. The positive estimates for the interaction terms persist, and the results show that the positive effect is driven by the most vulnerable properties in the $D_i^{Flood\_House}$ group, where the coefficient of the interaction term is positive and significant across all models. Depending on specification, the positive effect on vulnerable properties after the flood ranges from 32 to 34 % excluding control variables. Controlling for structure and location attributes reduces the effect to 21 to 26 %, but it still remains both significant and positive. The fact that there is an increase in price for the most flood-vulnerable properties after the flood is contradictory, especially since the price trend for waterfront properties are controlled for. Introducing structure and location controls reduced the positive effect, and it might be that the distance to nearest water body variable not adequately explains preferences for living near water (see below for a discussion on the specification of this variable), or possibly that the flood-vulnerable properties were indeed flooded and rebuilt or renovated afterwards, which could positively affect the prices. The latter explanation would, however, only be valid for those properties with a sufficiently large lag in the taxation value and value points variables.

Estimate for the structure and location control variables are reported in Table A1 in the Appendix. Covariates were analyzed with respect to their functional form using specification tests. Distance to nearest water body was found to have a non-linear effect on price, and was found most informative when using the log of distance. Both age and lot size were found to be best specified when adding a squared term to the regressions, implying a non-linear effect of age and lot size on price. Living area, however, did not show non-linear effects and is thus included as a linear term, along with the rest of the control variables. The fact that lot size is non-linear while living area is not, could be due to the fact that for many of the properties, lot sizes are very large compared to living area. In fact, the estimates show a concave relationship between lot size and price, implying diminishing returns to lot size as lot size increases.

Of the location variables, the coefficient for Waterfront is found positive and significant for both Model 2 and 4, and indicates a price premium of 26 % for the unrestricted sample, and 13 % for the sample within 300 m from nearest water body. The coefficient of ln(distance) to nearest water body is significant and negative, indicating that it is not only location directly by the water that is valued highly, but also relative proximity to water. The coefficient is estimated to -0.101 to -0.111 for the both models, implying that a 1 % increase in distance to nearest water body will lower price by approximately 0.1 %. The coefficient of Elevation is found insignificant, and this variable is already partly controlled for with the floodplain indicator variable.

Of the structure variables, building and lot taxation values are incredibly strong predictors of price. Depending on model, their estimates range from 0.52-0.56 for building taxation value, and 0.17-0.34 for lot taxation value, implying that a 1 % increase in building and lot taxation value increases price by 0.52-0.56 % and 0.17-0.34 % respectively. Since the taxation values are calculated based on an assessed value of the structure and property, including these likely means that other structural variables become less significant. Notably, living area, one of the most commonly used price indicators, is found insignificant and even negative in sign, when including taxation values. Despite that lot size, value points and age of the house are used in the calculation of the taxation values, these variables are all found to be significant in Model 2. As noted before, lot size shows a positive effect on price, with diminishing marginal effects as the lot size becomes bigger. Value points are found to have a positive effect on price, as expected, with a 1 point increase leading to a 0,4 % increase in price in Model 2, while the effect is positive but insignificant for Model 4. Lastly, age has a significant and negative effect on price in both models, with an expectedly diminishing marginal effect of as age as age increases. In the remainder of the results section, all control variables are included in the regressions presented, but left out of the results tables and discussion, in order to focus on the price effect of floodplain location.

6.1.1 Event study with flexible year binary variables

To investigate whether results are sensitive to the defined time period, the DD-estimation models with full structural and neighborhood control variables are run with yearly dummy variables from the period during the flood and onwards, indicating whether or not a property is sold that year. Since the flood occurred during November 2000 to May 2001, this period is considered a special case and complements the yearly dummy variables as the variable "DURING". Thus, the year 2001 dummy variables is adjusted so that it covers the period from June 2001 and onwards.

Parameter estimates and standard errors are reported in Table 4. As expected, there is a very strong effect for the most flood-sensitive properties during the flood, as can be seen by the estimate for $D_i^{Flood\_House} x\ DURING$, which ranges from -0,472 to -0,662 depending on whether the full or the restricted sample is used, implying that the most vulnerable properties ($D_i^{Flood\_House} = 1$) dropped 47-66 % in value due to their location in the floodplain during the flood compared to before. Furthermore, for this group of properties, the estimates are all lower for the first year after the flood (2001) compared to later years. Interestingly, there is a strong and significantly positive spike in prices 2002, just the year after the flood ended, which partly seems to drive the positive effect seen in the post-flood period as a whole.

For the less-vulnerable properties ($D_i^{Flood\_Prop} = 1$) there is no significant effect on price neither during the flood nor in the years after, except for the year 2012 where there was a positive and significant increase in prices of 34 %.

Overall, the yearly estimates suggest that while there was a strong and negative effect of flood risk on prices for the most vulnerable properties during the flood, this effect rebounded immediately after the flood, and for most years even was found positive.

6.2 Graphical analysis of price trends

To see whether the effect of floodplain on price shows variation over time, as well as to legitimize the use of DD estimation, I graphically analyze price trends for properties inside the floodplain (the treatment group) compared to those outside (the control group). Since the descriptive statistics show that the two groups differ widely in terms of lying e.g. lakefront, and the baseline models show that such characteristics are highly valued, simple price trend comparisons between the groups become too noisy. Instead, I compare the residuals from the pooled sample regressions in the baseline models. Using the residuals, I control for all observed time-invariant differences and thus it makes for a clearer starting point for visual comparisons. Figure 5 shows the by-month average of residuals separately for the treatment and control group, for the period January 1998 to December 2003. For DD estimation to be valid, the trends in the dependent variable should be similar between the treatment and control groups before the treatment was applied. As can be seen in Figure 5, there is seemingly no difference in trends before the flood. Further, apart from the sharp drop in during the middle of the flood, and two later spikes, no overall change in trend can be seen after the flood occurred either. A positive residual means that the model overestimates the price, while a negative residual means that the price is underestimated. Thus, the negative spikes during and after the flood signal that the baseline model underestimates prices specifically for a specific few of the floodplain properties.

In order to further disentangle the variation in price, I constrain the treatment group to properties where the house itself lies in the floodplain, which reduces the treatment sample from 399 to 60 observations. The trends for the limited sample and the control group are seen in Figure 6. While the noise increases due to the smaller sample size, a short but negative trend occurring during the flood can more clearly be seen. However, the negative trend is broken up by both positive and negative spikes following the flood. The apparent heterogeneity in residuals for “House in floodplain” might be explained by heterogeneity in variables such as distance to Lake Vänern, which may affect both the likelihood of a property getting flooded, as well as the perception of flood risk from the buyer’s perspective. This potential heterogeneity is further analyzed in the following section.

6.3 Robustness tests

In this section, the main results from the DD estimation are analyzed with regards to their sensitivity to spatial autocorrelation parameter values, the time variable definitions and floodplain variable definitions. Since the models with structure and location controls give the highest precision and the models that differentiate property’s sensitivity to flood seem the most relevant, Model 6 and 8 both are subject to the robustness tests.

6.3.1 Testing for spatial autocorrelation

One way to formally test whether the data show signs of spatial autocorrelation is to calculate Moran’s I test statistic and see whether or not the null hypothesis of no spatial autocorrelation is rejected. In order to calculate the statistic, a spatial weights matrix (SWM) that describes the spatial relationship between observations must be determined a priori. As the SWM, I use an inverse distance matrix, described previously in Section 6. I create a large number of SWM’s using cutoff value that range from 200 m up to 25 km. Moran’s I is calculated for both the dependent variable ln(price), and the residuals from an OLS model (similar to Model

1, but with homoscedastic standard errors) across this range, and shown as a function of the cutoff distance in Figure 7.

To reiterate, Moran's I is a measure of spatial correlation and takes a value between -1 to 1, where 0 signals that there is no spatial correlation, similar to the regular correlation coefficient. If the dependent variable shows spatial dependence, a model that completely explains the spatial dependence processes would yield a Moran's I close to 0 for the model's reisduals. In Figure 7 Moran's I for ln(price) ranges from 0,51 to 0,34, while for the residuals it ranges from 0,19 to 0,09, as the cutoff distance increases from 200 m to 10 km. Thus, property prices show a relatively high degree of spatial correlation, which increases as the distance between properties is reduced. Meanwhile, spatial correlation between the residuals is considerably smaller than that of ln(price) over the whole cutoff range, meaning that the model does explain a lot of the spatial dependence between the observations. However, spatial autocorrelation is still significant and positive. Regardless of whether Moran's I is calculated for ln(price) or the residuals, all of the statistics are significant at p-values that are too low to write out. The Z-values for the residuals range between 11.20 and 12.82. Thus, the null hypothesis of no spatial autocorrelation in the error term is clearly rejected.

Because of spatial autocorrelation, estimates of the standard errors from OLS estimation will be inconsistent, which is why the Spatial HAC estimator is used instead for all the models.

I use the results from the Moran's I calculation to make the selection of the SWM used in Spatial HAC less arbitrary. As seen in Figure 7, spatial dependence decreases steeply when the cutoff is increased from 200 to 1-2 km. Thereafter, it seems to level out. Thus, spatial dependence seems to be the strongest within a radius of 1-2 km. As a starting point for the Baseline models, 2 km is chosen as the cutoff distance. Since OLS is a special form of GMM when the model is just-identified (Conley, 1999), point estimates for models estimated using Spatial HAC are identical to those estimated with OLS, while the standard errors differ.

To see whether the standard errors are sensitive to different distance cutoff values, I estimate the model using 500 m and 5 km as cutoffs, which are reported in Table 5. Regardless of whether the distance cutoff value is changed for the model with the full or the distance-restricted sample, the standard errors only slightly change, and not enough to change significance status of any of the models' point estimates. Testing for trend differences prior the flood event

6.3.2 Testing pre-treatment differences in trends

In order to test whether there is any difference in trends between the treatment and control group prior to the flood I interact the floodplain indicator variable with year dummies for the period before the flood, 1998-2000, where $D_i^{Flood\_House}$ signifies the price discount for houses in the floodplain in year 1998, $D_i^{Flood\_House} \times Y1999$ significes the price discount in 1999 and $D_i^{Flood\_House} \times Y2000$ signifies the price discount in 2000. The Y2000 dummy only covers January-September 2000, since the flood began in October that year. If the estimates for the pre-flood interaction terms are significant, this would indicate a difference in trends prior to the flood. A difference in trends prior to the treatment might invalidate the parallel trends assumption, since any differences over time might be attributed to pre-treatment causes. The results from the regressions are shown in Table 6. Judging by the insignificant parameter estimates for the interaction terms, I find no significant trend difference between the treatment and control group before the flood.

6.3.3 Sensitivity to elevation cutoff values for the floodplain variables

I conclude the robustness tests by investigating whether the results are sensitive to modest adjustments of the floodplain variables. For $D_i^{Flood\_House}$, I define $D_i^{Flood\_House\_Adj}$ as houses that lie -0,5 to +0.5 m above the maximum water level, and for $D_i^{Flood\_Prop}$ I define $D_i^{Flood\_Prop\_Adj}$ as houses that lie +0.5 to +1.5 m above the maximum water level. If the results are robust, the estimates for the first group, $D_i^{Flood\_House\_Adj}$, should not change significantly compared to the second group $D_i^{Flood\_Prop\_Adj}$. As before, I use model 6 and 8 to investigate sensitivity to elevation cutoff values. Results are reported in Table 7.

Increasing the elevation cutoff values by 0,5 m for both groups reduces both the pre-flood discount, making it insignificant, and reduces the increase seen for the $D_i^{Flood\_House}$ group in Table 3, making also this estimate insignificant. Thus, the estimated effects are very sensitive to the specified elevation cutoff that defines the floodplain variables. The fact that the pre-flood discount becomes smaller and insignificant for the $D_i^{Flood\_House}$ group in Table 6 can be explained by the fact that the lowest, and thus most vulnerable, properties are omitted, however, it does not explain why the estimated effect is much lower after the flood, relative to the estimates in Table 3.

7 Discussion

7.1 Comparative analysis of results

My empirical analysis fails to show a negative effect of the flood on house prices, and for the most vulnerable properties specifically, the effect is even positive and significant, even when controlling for changing preferences in living near water. This may suggest that the value of non-insurable losses in the theoretical model is much lower than insurance costs (alternatively insurable losses, if insurance is not purchased), since the insurance costs in Sweden are negligible. To investigate if the flood price discount in Sweden differs from countries where flood insurance is not free, results from five previous similar studies are summarized in Table 8. Property markets in the US and the Netherlands have reacted more strongly to flood risk after a flood, relative to Sweden. Notably, there seems to be little evidence of property prices adjusting to flood risk before the flood, independent of the insurance system of the studied countries.

7.2 Policy implications

The fact that there are signs that the property market in Sweden does not adjust to flood risk neither before nor even after a traumatic, recent flood event has important implications for city planning, construction regulation and insurance policy. If individuals do not take flood risk into account when purchasing homes, and location near water continues to be an attractive feature, it could lead to overexploitation (from a societal point of view) in the floodplain, leading to moral hazard and social dead-weight loss. Regulation of where homes can be constructed may reduce this dead-weight loss, as individuals are restricted from building new homes in flood-prone areas. In fact, municipalities of Sweden already regulate where individuals may or may not construct new buildings depending on flood risk. This potentially lowers both insurable and non-insurable costs of floods to society, while also limiting people’s ability to live near water. The findings of this study seem to partly justify current policy, given the design of the insurance system. Optimal policy should arguably take into account preferences for living near water, as well as potential damages from floods, so as to regulate construction of new homes in a way that is socially optimal. Should a premium be put on flood insurance similar to the US, it is possible that the property market also will react similarly. In that case, price reductions of properties in the floodplain could lead to reduced willingness to construct and purchase further properties there, which would diminish the importance of municipal regulation. Although previous studied showed a significant flood risk discount, this only occurred after a flood event, and only temporarily, meaning that even for these policy settings regulation may be justified in order to prevent overexploitation in the floodplain.

7.3 Potential limitations

An important limitation that is shared with most previous studies is that it has not been possible to differentiate flooded from non-flooded properties. Thus, it is not known whether changes in prices were driven by changes in flood risk perception or property flood damages. To the author's knowledge, the only study that has been able to differentiate flooded- with non-flooded properties in the floodplain after a flood is Atreya and Ferreira (2015), in which it is found that only those properties that lie in the flood's inundated area were subjected to flood risk discount. In principle, this thesis uses a similar approach in which the actual flood water level is used. However, due to the slow rise in water levels in the Lake Vänern flood, authorities could successfully protect most of the vulnerable properties. This prohibits me from knowing which properties were flooded, and it may have made it harder for locals to get an understanding of which properties were most at risk of being flooded. Nevertheless, including properties that were sold at lower prices due to flood damage would only bias the flood risk discount upwards, and since no significant flood risk discount is found after the flood, the implications of the findings remain the same.

For the estimated price discount during the flood, which was found statistically significant at the 1 % level when restricting distance to nearest water body, it is worth to mention that the estimations is based on only 5 properties in the treatment group for this period.

One potential limitation that could be unique to this study, is that there may be a difference in how the flood event was perceived, relative to those previously studied. In the US, some of the floods were catastrophic, some happened due to hurricanes, and some involved the loss of human lives. While the Lake Vänern flood is considered the worst flood in modern history in Sweden, no loss of lives occurred, and due to the relatively successful emergency flood defences, it is possible that it left local inhabitants less traumatized, and through the availability bias mechanism had a relatively smaller impact on the property market. Thus it is possible that the differences seen in Table 8 are not only due to differences in insurance policy, but also differences in flood severity. However, since the pre-flood discount is also insignificant (similar to previous studies), this potential limitation does not affect policy implications, since policies are implemented regardless of recent flood events.

Although I use detailed elevation data to separate the treatment group from the control group, it is possible that price changes due to flood risk spilled over to nearby properties in the control group through property market dynamics, where the price of a property affects other properties nearby, and vice versa. This “spatial lag” effect may be investigated using a spatial autoregressive specification with autoregressive and heteroskedastic disturbances (SARAR). However, this has been outside the scope of this thesis.

8 Conclusions

In this thesis I study the effect of flood risk on price for properties around Lake Vänern, before and after the lake flooded in 2000-2001. I find no significant effect of flood risk on price, neither before nor after the flood. The results stand in contrast to previous work from the US and Netherlands which show significant price discounts for floodplain properties after, but not before, major flood events. One important explanatory mechanism could be that the incentive structures differs between the countries; in Sweden flood insurance is included in one’s home insurance free of surplus charge, whereas it is paid for in the US and Netherlands. The fact that monetary flood damages are covered by insurance seems thus to offset the effect that a dramatic flood has on the MWTP for flood risk. The finding that the property market does not “price in” flood risk neither before nor after a major flood event adds support to current regulatory policies, which prohibit the construction of homes in areas susceptible to floods. However, the potential limitations of the study, as discussed above, should be taken into account if these results are to be used for future policy design.

Suggestions for future studies could be to complement this empirical analysis with a survey to homeowners in the same region, in order to see if stated flood risk preferences are in line with the revealed ones. The empirical analysis could also be extended to include specifications using a SARAR model to take into account the possibility of a spatial lag effect.

The results from this study could serve as input to future cost-benefit analyses (CBA) on e.g. investments in flood control measures. Current CBA models face difficulties with both intangible costs, such as the cost uninsurable flood damages (Messner and Meyer, 2005), and benefits, such as the value of living near water. The results from this study, as well as previous ones on the topic[11], seem to suggest that uninsurable flood damages at least are significantly less valued by homeowners than insurable damages, while living near water is highly valued.

[11] Bin and Landry (2013) and Atreya et al. (2013) both find that the property price discount due to flood risk approximately equals the present value of flood insurance costs.

References

Andersson, J-O., Blumenthal, B., Nyberg, L., 2013. *Kartering av översvämningsrisker vid Vänern*. Karlstad: Universitetstryckeriet.

Anselin, Luc, and Anil K. Bera. 1998. Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. In: A. Ullah, D.E.A. Giles, ed. 1998. *Handbook of Applied Economic Statistics*. New York: Marcel Dekkers pp.237-290

Atreya, A., Ferreira, S., Kriesel, W., 2013. Forgetting the flood?: An Analysis of the Flood Risk Discount over Time. *Land Economics*, 89(4), pp.577-596.

Atreya, A., Ferreira, S., 2015. Seeing is believing? Evidence from Property Prices in Inundated Areas. Risk Analysis, 35(5), pp.828-848.

Bin, O., Landry, C.E., 2013. Changes in implicit flood risk premiums: Empirical evidence from the housing market. Journal of Environmental Economics and Management, 65(3), pp.361-376.

Blumenthal B., 2010. *När Vänern svämmade över*. [pdf] Karlstads universitet. Available at: <http://www.kau.se/sites/default/files/Dokument/subpage/2009/08/n_r_v_nern_sv_mmade_ver_final_version_pdf_78126.pdf> [Accessed 20 November 2015].

Chivers J., Flores, N.E., 2002. Market Failure in Information: The National Flood Insurance Program. *Land Economics*, 78(4), pp.515-521.

Conley, T., 1999. GMM Estimation with Cross Sectional Dependence. *Journal of Econometrics*, 92(1), pp.1-45.

Daniell, J., Wenzel, F., Schaefer, A., 2016. The economic costs of natural disasters globally from 1900-2015: historical and normalised floods, storms, earthquakes, volcanoes, bushfires, drought and other disasters. Geophysical Research Abstracts, Vol. 18 EGU2016-1899.

Dubin, R.A. 1998. Spatial Autocorrelation: A Primer. *Journal of Housing Economics*, 7(1998), pp.304-327.

EM-DAT, 2015. EM-DAT Database. [online] Available at: <http://www.emdat.be/advanced_search/index.html> [Accessed 29 November 2015].

FEMA, 2015a. About The National Flood Insurance Program: Overview. [online] Available at: <https://www.floodsmart.gov/floodsmart/pages/about/nfip_overview.jsp> [Accessed 15 December 2015].

FEMA, 2015b. About The National Flood Insurance Program: When Insurance Is Required. [online] Available at: <https://www.floodsmart.gov/floodsmart/pages/about/when_insurance_is_required.jsp> [Accessed 15 December 2015].

Freeman, A.M., 1974. On estimating air pollution control benefits from land value studies. *Journal of Environmental Economics and Management*, 1(1), pp.74-83.

Gibson, J., Bonggeun, K., Olivia, S., 2010. Can we trust cluster-corrected standard errors? An application of spatial autocorrelation with exact locations known. Working paper…. http://researchcommons.waikato.ac.nz/handle/10289/4945

Hallstrom, D.G., Smith, V.K., 2005. Market responses to hurricanes. *Journal of Environmental Economics and Management*, 50(3), pp.541-561.

IPCC, 2014: *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

Länsstyrelserna, 2006. *Översvämningsrisker i fysisk planering*. [pdf] Länsstyrelserna. Available at: <http://www.lansstyrelsen.se/uppsala/SiteCollectionDocuments/Sv/publikationer/2006/oversvam-fysisk-planering.pdf> [Accessed 19 November 2015].

MacDonald, D.N., Murdoch, J.C., White, H.L., 1987. Uncertain Hazards, Insurance, and Consumer Choice: Evidence from Housing Markets. *Land Economics*, 63(4), pp.361-371.

Mays, L.W., Koutsoyiannis, D., Angelakis, A.N., 2007. A brief history of urban water supply in antiquity. *Water Science and Technology: Water Supply*, 7(1), pp.1-12.

Messner, F., Meyer, V., 2005. http://cretscmhd.psych.ucla.edu/nola/volunteer/FoundationReports/DP_13-05_Flood_damage_and_vulnerability_Messner_Meyer_2005.pdf

Metria…. http://www.metria.se/Global/Produktblad/Infoblad_Nationell_Hojdmodell_web.pdf

Samarasinghe, O., Sharp, B., 2010. Flood prone risk and amenity values: a spatial hedonic analysis. *The Australian Journal of Agricultural and Resource Economics*, 54, pp.457-475.

Schoemaker, P.J.H., 1982. The Expected Utility Model: Its Variants , Purposes, Evidence and Limitations. *Journal of Economic Literature*, 20(2), pp.529-563.

Shilling, J.D., Sirmans, C.F., Benjamin, J.D., 1989. Flood Insurance, Wealth Redistribution, and Urban Property Values. *Journal of Urban Economics*, 26(1989), pp.43-53.

Siegrist, M., Gutscher, H., 2008. Natural hazards and motivation for mitigation behavior: people cannot predict the affect evoked by a severe flood. *Risk Analysis*, 28(3), pp.771-778.

SMHI, 2015a. Hög vattennivå i Vänern 2000-2001. [online]Available at: <http://www.smhi.se/kunskapsbanken/hydrologi/hog-vattenniva-i-vanern-2000-2001-1.26795> [Accessed 20 November 2015]

SMHI, 2015b. Fakta om Vänern. [online]Available at: <http://www.smhi.se/kunskapsbanken/hydrologi/fakta-om-vanern-1.4732> [Accessed 20 November 2015]

SOU, 2006. SOU 2006:94 Översvämningshot – Risker och åtgärder för Mälaren, Hjälmaren och Vänern. Stockholm: Edita.

Speyrer, J.F., Ragas, W.B., 1991. Housing Prices and Flood Risk: An Examination Using Spline Regression. *Journal of Real Estate Finance and Economics*, 4(1991), pp.395-407.

Thorsteinsson D., Semadeni-Davies, A., Larsson, R., 2007. Planning for river induced floods in urban areas – experiences and key issues for Sweden. In: S. Begum, M.J.F. Stive, J.W.Hall, ed. 2007. *Flood Risk Management in Europe – Innovation in Policy and Practice*, Amsterdam: Springer.

Tversky, A., Kahneman, D., 1973. Availability: a heuristic for judging frequency and probability. *Cognitive Psychology* 5(1973), pp.207-232.

Tables

Table 1: This table contains the descriptive statistics disaggregated over the treatment and control groups.

Table 1: Descriptive statistics

	$D_i^{Flood\_House} = 1$ House in the floodplain (n = 60)	$D_i^{Flood\_Prop} = 1$ House near the floodplain (n = 339)	$D_i^{Flood} = 0$ House above the floodplain (n = 8,613)
Variable name	Mean (sd)	Mean (sd)	Mean (sd)
Price, kSEK	1039 (611.4)	1335 (941.4)	873.9 (685.7)
Structure			
Living area, m^2	97.85 (48.60)	116.9 (50.31)	117.7 (42.65)
Additional area, m^2	32.42 (41.16)	34.92 (37.98)	38.31 (42.50)
Lot size, m^2	1621 (1296)	1491 (2002)	1372 (1894)
Building taxation value, kSEK	287 (217)	442 (362)	384 (301)
Lot taxation value, kSEK	232,1 (194)	294,2 (278)	174 (0.157)
Value points	24.55 (7.77)	26.56 (7.00)	27.21 (5.96)
Detached house (=1 if detached)	0.967 (0.181)	0.932 (0.252)	0.852 (0.355)
Vacation property (=1 if vacation property)	0.367 (0.486)	0.195 (0.397)	0.0974 (0.297)
Age, years	44.12 (19.53)	51.17 (21.18)	43.26 (20.05)
Location			
Elevation of house	45.91 (0.19)	46.74 (0.26)	58.03 (11.80)
Waterfront (=1 if property lies waterfront, else = 0)	0.400 (0.494)	0.100 (0.301)	0.0182 (0.134)
Distance to nearest waterbody, m	159.5 (144)	269.1 (340)	889.3 (642)
Period			
Year sold	2003 (4,7)	2004 (5,5)	2004 (5,4)
Sold before flood	0,42	0.0177	0.0408
Sold during flood	0.383	0.316	0.338
Sold after flood			

Table 2: This table contains the results from the baseline model using D^{Flood} as the explanatory variable for flood risk. The period of study is 1998-2003 where May 2001 – December 2003 constitutes the post-flood period.

Table 2: Regression using D_i^{Flood} as an explanatory variable

Dependent variable.: *ln (price)*

	Model 1	Model 2	Model 3	Model 4
Sample	Full	Full	Distance to water body < 300 m	Distance to water body < 300 m
D_i^{Flood}	-0.0429 (0.0480)	-0.0614** (0.0315)	-0.0149 (0.0632)	-0.0666 (0.0436)
D_{it}^{post}	-0.219 (0.139)	-0.164 (0.117)	-0.261 (0.243)	-0.186 (0.183)
$D_i^{Flood} \times D_{it}^{post}$	-0.00442 (0.0672)	0.0278 (0.0465)	-0.00210 (0.0757)	0.0377 (0.0520)
Structure and location controls	No	Yes	No	Yes
Municipality fixed effects	Yes	Yes	Yes	Yes
Zip code fixed effects	Yes	Yes	Yes	Yes
Year-month fixed effects	Yes	Yes	Yes	Yes
Waterfront time trend	Yes	Yes	Yes	Yes
Standard errors adjusted for spatial and temporal autocorrelation	Yes	Yes	Yes	Yes
Observations	9012	9012	1625	1625
R-squared	0.4161	0.6500	0.5497	0.7906

Table 3: This table contains the results from the models with D^{Flood_House} disentangled from D^{Flood_Prop}. The period of study is 1998-2003 where May 2001 – December 2003 constitutes the post-flood period.

Table 3: Regression using $D_i^{Flood\_House}$ and $D_i^{Flood\_Prop}$ as explanatory variables

Dependent variable.: *ln (price)*

	Model 5	Model 6	Model 7	Model 8
Sample	Full	Full	Distance to water body < 300 m	Distance to water body < 300 m
$D_i^{Flood\_House}$	0.0134 (0.116)	-0.112* (0.0631)	-0.0210 (0.123)	-0.102 (0.0781)
$D_i^{Flood\_Prop}$	-0.0549 (0.0494)	-0.0536* (0.0313)	-0.0132 (0.0692)	-0.0617 (0.0448)
D_{it}^{post}	-0.218 (0.139)	-0.165 (0.117)	-0.262 (0.244)	-0.176 (0.182)
$D_i^{Flood\_House} \times D_{it}^{post}$	-0.0306 (0.146)	0.216** (0.100)	-0.00448 (0.167)	0.267** (0.121)
$D_i^{Flood\_Prop} \times D_{it}^{post}$	0.000321 (0.0718)	-0.00643 (0.0493)	-0.00134 (0.0835)	-0.0152 (0.0551)
Structure and location controls	No	Yes	No	Yes
Municipality fixed effects	Yes	Yes	Yes	Yes
Zip code fixed effects	Yes	Yes	Yes	Yes
Year-month fixed effects	Yes	Yes	Yes	Yes
Waterfront time trend	Yes	Yes	Yes	Yes
Standard errors adjusted for spatial and temporal autocorrelation	Yes	Yes	Yes	Yes
Observations	9012	9012	1625	1625
R-squared	0.4161	0.6425	0.4276	0.7499

Table 4: The table contains results for the models run for the period 1998-2003 as well as 2012-2013 using year dummies instead of a single post-flood period dummy.

Table 4: Regressions with year dummy variables

Dependent variable.: *ln (price)*

	Model 6 with year dummies	Model 8 with year dummies
Sample	Full	Distance to water body < 300 m
$D_i^{Flood\_House}$	-0.152** (0.0662)	-0.0131 (0.0785)
$D_i^{Flood\_Prop}$	-0.0845** (0.0396)	-0.0356 (0.0489)
$D_i^{Flood\_House} \times DURING$	-0.472** (0.231)	-0.662*** (0.218)
$D_i^{Flood\_House} \times Y2001$	0.0237 (0.138)	-0.186 (0.183)
$D_i^{Flood\_House} \times Y2002$	0.408** (0.158)	0.332* (0.183)
$D_i^{Flood\_House} \times Y2003$	0.139 (0.141)	0.176 (0.154)
$D_i^{Flood\_House} \times Y2012$	0.328 (0.251)	-0.0596 (0.220)
$D_i^{Flood\_House} \times Y2013$	0.194* (0.103)	0.243 (0.285)
$D_i^{Flood\_Prop} \times DURING$	0.114 (0.0869)	0.0568 (0.101)
$D_i^{Flood\_Prop} \times Y2001$	-0.0336 (0.0872)	0.0299 (0.0826)
$D_i^{Flood\_Prop} \times Y2002$	0.0931 (0.0576)	0.0307 (0.0724)
$D_i^{Flood\_Prop} \times Y2003$	-0.00226 (0.102)	-0.0518 (0.112)
$D_i^{Flood\_Prop} \times Y2012$	0.342** (0.141)	0.0434 (0.128)
$D_i^{Flood\_Prop} \times Y2013$	0.0743 (0.0604)	0.0740 (0.0691)
Structure and location controls	Yes	Yes
Municipality fixed effects	Yes	Yes
Zip code fixed effects	Yes	Yes
Year-month fixed effects	Yes	Yes
Waterfront time trend	Yes	Yes
Standard errors adjusted for spatial and temporal autocorrelation	Yes	Yes
Observations	9012	1625

R-squared	0.6013	0.7781

Table 5: This table contains results from the robustness test for choice of distance cutoff value for the inverse-distance matrix used by the SpatHAC estimator, using the specification in Table 3.

Table 5: Sensitivity to choice of distance cutoff value (dist_co)

Dependent variable.: *ln (price)*

	dist_co = 0.5 km	dist_co = 5 km	dist_co = 0.5 km	dist_co = 5 km
Sample	Full	Full	Distance to water body < 300 m	Distance to water body < 300 m
$D_i^{Flood\_House}$	-0.136** (0.0647)	-0.136** (0.0639)	-0.111 (0.0855)	-0.111 (0.0854)
$D_i^{Flood\_Prop}$	-0.0658** (0.0300)	-0.0658** (0.0322)	-0.0803* (0.0416)	-0.0803* (0.0434)
D_{it}^{post}	-0.00419 (0.0247)	-0.00419 (0.0238)	-0.0155 (0.0454)	-0.0155 (0.0422)
$D_i^{Flood\_House} \times D_{it}^{post}$	0.209** (0.0997)	0.209** (0.102)	0.257** (0.129)	0.257** (0.130)
$D_i^{Flood\_Prop} \times D_{it}^{post}$	0.00157 (0.0516)	0.00157 (0.0518)	-0.0115 (0.0583)	-0.0115 (0.0566)
Structure and location controls	Yes	Yes	Yes	Yes
Municipality fixed effects	Yes	Yes	Yes	Yes
Zip code fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Waterfront time trend	Yes	Yes	Yes	Yes
Standard errors adjusted for spatial and temporal autocorrelation	Yes	Yes	Yes	Yes
Observations	9012	9012	1625	1625
R-squared	0.6425	0.6425	0.7499	0.7499

Table 6: This table contains results from the pre-flood trend differences run for the period January 1998 to December 2000.

Table 6: Testing for pre-flood trend differences between floodplain- and non-floodplain properties

Dependent variable.: *ln (price)*

	Model 6 with year dummies	Model 8 with year dummies
Sample	Full	Distance to water body < 300 m
$D_i^{Flood\_House}$	-0.164 (0.115)	-0.0408 (0.109)
$D_i^{Flood\_Prop}$	0.0107 (0.0502)	-0.0108 (0.0629)
$D_i^{Flood\_House} \times Y1999$	0.129 (0.135)	0.0979 (0.157)
$D_i^{Flood\_House} \times Y2000$	0.0611 (0.168)	-0.0613 (0.151)
$D_i^{Flood\_Prop} \times Y1999$	-0.113 (0.0873)	-0.00634 (0.0825)
$D_i^{Flood\_Prop} \times Y2000$	-0.0675 (0.0663)	-0.0252 (0.0716)
Structure and location controls	Yes	Yes
Municipality fixed effects	Yes	Yes
Zip code fixed effects	Yes	Yes
Year-month fixed effects	Yes	Yes
Waterfront time trend	Yes	Yes
Standard errors adjusted for spatial and temporal autocorrelation	Yes	Yes
Observations	9012	1625
R-squared	0.6013	0.7781

Table 7: This table contains results from the robustness test using adjusted elevation cutoffs in the specification, using the specification in Table 3.

Table 7: Adjusted elevation cutoffs Dependent variable.: *ln (price)*		
	Model 6 with adjusted elevation cutoffs	Model 8 with adjusted elevation cutoffs
Sample	Full	Distance to water body < 300 m
$D_i^{Flood\_House}$	-0.0652 (0.0432)	-0.0634 (0.018)
$D_i^{Flood\_Prop}$	-0.0185 (0.0316)	-0.030 (0.0382)
D_{it}^{post}	0.105*** (0.0223)	0.135*** (0.0330)
$D_i^{Flood\_House} \times D_{it}^{post}$	0.0467 (0.0777)	0.0553 (0.0694)
$D_i^{Flood\_Prop} \times D_{it}^{post}$	0.0329 (0.0393)	0.0471 (0.0507)
Structure and location controls	Yes	Yes
Municipality fixed effects	Yes	Yes
Zip code fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Waterfront time trend	Yes	Yes
Standard errors adjusted for spatial and temporal autocorrelation	Yes	Yes
Observations	9012	1625
R-squared	0.6013	0.7781

Table 8: Comparison of previous findings relative to this stud. Note that a positive discount value implies a negative effect of flood risk on price.

Authors (year)	Country of study	Flood event	Pre-flood discount	Significant?	Post-flood discount	Significant?	Time to normalization
Atreya et al. (2013)	USA	Hurricane Floyd, 1999	9 %	Weakly	22-23 %	Yes	4-9 years
Bartosova et al. (2000)	USA	Menomonee River, 1997	5 %	No	19 %	Yes	Not studied
Bin and Landry (2013)	USA	Flint River, 1994	-4 %	No	9-13 %	Yes	5-6 years
Daniel et al. (2007)	Netherlands	River Meuse, 1993	-2 %	No	7-13 %	Yes	At least 10 years
Hallstrom and Smith (2005)	USA	Hurricane Andrew, 1992	-3 %	No	14-19 %	Yes	Not studied
This study (2016) – using Model 4	**Sweden**	**Lake Vänern flood 2000-2001**	**4 %**	**No**	**-2 %**	**No**	**Immediately**

Figures

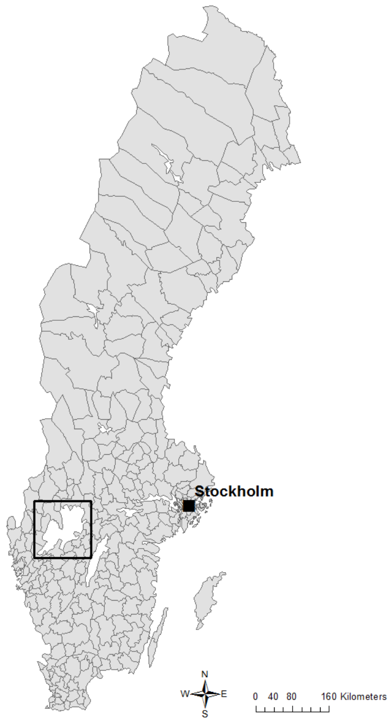

Figure 1: Map of Sweden with the black rectangle marking Lake Vänern and the study area.

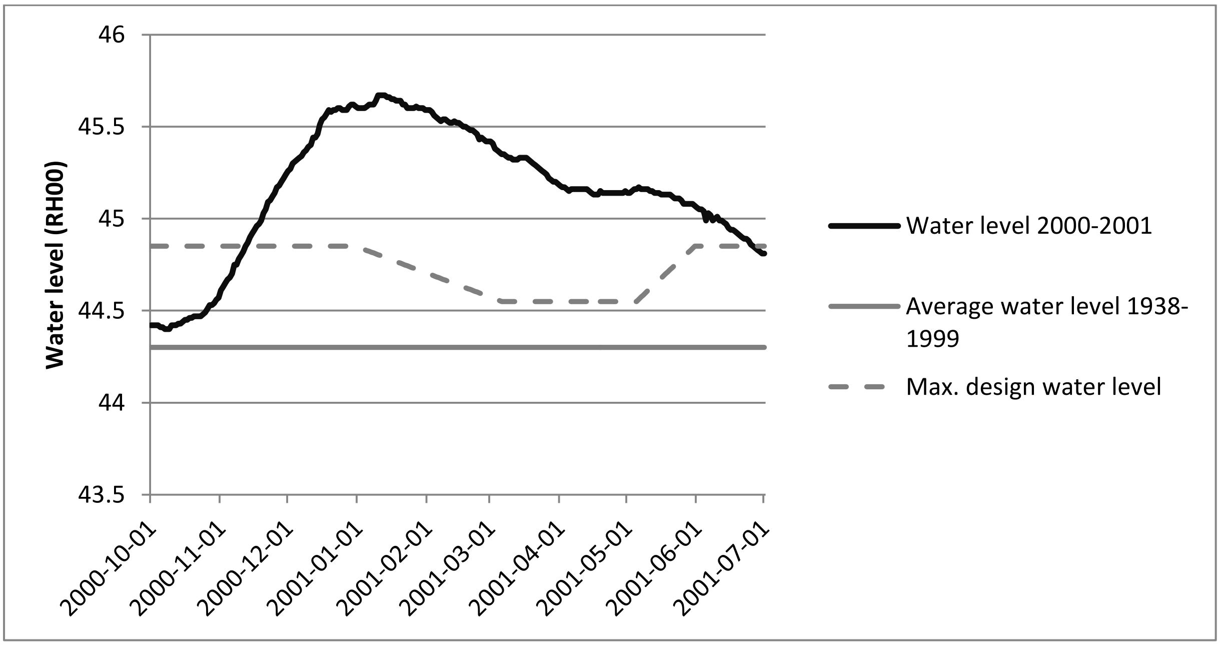

Figure 2: Time series of the water level in Vänern from October 2000 to June 2001.

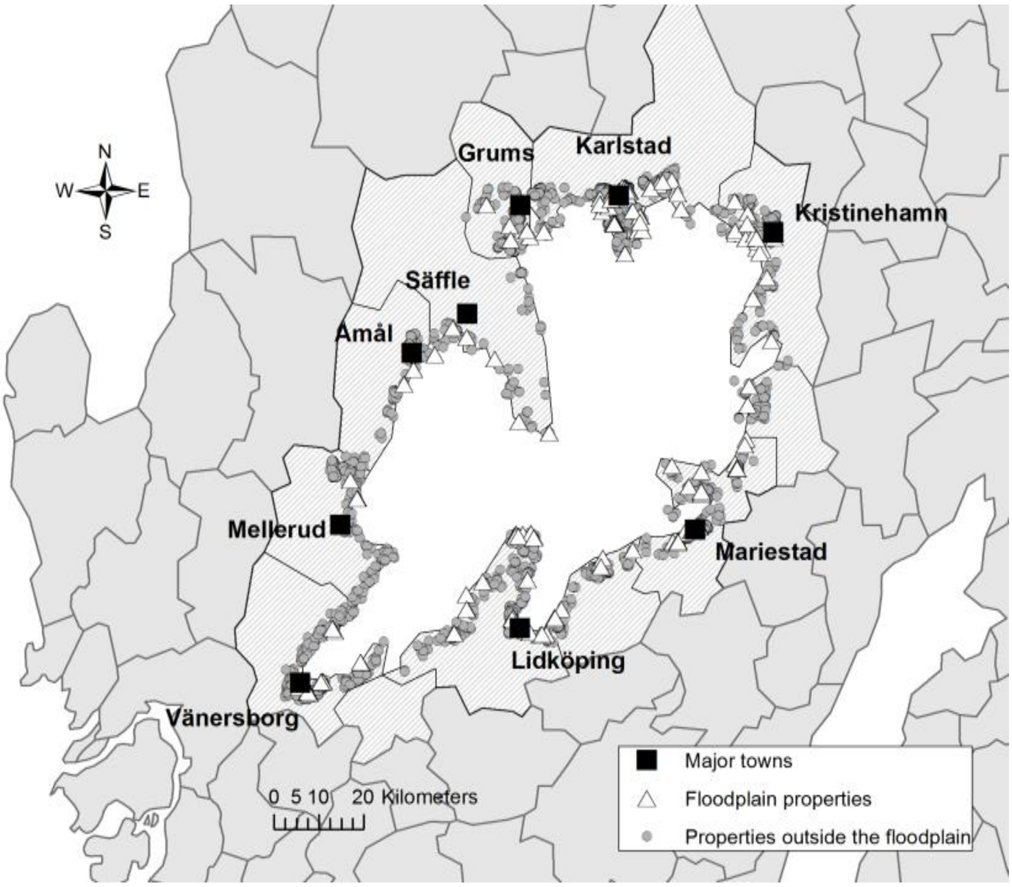

Figure 3:Map showing Lake Vänern, property transaction observations, bordering municipalities and major towns. Bordering municipalities are shaded in lighter gray than non-bordering municipalities.

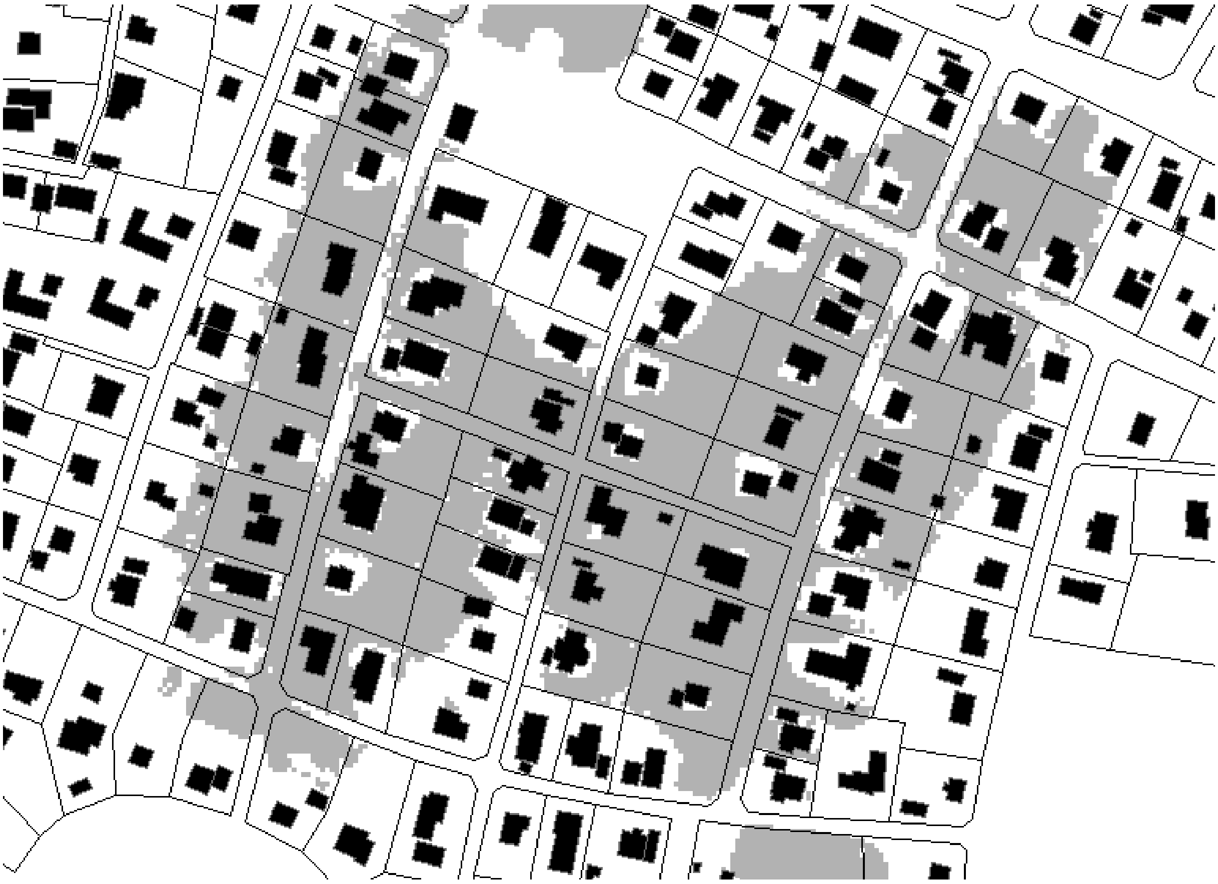

Figure 4: The figure shows property parcels and buildings overlaid on a DEM where gray indicates the floodplain (the area that lies below the maximum flood level). Properties where the house intersects with the floodplain are indicated by $D^{Flood\_House} = 1$. Properties where only the property but not the house intersects with the floodplain, but where the house lies not more than 1 m above the maximum flood level are indicated by $D^{Flood\_Prop} = 1$. The two groups aggregated are indicated by $D^{Flood} = 1$ in order to separate them from properties not at risk of flood ($D^{Flood} = 0$).

Figure 5: Time trend of residuals from the baseline model covering all floodplain properties for ln(price) from January 1998 to December 2003. Red lines mark the beginning and end of the flood.

Figure 6: Time trend of residuals from the baseline model covering properties where the house is in the floodplain, for ln(price) from January 1998 to December 2003. Red lines mark the beginning and end of the flood.

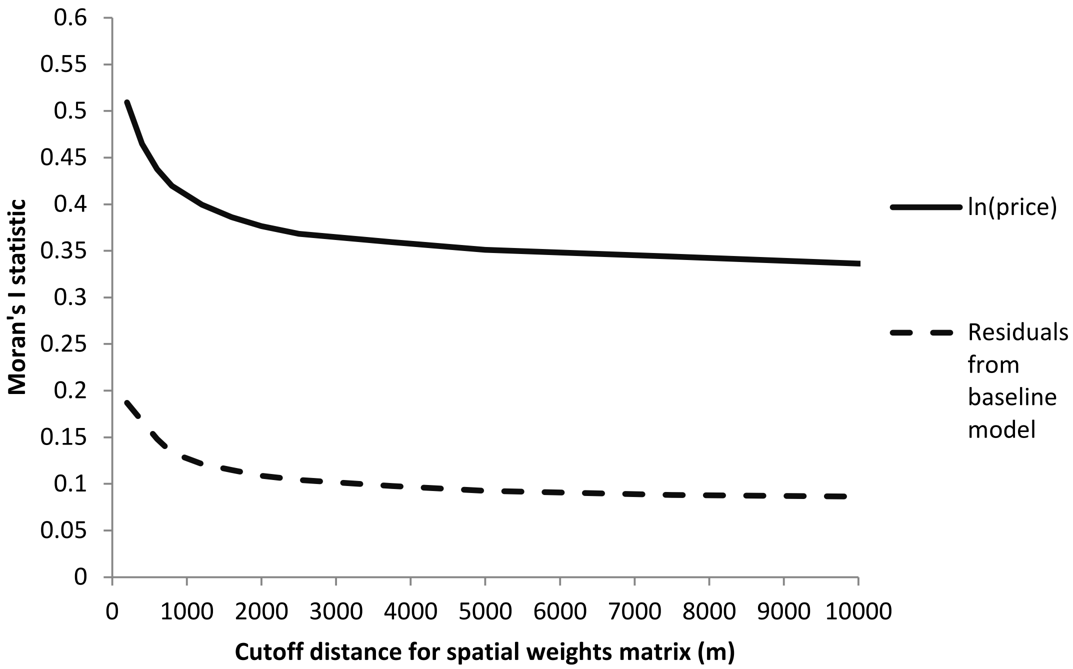

Figure 7: Measure of spatial autocorrelation in ln(price) and residuals from Model 1 as a function of cutoff distance, using an inverse distance matrix.

Appendix

Tables

Table A1: This table contains the results for all covariates excluding zip code and year-month coefficients.

Table A1: Main results table with full list of structure and location controls

Dependent variable.: *ln (price)*

	Model 1	Model 2	Model 3	Model 4
Sample	*Full*	*Full*	Distance to water body < 600 m	Distance to water body < 600 m
D_i^{Flood}	-0.0504 (0.0477)	*-0.0758** (0.0312)*	-0.0604 (0.0649)	-0.0424 (0.0400)
D_{it}^{post}	-0.0448 (0.0349)	*-0.00437 (0.0246)*	-0.137* (0.0793)	0.00224 (0.0427)
$D_i^{Flood} \times D_{it}^{post}$	0.0917 (0.0688)	*0.0334 (0.0478)*	0.0995 (0.0894)	0.0164 (0.0627)
Elevation of house		0.00117 (0.000881)		0.00107 (0.00168)
Waterfront (=1 if property lies waterfront, else = 0)		0.255*** (0.0639)		0.132** (0.0602)
ln Distance to nearest waterbody, m		-0.111*** (0.0107)		-0.101*** (0.0175)
Living area, m2		0.00135 (0.00262)		-0.000765 (0.00435)
Additional area, m2		0.000402 (0.000431)		-0.000214 (0.000705)
Lot size, m2		4.99e-05*** (1.08e-05)		2.90e-05** (1.39e-05)
Lot size squared		-1.23e-09*** (3.86e-10)		-6.15e-10 (6.31e-10)
ln Building taxation value, kSEK		0.560*** (0.0752)		0.520*** (0.0621)

ln Lot taxation value, kSEK		0.169*** (0.0393)		0.335*** (0.0583)
Value points		0.00412* (0.00214)		0.00203 (0.00238)
Vacation property (=1 if vacation property)		0.0723** (0.0329)		0.164*** (0.0371)
Age, years		-0.00625*** (0.00239)		-0.00807** (0.00344)
Age squared		7.06e-05*** (2.14e-05)		0.000102*** (3.40e-05)
Structure and location controls	No	Yes	No	Yes
Municipality fixed effects	Yes	Yes	Yes	Yes
Zip code fixed effects	Yes	Yes	Yes	Yes
Year-month fixed effects	Yes	Yes	Yes	Yes
Standard errors controlled for spatial and temporal autocorrelation	Yes	Yes	Yes	Yes
Observations	9012	9012	1625	1625
R-squared	0.4161	0.6424	0.4276	0.7499