IS FLOOD RISK ALREADY AFFECTING HOUSE PRICES?

Lessons learned from an impact assessment study in the Netherlands

ABN·AMRO

11 February 2022

- Global warming increases the probability of flooding. If this risk is not reflected in house prices this raises concerns about a sudden risk re-pricing that may prove financially destabilizing.
- Our research findings are inconclusive but point towards current flood risks being not (fully) priced in to house valuations
- The key methodological challenge in providing a robust estimate of flooding risk on house prices is finding omparable houses for the exercise, while still keeping the sample size sufficiently large.

Author:

Jeannine van Reeken
Economist
(contact for press 06-55 52 28 80)
jeannine.van.reeken@nl.abnamro.com

Sandra Phlippen
Chief Economist
sandra.phlippen@nl.abnamro.com

INTRODUCTION

Global warming increases the probability of flooding from either breaking- or flooding barriers as the sea level rises or from more frequent and harder rainfall causing damage directly or via flooding rivers. The real estate sector in particular is vulnerable to this type of physical climate risk. The risks of flooding is already present in the Netherlands today and the government has given estimations of increasing probabilities by location in 2050.
Home owners, who commit themselves financially via a mortgage to the long term value of their home, should in theory be fully aware of the value reducing potential of current and future flooding probabilities. The question is: are they?

If they are, we would expect that the value of Dutch houses that are located in areas that are already today more likely to flood are lower than houses that are situated on land that is less likely to flood. Keeping all other price determinants constant.

If they are not, this means that current and future potential damage is not (completely) taken into account. This is of great interest to our bank, as a sudden risk re-pricing could occur that may prove financially destabilizing. Knowing how physical risks of climate change such as flooding probabilities can affect the financial stability of banks and thereby the financial system will become part of a regulatory requirement from the ECB.

Our research findings are intended to share the lessons learned in how to approach a question like this. We believe that an open exchange of trials and errors in research aimed at estimating the economic impact of climate change contributes to financial resilience of our clients, our bank and our economy.

In this article we share our findings and the different empirical approaches we considered, we explain the limitations and recommendations for further research on this topic.

FINDING THE RIGHT COMPARISON

To assess the impact of flood risk probabilities on a house price we need to correctly attribute a difference in sales price to the difference in flood risk. This is a daunting task, as our observations of the varying house prices could well be merely a correlation and not caused by differences in flood risk probabilities. Correlation is just a relationship but doesn't imply that one event is the result of another event. Correlation could be coincidental, or a third factor may be causing both variables to change.

The main challenge when identifying causation is finding the right comparison. In our research setting, we only observe the sales price of a house with a given flood probability. We don't know what the price of that house in that moment would have been had the flood probability been different. In other words, we don't have a counterfactual. And this comparison of the observed house with it's counterfactual would be the causal effect we are looking for. There are basically two ways to estimate causal impact and construct a counterfactual. First approach is a randomized experiment and the second approach is causal inference using econometrics. Regarding our research question, a randomized experiment is not a feasible option and we will use econometrics.

DATA

We construct a dataset using multiple sources. Our main source is the ABN AMRO Mortgage internal data. Here we selected sales price and house characteristics of residential real estate bought with an ABN AMRO mortgage from January 2018 till October 2021. We used publicly available BAG data to add the geolocation, function and size of the building. Based on the geolocation we enriched our dataset with >50cm flood probabilities in 2021 using the >50cm flood map published by the Landelijk Informatiesysteem Water en Overstromingen (LIWO). The LIWO together with the KNMI (Dutch Weather Institute) and Wageningen Environmental Research (WENR) create flood maps and make them publicly available. For the flooding depth > 50cm the chance is defined as a yearly chance. Lastly, we used CBS data to enrich our dataset with neighborhood characteristics, such as percentage of neighborhood inhabitants with high income, - with social benefits, - level of urbanity, - distance to amenities, - distance to green and - distance to water.

FILTERS AND SELECTION

We had to make some selections when building the complete dataset. First, we started by selecting only properties in the Mortgage dataset with a known location (without a location we are not able to match to different datasets). Second, we only keep houses we were able to match with the BAG dataset. Next, we applied filters to the total set in order to deal with data quality issues. We remove outliers in terms of transaction price, area and income. We removed properties in districts with less than 50 inhabitants because CBS does not publish statistics for these areas (8 properties). We removed properties that are located in water regions according to the CBS data (5 properties). 2 properties are removed because they do not have a known energy label. Since we are interested in the residential real estate we only kept properties with a 100 percent housing function according to BAG, excluding properties with a (partial) commercial purpose. We exclude all apartments from the analysis. Other properties removed from the dataset are: garages, holiday houses and farm houses.

METHODOLOGY

We made use of a hedonic price model. A hedonic price model allows us to study the relationship between the sales price of houses and their characteristics. A hedonic price refers to the implicit price of a certain attribute based on the association of the sales price and each of its attributes. A hedonic price function is typically used for products like computers, cars and houses.

In this study we compare houses with different flooding probabilities (yearly probabilities of a flood >50cm) to each other, controlling for a suite of location- and property-specific characteristics. This cross-sectional method is widely used in historical literature on flood risk. However, recent studies point out that this approach is likely to suffer from omitted variable bias. This happens when there are unobserved characteristics that are correlated with both the flooding probabilities and sales price. This means dissimilar houses are being compared and a reliable counterfactual cannot be constructed. We will come back to this issue later.

This analysis pools all houses within the same flood risk category to estimate the flood discount per flood risk category. We observe five different flood risk categories: no significant risk, extremely small risk (<1/30.000), very small risk (1/3000 to 1/30.000), small risk (1/300 to 1/3000) and medium risk (> 1/300). Our hedonic model takes the following form:

$$\ln(price_{itrn}) = \beta_1 ExtremelySmallRisk_i + \beta_2 VerySmallRisk_i + \beta_3 SmallRisk_i + \beta_4 MediumRisk_i + \delta \boldsymbol{Z}_i + \theta \boldsymbol{C}_n + \alpha_t + \gamma_r + \varepsilon_{itrn}$$

Where the dependent variable, $\ln(price_{itrn})$, is the natural logarithm of the sales price of property *i* at time *t* in region *r* and neighborhood *n*. The *Risk* variables are the variables of interest. Where ExtremelySmallRisk is a binary variable equal to 1 if property *i* has an extremely small >50cm flood probability (< 1/30.000). β_1 is the estimated effect of having an extremely small >50cm flood probability compared to no significant flood probability. Interpretation of the other risk variables with its coefficients is similar. We account for property-specific characteristics in $\delta \boldsymbol{Z}_i$, where $\boldsymbol{Z}_i$ is a vector of property characteristics: building year, energy efficiency label, property type, property area, ground lease indicator and new construction indictor. We account for neighborhood-specific characteristics in $\theta \boldsymbol{C}_n$, where $\boldsymbol{C}_n$ is a vector of neighborhood characteristics from the Dutch Central Bureau of Statistics: population characteristics, income & social security characteristics, level of urbanity and distance variables. α_t is a fixed effect for the quarter of sale to account for seasonal market changes, and γ_r is a fixed effect for each COROP plus region which absorbs regional differences. A COROP plus region is a division of the Netherlands for statistical purposes, used by Central Bureau of Statistics Netherlands. There are 52 COROP plus regions in the Netherlands.

RESULTS

The results are presented in Table 1. Control variables are omitted from the table. When including the flood risk categories as sole regressors (Model I), we do not find a price discount for the properties with a medium >50cm flood probability (> 1/300) compared to properties with no significant flood probability. Only after including location and time fixed effects (model III) we find a discount for the highest flood probability.

Table 1:OLS Estimation Results

	I	II	III	IV	V
Intercept	12.7496***	8.6286***	8.7216***	9.0790***	9.0790***
	(0.0027)	(0.0268)	(0.0209)	(0.0211)	(0.0418)
ExtremelySmallRisk	0.0915***	0.1015***	-0.0199***	-0.0128***	-0.0128
	(0.0075)	(0.0056)	(0.0046)	(0.0044)	(0.0081)
VerySmallRisk	-0.1373***	-0.0594***	-0.0585***	-0.0348***	-0.0348***
	(0.0097)	(0.0071)	(0.0054)	(0.0052)	(0.0075)
SmallRisk	-0.1025***	-0.0393***	-0.0330***	-0.0046	-0.0046
	(0.0066)	(0.0049)	(0.0041)	(0.0040)	(0.0055)
MediumRisk	0.0223***	0.0710***	-0.0528***	-0.0225***	-0.0225***
	(0.0057)	(0.0042)	(0.0035)	(0.0034)	(0.0062)
R-squared Adj.	0.0122	0.4730	0.7486	0.7767	0.7767
Covariance type	nonrobust	nonrobust	nonrobust	nonrobust	cluster
N	55115	55115	55115	55115	55115
Property characteristics	No	Yes	Yes	Yes	Yes
Location FE	No	No	Yes	Yes	Yes
Time FE	No	No	Yes	Yes	Yes
CBS neighborhood characteristics	No	No	No	Yes	Yes

Notes: This table reports the OLS estimation results. The dependent variable is the natural logarithm of the sales price. In model V standard errors are clusetered by COROPplus and quarter. Standard errors in parentheses. * p<.1, ** p<.05, ***p<.01

To interpret the Log-Linear model dummy coefficients (β_1, β_2, β_3, β_4) we calculated the percent change in house price associated with each coefficient of our most complete hedonic model (model V). These results are presented in Figure 1. Properties with a medium >50cm flood probability have about 2.2% house price discount compared to similar properties without a significant flood probability. If people take flood risk into account when buying a house, one would expect higher flood risk properties to have a bigger price discount. However, this is not what our results show. The biggest discount is found for the very small >50cm flood probability (1/3000 to 1/30.000) properties. This result seems unlikely. Although we have a large sample and we control for property and neighborhood characteristics and make use of location and time fixed effects, it's likely that these results suffer from omitted variable bias.

1. Percentage change in house price associated with flood risk

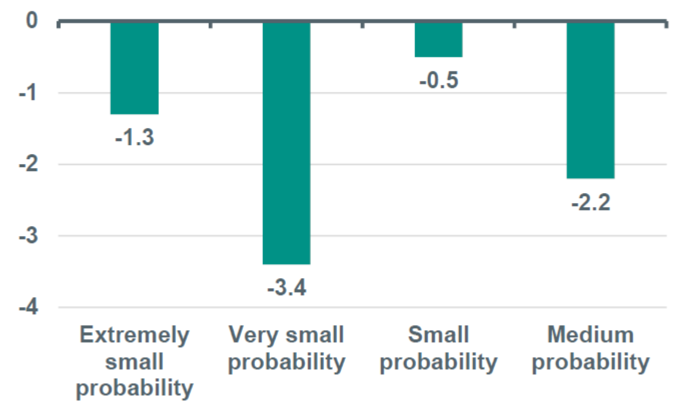


Table 2: OLS Estimation Results by Province

	Drenthe	Flevoland	Friesland	Gelderland	Groningen	Limburg	Noord-Brabant	Noord-Holland	Overijssel	Utrecht	Zeeland	Zuid-Holland
Intercept	9.7954***	10.0403***	9.6390***	9.7783***	9.7810***	9.4097***	9.3169***	9.3952***	9.5594***	9.2134***	9.2269***	8.9480***
	(0.1409)	(0.3012)	(0.1518)	(0.1101)	(0.2519)	(0.1273)	(0.0616)	(0.0753)	(0.0887)	(0.0951)	(0.1307)	(0.0701)
ExtremelySmallRisk	-0.0000***	0.0000	0.0189	0.0508	0.0303	0.0000	0.0000***	0.0009	0.0000**	0.2010***	-0.0392	-0.0334***
	(0.0000)	(0.0000)	(0.0312)	(0.0357)	(0.0374)	(0.0000)	(0.0000)	(0.0129)	(0.0000)	(0.0368)	(0.0254)	(0.0098)
VerySmallRisk	-0.0000***	-0.0783	-0.0113	-0.0377**	-0.0443	-0.0021	-0.0159	-0.0147	0.0068	0.0161	0.0217	-0.0314**
	(0.0000)	(0.0633)	(0.0238)	(0.0152)	(0.0539)	(0.0234)	(0.0187)	(0.0143)	(0.0273)	(0.0252)	(0.0250)	(0.0124)
SmallRisk	-0.0000***	-0.0575	-0.0216	0.0189	-0.1021**	0.0381	0.0039	-0.0333***	0.0365*	-0.0162	-0.0564**	-0.0046
	(0.0000)	(0.0374)	(0.0231)	(0.0124)	(0.0432)	(0.0253)	(0.0126)	(0.0101)	(0.0199)	(0.0106)	(0.0240)	(0.0124)
MediumRisk	0.1001***	-0.0000	0.0604	-0.0242*	-0.0180	0.0269	-0.0245*	-0.0293***	0.1163***	-0.0658***	-0.0715**	-0.0452***
	(0.0374)	(0.0000)	(0.0647)	(0.0129)	(0.0266)	(0.0366)	(0.0133)	(0.0091)	(0.0201)	(0.0093)	(0.0363)	(0.0098)
R-squared Adj.	0.7797	0.8059	0.7327	0.7398	0.7492	0.7049	0.7245	0.8127	0.7518	0.7792	0.7450	0.7982
Covariance type	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
N	1560	1597	1697	6425	840	2986	9538	9397	3192	4863	1449	11571
Property char	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CBS char	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

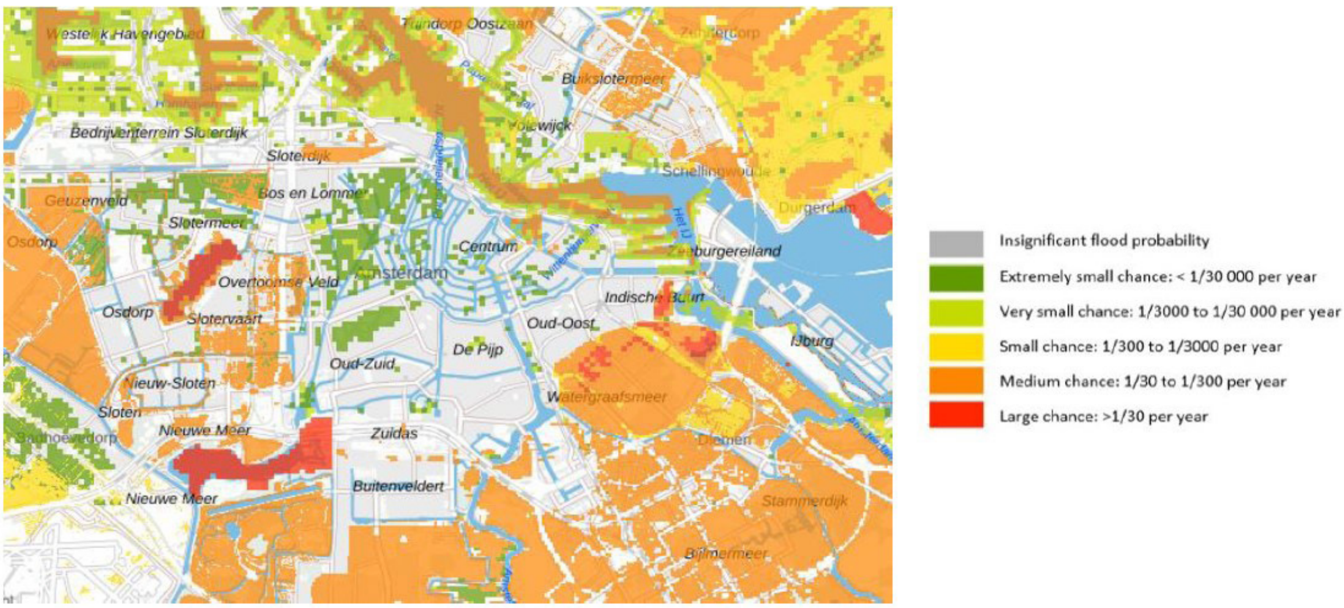
Notes: This table reports the OLS estimation results. The dependent variable is the natural logarithm of the sales price. Standard errors are clusetered by COROPplus and quarter. Standard errors in parentheses. * p<.1, ** p<.05, ***p<.01

ROBUSTNESS AND LIMITATIONS

Given that we do not find a larger discount for higher flooding probabilities, in fact, we find the biggest discount for the risk category with a 1/3000 to 1/30000 yearly flood chance, we believe these results are capturing something else. As a robustness check, we look at whether the results are similar for the different provinces of the Netherlands, see Table 2. From Table 2 it's clear that results change quite a bit depending on which province we look at. There are even provinces where a bigger than 1/300 flood probability leads to a price mark-up instead of discount, see Drenthe and Overijssel.

To understand the omitted variable bias better, it's useful to have a closer look at the flood probability map. Figure 2 shows the flood probabilities for Amsterdam. Within Amsterdam there are differences in house prices depending on the area, with Amsterdam city center being more expensive than the outer part. So, for this area we are capturing something else besides flood risk in our regressions, namely houses in the outer area have a discount compared to the city center. Our regressions suffer from omitted variable bias.

Figure 2:Flood probabilities >50cm Amsterdam (source LIWO)



IMPROVING THE COMPARISON

Apparently, we are not comparing similar houses. In order to deal with this omitted variable bias we have to make use of location fixed effects on a more local level than COROP plus region. Therefore, we did a second analysis on a subset of the data on a more local level. For this analysis we only compared the medium flood probabilities (> 1/300) with no significant flood probabilities, leaving out all other risk categories. In order to make the omitted variable bias as small as possible we only include terraced houses, making use of the fact that terraced houses are a more homogenous group than if we also include detached houses. We further restrict our sample to only those districts (in Dutch "wijken") with at least 5 properties with a medium flood probability and at least 5 properties with no significant flood probability. This reduces the sample to 1044 properties without a significant flood probability and 981 properties with at least 1/300 flood probability. Figure 3 shows the location of the 53 remaining districts in scope.

Figure 3: Regions in scope

This time, to deal with the increase in house prices over time we make use of the house price index by province as published by CBS as opposed to time fixed effects. In the regression we include controls for property characteristics and add district fixed effects. See Table 3 for the results. We still find a discount of about 2.5% for properties with a flood probability higher than 1/300 compared to no significant chance. However, this time the estimate is not significant on a 1% level, but still significant on the 5% level.

Table 3: OLS Estimation Results with district FE

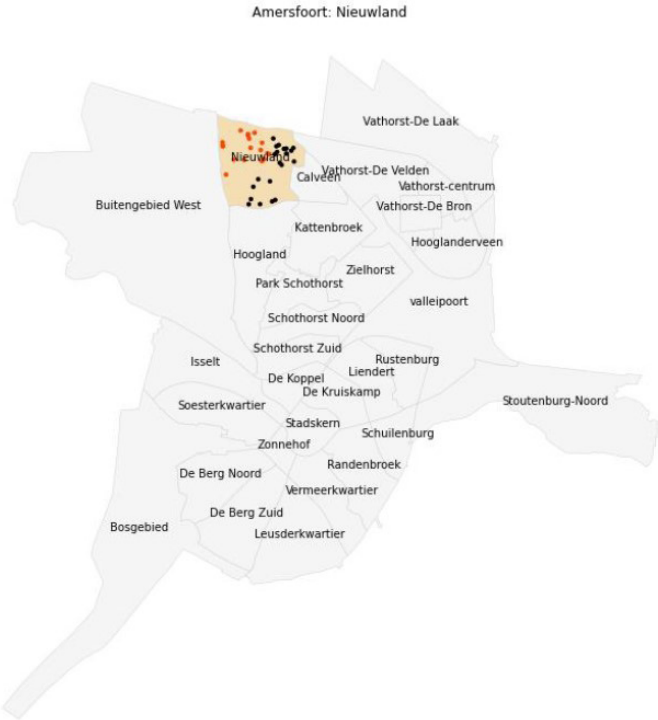
	I
Intercept	12.2926***
	(0.0559)
MediumRisk	-0.0253**
	(0.0114)
R-squared Adj.	0.7625
Covariance type	cluster
N	2025
Property characteristics	Yes
District FE	Yes

Notes: This table reports the OLS estimation results. The dependent variable is the natural logarithm of the sales price. Standard errors are clustered by COROPplus and quarter. Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01

By only including terraced houses and using fixed effects on district level we addressed some of the omitted variable bias. However, we don't know all property characteristics and districts can still be heterogeneous with preferred areas. We will illustrate these concerns with an example case study.

Case study: Nieuwland (part of municipality Amersfoort)
Nieuwland is part of the Dutch city Amersfoort, see Figure 4. The area has a total of about 4500 houses. The project started in 1995 and finished in 2001. This means that all houses have a similar building year and style and houses are comparable in energy efficiency label (all houses in our sample have label A or B). The black dots in Figure 4 reflect the houses with no significant flood probability and the red dots properties with a medium flood probability. Nieuwland can be divided into 5 different neighbourhoods. One of these parts is named "citygarden" and is located in the upper west part of Nieuwland. The main defining characteristic of this part of Nieuwland is the fact that the houses do not have an own garden, but all the area is shared property. As Figure 4 shows, a large portion of the houses with a flood probability are located in this part. Although we include some property characteristics, we do not include if the property has a garden. Hence, for this area it's not clear what we measure. Do we measure a price discount because of a flood risk? Or do we actually look at a price difference because a large portion of these houses do not have a garden?

Figure 4: Nieuwland and flood probabilities



So, one could argue that district level fixed effects is not local enough to compare similar houses. Therefore, we zoom in even further to neighbourhood level (in Dutch: "buurt"). We only select those neighbourhoods with at least 5 properties with a medium flood probability and at least 5 properties with no significant flood probability. This reduces the sample to only 294 properties, 131 properties without a significant flood probability and 163 properties with at least 1/300 flood probability, located in 11 different neighbourhoods. To correct for different sizes we look at price per square meter instead of absolute sales price. Here we find in 7 out of the 11 neighbourhoods a lower average sales price for houses with a flood probability compared to houses without a flood probability, see Table 4. However, most of these differences are not statistically significant. Only three differences are statistically significant, two times a lower average sales price for houses with a flood risk and one time an average higher price for the houses with a flood risk.

Table 4: Average sales price per neighbourhood and flood probability

Neighbourhood	Number of housholds	Flood probability	Number of observations	Mean salesprice per m2 (time index corrected)	Welch's t-test Statistic	P-value	Mean significantly different
Neighbourhood 1	850	No significant probability	6	€ 1.688	-1,91	0,069	No
		Medium probability (>1/300)	18	€ 1.780			
Neighbourhood 2	1520	No significant probability	5	€ 1.603	-2,45	0,03	Yes
		Medium probability (>1/300)	22	€ 1.816			
Neighbourhood 3	1250	No significant probability	5	€ 2.059	2,29	0,042	Yes
		Medium probability (>1/300)	11	€ 1.828			
Neighbourhood 4	3410	No significant probability	8	€ 3.431	1,71	0,114	No
		Medium probability (>1/300)	12	€ 2.953			
Neighbourhood 5	2895	No significant probability	7	€ 1.983	0,68	0,51	No
		Medium probability (>1/300)	12	€ 1.924			
Neighbourhood 6	2395	No significant probability	15	€ 2.362	2,45	0,024	Yes
		Medium probability (>1/300)	12	€ 2.095			
Neighbourhood 7	3460	No significant probability	13	€ 2.468	1,73	0,10	No
		Medium probability (>1/300)	21	€ 2.252			
Neighbourhood 8	2910	No significant probability	14	€ 1.801	1,54	0,149	No
		Medium probability (>1/300)	7	€ 1.648			
Neighbourhood 9	1105	No significant probability	8	€ 2.406	-0,12	0,91	No
		Medium probability (>1/300)	5	€ 2.457			
Neighbourhood 10	3135	No significant probability	42	€ 1.931	-1,23	0,254	No
		Medium probability (>1/300)	5	€ 2.033			
Neighbourhood 11	5150	No significant probability	8	€ 2.034	1,12	0,19	No
		Medium probability (>1/300)	38	€ 1.920			

LESSONS LEARNED: THE CHALLENGE OF FINDING THE RIGHT COMPARISON WHILE KEEPING SUFFICIENT OBSERVATIONS

Main challenge in impact evaluation is constructing the "right" counterfactual. We used a cross-sectional method to analyse the impact of flood risk on house prices. In order to construct a convincing counterfactual using a cross-sectional method all confounding factors need to be included in the regression model. If there are unobserved characteristics that are correlated with both the flooding probabilities and sales price the results suffer from an omitted variable bias. And indeed, we showed that our initial model suffered from omitted variable bias.

In order to overcome this issue we improved our comparison by exploring flood risk variation on a more local level. When only looking at houses with different flood risk probability within the same neighborhood we do not find that flood risk is priced into the residential real estate market. This could be concerning because it would tell us that home owners who do not consider current potential losses from flooding are bound to not take future losses from flooding into account at all. For locations where future losses are estimated to increase through increasing flood risk, there is the risk of a sudden price correction. Whether and how this affects client's-, the bank's and the economy's financial resilience is not known today.

These research findings are not allowing for any firm conclusions. More research is need, given the importance of the topic. The key challenge to overcome is finding the right comparison of houses without losing too many observations. As location is extremely important for valuation, the right comparison comes down to comparing properties that are as close to each other as possible. One way would be to compare a specific property's sales price over time while the flood probability changes. This would require a repeated sales data set with varying flood probabilities over time. We do not have these data. Another option is a comparison of properties using a border discontinuity design. If one has enough observations on a high proximity to each other, this is feasible. In general, including more property and location specific characteristics would improve the comparison.