Cumbria's Property Market Under the Surge: A Flood Risk Examination

A case study on how flood risk in flood-prone area of Cumbria, England affects residential real estate property prices

Master Thesis

MSc Real Estate Studies

Angelos Blantis

S5406714

29th December 2023

COLOFON

Title	Cumbria's Property Market Under the Surge: A Flood Risk Examination
Date	Dec. 29, 2023
Version	Final version
Author	Angelos Blantis
Student number	s5406714
E-mail	a.blantis@student.rug.nl
Supervisor	Dr. Xiaolong Liu
Email	xiaolong.liu@rug.nl
Assessor	Prof. Dr. Arno van der Vlist
E-mail	a.j.van.der.vlist@rug.nl
Master	MSc Real Estate Studies
Faculty	Faculty of Spatial Sciences

ABSTRACT

This study examines the relationship between flood risk and property prices in the area of Cumbria, England. Natural disasters have become a global menace due to global warming, and, in particular, Europe has experienced numerous storms and flooding in recent years. Understanding the diverse effects of these dangers has thus become critical in order to adapt preventative as well as urban planning of infrastructure and real estate. I explore the relationship between real estate property sales prices and three floods that occurred in January 2005, November 2009, and December 2015, using a hedonic pricing model. For my research model, I utilized a sample of 184,048 real estate transactions in Cumbria, UK's third-largest county, from 2003 to 2022. I find evidence indicating a statistically significant negative correlation between the November 2009 flood and real estate property prices, whereas there is no statistical negative evidence on the consequences of the January 2005 and December 2015 flood. I contribute to the literature by expanding related research to other regions and cities in England, aiming to close the gap of empirical evidence on market behavior and reaction patterns in the case of flooding and natural disasters in general. Specifically, this thesis focuses on flood events that occurred not in one specific place but in several areas and cities of the county that have been affected. Additionally, results show that the discount differs after the effects of flooding among residential properties, considering their property types and price segments, ranging from low-end, middle-end, and high-end housing.

“Master thesis are preliminary materials to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the author and do not indicate concurrence by the supervisor or research staff”

PREFACE

Dear reader,

Before delving into the master thesis “Cumbria's Property Market Under the Surge: A Flood Risk Examination” as my graduate document for the Master of Science Real Estate Studies at the University of Groningen.

I was engaged in researching and writing for this thesis with the primary aim of contributing knowledge about the urgent consequences of climate change and its implications on the real estate market. I would not be able to finish this thesis without the help of others. First of all, I would like to thank my thesis supervisor, Dr. Xiaolong Liu. Without his support, guidance throughout the research process, and great cooperation, this analysis would not have been achievable.

Furthermore, a special thanks to all my family members, friends and colleagues whom I bored with for a long time with my thesis topic. Their understanding of my state of mind when the writing process did not immediately go the way I wanted should not be mentioned unnoticed. They always kept me motivated for making progress.

Last but not least, I hope that this master thesis will serve as a valuable resource. If fellow scholars, researchers, practitioners, or anyone interested in the field of real estate studies find assistance in the presented research, it would be an honor for me. My ultimate goal is that the findings presented within these pages can contribute to the ongoing discourse and enhance understanding of real estate dynamics. Ultimately, this knowledge can guide decision-making and promote sustainable development, within the real estate sector.

In loving memory of my beloved cousin Kostas.

I hope you enjoy reading.

Angelos Blantis

Groningen, December 29, 2023

Table of Contents

1.INTRODUCTION

1.1 Motivation

The world's most damaging natural hazard affecting millions of people each year, is extreme flooding (Zhang, 2016). In England, flooding represents the largest natural disaster risk, and this risk is expected to increase further due to climate change, putting more people and their properties in danger (Surminski and Eldridge, 2015). In order to control this growing flood risk, a comprehensive portfolio of measures and actions is essential to reduce the probability of flooding. Following such an approach will not only minimize the impact and damages but also provide financial and economic insights for the residual risk. In this thesis, the primary focus is to explore the correlation between floods and housing prices, with a special emphasis on flood-prone areas in England.

Researchers have delved into perspectives to understand how flood risks affect real estate value, especially considering that floods don't just cause immediate damage to homes but also have long-term effects on the housing market. Thus, throughout the years, there has been an increasing scholarly attention to natural hazards and their effects on real estate property values. According to (Lamond, Proverbs and Hammond, 2010) properties located in flood affected areas often experience a decrease in value due to people's concerns about potential flooding risks. This situation not only affects homeowners, who may witness a decline in their property values over time, but also impacts potential buyers who are hesitant to invest in vulnerable areas, as well as investors who need to consider and rethink the uncertain future of such properties. Therefore, researchers are examining three aspects when studying how natural disasters affect property values: understanding consumer behavior when challenged with such risks, assisting governments in preparing for disasters, within high risk areas (Gharbia et al., 2016), and helping insurance companies set fair premiums based on risk assessments (Harrison, T. Smersh and Schwartz, 2001). In a world where climate change is making floods more common and unpredictable, there is a need to understand how these floods impact both communities and housing markets. Such understanding can provide insights that help in dealing more effectively with flooding.

Numerous academic papers have been published, focusing on how natural disasters and especially flooding, affects real estate prices. However, it is worth noting that there is a lack of research, specifically examining the area of Cumbria UK, and only limited research exists investigating in particular cities in the area rather than the wider region as a whole. Why in particular the area of Cumbria? Cumbria, a county in northwest England (Spencer et al., 2017), remains relatively unknown to many, but its hidden story is compelling. A story that has not been extensively explored from an economic perspective, despite the fact that it's a place where recurrent flood events have a significant influence on the local housing market. Studying Cumbria's situation not only reveals the challenges faced by the area but also adds valuable insights on managing flood risks and understanding housing markets more broadly. Cumbria's significance as the UK's third-largest county and its diverse economic activities, including tourism, agriculture, manufacturing, and nuclear industries, provides an ideal opportunity to study the economic consequences of natural disasters across various sectors (Cumbria County Council, 2023). Therefore, the primary objective is to examine this particular area and, in doing so, to produce results that will explain how real estate market in the area of Cumbria operates. Hence, this thesis will examine the relationship between these variables by conducting a thorough analysis using statistical models to determine whether floods correlate with residential housing prices.

1.2 Academic relevance

Previous studies address the effect of natural hazards such as floods, earthquakes and hurricanes on residential property prices. Specifically, the impact of flood risk on property values has been thoroughly studied, particularly whether a discount or a premium on housing prices exists and how the real estate market is responding. While some scholars agree that the risk of flooding leads to a price discount (Bin and Landry, 2013), other scholars state that housing prices do not decline right away after the flood and do not increase or recover later (Skantz & Strickland, 1987), or even that properties located in a flood zone sell with a premium due to their locational attractiveness (Atreya and Czajkowski, 2016).

In general, most of the research state that real estate property sells with a discount if the property is affected by a flood event or is located within a natural hazard risk zone. A study that examine floods finds that residential properties located within floodplains have lower market values compared to equivalent houses located outside floodplains (Gharbia et al., 2016). The study also demonstrates that the price discount from locating within a floodplain is significantly larger after a flood event than before. Research conducted by (Belanger & Bourdeau-Brien, 2017; Beltrán, Maddison and Elliott, 2019; Bui, Wen and Sharp, 2022; Votsis & Perrels, 2015) find a significant price drop for properties affected by flooding, influencing generally negative the housing market. It is also noted that the type of properties has an even greater effect, with larger price reductions in particular for coastal or waterfront properties, having a magnitude of more than 10% to even 30%. Findings also suggest that there is a variation of the effect of flood risk depending on the housing market cycle, with an even greater discount being observed in down markets.

However, more than few academics claim that although a negative impact on property prices with lowering values can be seen, the prices recover or even exceed their pre-flood levels due to property recovery work and improvements (Hirsch & Hahn, 2018). The results from (Atreya, Ferreira and Kriesel, 2013) indicate that flood risk discounts are transitory, and the perception of flood risk diminishes over time, while (Nguyen et al., 2022) show that the added discount progressively disappears over time just after the flood event, indicating that the concerns of homebuyer is short-lived. Additionally, literature considering the impact of flooding on housing prices examine that although a singular flood event reduces property values, repeated flooding does not necessarily result in further declines (Speyrer & Ragas, 1991). Research conducted by (Clayton et al., 2021) and (Lamond et al., 2019) also mention that there is evidence suggesting that repeated experience of hazard impact do influence risk assessment, but only for limited a time, unless the frequency remains high. Furthermore, in the case of repeated events, “people tend to forget” that a property is located in a high-flood zone. A resilient pattern that can be also attributed to several factors such as economic factors, market dynamics and investor behavior since the underlying value of real estate is influenced by a wide range of factors beyond the immediate impact of a flood event. Moreover, several analyses reveal that houses situated in flood areas experience a noticeable price variation due to their proximity to numerous amenities. These amenities include Central Business Districts (CBDs), green spaces, coastline, rivers, lakes, schools and even rail facilities. Thus, another intriguing pattern can be observed: even though a property is located in a high-flood risk zone, they sell with a premium due to their locational amenities (Gharbia et al., 2016; Atreya and Czajkowski, 2016).

Unveiling several articles, the effect of flooding risk on housing prices has been repeatedly examined. However, there is a notable distinction between areas that have experienced a flooding event and those that have not. Additionally, most of the research has been conducted in the United States or Asian countries, in which city designs are in a way different compared to European ones. Furthermore, lots of research mostly emphasizes on events of 10 to 20 years ago, in a period when climate change was not as crucial as it is today. It is useful to investigate if flooding risk does impact the residential property value and contribute new knowledge to the current literature.

1.3 Research problem statement

The main objective of this master thesis is to explore the relationship between flooding and residential real estate property prices, in the area of Cumbria, United Kingdom. Investigating the association of floods with housing prices in this region presents a research opportunity due to its unique geographical and historical characteristics. Cumbria is as a strategically unique area within the UK being the third largest county and hosting a notable concentration of civil and military nuclear industries along with various other economic activities. However, it is also an area prone to flooding due to its topographical landscape and coastal proximity that brings warmer mid Atlantic air, making it the wettest area in England. The area experiences the highest levels of total rainfall in the country and has rivers converging with some of Europe’s largest tidal ranges making it vulnerable to flooding incidents. Thus, the region has faced record-breaking floods, with one record surpassing the other (Environment Agency, 2023; Barker, Wilby and Borrows, 2004). Notably, the floods in Cumbria have occurred approximately every five years for the past half-century, including the devastating January 2005 floods, the November 2009 floods, and the catastrophic flood of December 2015. Floods described as the worst recorded in over 550 years. These repeated flooding’s have not only caused enormous suffering to the people and the infrastructure but have also significantly impacted and damaged the local real estate market. Investigating the implications and the resilience of Cumbria's real estate is worthwhile, as it can help improve flood risk management, promote economic stability, and deepen our understanding of how housing markets function in flood prone areas. An investigation that can provide insights for managing the correlation of natural disasters with housing markets and communities. It is crucial to recognize that Cumbria's housing market differs from that of a typical urban area.

The aim of this thesis is to contribute by expanding related research on another region and cities in England to close the gap of empirical evidence on market behavior and reaction patterns in the case of flooding and generally natural disasters. In particular, since this thesis will focus on flood events that did not occur in one specific place but in several areas and cities of the county that have been affected. The goal is to analyze three distinct flood events that took place in various areas, impacting the entire region of Cumbria in three different years, examining how housing prices in flood-affected areas have been influenced by these floods. Additionally, since the floods that occurred in Cumbria in 2005, 2009, and 2015 were primarily a result of intense and extended rainfall, which makes them predominantly pluvial, I contribute to the literature by examining the economic impact of pluvial floods. This area has received less attention compared to other type of floods, particularly in the UK area. The impact of pluvial floods risk may differ from other type of floods such as those caused by hurricanes or coastal floods, as the damage might not be as immense, resulting in lower discount of the property prices. The following research questions have been conducted in order to examine this relationship, namely:

Central Research Question: How is the market value of residential real estate in Cumbria, England associated with the risk of flooding?

To answer this research question, the topic is divided into three sub questions, which will be answered individually in within this thesis.

Research Question 1: What is the impact of flood risk on the prices of residential real estate properties based on literature?

What is the relationship between flood events and residential property prices based on literature? Is there a correlation to be seen or does a relationship between them not exist?

This question seeks to investigate whether a relationship exists between flooding and property values. The above question will be answered via a comprehensive review of the existing literature. The answer will be given through the construction of the theoretical framework of this thesis in chapter two. In this chapter the effects of natural hazards, especially flooding, on residential properties as described in the literature, are examined. The approach will be to first, explore the prices of residential properties, after which the effect of natural hazards is examined. Findings in these two areas lead to the formulation of the hypotheses of this study.

Research Question 2: What is the extent of the discount or premium applied to residential property in flood-affected areas of Cumbria?

Is there empirical evidence to support that residential properties in flood-affected zones within Cumbria experience larger discounts or premiums compared to properties in non-flood-affected areas?

The following question will be answered through statistical analysis of the empirical data using the hedonic pricing method. A method that is further specified below in the methodology part.

Research Question 3: To what extent do the effects of a flooding differ among residential properties, considering their property types and price segments, ranging from low-end, middle-end to high-end housing?

The above question will be addressed in detail using the Chow-test econometric technique. Findings will be obtained through the inclusion of heterogeneity tests for each target individually.

1.4 Conceptual Model

Based on theory about flooding risk affecting property prices a conceptual model is constructed as seen in Figure 1. The aim is to find a quantitative relationship between flooding and housing prices, moderated by flood affected areas.

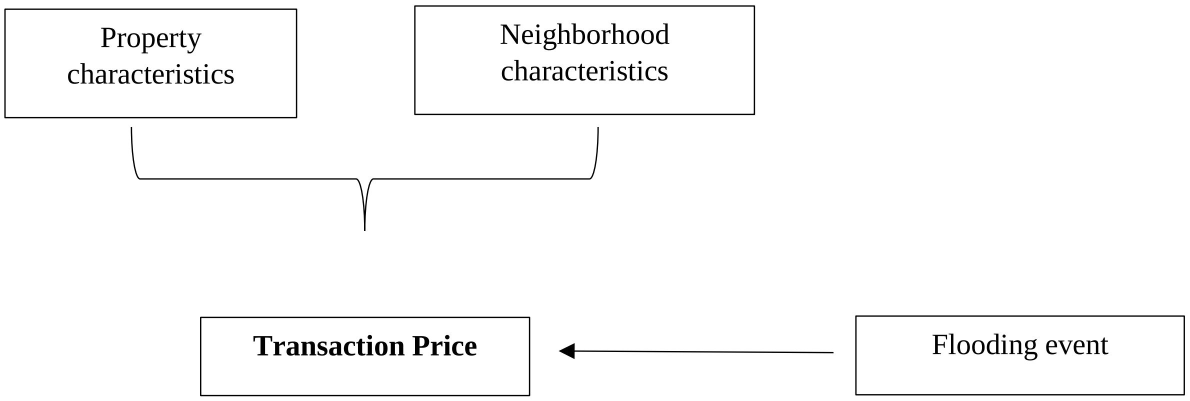


Figure 1*: Conceptual model explaining the effect of flooding on residential property price*

1.5 Outline

The remainder of this thesis is structured as follows. Chapter 2 consists out of the theoretical framework and hypotheses. The third chapter will examine data as well as the methods of the regression analysis, and also a more detailed explanation of the case study region data. The fourth chapter will include regression results and findings interpretation of the study while Chapter 5 discusses limitations of the study, main conclusion and some suggestions for future research and policy implications.

2. LITERATURE REVIEW & HYPOTHESES

2.1 Residential Property Prices

The literature review concentrates on research examining the impact of natural hazards on real estate prices. Considering the frequent and unpredictable nature of natural hazards, which have the potential to devastate people's belongings, it is logical for scholar in real estate to explore how these events influence real estate properties. In general, the housing market and property prices are affected by a complex interaction of economic, governmental, environmental, and social factors. Understanding these factors is critical for homeowners, investors, and regulators as they navigate the changing landscape of real estate.

Real estate in general and residential properties specifically are quite distinct from products traded in the market due to its characteristics, namely immobility and heterogeneity. Each property has its uniqueness since the possibility of two properties being exactly the same is minimal. However residential properties also share features such as living spaces, bathrooms and specific locations of land. The combination and variation of these features contribute to their heterogeneity, which has been extensively studied in literature, and examined mainly through the hedonic pricing method (Cheshire and Sheppard, 1995; Rosen, 1974). These characteristics can be broadly grouped into three categories; neighborhood, location and structural features. Furthermore, researchers like (Dubin, 1988; Stamou, Mimis and Rovolis, 2017), have demonstrated that these factors significantly impact and determine the prices of properties. The pricing dynamics of properties unfold in the marketplace where supply meets demand. It's worth noting that besides their heterogeneity the supply side of the real estate market remains stable as pointed out by (DiPasquale and Wheaton, 1992). As a result, short term price fluctuations are primarily influenced by the demand side of the market. Households acting as economic agents represent this demand side and therefore have a crucial role in determining residential property prices.

2.2 Impact of flood hazards on property prices

When studying risk factors like natural hazards we often view these hazards as geographic characteristics. The main finding in studies that investigate the effects of flooding risk is that the event of a large flood increases the perceived risk (Atreya and Ferreira, 2014). This means that the likelihood of events happening and the potential economic losses associated, are capitalized into the value of residential properties in areas prone to hazards. People assess the probability of an event occurring based on how examples or information about similar events can be recalled. To raise awareness about risks among citizens and help local governments formulate land use policies it is crucial to implement strategies such as hazard mapping and sharing hazard related information with the public as suggested by (Grothmann and Reusswig, 2006). These maps can also enhance disaster response capabilities and encourage individuals to adopt self-protective behaviors during natural disasters. However, when considering techniques, it's important to understand how risk discounts correlate with access, to hazard information since asymmetric information may impact housing prices according to (Akerlof, 1970). If there is a difference in the information, between the buyer and the seller it becomes difficult for the buyer to assess whether the price of a house accurately reflects its quality. This lack of certainty can lead to properties being overvalued or undervalued and ultimately lead in market inefficiencies. In comparison to buyers, sellers typically have the advantage of possessing information regarding potential flood risks and past experiences (Pope, 2008). However, when hazard information is publicly available, it is likely that the flooding risk discount is reflected in the prices of properties in hazardous zones and improves market efficiency. Understanding the contrast between market efficiency and inefficiency is crucial for stakeholders in the housing market, including buyers, sellers, and investors. An efficiently competitive market provides trustworthy and transparent prices. On the other hand, market inefficiency is a challenge as well as an arbitrage opportunity for investors who price abnormalities and information gaps.

In an efficient housing market, which is characterized by a perfect, complete, costless and instant transmission of information, property values in vulnerable areas tend to be lower than the risk-free areas, since both sellers and buyers identify hazard information which affects the acceptance of a lower price for a property (Beltrán, Maddison and Elliott, 2019). Several scholars acknowledge that flood hazards have a negative impact on housing prices by using various statistical models such as hedonic pricing models, repeat sales analysis and difference-in-difference estimation. For example, (Zhang, 2016) finds that being located within a floodplain reduces property value, with the negative impact being stronger among lower-priced homes and weaker among higher-priced homes. In another paper, (Shr and Zipp, 2019) investigate the impact of flood zone designation on property values. The results show that when a property is assigned to flood zone status, housing values decrease by more than 11%. Analyzing a large dataset of over 4.8million houses in the UK (Beltrán, Maddison and Elliott, 2019) reveal that there is a reduction in property values following flooding, ranging for highest price quartile from 10.1% to 31.4% for lowest price quartile. These researchers also found that the magnitude of the impact and the origin of flooding affect the extent of price discounts. Another study of (Lamond et al., 2010) examines the effect of floods on the price of residential properties that are transacted in 13 different areas in the United Kingdom using a variation of the repeat sales approach. The results demonstrate that the impact of flood events on property prices is highly variable, without any notable impact of flood designation. Moreover, pluvial flooding and its economic effect on property values in Ho Chi Minh City, Vietnam, are examined by (Bui et al., 2022). The study integrates the hedonic property model in a difference-in-differences framework and a spatial econometric analysis. Researchers find that following a major flood in September 2017, values of the flood affected properties experienced a 9% drop. The study's conclusions imply that the housing market has suffered as a result of the increased danger of floods. Further investigation by (Belanger & Bourdeau-Brien, 2017) that concentrates on residential properties in England, discovers also that flood risk significantly lowers property values with averages of 1.5% and that waterfront properties have an even larger decrease. Findings also suggest that the impact of flood risk varies depending on the housing market cycle, with a greater discount observed in down markets. Subsequently, risk discounts vary between regions, depending on the amount of damage caused in each area and the timespan of each flood. Thus, the effect of flooding on the housing market varies both temporarily and spatially.

As previous researches have already shown, flood risk reduces property prices. Yet, when new information is obtained, the risk perception of consumers may change, and residential property values adjust accordingly and unexpected patterns are formed. Thus, many researchers argue that, while there is a negative influence on property prices with lower values, prices recover or even exceed pre-flood levels. (Atreya et al., 2013) examine the impact of a large flood event on residential property prices and the persistence of this effect over time. While the findings indicate that there is a significant decrease in price, this effect was short-lived since it disappeared between four and nine years after the disaster occur. According to the study, flood risk discounts are temporary, and as the time passes the perception of flood risk diminishes. Another research by (Nguyen et al., 2022) examine the effects of flood risk on housing prices in a coastal area of New Zealand. Specifically, the study focuses on the Minimum Floor Level (MFL) zone, which was imposed by the city council as a requirement for new construction in areas identified as at risk of flooding. Applying a 'diff-in-diff-in-diff' strategy in hedonic regression analyses, they found that houses in the MFL zone sell for a discount of about 5 per cent prior to the flood. However, the added discount that was noticed just after the flood event progressively disappears over time, indicating that the concerns of homebuyers are short-lived. Further, a study conducted by (Hirsch & Hahn, 2018) measures the impact of a 100-year flood risk on property rents and values in Germany. On average, purchasing properties situated in flood prone areas are priced at EUR299 lower per square meter when compared to those outside the risk zone. Rental prices also experience a decrease in flood affected regions. However, they eventually rebound and even surpass their pre-flood levels due to property recovery work and improvements. (Bin & Landry, 2013) find also that property values in flood plains significantly decline after hurricanes, falling from almost 5% to 8%. The risk premium for properties in flood zones does, though, reduce with time and essentially disappears roughly 5 or 6 years, according to the study. These results imply that if hazard events become more frequent, buyers and seller's perceptions of risk as well as insurance requirements may change.

Nevertheless, recent literature also reports contradictory results. As illuminated by the research findings, repeated flooding poses a complex challenge to property markets resulting into intriguing pattern. A study conducted by (Clayton et al. 2021), who analyze the impact of climate change on asset values found that although property prices initially experience a decline following a climate event, there is evidence that repeated experience of hazard does impact the influence of risk assessment, but only for a limited time, unless the frequency remains high. Building on this narrative (Speyrer & Ragas,

1991) examine the impact of flood risk and mandatory flood insurance on property values using a dataset of almost 2,000 homes sold in the New Orleans from 1971 to 1986. The analysis shows that being located in a floodplain reduces property values and also confirms that repeated flooding does not further reduce property values.

As discussed thus far, it becomes evident that in most cases, these hazards tend to have an influence on property prices, and mostly result in price discounts. However, delving deeper into the dynamics, a more nuance perspective is revealed. While flooding may lead to a short-term depreciation in property values, an interesting pattern arises over the long term. Contrary to immediate expectations, housing prices show resilience, often bouncing back to pre-flood levels or even surpassing them. Surprisingly, repeated occurrences of flooding do not necessarily worsen the negative impact on property prices. Something that could be attributed to high awareness among citizens regarding the potential dangers, government interventions employing flood mitigation techniques or economic factors such as supply and demand. Furthermore, it is also worth noting that, paradoxically, there is a rise to a premium in property values. Properties located in areas with significant flood risk, such as those near coastlines or major rivers and lakes, exhibit this trend. In that case, flood risk discounts co-exist with amenities derived from these bodies of water (Bin & Kruse, 2006) and homeowners enjoy various benefits from their proximity to these natural features. Influenced by their proximity to amenities such as Central Business Districts (CBDs) rivers, lakes, schools, green spaces, rail facilities and the coastline, homeowners are willing to pay for flood insurance to enjoy living near water while taking flood hazards into account in the valuation of their property (Atreya & Czajkowski, 2014**;** Bin, et al., 2008; Kim, et al., 2017). Consequently, natural amenities can create a price premium leading property values in hazardous zones being valued higher than in a safe-spaces, a concept supported by (Rajapaksa, et al., 2017). The relationship between floods and house prices is multifaceted and not straightforward. It involves short-term drops and long-term recoveries, influenced by people's awareness, government actions, and economic factors. By understanding these dynamics, it allows for a broader comprehension on how property values in flood prone areas are shaped.

2.3 Hypotheses

The literature review allows us to derive theoretical predictions, and answer *"**What is the impact of flood risk on the prices of residential real estate properties based on literature?**"*. As research shows properties located in a flood-prone area sell with a discount than those that are likely to remain unaffected. Flood events impact the way consumers asses associated risks by the magnitude of their impact, which thereafter is reflected in property valuation. Consequently, consumers are pricing their perceived risk into the valuation of a specific property. However, price premiums could occur caused by the amenity of living near coastlines, green spaces or rivers. In addition, risks are assessed more strongly based on historical events than based on future flood models.

Based on the literature findings three hypotheses can be stated and specified to answer the composing research questions.

H1: "A flood has a significant negative effect on housing prices in flood affected areas."

H2: "Residential properties located within areas affected by recent floods of January 2005, November 2009, and December 2015 in the area of Cumbria sell at a discount compared to similar properties un-affected by these floods."

H3. "The discount of a residential property, as a result of a flood, varies with heterogeneity in property types and prices segments across property."

3. DATA & METHODS

3.1 Context

Cumbria situated in the northwest of England, is a ceremonial county that shares its borders with Scotland to the north, and the counties of Yorkshire and Durham to the southeast. The county is predominantly rural, with an area of 6,769 km2 (2,614 sq.m mi) and is home to a population of 500,012. This makes it the third largest ceremonial county in England in terms of area, yet the eighth smallest in population size (Cumbria County Council, 2023). The area of Cumbria includes districts such as Allerdale, Carlisle, Eden, South Lakeland, Copeland, Borrow-in-Furness and cities like Carlisle, Workington, Cockermouth and etc. Noteworthy settlements after Carlisle (74,281) which is the capital city of the county, include Barrow-in-Furness (56,745), Kendal (29,593), and Whitehaven (23,986). Cumbria is well-known for its natural beauty and much of its landscape is protected. Nestled in the mountainous heart of the county, the Lake District which is primarily famous for its mountainous range and greenery, it includes England's tallest peaks and biggest lakes, all within the protected confines of the Lake District National Park. Figure 2 depicts a map illustrating the various areas of the county, and how they all combine to make the county of Cumbria. It also highlights the location of the Lake District, the mountainous areas as well as the greenery of the region (Cumbria Tourism, 2023).

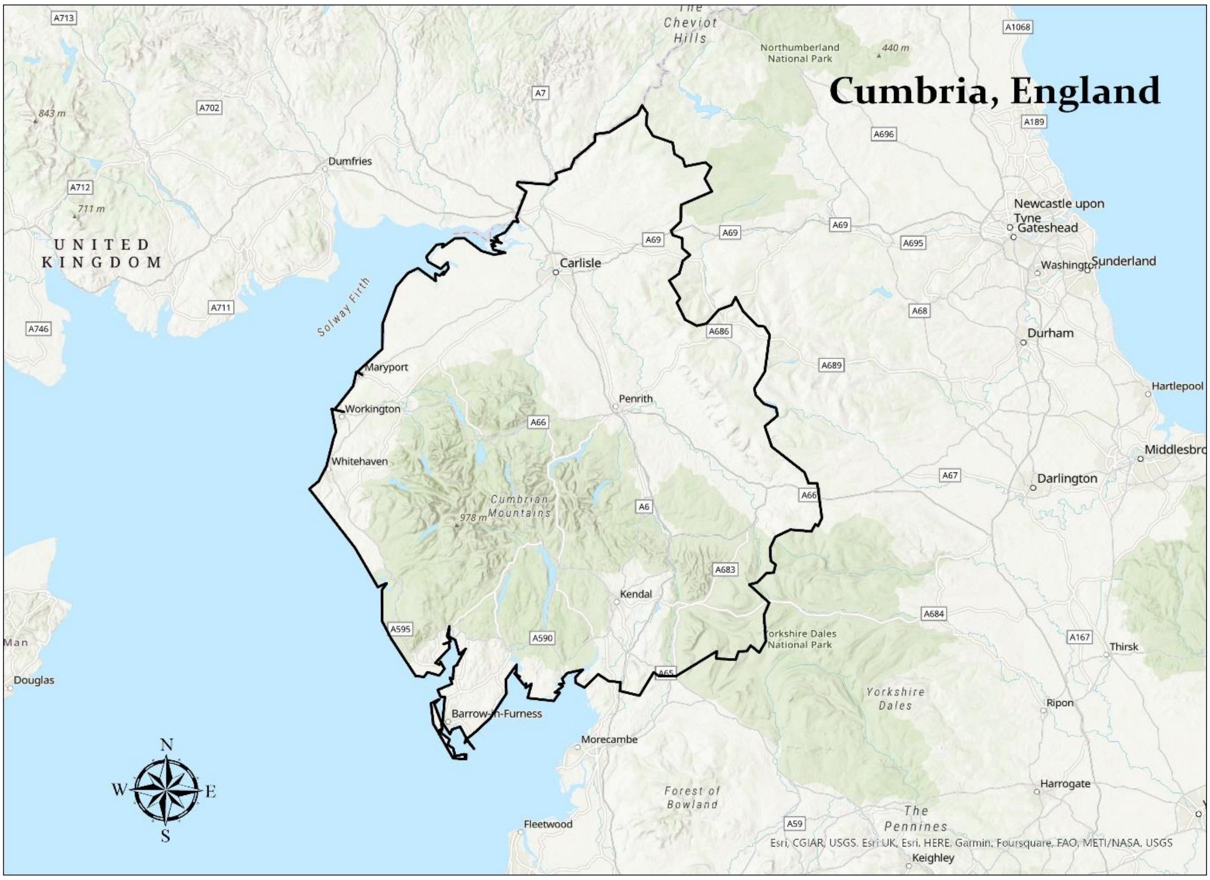


Figure 2: Map depicting the county of Cumbria and the North of England

Due to its proximity to the Irish sea, Cumbria is exposed to the warmer air of the mid-Atlantic. The warmer the air, the more moisture the clouds produce. This, coupled with its mountainous landscape, give the area an enormous amount of rainfall. In fact, Cumbria and its Lake District is the wettest region in the whole of England and prone to an elevated risk of flooding (Environment Agency, 2023). While Cumbria experiences frequent rainfall, its valleys tend to drain quite quickly through its well-established river courses. Unfortunately, communities in Cumbria have suffered several incidents of severe flooding in recent years most notably in 2005, 2009 and 2015, the worst floods in more than 550 years. Over 50 flood events occurred from 1800 to 1979, with flooding occurring approximately every 11,4 years and major floods happening roughly every 42,7 years. Furthermore, Cumbria has the highest rainfall totals in England, streams with the steepest slopes, and the meeting of rivers with some of the highest tidal ranges in Europe (Barker, Wilby and Borrows, 2004).

Floods Background

In January 2005, an exceptional catastrophe hit Cumbria (Harper, 2005), England, when 15% of the region's average annual rainfall poured down in just 36 hours, an event expected once in 200 years. The consequences were staggering: over 2,000 properties flooded up to two meters, leaving over 3000 people homeless for up to a year, and the economic impact was severe, with estimated losses exceeding £450 million (Pryce, Chen and Mackay, 2009)**.** The disaster began with severe weather warnings issued on January 6th and 7th, as relentless rain and strong winds persisted for three days. Flood Watch issued flood warnings for certain areas, but on 8^{th} and 9^{th} January 2005 a month’s rain fell within thirty-six hours on the Eden valley, resulting into flooding and destroying Carlisle and the Eden Valley. The extended rainfall occurred over the high ground of the nearby Lake District and Pennines, draining into the River Eden on which Carlisle is situated. The impact was particularly severe in the Eden valley, surpassing the catastrophe of 1822 and the city of Carlisle alone evacuated 10,000 people, with over 1900 properties and nearly 300 businesses damaged. Further, power loss affected 76,000 homes due to a flooded electricity substation (Whyte, 2009). Historically, the Eden catchment experienced several floods, but the January 2005 event was incomparable. The flood peak, over 1 meter higher than any in the previous 200 years, submerged Carlisle city center and neighboring cities (Spencer et al., 2017). The highest rainfall (180.4 mm) was recorded at Rydal Hall in Cumbria different places had over 100 mm of rain in one day. This is estimated as likely to occur less than once in 200 years (Met Office, 2005).

On 19-20 November 2009, as a result of a prolonged period of record-breaking rainfall over the mountains of the central Lake district in north-west England, many of the rivers and lakes within the region experienced exceptionally high flows, with the greatest devastation occurring along the River Derwent and the nearby areas. This exceptional event followed several days, and caused widespread flooding in Cumbria, which particularly affected the communities of Cockermouth and Workington in

Allerdale district (Stewart et al., 2010). Earlier in the week the rainfall ensured that the ground was saturated in many areas and a warm, moist south-westerly airstream, associated with a deep Atlantic depression, contributed to the flood eastwards (Met Office, 2009). The rainfall averaged over 10 mm/hour for over 36 hours, and the rain scale at Seathwaite in the headwaters of the Derwent river established a UK record of 316.4 mm of rainfall over 24 hours (Miller et al., 2013; Sibley, 2010). This exceeded the previous record of 279 mm, which was the recognized 24-hour maximum rainfall, during the Martinstown storm of July 1955. The impact was devastating, with flooding affecting five out of six boroughs, the only one escaping being Barrow-in-Furness. The effect on properties was concentrated in Allerdale and South Lakeland with the most significant infrastructure damage occurring in Allerdale. Tragically, one person lost his life, and 2,239 properties, 250 farms, and 25 bridges were affected. The Port of Workington closed, and 3,057 businesses suffered, with an estimated weekly cost of £2 million due to travel disruptions. Cockermouth a city in the district of Allerdale, absorbed a major impact with over 80% of the business located there being affected. The aftermath highlighted the vulnerability of the region to extreme weather events, leaving a lasting impact on communities, businesses, and infrastructure (Cumbria County Council, 2023; Wedawatta, et al., 2013).

In December 2015, Cumbria was subjected to its third extreme flood event in a decade. The county was impacted by the Storm Desmond which was unparalleled in many aspects. Not, did it bring record breaking rainfall and river flows but it also caused extensive flooding and affected a large number of people and properties. In specific detail, the measurements taken at Honister Pass documented 341.4mm of rainfall in the 24 hours leading up to December 5, setting a new record for any 24-hour rainfall period in the UK. Furthermore, an additional record for the UK emerged at Thirlmere, with 405.0mm of rainfall recorded in just 38 hours, establishing another record for a 48-hour rainfall period (Met Office, 2015). Cockermouth, Appleby, Carlisle, Keswick, Kendal, and surrounding areas reported severe flooding, with Carlisle's water levels rising to 7.9m in the early hours of Sunday, exceeding the previous record level of 7.2m recorded in 2005 (Environment Agency, 2016). The scale of flooding in December 2015 was unprecedented in Cumbria. The communities most affected by this event were Carlisle, Allerdale, Eden, Barrow-in-Furness and South Lakeland. In total 7,465 properties were flooded equating to an estimated 14,694 persons. Carlisle district shows a higher proportion of households flooded in more deprived areas. During the peak flooding period around 17,911 customers were left without electricity. Initially about 667 properties faced water supply issues, particularly in Allerdale and Eden as well as the Network Rail encountered 127 incidents due to the floods. Also, it impacted 45 schools resulting in approximately 3,034 children not being able to attend school until the end of the autumn term (Cumbria County Council, 2016). In addition, the estimated cost for damages, across the UK ranged from £5 billion to £5.8 billion with insurance payouts estimated at £1.3 billion. The aftermath was characterized by extensive flooding leading to two fatalities and widespread inundation of homes and businesses (Spencer et al., 2017).

Mapping and visualizing flood affected areas

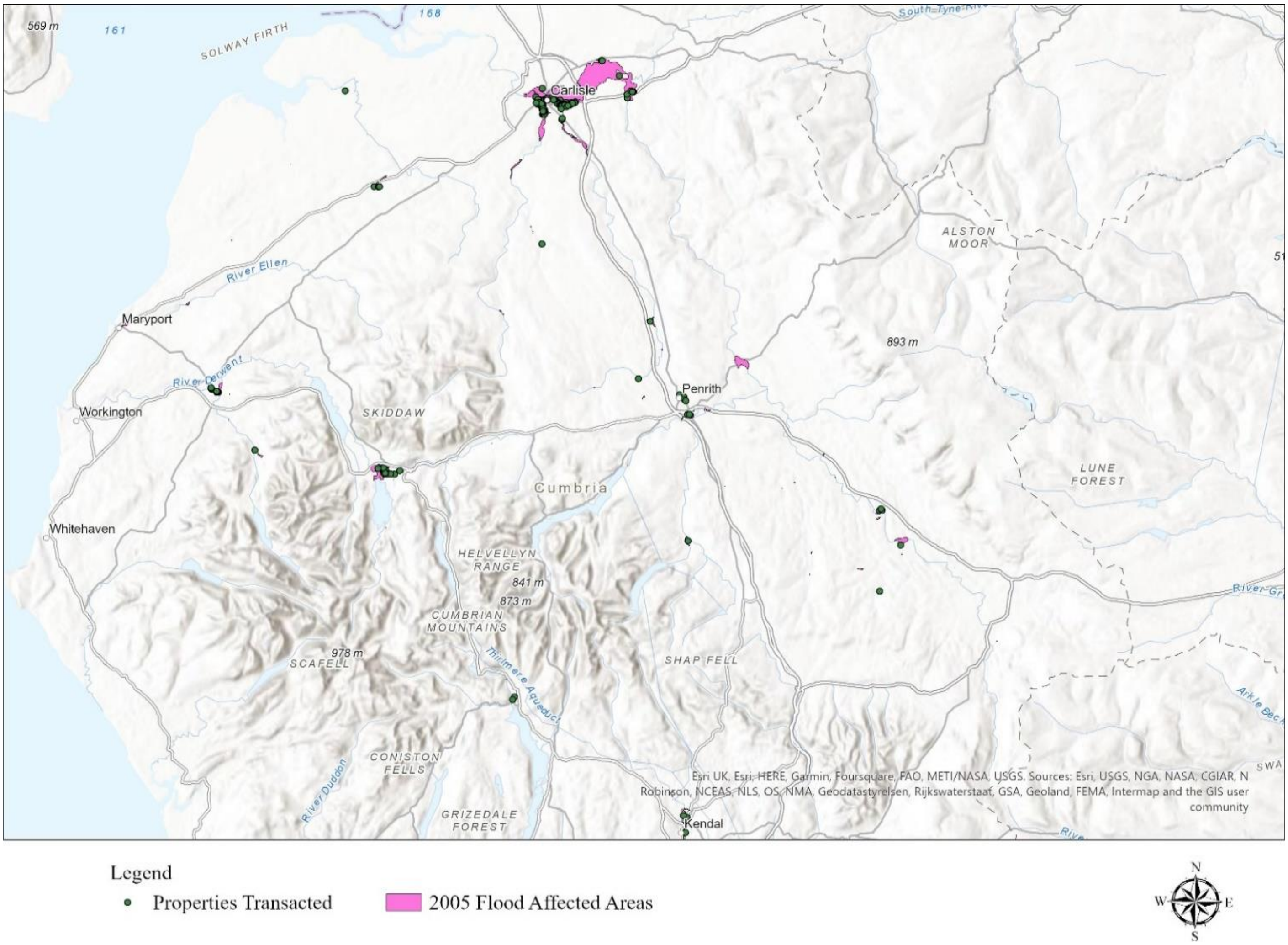


Figure 3: The pink outlines are the areas affected by the historic flood of 2005 in the area of Cumbria. The green circles represent the properties, which are affected by the floods and transacted later.

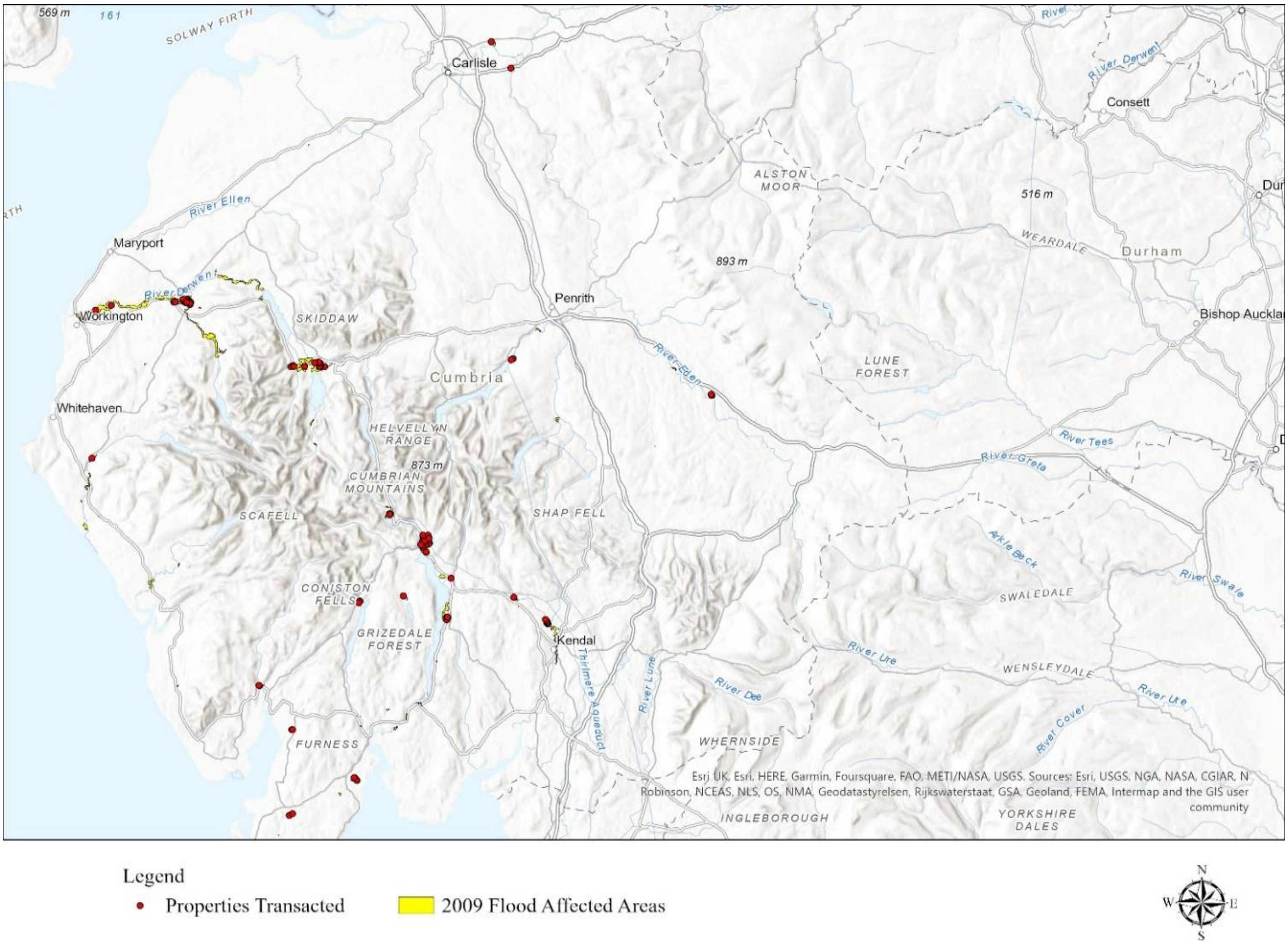


Figure 4: The yellow outlines are the areas affected by the historic flood of 2009 in the area of Cumbria. The red circles represent the properties, which are affected by the floods and transacted later.

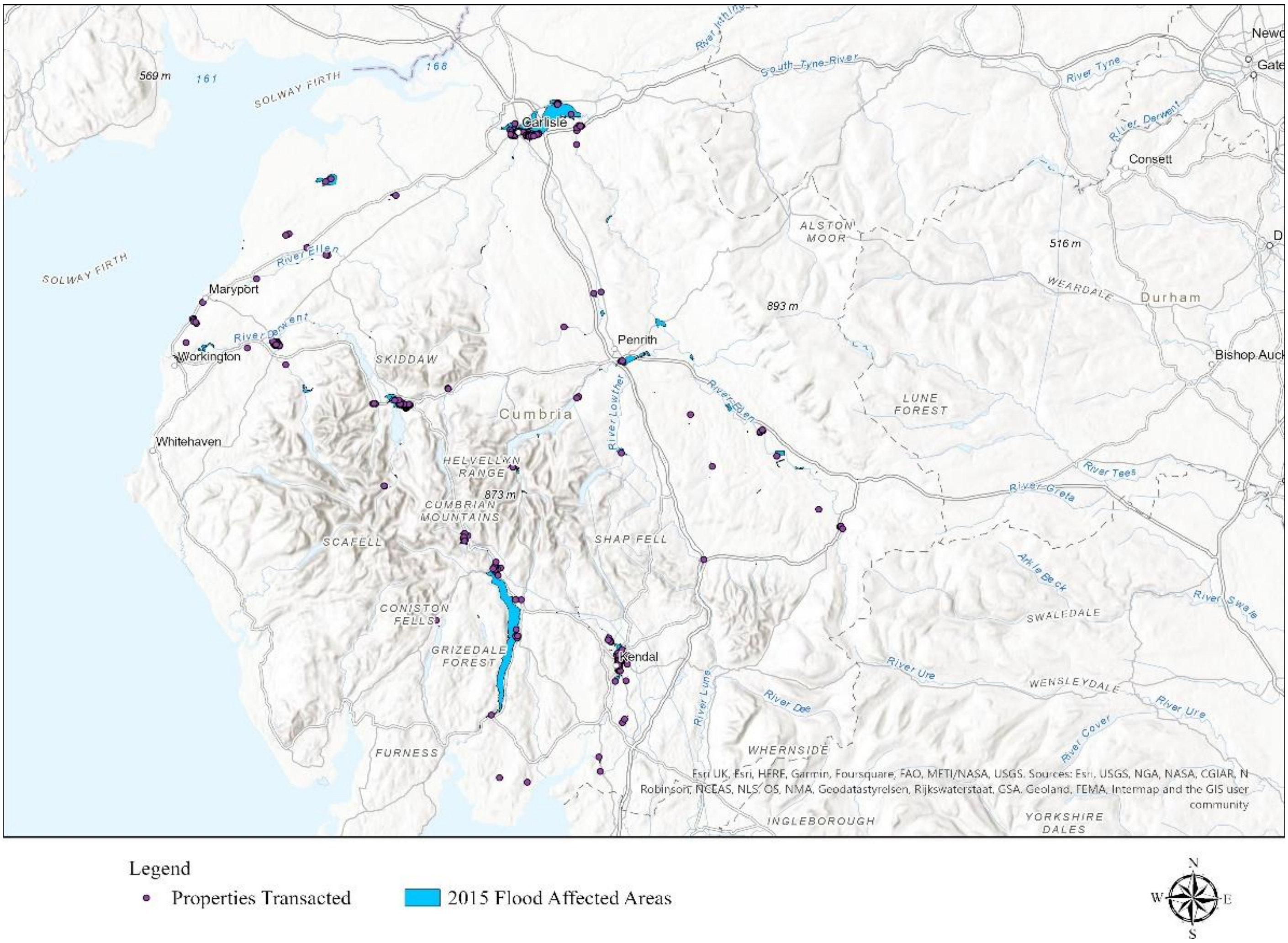


Figure 5: The blue outlines are the areas affected by the historic flood of 2015 in the area of Cumbria. The purple circles represent the properties, which are affected by the floods and transacted later.

3.2 Data collection

The analysis relies on two datasets from different sources. The transaction prices of the area of Cumbria from 2003 to 2022 that have been obtained from the HM Land Registry (HM Land Registry, 2023), which registers all transaction prices for the given location and time period. For the data about the flooded areas, I used the recorded flood outlines from the UK government website in form of a GIS layer (Environment Agency, 2023). The transaction price data has a total of 184,048 observations in the area of Cumbria in the time period of 2003 to 2022. Transaction prices below or above a threshold may represent outliers, hence any transactions outside a price range from 10,000 to 2,500,000 pounds have been excluded.

The recorded flood outline offers a precise geographical map outlining the floods in 2005, 2009 and 2015. With the ArcGIS Pro system, we can create a layer of the transactions with the exact location. In order to indicate, which properties are located within a flood affected area and which properties are located in a non-flooded area, I have to include the transaction dataset and the recorded outline map in GIS. Then, I can generate x and y coordinates for each transacted property. By intersecting the recorded flood outlines with the layer of transactions, which are both based on x and y coordinates, I created an additional layer showing which properties have been affected by the respective floods. By merging the datasets by the coordinates, I have exact determination whether a property was affected by either flood or is located in a non-flooded area. This allows us to be exact with the statements, which transaction has been affected by the floods. It is important for the study to run the analysis with the exact positions in order to get representative results.

The recorded flood outline map shows the recorded flood outlines of recent floods, recorded since 1969 (Environment Agency, 2023). This layer gives detailed information about the cause and time period of the floods. The recorded flood outline includes observations of several flood events: the January 2005 flood, the November 2009 flood, and the December 2015 flood, are the particular incidents that are of interest. After dropping the observations from the floods that are out of interest, there are 2,828 properties, which have been affected by the flood in 2005, 1,453 properties affected by the November 2009 flood, and 5,283 properties affected by the December 2015 flood. These floods are predominantly same in their origin. While the 2005 flood was caused by heavy rainfall, and the main river flooded due to an overtopping of the existing defences, the 2009 flood was also caused by heavy rainfall which surpassed the previous record with a rainfall of 316.4mm in 24 hours. Consequently, the main rivers overflowed and flooded due to the absence of raised defenses as well as overtopping these defences. This amount of rainfall in 24 hours set a new record for any 24-hours rainfall period in the UK. The source of the 2015 flood was also heavy rainfall, reaching 341.4mm in 24 hours and setting a new high record for any 24-hour rainfall in the general area of the UK. The main rivers in the areas overflowed due to channel capacity being exceeded and the overtopping of defences. Therefore, in our case, properties close to the tidal area were not as much affected by the flood. Instead the floods seemed to spread across the whole cities and areas near rivers. Thus, the floods that occurred in Cumbria in 2005, 2009, and 2015 were primarily a result of intense and extended rainfall, which makes them predominantly pluvial. However, they can also be described as fluvial since the heavy rainfall led to the overflow of the rivers.

Flood-affected properties in Figures

The number of properties affected, as mentioned by the impact assessment conducted by Cumbria County Council, differs from the transactions directly influenced by floods. In Table 6, you can see a subset of properties that were directly involved in transactions impacted by floods. The variation in numbers stems from the fact that a property, although impacted by a flood, may not necessarily undergo a transaction during or after the event. Some property owners choose to retain their homes despite the flood impact, while others may face challenges in finding a buyer. So, it is essential to recognize that the reported numbers of flood-affected properties differ from the transactions made during or after the flood event. Additionally, it is notable that floods may affect specific properties up to a certain point, leaving neighboring properties unaffected. Consequently, these neighboring houses are classified as non-flooded (Recorded Flood Outlines, 2023). The dynamic nature of flood events and potential changes in flooding patterns contribute to the divergence in figures, highlighting the complexity of assessing and understanding flood impacts. Thus, the number of flooded properties varies depending on methodology.

Table 6: Flood-Affected Properties

Floods	**Environment Agency** *(Flood-Affected Properties)*	**ArcGIS Pro intersection** *(Flood-Affected Transactions)*
January 2005 Homes Flooded	Over 2,000	2,828
November 2009 Homes Flooded	2,239	1,453
December 2015 Homes Flooded	7,465	5,283
Non-Flood-Affected Homes	-	174,484

Note: Table 6 depicts in column 1 the total figures based on data provided by (Cumbria County Council, 2023; Environment Agency, 2023). In column 2 the total figures based on the spatial merging in ArcGIS Pro of the two basic datasets, the property transactions dataset and the recorded flood outlines layer are stated.

3.3 Descriptive statistics

As shown in the table 7, there are 184,048 total observations after merging the datasets. I separated the data into four different groups: flood-affected transactions of 2005, 2009 and 2015 and non-flood-affected transactions. The Descriptive Statistics tables below show that the flood in December 2015 affected a lot more properties in the area of Cumbria than the January 2005 and November 2009 floods did. During the flood in 2015, 5,283 properties were located in the flood zone, compared to 2,828 properties in 2005 and the 1,453 properties in 2009. There are 174,484 properties located in the non-flood zone, meaning that these properties were neither affected by the flood of 2005 nor in 2009 or 2015. For all observations, the transaction prices range from 10,000 pounds to 2,5 million pounds, which stays constant across the three flood groups as well, with minor changes. The mean varies to some extent across the groups, the flood 2009 has the highest mean of 226,284.92 pounds and the flood 2005 has the lowest mean of 146,335.9 pounds.

The following variables represent the control variables during the regression: property type, new build and estate type. The control variable of property type is separated by detached, semi-detached, flat/maisonettes, terraced, or other. In all flood groups and the non-flood area, the property types are distributed somewhat evenly, with the terraced properties being the most present property type. Interestingly, the property types between the flood zone 2005 and the non-flood area behave very similar, while the flood zone in 2009 and 2015 shows some differences. The flat and maisonettes are the second most occurred property types in the December 2015 and November 2009 flood affected area, while the property type of semi-detached real estate occurs second most often in the January 2005 affected flood area and the non-flood area. The variable for new build defines whether the property sold has been newly built. The number for new built properties is higher in the non-flood affected areas than in the other ones. In the non-flooded zone, there are 7.7% properties new built, while in the flood zone of 2005 there are 5.5%, in the flood zone of 2009 there 5.6%, and in the flood zone of 2015 there are 4.2% new built properties. For the control variable of the estate type, which describes whether the properties are sold as freehold or leasehold we find similar behaviour. In the UK, freehold describes the ownership of the building and the land, while leasehold is considered as leasing the property from the freeholder for a prespecified time period, which last usually around 90 to 120 years. In the flood zone 2005 and the non-flood area the estate for freehold is with 77.8% and 88.6% dominant, while in the flood zone 2009 and 2015 the estate type for freehold and leasehold splits pretty evenly, the freehold being slightly higher with 61.5% and 74%.

Table 7 Descriptive Statistics

Non-Flooded Area

Variable	**Observations**	**Mean**	**Std. Dev.**	**Min**	**Max**
Price Paid	174,484	176,224.4	140,516.5	10,000	2,500,000.00
Property Type	174,484	.	.	.	.
Detached	174,484	0.245	0.43	0	1
Flat/Maisonettes	174,484	0.077	0.267	0	1
Other	174,484	0.029	0.167	0	1
Semi-detached	174,484	0.286	0.452	0	1
Terraced	174,484	0.363	0.481	0	1
Newbuild (1=Yes)	174,484	0.077	0.267	0	1
Estate Type (1=Freeh.)	174,484	0.886	0.317	0	1
Year	174,484	2012.477	6.146	2003	2022

Flood January 2005

Variable	**Observations**	**Mean**	**Std. Dev.**	**Min**	**Max**
Price Paid	2,828	146,335.9	112,973	10,545	2,100,000.00
Property Type	2,828	.	.	.	.
Detached	2,828	0.084	0.278	0	1
Flat/Maisonettes	2,828	0.170	0.384	0	1
Other	2,828	0.032	0.177	0	1
Semi-detached	2,828	0.180	0.376	0	1
Terraced	2,828	0.532	0.499	0	1
Newbuild (1=Yes)	2,828	0.055	0.229	0	1
Estate Type (1=Freeh.)	2,828	0.778	0.415	0	1
Year	2,828	2012.352	6.259	2003	2022

Flood November 2009

Variable	Observations	Mean	Std. Dev.	Min	Max
Price Paid	1,453	226,284.92	172,579.13	10,138	1,983,115.00
Property Type	1,453	.	.	.	.
Detached	1,453	0.122	0.327	0	1
Flat/Maisonettes	1,453	0.251	0.434	0	1
Other	1,453	0.066	0.248	0	1
Semi-detached	1,453	0.122	0.327	0	1
Terraced	1,453	0.439	0.496	0	1
Newbuild (1=Yes)	1,453	0.056	0.231	0	1
Estate Type (1=Freeh.)	1,453	0.615	0.487	0	1
Year	1,453	2012.758	6.352	2003	2022

Flood December 2015

Variable	Observations	Mean	Std. Dev.	Min	Max
Price Paid	5,283	184,230.68	142,205.0	10,000	2,100,000.00
Property Type	5,283	.	.	.	.
Detached	5,283	0.113	0.316	0	1
Flat/Maisonettes	5,283	0.213	0.409	0	1
Other	5,283	0.042	0.201	0	1
Semi-detached	5,283	0.179	0.383	0	1
Terraced	5,283	0.453	0.498	0	1
Newbuild (1=Yes)	5,283	0.042	0.201	0	1
Estate Type (1=Freeh.)	5,283	0.74	0.439	0	1
Year	5,283	2012.37	6.202	2003	2022

Note: The descriptive statistics show the number of observations, the mean, the standard deviation, the minimum, and the maximum for each variable of property characteristics, separated by the flood affected area of the January 2005 flood, the November 2009 flood, the December 2015 flood, and the non-flooded area.

3.4 Methodology

This research aims to find the relationship between flood risk and property prices. Therefore, the model shows the impact of the key independent variable, the January 2005 flood, the November 2009 flood and the December 2015 flood, on the dependent variable, the transaction price. Since the distribution of the transaction prices is skewed, I take the natural logarithm of the transaction price so I can obtain a normal distribution. The natural logarithm of the dependent variable helps with the linearity between the dependent and independent variables and improves the results. I include control variables describing property characteristics like property type, estate type, and new build, time characteristics like sales month and sales year, and location effects like postcode and district to reduce bias. The estate type and new build variables are dummy variables determining whether the property is a freehold or a leasehold, and whether the property is newly built. In regression equations, dummy variables typically take binary values 0 or 1 to signify the absence or presence of specific categories. Occasionally we must introduce a factor that has two or more distinct levels (He, Renshaw and Szelest, 1998). In our empirical econometrical analysis if the value of estate type is equal to 1 the observation is a freehold and if the value is equal to 0 the observation is a leasehold. The same applies to the dummy variable of new build where if the variable is equal to 1 the observation is a newly built and if its equal to zero the observation isn't a newly built property. The property type which will be in form of a categorical variable determines whether the property is a detached, semi-detached, terraced, flat/maisonettes, or other property. The reference property type is set to be the detached type. As for the location effects the independent variables included are: district which describes the district where the property sold is located as well as the variable of postcode which characterizes the postcode where the property sold is located. Both of the spatial characteristics will be represented as categorical variables. Regarding the time characteristics we have the independent variables of sales year which is a categorical variable describing the transaction year, while the variable of sales month will be also a categorical variable reporting the transaction month in our models.

To determine the correlation of flood risk and property prices, I employ a hedonic pricing model to see the price change after a flood in a flood affected area compared to the non-affected area. Hedonic models are commonly used to find the relationship between house prices and external effects. In order to examine the research question 2 and 3 and test the hypothesis made that *"A flood has a significant negative effect on housing prices in flood affected areas"* and the second hypothesis that *"Residential properties located within areas affected by recent floods of January 2005, November 2009, and December 2015 in the area of Cumbria sell at a discount compared to similar properties un-affected by these floods"* , I utilize the following statistical models to explain the relationship between flooding and residential real estate property prices. The equation of the primary model is stated as follows:

$$ln(P_{it}) = cons + \boldsymbol{\beta_1 Floodarea2005} + \boldsymbol{\beta_2 Floodarea2009} + \boldsymbol{\beta_3 Floodarea2015} + \beta_4 EstateType + \beta_5 NewBuild + \beta_6 PropertyType + \beta_7 District + \beta_8 PostCode + \beta_9 SalesYear + \beta_{10} SalesMonth + \varepsilon_{i,t} \quad (1)$$

where $ln(P_{it})$ represents the natural logarithm of transaction price, β is the coefficient for the independent variables, and $\varepsilon_{i,t}$ is the error term, which captures the factors not included in the model that affect the dependent variable Y. Additionally, the error term represents the unobservable and potentially unpredictable factors that can affect the dependent variable but are not explicitly model by the included independent variables. It captures the effects of these unobserved factors and the fundamental uncertainty in the relationship between Y and X (Brooks and Tsolacos, 2010). The notation of *i* represents the transactions being made: i = 1, 2, 3, 4…., N, while the notation of *t* formulates the time the transactions occurred: t = 1, 2, 3, 4…., N.

Next, the basic hedonic equation is further developed into two different models. In the second model, described as my base model and serving as the primary focus of this analysis, I exclude specific observations from my property transactions dataset in order to address potential spatial correlation between the dummy variables of flood. If the dummies are highly correlate and if the events relate to the same area, then basically multicollinearity may appear, which is a statistical concept where several independent variables in a model are correlated (Brooks and Tsolacos, 2010). Since the majority of the cities and areas affected by the floods included in the investigated area are not the same, we create three different flood dummies. The only exception to this idea is the capital city of Cumbria, named Carlisle, and some other minor insignificant villages that might strengthen the correlation issue. This only happens in two out of the three floods we are investigating, specifically the 2005 flood and the 2015 flood, both of which include the capital city of Carlisle. In order to tackle this issue and avoid multicollinearity, I will exclude the capital city of Cumbria and the other related areas. Thus, the equation of our base model is stated as followed:

$$ln(P_{it}) = cons + \boldsymbol{\beta_1 Floodarea2005(capitalex)} + \boldsymbol{\beta_2 Floodarea2009} + \boldsymbol{\beta_3 Floodarea2015(capitalex)} + \beta_4 EstateType + \beta_5 NewBuild + \beta_6 PropertyType + \beta_7 District + \beta_8 PostCode + \beta_9 SalesYear + \beta_{10} SalesMonth + \varepsilon_{i,t} \quad (2)$$

where all variables remain the same as in the previous model. The only difference is the addition of the key independent variable of 2005 flood and 2015 flood, where in this case are describe as followed: $\beta_1 Floodarea2005(capitalex)$ which is a dummy variable that equals 1 if property sold is located within the area affected by the flood of 2005, excluding the area of the capital city of the county, and $\beta_3 Floodarea2015(capitalex)$ that is also a dummy variable that equals 1 if property sold is located within the area affected by the flood of 2015, excluding the area of the capital city Carlisle.

However, considering the significance of Carlisle as the county's capital and its unique and influential features in our study area, I believe it is essential to include it in a model. As the capital city, not only encompasses a substantial portion of the population, but also serves as an economic, administrative and cultural center. These distinct characteristics can have an impact on the variables we are investigating. By including Carlisle in our analysis, I aim to capture the nuanced effects and contributions of this city on the dependent variable. This decision is driven by recognizing that excluding such an area in the base model may result in an incomplete representation of the overall dynamics within the county. Including Carlisle ensures a comprehensive analysis that encompasses the broader dynamics of the entire county. Thus, we run a third model in which the city of Carlisle will be included and we acknowledge that the flooding events in 2005 and 2015, excluding the 2009 flood, will be the primary focus in this particular model. The equation of the third model is stated as followed:

$$ln(P_{it}) = cons + \boldsymbol{\beta_1 Floodarea2005(capitalincl)} + \boldsymbol{\beta_2 Floodarea2015(capitalincl)} + \beta_3 EstateType + \beta_4 NewBuild + \beta_5 PropertyType + \beta_6 District + \beta_7 PostCode + \beta_8 SalesYear + \beta_9 SalesMonth + \varepsilon_{i,t} \qquad (3)$$

where all variables remain the same as in the previous model. The difference in this particular case is the addition of the key independent variable of 2005 and 2015 floods, where in this case are describe as followed: *β_1Floodarea2005(capitalincl)* which is a dummy variable that equals 1 if property sold is located within the area affected by the flood of 2005, including the area of the capital city of the county, and *β_3Floodarea2015(capitalincl)* that is also a dummy variable which equals 1 if property sold is located within the area affected by the flood of 2015 and includes the city of Carlisle.

Table 8: Overview of Regression Models

Variable	Model 0	Model 1	Model 2	Model 3
Flood area	X	X	X	X
Property Characteristics		X	X	X
Time Characteristics			X	X
Location effects				X

Note: The table shows which variables have been used in the different regression models. The X indicates whether the variables have been included.

The table 8 above shows a summary of the different regression models when performing the Hedonic Pricing model. I start with the regression of my key variables of interest. In the following models I keep adding the control variables of property characteristics, which are the variables for new build, property type, and estate type, time characteristics, which differentiate between sales month and sales year, and location effects, which control for district and postcode. By adding more control variables to each regression model, I can compare the results and use the best fitting model, which can be identified through the highest R-squared.

Sensitivity Analysis

There are a number of assumptions that must be satisfied in order to use the ordinary least squares (OLS) regression model including (1) the sum of errors has a mean of zero, (2) the variance of errors is constant, (3) error observations are independent of each other, (4) no correlation exists between the error terms and independent variables and (5) the error term is normally distributed (Brooks & Tsolacos, 2010). While running a simple regression model with the dependent variable and independent variables of interest tests, the aim is to ensure that the OLS model’s conditions are satisfied and the robustness of the primary model is verified. It is observed, through conducting several checks including a Breusch Pagan, Shapiro-Wilkes, Linktest, Regcheck (Stata command) and others, that there are concerns with heteroskedasticity, where the error variance is not constant, as well as issues with the non-normality of the residuals. In order to tackle the issue of heteroscedasticity, I used the natural logarithm of the dependent variable of transaction prices, as it can be seen in graph 9 in the Appendix. This transformation is clearly depicted in the respective histograms, contrasting the distribution before and after the logarithmic transformation, where a normal distribution is evident. Additionally, in order to contribute further to reliability of the regression results, I used the technique of robust standard errors to further address heteroscedasticity and ensure a constant variance of errors. Furthermore, I conducted additional checks to verify the appropriate functional form in my model, and check the non-normality of the residuals (Appendix Graph 10)

When constructing regression models, it is also important to test for multicollinearity, in which the independent variables are found to be correlated. This can lead to skewed or misleading results and therefore the regression results would not be accurate. So, to examine the robustness of my dataset, I start by checking the data for multicollinearity using the correlation matrix table 12 (Appendix) which is showing the correlation between the independent variables. Based on the correlation values, there doesn't seem to be a strong linear relationship among the independent variables chosen. The absence of high correlation suggests that collinearity is unlikely to unduly bias my results. Moreover, I check collinearity using the VIF statistics where it computes a value for each variable starting at one and increasing indefinitely. To interpret this value the rule states that a VIF value of 1 indicates that two variables are not correlated, a value between 1 and 5 indicates a moderate correlation and a value above 5 indicates a high correlation (Aiken et al., 1991). The VIF results, which can be found in Table 13 demonstrate that the categorical variable "property type" and the dummy variable of "estate type" have correlations, with VIF values between 2 and 3. To address multicollinearity we could potentially drop variables that are highly correlated with others. However, after considering both the correlation matrix and the VIF statistics together, I conclude that multicollinearity does not significantly affect the validity of our results. Additionally, we observe that none of the other independent variables exhibit levels of correlation, thus, no multicollinearity is detected. Therefore, I have decided to retain all variables in the regression model.

The reported findings mentioned, specifically focus on the primary model. As we stated before, we were expecting multicollinearity in our primary model since some of the cities investigated are the same. Surprisingly the VIF statistics provided in the table above indicate that there is no correlation between our variables. However, to ensure the reliable and robust results and eliminate any potential influence of multicollinearity on our findings, a strategic decision was made to exclude those cities from our analysis. Subsequently we created the base model following the previously outlined methodology, and the VIF statistics of this model are provided in Appendix Table 14 for complete transparency and validation of our model refinements. The outcomes derived from this statistical analysis also confirm the absence of multicollinearity.

Table 13: Variance Inflation Factor (Primary model)

	VIF	1/VIF
Flood area 2005	1.023	.977
Flood area 2009	1.007	.993
Flood area 2015	1.014	.987
Estate Type	2.433	.411
Newbuild	1.04	.961
Property Type (Detached)		
Flat/Maisonettes	2.741	.365
Other	1.135	.881
Semi-Detached	1.587	.63
Terraced	1.705	.586
Sales Year	1.035	.967
Sales Month	1	1
Postcode	1.432	.698
District	1.436	.696
Mean VIF	1.43	.

Note: This table denotes the variance inflation factor per independent variable. The variable Detached is used as reference variable and therefore is not included. A VIF value of 1 indicates that two variables are not correlated, a value between 1 and 5 indicates a moderate correlation and a value above 5 indicates a high correlation.

Chow test - Heterogeneity

In order to test for the third hypothesis of the analysis of whether *"the discount of a residential property, as a result of a flood, varies with heterogeneity in property types and prices segments across property"* I test the data for heterogeneity by performing a Chow-Test. The Chow test is a statistical test that is used to determine whether there is a structural break in the coefficients of a regression model, meaning whether the relationship between the dependent variable and the independent variables changes significantly at some point. By including a Chow-Test, it involves dividing the data into two or more subsamples and comparing the regression coefficients for each subsample. Comparing the data when regressing together vs separately (Nielsen and Whitby, 2015). In this case, I check for a structural break, analyzing first the property types and after the price segments. Therefore, in the first Chow test conducted, I test for a break in property types to determine whether the coefficients stay constant or differ across property types. Similarly, in the second Chow test conducted, I apply the same approach to asses price segments. The null hypothesis of the Chow tests is that there is no significant difference between the coefficients of two subgroups in a regression model, which means that there is no structural break *"coefficients stay constant across the sample"*. Whereas the alternative hypothesis refers to that there is a structural break in the relationship between the independent and dependent variables across the groups *"the coefficients differ across the sample"*. Thus, in practice, for the first Chow test, I separate property types by generating a new variable and redefining each type, detached, semi-detached, flat/maisonettes, terraced, and other. Subsequently, in the second Chow test, I distinguish the price segments by creating a new variable and redefining each segment, low-end, middle-end and high-end. After running a pooled regression for both tests, as well as for each property type and each price segment, I can use the F-value formula to determine the F-value for the Chow-tests. The Chow test statistic is given by the formula:

$$Chow = ((SSR1 + SSR2) - SSR) / (k * SSR / n)$$

where SSR1 represents the sum of squared residuals from the first subset of the data, SSR2 is the sum of squared residuals from the second subset of the data, SSR is the sum of squared residuals from the regression of all the data, k denotes the number of independent variables in the regression, and n signifies the sample size.

4. RESULTS AND DISCUSSION

4.1 Results

In this chapter, the regression results of the hedonic pricing models and the Chow-Tests performed are presented. This thesis aims to explore the relationship between flood risk and property prices in flood-affected areas, investigating whether they sell at a discount or premium compared to similar properties un-affected by these floods. Additionally, the analysis examines whether heterogeneity exists concerning the discount or premium adjusted to residential property prices. To assess these associations, the analysis focuses on specific flood events in January 2005, November 2019 and December 2015, three record-breaking events that occurred in the area of Cumbria during the period from 2003 to 2022. First, the results for testing the first and second hypotheses H1, H2 are discussed, followed by the presentation of Chow-test results, examining heterogeneity to test the third hypothesis H3. To conclude whether H1 as well as H2 can be justified, it is essential to look at key variables of Flood area 2005, Flood area 2009 and Flood area 2015. To test the significance of heterogeneity and address H3, it is important to analyze the independent variable of property type and the created price segments.

To avoid biasing my results, I ran three different models. The first one is the primary model, including all the variables to be explored. The base model consists of the same variables, with the only difference being the exclusion of the capital city of Cumbria, Carlisle. This exclusion is because two out of the three floods of interest occurred there. The third model includes the capital city again, but this time, we exclude the variables related to Flood area 2009, aiming for more specific view of the flooded areas of Carlisle and its surroundings. In building up my models, I conducted four separate regressions for each model, initially focusing on flood related variables only. In the second regression and beyond, I incorporated control variables, encompassing property characteristics such as Estate Type, Newbuild and Property type. In the third regression, time characteristics such as year and month, were integrated, and in the fourth regression, location characteristics like postcode and district were added. While running these different regressions, it is evident, based on the R-squared values, that the regression results successively improve.

In the table below, the regression results are presented, using the natural logarithm of housing transaction prices as the dependent variable due to the skewed distribution of transactions prices. Moreover, to account for heteroscedasticity, where the variance of the errors is not constant, robust standard errors were applied to the models during regression runs. The regression results provided by the tables of 15, 16 and 17 below provide an overview of the coefficients for the regression equation, illustrating the effect of independent variables on explaining the variance of the dependent variable, the transaction prices. Hence, the coefficients describe the relationship between property prices and flood risk, after controlling for other effects of property characteristics, time and location effects. To interpret the resulting regression coefficients, they need to be translated back to the level of transactions prices. Following the interpretation of dummy variables as outlined by (Brooks and Tsolacos, 2010), the impact of the coefficients on transaction prices can be calculated using the following equation:

$$\%\Delta \text{ in transaction price} = (e^{\text{coefficient}} - 1) * 100$$

Table 15: Regression Results of the Primary Model (Dependent Variable: Transaction prices)

	Model (1) Variables of Interest	Model (2) Property Characteristics	Model (3) Time Effects	Model (4) Location Effects
Variables	*Coefficients S.E.*	*Coefficients S.E.*	*Coefficients S.E.*	*Coefficients S.E.*
Flood area 2005	-0.137***	0.0384***	0.0354***	0.0452***
	(0.0109)	(0.00995)	(0.00979)	(0.00894)
Flood area 2009	0.226***	0.367***	0.365***	-0.0545**
	(0.0202)	(0.0197)	(0.0192)	(0.0181)
Flood area 2015	0.0685***	0.208***	0.208***	0.0246**
	(0.00902)	(0.00859)	(0.00841)	(0.00763)
Property Characteristics				
Estatetype (1=Freehold)		0.128***	0.116***	0.158***
		(0.00715)	(0.00685)	(0.00671)
Newbuild (1= Yes)		0.0840***	0.0924***	0.154***
		(0.00406)	(0.00387)	(0.00361)
Property Type (*Base: Detached*)				
Flat/Maisonette		-0.680***	-0.688***	-0.760***
		(0.00854)	(0.00822)	(0.00767)
Other		-0.268***	-0.394***	-0.431***
		(0.0197)	(0.0196)	(0.0193)
Semi-detached		-0.554***	-0.553***	-0.467***
		(0.00311)	(0.00291)	(0.00256)
Terraced		-0.903***	-0.894***	-0.759***
		(0.00322)	(0.00303)	(0.00263)
Time Effects				
Sales Year			Yes	Yes
Sales Month			Yes	Yes
Location Effects				
Postcode			No	Yes
District			No	Yes
Constant	11.86***	12.29***	11.74***	11.65***
	(0.00162)	(0.00739)	(0.00996)	(0.0186)
N	184048	184048	184048	184048
R^2	0.002	0.282	0.354	0.527
Adj. R^2	0.002	0.282	0.354	0.527

** $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

*Note: The table shows the regression results of the Base Model. The dependent variable is the natural logarithm of transaction prices. Parentheses include standard errors, *** indicates that the variable is statistically significant at the 1% significance level,** at the 5% significance level and * at the 10% significance level.*

At first glance, we can conclude from the table that the chosen independent variables for examination exhibit statistical significance and seem to have an impact on the dependent variable examined. Based on the findings, every Flood variable is statistically significant in all four models. In the first model, where only the variables of interest are included, it can be seen that each key independent variable is statistically significant. For example, the variable of Flood2005 has a coefficient estimated at -0.137 and is statistically significant at the 1% level of significance. This means that, holding all other independent variables constant (ceteris paribus), a property located in an area affected by the Flood of 2005 is transacted with a discount of -12.8%. Admittedly, we can see from the results of model (1) that the coefficients of the regression, using the hedonic pricing model, are statistically significant at 1% level of significance. This implies an impact on the dependent variable, and there is a significant correlation when regressing only with the key independent variables. Yet, it is observed that the R-squared is extremely low, at around 2%, indicating that it explains only a small portion of the variation in the dependent variable. This suggests that approximately 2% of the variation in the dependent variable is explained by the independent variables in the regression model. Further, it indicates that the model doesn't provide a "good" fit for the data, and therefore, it is believed that the results are inaccurate and not in line with the expectations of this research. The interpretation of these results, therefore, raises suspicion regarding the effect of the independent variable of interest on our model.

In order to reach better and more reliable results, the regression method of the hedonic pricing model is then tested by adding control variables. While adding property characteristics, it can be seen that there is a change in the variables related to flooded areas, particularly in the Flood area 2005 variable. In the first model, it had a negative sign, and now it has a positive sign in the second model. Every sign of these variables in the second model continue to be positive and statistically significant at the 1% level of significance. As for the variables related to housing characteristics, it is observed that all of them are statistically significant and seem to affect the dependent variable under examination. In the third model, where time characteristics are added with variables such as year and months, there are no significant changes other than small adjustments in the values of the coefficients. In the regression of the fourth model, after including all independent variables of interest, it is observed that every independent variable is statistically significant. Column 4 reports the results from the most complete and preferred specification, since it includes all relevant variables and also controls for time and location effects. The striking feature is that, compared to the first model where the variable Flood area 2005 has a negative sign, in the fourth model that we run, the sign becomes positive. Furthermore, it is noted that the variable Flood area 2009, while having a positive sing in each model, becomes negative in the last model of our analysis. This means that if a property is located in an area affected by the 2009 Flood, holding all other independent variables constant (ceteris paribus), at the 10% significance level, it will result in a discount of 5,30% of the transaction price. In contrast, the variable Flood area 2015 has a positive sign observed in all four models. Therefore, if a house is located in the flood-affected area of

2015, holding all other independent variables constant, it is concluded that it would sell at an increased price of 2.49% at the 10% significance level, resulting into a premium. Additionally, the coefficients for property types, including flat/maisonettes, other, semi-detached and terraced, are all negative in comparison to the reference category, which is the type of detached. This emphasizes the lower prices associated with the property types, supporting the need to consider diverse property characteristics in understanding housing prices.

In addition, what is observed is that adding every control variable to the model increases the R-squared, indicating an improvement in the regression results. The R-squared value of 0.527 in the fourth model explains a large portion of the variation in the dependent variable, in comparison to the 0.002 value of the first model. The value of 0.527, concludes that approximately 52.7% of the variation in the dependent variable is explained by the independent variables in the regression model. This suggests that the model provides a "good" fit for the data, and a substantial proportion of the variability in the dependent variable can be accounted for by the independent variables. As a result, based on the base model, I do find evidence supporting a significant relationship between flood risk and housing prices, as all of our key independent variables are statistically significant. Yet, based on the given signs, a conclusion can't be derived for our second hypothesis that "*Residential properties located within areas affected by recent floods of January 2005, November 2009, and December 2015 in the area of Cumbria sell at a discount compared to similar properties un-affected by these floods",* since Flood Area 2005 in model (1) and Flood area 2009 in model (4) are associated with decreased prices, while Flood Area 2015 unexpectedly leads to increased prices in each model, as well as Flood Area 2005 in the fourth model. This could be caused by multicollinearity, which might be present, and the results are likely to be not reliable. That's why we included the following two models: in the first one, we exclude the capital city of Carlisle, and in the second one, we run the regression including it but focus only on that area.

Table 16: Regression Results of the Base Model (Dependent Variable: Transaction prices - Capital Excluded)

	Model (1) Variables of Interest	**Model (2) Property Characteristics**	**Model (3) Time Effects**	**Model (4) Location Effects**
Variables	*Coefficients S.E.*	*Coefficients S.E.*	*Coefficients S.E.*	*Coefficients S.E.*
Flood area 2005excl.	0.207^{***}	0.311^{***}	0.308^{***}	-0.0255
	(0.0256)	(0.0232)	(0.0229)	(0.0200)
Flood area 2009	0.203^{***}	0.338^{***}	0.337^{***}	-0.0539^{**}
	(0.0203)	(0.0196)	(0.0191)	(0.0181)
Flood area 2015excl.	0.190^{***}	0.300^{***}	0.304^{***}	-0.00221
	(0.0116)	(0.0112)	(0.0109)	(0.0102)
Property Characteristics				
Estatetype (1=Freehold)		0.123^{***}	0.117^{***}	0.181^{***}
		(0.00831)	(0.00797)	(0.00802)
Newbuild (1= Yes)		0.0921^{***}	0.0967^{***}	0.146^{***}
		(0.00522)	(0.00497)	(0.00455)
Property Type (*Base: Detached*)				
Flat/Maisonette		-0.651^{***}	-0.658^{***}	-0.741^{***}
		(0.0100)	(0.00967)	(0.00912)
Other		-0.313^{***}	-0.451^{***}	-0.472^{***}
		(0.0215)	(0.0214)	(0.0208)
Semi-detached		-0.538^{***}	-0.538^{***}	-0.434^{***}
		(0.00370)	(0.00346)	(0.00299)
Terraced		-0.932^{***}	-0.922^{***}	-0.735^{***}
		(0.00381)	(0.00357)	(0.00309)
Time Effects				
Sales Year			Yes	Yes
Sales Month			Yes	Yes
Location Effects				
Postcode			No	Yes
District			No	Yes
Constant	11.88^{***}	12.33^{***}	11.76^{***}	11.79^{***}
	(0.00191)	(0.00861)	(0.0118)	(0.0413)
N	140447	140447	140447	140447
R^2	0.003	0.281	0.355	0.558
Adj. R^2	0.003	0.281	0.355	0.557

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

*Note: The table shows the regression results of the Base Model. The dependent variable is the natural logarithm of transaction prices. Parentheses include standard errors, *** indicates that the variable is statistically significant at the 1% significance level,** at the 5% significance level and * at the 10% significance level.*

After regressing model (2), in which we exclude areas that have been affected by more than one flood event to address potential multicollinearity issues, it becomes evident that signs, coefficients values, and significance changes drastically. Each key independent variable appears to be statistically significant in the first three models that we run, at 1% level of significant, indicating a positive sign. However, in model (4), incorporating all variables that best fit the data and boasts the highest R-squared with a value of 0.528, meaning that the independent variables account for 55,8% of the variance of the dependent variable. Notably, the coefficients of the regression in model (4) using the hedonic pricing model have values that may be in line with reasonable expectations and the literature. The values and the signs of the coefficients are considered reasonable, indicating that the results are accurate and in line with the expectations of this research. The interpretation of the results suggests that if a property is located in a flood-affected area during the event in November 2009, holding all independent variables constant (ceteris paribus), at 10% level of significance will result in a decrease in property prices by -5,24%. The same trend is observed for the variables Flood area 2005excl. and Flood area 2015excl., both showing a negative coefficient, depicting that being located in areas affected by those events would lead to a fall in property prices. Yet, both of these variables are statistically insignificant in this regression, suggesting that, based on the observed data, there is insufficient evidence that these variables have a significant impact on the dependent variable. Besides our key independent variables, it is also seen that the property characteristics play a crucial role in influencing transactions prices. There is a diverse impact where the dummy variables of estate type and newbuild have positive sings, while the categorical variable of property type accounts for negative values. So, if the estate type is a freehold, at 1% level of significance and holding all other independent variables constant, it will cause an increase of 19,84% in the transaction prices. Similarly, if the property is a newbuild, it will result in a price increase of about 15,71%. Regarding the categorical variable of property type, where each type is compared to our base category which is detached, they imply varying impacts on transactions prices. For instance, the type of flat/maisonettes has a coefficient of -0,741, signifying a statistically significant negative association, indicating lower transaction prices compared to detached properties, holding all other independent variables constant at 1% level of significance by 52,33% . The same applies to the property type of terraced ,which has a significant negative correlation compared to the base category, implying a decrease in transaction prices by 35,20%, holding all other independent variables constant and at 1% significance level.

In general, on the one hand, it is observed that most of the variables seem to affect the dependent variable of the analysis in a statistically significant way. Variables such as property characteristics and the 2009 Flood area variable appear to have an important influence on our dependent variable, with the desired signs and in line with our expectations, as well as in accordance with general consensus and consistent with the literature. However, on the other hand there are variables such as the dummy variables of Flood area 2005excl. and Flood area 2015excl. Although they have the desired signs and coefficient values, those variables are not statistically significant and seem to not influence the transactions prices, the dependent variables of the analysis. Consequently, what can be derived from the regression results is that there is a partial support to our second hypothesis that *"residential properties located within areas affected by floods sell at a discount compared to similar properties un-affected by these floods".* Partially, and not a fully support, since even though all key independent variables of flooded areas have the desired sign, only the Flood area 2009 is statistically significant. Thus, this difference highlights the nature of how floods impact areas, emphasizing the need for a detailed examination of the timing and location aspects of flood events.

Table 17: Regression Results of the Third Model (Dependent Variable: Transaction prices - Capital included)

	Model (1) Variables of Interest	**Model (2) Property Characteristics**	**Model (3) Time Effects**	**Model (4) Location Effects**
Variables	*Coefficients S.E.*	*Coefficients S.E.*	*Coefficients S.E.*	*Coefficients S.E.*
Flood area 2005incl.	-0.139*** (0.0109)	0.0342*** (0.00995)	0.0312** (0.00979)	0.0461*** (0.00893)
Flood area 2015incl.	0.0666*** (0.00902)	0.204*** (0.00859)	0.204*** (0.00841)	0.0259*** (0.00762)
Property Characteristics				
Estatetype (1=Freehold)		0.121*** (0.00717)	0.109*** (0.00688)	0.158*** (0.00670)
Newbuild (1= Yes)		0.0830*** (0.00407)	0.0914*** (0.00388)	0.154*** (0.00361)
Property Type (*Base: Detached*)				
Flat/Maisonette		-0.679*** (0.00858)	-0.687*** (0.00826)	-0.760*** (0.00767)
Other		-0.264*** (0.0197)	-0.390*** (0.0196)	-0.431*** (0.0193)
Semi-detached		-0.554*** (0.00311)	-0.553*** (0.00292)	-0.467*** (0.00256)
Terraced		-0.901*** (0.00323)	-0.892*** (0.00303)	-0.760*** (0.00263)
Time Effects				
Sales Year			Yes	Yes
Sales Month			Yes	Yes
Location Effects				
Postcode			No	Yes
District			No	Yes
Constant	11.86*** (0.00162)	12.29*** (0.00740)	11.75*** (0.00998)	11.65*** (0.0185)
N	184048	184048	184048	184048
R^2	0.001	0.279	0.352	0.527
Adj. R^2	0.001	0.279	0.352	0.527

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

*Note: The table shows the regression results of the Base Model. The dependent variable is the natural logarithm of transaction prices. Parentheses include standard errors, *** indicates that the variable is statistically significant at the 1% significance level,** at the 5% significance level and * at the 10% significance level.*

In the presented table, following the regression for model (3), where we include the city of Carlisle, which is the county's capital city, recognizing its significance with the unique and essential influential features that it might have. The reason for including the capital city is to capture even the nuanced effects and contributions this city might have on our dependent variable. Yet, the primary focus in this case are the dummy variables of Flood area 2005 and Flood area 2015, excluding the 2009 Flood area, as these areas correspond only to these two particular flood events. From the results, it can be seen that both of the independent variables of interest are statistically significant in all four of the different models ran. It is observed that in the first model where only the variables of interest are included, the variable Flood area 2005incl. has a negative sign, still, when all the variables that we consider to affect our dependent variable such as housing characteristics, as well as time and location characteristics, are included, it is noted that the sign changes and becomes positive. Thus, an interpretation of Flood area 2005incl., which has a coefficient of 0.0461 and is statistically significant at 1% level of significance, would be that if a residential property is located in the flood-affected area of 2005, holding all other independent variables constant, it would be transacted at an increased price of 4,71%, resulting into a premium. The same applies to the variable of Flood area 2015incl. with a coefficient of 0.025 and statistical significance at 1% level, which will lead to an increase in the transaction price of 2.62% if the property transacted is located and affected by the Flood of 2015. As for the additional independent variables included in the 4^{th} model, a relatively similar behaviour is observed as in the two previous regressions, being statistically significant and impacting the dependent variable. In addition, Model (4) includes all variables which fit the data best and has the highest R-squared, which means that the independent variables account for roughly 52,7% of the variance of the dependent variable, the property transactions prices. Therefore, we can conclude that based on the regression of the model (3), there is no support of hypothesis (2) that the properties affected from a flood sell at discount. On the other hand, it is observed that in this case, the properties sell at a premium instead of a discount.

Chow test – Heterogeneity for Property Types

In order to test for heterogeneity between subsamples a Chow test is performed between the five subsamples of the property types, detached, semi-detached, flat/maisonettes, terraced and other. The Chow test determines whether the parameters of the subsamples differ from each other in such a way that they are better estimated in five separate regressions instead of a pooled regression (Chow, 1960). The Chow test can be formulated as:

$$F = \frac{\left(RSSp - (RSS1 + RSS2 + RSS3 + RSS4 + RSS5)\right)}{(RSS1 + RSS2 + RSS3 + RSS4 + RSS5)} \times \frac{(n - 5k)}{(5k - k)}$$

Where RSSp is the sum of residuals for the pooled model. RSS1, RSS2, RSS3, RSS4 and RSS 5 are the sum of residuals of the individual models of the five subsamples, respectively, detached, semi-detached, flat/maisonettes, terraced and other, k indicates the number of parameters and n the total number of observations. Since the base model is the one which this analysis focuses on the most, the chow-test is performed based on that model, presenting the following equation:

$$F = \frac{\left(40496.01 - (4748.40 + 5101.97 + 2203.19 + 8500.47 + 7892.85)\right)}{(4748.40 + 5101.97 + 2203.19 + 8500.47 + 7892.85)} \times \frac{(140447 - 5 \times 117)}{(5 \times 117 - 117)}$$

$$F = 125.48$$

The critical value for F, given 140,447 and 117 degrees of freedom, is 1. 32 at the 1%-level. The F-value obtained from the Chow test is larger with 125.48. This means that H0 can be rejected at the 1%-level and the estimated parameters between properties with the property types are not equal. So, referring to the results of the Chow-Test, the null hypothesis is rejected, meaning that there is significant evidence for heterogeneity across property types. The table 18 presents the results conducted for each property type individually, revealing insights into the heterogeneity among property types. The dependent variable is the natural logarithm of transaction prices, and the models include time effects like sales month and sales year as well as location effects controlled for postcode and district. The coefficients for Flood area 2005excl., Flood area 2009, Flood area 2015excl., Estate Type, and Newbuild are reported for each property type category detached, semi-detached, flat/maisonettes, terraced, Other as well as for the pooled regression. The findings show that there are differences in how flood areas, estate type and whether they are newly built or not, affect property values. Specifically, the coefficients for Flood area 2005excl., Flood area 2009 and Flood area 2015excl. have effects of varying degrees indicating the influence of these flood events on property prices. Interestingly the association between flood risk and property prices varies across types of buildings with coefficients ranging from positive to negative and mostly being statistically significant. Findings from the regression results indicate that when the property type is a flat/maisonette, it commands a modest yet notably significant premium in comparison to the other four property types the detached, semi-detached, terraced and other, which each exhibits a statistically significant negative coefficient, except the property type of semi-detached which is statistically insignificant. This implies that these alternative property types are sold at a discount rate. Similarly, the variables of estate type and whether a property is a newbuild or not also have distinct effects on property prices across different categories. This is in line with my previous findings, indicating a negative correlation between floods and property prices in the flood affected areas. Moreover, since more of the property types show significant coefficients for the three different Flooded area variables, I have to assume that there is significant difference on the impact of flood risk on property price across property types. Additionally, the rejection of the null hypothesis in the Chow- test highlights that there is heterogeneity among property types. Therefore, a conclusion can be derived, based on the results, that there is support on our third hypothesis H3 that *"the discount of a residential property, as a result of a flood, varies with heterogeneity in property types and prices segments across property".* Overall this table provides insights that contribute to a comprehensive understanding of how various factors impact property values while acknowledging the diverse nature of the real estate market.

Table 18: Heterogeneity for Property Type (Base Model)

Variables	Pooled	Detached	Semi-Detached	Flat/Maisonettes	Terraced	Other
Flood area 2005excl.	-0.133***	-0.113**	0.0738	0.00756	-0.111***	-0.0741
	(0.000)	(0.007)	(0.103)	(0.841)	(0.000)	(0.696)
Flood area 2009	-0.157***	-0.0103	-0.0309	0.117***	-0.127***	-0.635***
	(0.000)	(0.722)	(0.283)	(0.000)	(0.000)	(0.000)
Flood area 2015excl.	-0.0900***	-0.0348	-0.0285	0.0880***	-0.0405***	-0.280**
	(0.000)	(0.063)	(0.079)	(0.000)	(0.000)	(0.008)
Estate Type (1=Freeh.)	0.462***	0.311***	0.207***	0.125***	0.0900***	0.695***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Newbuild (1=Yes)	0.290***	0.0535***	0.116***	0.301***	0.169***	-0.135
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.408)
Time Effects	Yes	Yes	Yes	Yes	Yes	Yes
Location Effects	Yes	Yes	Yes	Yes	Yes	Yes
_cons	11.13***	11.72***	11.44***	11.27***	11.16***	4.634**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
N	140447	33509	37410	11677	53392	4459
R^2	0.415	0.397	0.450	0.480	0.574	0.130
Adj. R^2	0.414	0.395	0.449	0.476	0.573	0.111
RSS	40496.01	4748.40	5101.97	2203.19	8500.47	7892.85

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: This table presents the regression results for each property type individually. The dependent variable is the ln (transaction price). Time effects are sales month and sales year, and location effects control for postcode and district. Parentheses include the p-values.

Chow test – Heterogeneity for Price segments

Initially, in order to proceed with the implementation of the Chow-test, each price segment must be determined and characterized. Given the wide range of property prices in the dataset, with a substantial number of lower priced properties and a comparatively smaller number of higher-priced properties, I considered an approach that explicitly takes into account the distribution of prices. This approach employs specific price thresholds, based on the distinctive market dynamics. As a guideline is the average property price of the county of Cumbria taken, ranging from approximately 212,000 pounds to 222,000 pounds (Howard, 2023; Data.gov.uk, 2022; Plumplot, 2023). Therefore, considering the average selling price of a house and common sense, it was deemed reasonable to split and segment the sample. It is assumed that the prices for the properties defined as low-end range up to 150,000 pounds, the middle-end is determined from 150,000 to 300,000 pounds, while the high-end properties are characterized as those selling above 350,000 pounds.

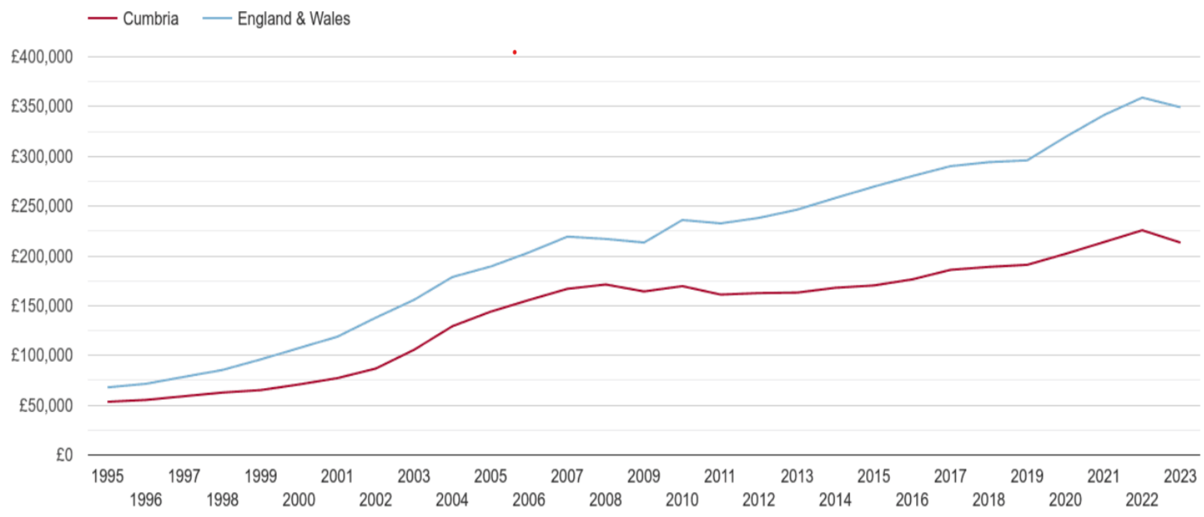


Figure 19:The graph shows the average houses prices of Cumbria, yearly average nominal prices compared to England (Plumplot, 2023)

Following a similar approach as described in the example above and to test for heterogeneity between subsamples, a second Chow test is performed between the three subsamples of the price segments: low-end, middle-end and high-end housing. The Chow test can be formulated as:

$$F = \frac{RSSp - (RSS1 + RSS2 + RSS3)}{(RSS1 + RSS2 + RSS3)} \times \frac{(n - 3k)}{(3k - k)}$$

Where RSSp is the sum of residuals for the pooled model. RSS1, RSS2, RSS3 are the sum of residuals of the individual models of the three subsamples, respectively, low-end, middle-end and high-end, k indicates the number of parameters and n the total number of observations. Similar to the earlier case, in this particular instance, the base model is used as it is the one which this analysis focuses on the most. The chow-test is performed based on that model

$$F = \frac{\left(30618.93 - (11021.16 + 1575.64 + 1710.41)\right)}{(11021.16 + 1575.64 + 1710.41)} \times \frac{(140{,}447 - 3 \times 121)}{(3 \times 121 - 121)} = 659.960$$

The critical value for F, given 140,447 and 121 degrees of freedom, is 1.32 at the 1%-level. The F-value obtained from the Chow test is larger with 659.960, which means that H0 can be rejected at the 1%-level, and the estimated parameters between properties with different price segments are not equal. Referring to the results of the Chow-Test, there is significant evidence for heterogeneity across price segments. The results obtained for each price segment separately are shown in Table 20, providing details regarding the heterogeneity among price ranges. These models also take into account location effects that are adjusted for postcode and district, as well as time effects like sales month and sales year. For each price category, low-end, middle-end, and high-end, as well as for the pooled regression, the coefficients for Flood area 2005excl., Flood area 2009, Flood area 2015excl., Estate Type, Newbuild, and Property type are presented. The findings show differences in how flood areas, property types, estate type and whether they are newly built or not, affect property values. In particular, the effects of the Flood area 2005excl., Flood area 2009, and Flood area 2015excl. differ, suggesting a correlation between these flood occurrences and real estate values. It is noteworthy that the influence of flood risk on property prices varies across different property price segments, with coefficients ranging from positive to negative and partially being statistically significant. Regression results indicate that the low-end and high-end price segments command a modest yet notably significant discount compared to the middle-end segment, which exhibits insignificantly positive coefficients with Flood area 2009 being the only statistically significant variable among the key independent variables. Similarly, the variables of estate type, property types and whether a property is a newbuild or not, have also distinct effects on property prices across different categories. This aligns with my previous findings, which resulted into a negative impact on property prices in the flood affected areas. Moreover, since balanced significance results for the price segments can be observed for the three different Flooded area variables, it can be assumed that there is a partially significant difference in the correlation of flood risk and property values across price segments. Additionally, the rejection of the null hypothesis in the Chow test highlights heterogeneity among different price ranges. Therefore, based on the results, a conclusion can be deducted that there is support for our third hypothesis H3 that *"the discount of a residential property, as a result of a flood, varies with heterogeneity in property types and prices segments across property*", regarding price segments.

Table 20: Heterogeneity for Price Segments (Base Model)

Variables	Pooled	Low-end properties	Middle-end properties	High-end properties
Flood area 2005excl.	-0.0255	0.0207	0.0122	-0.0601*
	(0.189)	(0.420)	(0.287)	(0.046)
Flood area 2009	-0.0539***	-0.0973***	0.0176*	-0.0684***
	(0.000)	(0.000)	(0.029)	(0.000)
Flood area 2015excl.	-0.00221	-0.0245*	0.00339	-0.0231
	(0.795)	(0.034)	(0.491)	(0.081)
Estate Type (1=Freeh.)	0.181***	0.0620***	0.0264***	0.00792
	(0.000)	(0.000)	(0.000)	(0.630)
Newbuild (1=Yes)	0.146***	0.152***	0.0440***	-0.0634***
	(0.000)	(0.000)	(0.000)	(0.000)
Property Type (Base: Detached)				
Flat/Maisonettes	-0.741***	-0.469***	0.168*	-0.190***
	(0.000)	(0.000)	(0.029)	(0.000)
Other	-0.472***	-1.127***	0.0589***	-0.353***
	(0.000)	(0.000)	(0.000)	(0.000)
Semi-Detached	-0.434***	-0.168***	0.138***	-0.130***
	(0.000)	(0.000)	(0.000)	(0.000)
Terraced	-0.735***	-0.408***	-0.159***	-0.164***
	(0.000)	(0.000)	(0.000)	(0.000)
Time Effects	Yes	Yes	Yes	Yes
Location Effects	Yes	Yes	Yes	Yes
_cons	11.79***	11.28***	12.21***	12.83***
	(0.000)	(0.000)	(0.000)	(0.000)
N	140447	71330	50441	18676
R^2	0.558	0.330	0.182	0.248
Adj. R^2	0.557	0.329	0.180	0.244
RSS	30618.93	11021.16	1575.64	1710.41

.

*$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: This table presents the regression results for each price segment individually. The dependent variable is the ln (transaction price). Time effects are sales month and sales year, and location effects control for postcode and district. Parentheses include the p-values.

4.2 Discussion

The existing literature examines natural hazards as geographical attributes that are in detail linked to specific areas. In areas prone to flooding, households facing high risks are likely to negotiate either a premium or a discount while purchasing a residential property. In a perfect market the trade-off between expected loss from a natural hazard and the premium for locating outside increased risk areas would be equal. However, the literature find that this is not the case and household either underestimate or overestimate the risk depending on the number of occurrences and the number of years since the last natural hazard. Following our regression results, first, I find that the risk of flood events can, but does not always, have a negative effect on house prices, confirming that natural hazard risks lead to a certain discount on residential property value. Second, similarly to previous studies I find evidence against the economic theory of efficient housing markets (Belanger and Bourdeau-Brien, 2018; Donahue and Tuohy, 2006). Residential property in the area of Cumbria seem to be traded at a discount as can be seen from the results of the base model regression. However, the third model indicates that the properties might also sell at a premium after we include the city of Carlisle.

This study tried to ascertain whether properties located in an area that were flooded by the Floods of 2005, 2009 or 2015 are affected in financial terms. Considering the first hypothesis: *"a flood has a significant negative effect on housing prices in flood-affected areas"* seems to be accurate based on the regression results. This can be observed in the statistical significance of the key independent variables, where the Flood areas variables of 2005, 20009 and 2015 are mostly significant throughout the regressions meaning that there is a positive correlation and an influential effect of the independent variables on the dependent variables, the transactions prices. This positive interrelationship follows the trend of literature examining natural hazards, depicting that Floods have a significant impact and affect housing transaction prices (Atreya, Ferreira and Kriesel, 2013; Gharbia et al., 2016).

Regarding our second hypothesis, what can be derived from the regression results is that there is a partially support that *"Residential properties located within areas affected by recent floods of January 2005, November 2009, and December 2015 in the area of Cumbria sell at a discount compared to similar properties un-affected by these floods"*. Partially, and not a fully support since although all key independent variables of Flooded areas have the desired sign, however only the Flood area 2009 is statistically significant. This can be derived from the conclusions considering our base model. There is a discount for residential properties sold in flood-affected areas of -5.24%. This is in line with reasonable expectations and the literature. However, the literature finds higher discounts in residential property prices for instance, (Beltrán et al., 2019) argue that there is a discount in property values following flooding, ranging from 10.1% to 31.4%. Compared to the dataset used on this article, the discount for residential properties might be higher because they analyze a larger dataset with over 4.8million houses in the UK, compared to this analysis where 184,048 property transactions are analyzed. Closer to our results, (Bin & Landry, 2013) find that property values in flood plains significantly decline, falling from almost 5% to 8%. On the other hand, if we consider the third model as the main one, then we can reject our hypothesis that residential properties affected by floods sell at a discount, since based on the regression results of this model, they would sell at a premium and not at a discount. A result that might originate from a multicollinearity issue, based on our expectation and logical consensus. Yet it might be also true that a premium has been adjusted as the proposed literature by (Atreya & Czajkowski, 2016; Bin, et al., 2008; Kim, et al., 2017). This rise in property values leading to a premium after a flood event can be attributed to location attractiveness since homeowners are willing to pay this premium in order to enjoy living near natural features such as coastlines, lakes, rivers, green spaces and others. Therefore, the positive signs observed in the third model run, could be due to the locational attractiveness of the area targeted in which it is the county's capital. Hence the demand and amenities are increased, which may be the reason why people are willing to pay a premium despite the fact that there is a high risk of flooding. Furthermore, according to the literature, a premium or generally the non-influence of flooding on property values is also likely to be due to the fact that when the frequency of such events remains high as well as repeated, there is no effect on property values (Clayton et al. 2021; Speyrer & Ragas, 1991). However, to recapitulate we are convinced that there is a partial support to our second hypothesis based on the results of our main regression of the base model.

As for the third hypothesis of this analysis, whether *"the discount of a residential property, as a result of a flood, varies with heterogeneity in property types and prices segments across property"*, seems to be accurate for property types such as detached, semi-detached, terraced and other, except of the type of Flat/maisonettes which results into a premium and not a discount. This is in line with my previous findings in the base model, which resulted in a negative correlation of property prices in the flood affected areas. Moreover, since most of the property types show significant coefficients for the three different Flooded area variables, I have to assume that there is a significant difference in the relationship between flood risk and property prices across property types. Furthermore, the Chow test indicates that the estimates for houses based on their property type are not equal. Thus, the rejection of the null hypothesis in the Chow- test also highlights that there is heterogeneity among property types. As far as whether heterogeneity in price segments exists, it appears from the findings that it is partially true. The low-end and high-end price segments seem to have a modest yet notably significant discount compared to the middle-end segment, which results in a statistically insignificant premium. Therefore, a conclusion to be draw is that there is a partially significant difference in the correlation between flood risk and property values across price segments, since balanced statistical significance is seen in the coefficients of the key independent variables throughout price segments. Moreover, rejecting the null hypothesis of the Chow test further strengthens the evidence of heterogeneity among low-end, middle-end and high-end properties. The literature also suggests that each property has its uniqueness and the possibility of two properties being exactly the same in negligible. The combination of various features such as structural, neighborhood and location characteristics of each property contributes to their heterogeneity and researchers like (Dubin, 1988; Stamou et al. 2017) have demonstrated that these factors significantly impact and determine the prices of properties. Therefore, the results of this specific analysis align with the pattern observed in the existing literature and further supports that residential property prices vary according to the inherent heterogeneity of each property.

5. CONCLUSION

The main objective of this master thesis was to explore the relationship between flooding and residential real estate property prices, in the area of Cumbria, United Kingdom. Thus, the aim of this thesis is to contribute by expanding related research on another region and cities in England, closing the gap of empirical evidence on market behavior and reaction patterns in the case of flooding and generally natural disasters. As flooding and other natural disasters have become a greater threat around the world as a result of climate change, affecting the entire infrastructure of power supply, roads, railways, and real estate. It is consequently critical to comprehend the association among flood risk and the real estate market in order to provide market information transparency, modify optimal insurance prices, and execute effective prevention measures. Existing literature has produced contradictory results regarding the relationship between flood risk and property prices, with some articles claiming that properties sell at a discount in a flood zone while other state that a premium can be adjusted. The central research question is***:*** *How is the market value of residential real estate in Cumbria, England associated with the risk of flooding?*

To answer the central research question, the topic is broken down into three sub questions, starting with the first research question: *"What is the impact of flood risk on the prices of residential real estate properties based on literature?"*, I reviewed existing literature extensively in order to reach any possible deductions. Research have shown that the risk of natural hazards leads to a discount on transaction prices for an affected property in comparison to those that are likely to remain unaffected. Flood disasters influence how consumers assess associated risks due to the scale of their impact, which is then reflected in property valuation. As a result, consumers factor their perceived risk into the valuation of a certain property. However, price premiums may also develop as a result of the convenience of living near coasts, green spaces, or rivers. Furthermore, it is worth noting that there is a lack of research, specifically examining the area of Cumbria UK, and only limited research exists investigating in particular cities in the area rather than the wider region as a whole. Therefore, this research contributes to answering, with the help of regression analysis, the question of: *To what extent is the discount or premium applied to residential property in flood-affected areas of Cumbria and whether there is a significant correlation?*

To determine the impact of flood risk on property prices of the three particular flood events of 2005, 2009 and 2015, I used a hedonic pricing model to see the price change after a flood in a flood-affected area compared to the non-affected area. So, in order to examine the research question 2 as well as 3 and test the hypothesis made, I use this empirical methodology to explain the relationship between flooding and residential real estate property prices. Regarding the second research question: *What is the extent of the discount or premium applied to residential property in flood-affected areas of Cumbria?* Firstly, the findings of the show contrasting results for the flood affected area variables. While regressing for the primary model, the key independent variables yielded mixed results, with the Flood area 2009 variable having a negative sign, and the additional two Flood area variables being positive. We considered these results as an initial phase and a preliminary encounter for the forthcoming execution of the additional regressions of interest, given the possibilities and suspicion of multicollinearity. While regressing for our base model, in which we excluded the capital city of Carlisle to avoid potential spatial correlation between the dummy variables of flood, as multicollinearity might have emerged, the results appear to have the expected signs, as well as the values of the coefficients, which are in line with consensus and the corresponding literature. The results indicate that the discount is adjusted to 5.25% if the flooded area was affected by the 2009 Flood, while for the Flood area of 2005, it is 2.5%, and for the 2015 Flooded area, the lowest rate is 0.22%. Yet, statistical significance was only evident for the Flood area 2009 variable, and was not apparent for any of the additional key independent variables. In contrast, looking on the results of the third regression analysis, it can be derived that the key independent variables have been impacted positively, resulting into a premium and not a discount. Findings indicate that a premium is adapted for the 2005 Flood area at 4.71% and for the 2015 Flood area at 2.62%, while the 2009 Flood area was not taken into account in this particular analysis. Additionally, statistical significance was noticed for both of the key independent variables. A premium that can be attributed either to locational attractiveness, to the fact that people tend to forget when events are frequent and repeated, or perhaps due to some improper econometric technique such as multicollinearity.

Additionally, in order to test the third hypothesis of the analysis and address the third question: *To what extent do the effects of a flooding differ among residential properties, considering their property types and price segments, ranging from low-end, middle-end to high-end housing?* , I assessed the data for heterogeneity through the execution of Chow-tests. Specifically, I checked whether the impact of flood risk on property prices differ across property types. The results of the Chow-test lead to the conclusion that I can reject the null hypothesis, indicating that there is heterogeneity across property types. After running the various regressions, most property do show significant coefficients for the three flood areas with the expected signs. Only the type of semi-detached lacks any statistical coefficient, although the sign is the anticipated one. The property type for Flat/Maisonettes is the only property type showing significant coefficients with the opposite positive sign. Therefore, I conclude that there is heterogeneity across property types, with most of the results being statistically significant and in line with the regression results of our base model. Additionally, it was also tested whether the influence of flood risk on property prices differs among price segments. In this case, I also rejected the null hypothesis of the Chow test and a structural break is observed, indicating heterogeneity across price segments. Yet, upon observing the results of the regressions, the coefficients of the key independent variables related to floods yield mixed results. While middle-end properties show positive signs, leading to a premium if the property has been affected by flooding and is sold, only the variable Flood area 2009 is statistically significant. Regarding low-end and high-end properties, most of them exhibit the anticipated signs, resulting in a discount, which is also consistent with logical expectations and the regression results of the base model. Thus, due to the mixed outcomes, it can be stated that there is heterogeneity across price segments, however, the hypothesis is partially supported.

In conclusion, the results of this research have been intriguing and partly support my hypotheses. To answer the central research question: *How is the market value of residential real estate in Cumbria, England associated with the risk of flooding?* , considering our base model, I conclude that flood risk does influence the market value of residential real estate in the area of Cumbria, England. The 2009 Flood caused by extremely heavy rainfall that set a record with 316.4mm in 24 hours, significantly impacts residential property prices. However, the 2005 and 2015 floods, also caused by intense record-breaking rainfall, show the expected signs but do not appear to affect our model in a statistically significant way. In addition, when running the third model that includes the capital city of the region and focuses on this particular area, they seem to have a statistically significant effect on house prices with a positive sign, leading to a premium rather than a discount. Based on these partly contradictory findings, the government may be able to implement a flood awareness strategy to promote market transparency. Flooding is becoming a more serious issue around the world, particularly due to climate change. When the general population is educated and informed about this hazard, they will be able to act on it, potentially resulting in a more competitive market due to increased transparency.

5.1 Limitations & Future Research

During the analysis, I identified several limitations that should be considered for future research. Firstly, it is recommended to include more property characteristics as control variables, as well as locational effects to improve the model fit. By incorporating these additional independent variables, transactions in the area examined could be described more efficiently, and potentially improving the model, as measured by the R-squared. The original dataset already included features like build status, property type and estate type, and adding individual characteristics such as property size, presence of a parking, garden presence, basement availability, number of rooms and renovations could provide a more detailed description of the properties involved in transactions. Moreover, additional locational characteristics could be included, such as specific street names, proximity to water, or proximity to flood-prone areas, allowing for a more accurate characterization of transacted properties. Additionally, the more of the dependent variable's variation is explained before adding the variable of interest and the better the overall model fit, the more accurate researchers might be able to distill the impact of the flood risk.

Furthermore, there is a limitation regarding the consideration of long-term effects. The absence of arguments linking flood events with their impacts over extended periods creates challenges in isolating the influence of flood risk on property prices. It becomes crucial to recognize biases introduced by external factors such as diverse market cycles including events like the financial crisis of 2008, booming real estate markets because of low interest rates in a global basis, and geopolitical events like Brexit for a more accurate interpretation of results. This research also suffers from a limitation regarding how damage levels are determined within flood zones. When categorizing properties based on their location in areas affected by floods or not, it would greatly improve the outcomes to consider the damage caused by the floods. It would be more insightful to distinguish between properties that suffered damage and those located in flood affected areas but remained undamaged. Additionally, gathering information about the extent of damage for each property would result in more accurate outcomes.

Lastly, the research focuses on the time period of 2003 to 2022, investigating three record-breaking flood events in January 2005, November 2009 and December 2015. Further research could be suggested to broaden the investigated period. When examining a larger period this offers the opportunity to investigate if trend effects occur. To broaden knowledge about the financial consequences of climate change it is interesting to study if depreciation still occurs after a longer period. Additionally, expanding the case study area would allow us to explore if these results can be generalized to regions with similar characteristics. Furthermore, in this study the case study areas are subjectively selected based on historical events and supporting reports. However, as found that the degree of impact had a moderating effect. Future research could investigate what the exact moderating effect is by for instance incorporating the amount of damage per region. Only including economic damage would not be sufficient, as residents could also suffer emotional damage as a result of experiencing a near flood, a challenging aspect to quantify in monetary terms.

6. REFERENCES

Aiken, LS, West, SG, Reno, RR & Publishing, S 1991, *Multiple regression : testing and interpreting interactions*, Sage Publications, Druk, Newbury Park ; London ; New Delhi.

Akerlof, GA 1970, 'The Market for "Lemons": Quality Uncertainty and the Market Mechanism', *The Quarterly Journal of Economics*, vol. 84, no. 3, pp. 488–500.

Atreya, A & Czajkowski, J 2014, 'Housing Price Response to the Interaction of Positive Coastal Amenities and Negative Flood Risks', *2014 Annual Meeting, July 27-29, 2014, Minneapolis, Minnesota*.

Atreya, A & Czajkowski, J 2016, 'Graduated Flood Risks and Property Prices in Galveston County', *Real Estate Economics*, vol. 47, no. 3, pp. 807–844.

Atreya, A & Ferreira, S 2014, 'Seeing is Believing? Evidence from Property Prices in Inundated Areas', *Risk Analysis*, vol. 35, no. 5, pp. 828–848.

Atreya, A, Ferreira, S & Kriesel, W 2013, 'Forgetting the Flood? An Analysis of the Flood Risk Discount over Time', *Land Economics*, vol. 89, no. 4, pp. 577–596.

Barker, PA, Wilby, RL & Borrows, J 2004, 'A 200-year precipitation index for the central English Lake District / Un indice de précipitation de 200 ans pour la Région des Lacs en Angleterre', *Hydrological Sciences Journal*, vol. 49, no. 5.

Belanger, P & Bourdeau-Brien, M 2017, 'The impact of flood risk on the price of residential properties: the case of England', *Housing Studies*, vol. 33, no. 6, pp. 876–901.

Beltrán, A, Maddison, D & Elliott, R 2019, 'The impact of flooding on property prices: A repeat-sales approach', *Journal of Environmental Economics and Management*, vol. 95, pp. 62–86.

Bin, O & Kruse, JB 2006, 'Real Estate Market Response to Coastal Flood Hazards', *Natural Hazards Review*, vol. 7, no. 4, pp. 137–144.

Bin, O, Kruse, JB & Landry, CE 2008, 'Flood Hazards, Insurance Rates, and Amenities: Evidence from the Coastal Housing Market', *Journal of Risk & Insurance*, vol. 75, no. 1, pp. 63–82.

Bin, O & Landry, CE 2013, 'Changes in implicit flood risk premiums: Empirical evidence from the housing market', *Journal of Environmental Economics and Management*, vol. 65, no. 3, pp. 361–376.

Brooks, C & Tsolacos 2010, *Real Estate Modelling and Forecasting*, Cambridge University Press.

Bui, N, Wen, L & Sharp, B 2022, 'House Prices and Flood Risk Exposure: An Integration of Hedonic Property Model and Spatial Econometric Analysis', *The Journal of Real Estate Finance and Economics*.

Cheshire, P & Sheppard, S 1995, 'On the Price of Land and the Value of Amenities', *Economica*, vol. 62, no. 246, p. 247.

Chow, GC 1960, 'Tests of Equality Between Sets of Coefficients in Two Linear Regressions', *Econometrica*, vol. 28, no. 3, p. 591.

Clayton, J, Devaney, S, Sayce, S & Van de Wetering, J 2021, 'Climate Risk and Real Estate Prices: What Do We Know?', *The Journal of Portfolio Management*, vol. 47, no. 10, pp. 75–90.

Cumbria County Council 2016, *Carlisle Flood Investigation Report*, 3 May, Environment Agency, pp. 1–52.

— 2023, *Welcome to Cumbria County Council*, www.cumbria.gov.uk.

Cumbria County Council Performance & Intelligence Team 2018, *Flooding in Cumbria, December 2015*, 2 July, Cumbria County Council, Cumbria, pp. 1–66.

Cumbria Tourism 2023, *An introduction to Cumbria*, Visit Lake District, viewed 23 November 2023, https://www.visitlakedistrict.com/information/an-introduction-to-cumbria.

DiPasquale, D & Wheaton, WC 1992, 'The Markets for Real Estate Assets and Space: A Conceptual Framework', *Real Estate Economics*, vol. 20, no. 2, pp. 181–198.

Donahue, AK & Tuohy, RV 2006, *Lessons We Don't Learn A Study of the Lessons of Disasters, Why We Repeat Them, and How We Can Learn Them*, https://calhoun.nps.edu/bitstream/handle/10945/25094/78.pdf?sequence=1&isAllowed=y.

Dubin, RA 1988, 'Estimation of Regression Coefficients in the Presence of Spatially Autocorrelated Error Terms', *The Review of Economics and Statistics*, vol. 70, no. 3, p. 466.

Environment Agency 2016, 'Reducing flood risk from source to sea', pp. 1–13.

— 2023, *Flood Risk Areas*, www.data.gov.uk.

Gharbia, SS, Naughton, O, Farrelly, V, Lyons, R & Pilla, F 2016, 'Attitudes to systemic risk: The impact of flood risk on the housing market in Dublin', *2016 18th Mediterranean Electrotechnical Conference (MELECON)*.

Grothmann, T & Reusswig, F 2006, 'People at Risk of Flooding: Why Some Residents Take Precautionary Action While Others Do Not', *Natural Hazards*, vol. 38, no. 1-2, pp. 101–120.

Harper, M 2005, *10 years on from the Cumbrian and Carlisle Floods of 2005 - Creating a better place*, environmentagency.blog.gov.uk, viewed 19 October 2023, https://environmentagency.blog.gov.uk/2015/01/08/10-years-on-from-the-cumbrian-and-carlisle-floods-of-2005/.

Harrison, D, T. Smersh, G & Schwartz, A 2001, 'Environmental Determinants of Housing Prices: The Impact of Flood Zone Status', *Journal of Real Estate Research*, vol. 21, no. 1-2, pp. 3–20.

He, J, Renshaw, E & Szelest, B 1998, 'Modeling the Outcome of Presidential Elections Using Economic Variables: The Smart Use of Dummy Variables', *Journal of Policy Modeling*, vol. 20, no. 2, pp. 251–259.

Hirsch, J & Hahn, J 2018, 'How flood risk impacts residential rents and property prices', *Journal of Property Investment & Finance*, vol. 36, no. 1, pp. 50–67.

HM Land Registry 2023, *HM Land Registry Open Data*, landregistry.data.gov.uk.

Howard, L 2023, *House Prices: Zoopla Sees Values Fall But Says Volume Of Sales Is Increasing*, Forbes Advisor UK, viewed 27 December 2023, <https://www.forbes.com/uk/advisor/personal-finance/2023/12/22/house-prices-updates/>.

Kim, J, Park, J, Yoon, DK & Cho, G-H 2017, 'Amenity or hazard? The effects of landslide hazard on property value in Woomyeon Nature Park area, Korea', *Landscape and Urban Planning*, vol. 157, pp. 523–531.

Lamond, J, Proverbs, D & Hammond, F 2010, 'The Impact of Flooding on the Price of Residential Property: A Transactional Analysis of the UK Market', *Housing Studies*, vol. 25, no. 3, pp. 335–356.

Lamond, JE, Bhattacharya-Mis, N, Chan, FKS, Kreibich, H, Montz, B, Proverbs, DG & Wilkinson, S 2019, 'Flood risk insurance, mitigation and commercial property valuation', *Property Management*, vol. 37, no. 4, pp. 512–528.

Lancaster, KJ 2019, 'A New Approach to Consumer Theory', *Journal of Political Economy*, vol. 74, no. 2, pp. 132–157.

— *Floods in Carlisle - January 2005 - Met Office*, www.metoffice.gov.uk.

— 2009, *Heavy Rainfall in Lake District November 2009 - Met Office*, https://www.metoffice.gov.uk.

— 2015, *Flooding in Cumbria December 2015*, https://www.metoffice.gov.uk.

Miller, JD, Kjeldsen, TR, Hannaford, J & Morris, DG 2013, 'A hydrological assessment of the November 2009 floods in Cumbria, UK', *Hydrology Research*, vol. 44, no. 1, pp. 180–197.

Nguyen, Q, Thorsnes, P, Diaz-Rainey, I, Moore, A, Cox, S & Stirk-Wang, L 2022, 'Price recovery after the flood: risk to residential property values from climate change-related flooding*', *Australian Journal of Agricultural and Resource Economics*, vol. 66, no. 3, pp. 532–560.

Nielsen, B & Whitby, A 2015, 'A Joint Chow Test for Structural Instability', *Econometrics*, vol. 3, no. 1, pp. 156–186.

Plumplot 2023, *Cumbria house prices in maps and graphs.*, www.plumplot.co.uk, viewed 12 December 2023, https://www.plumplot.co.uk/Cumbria-house-prices.html.

Pope, JC 2008, 'Do Seller Disclosures Affect Property Values? Buyer Information and the Hedonic Model', *Land Economics*, vol. 84, no. 4, pp. 551–572.

Pryce, G, Chen, Y & Galster, G 2011, 'The Impact of Floods on House Prices: An Imperfect Information Approach with Myopia and Amnesia', *Housing Studies*, vol. 26, no. 2, pp. 259–279.

Pryce, G, Chen, Y & Mackay, D 2009, 'Estimating the Impact of Floods on House Prices: An Application to the 2005 Carlisle Flood', *European Real Estate Society*, pp. 1–21.

Rajapaksa, D, Wilson, C, Managi, S, Hoang, V & Lee, B 2016, 'Flood Risk Information, Actual Floods and Property Values: A Quasi-Experimental Analysis', *Economic Record*, vol. 92, pp. 52–67.

Recorded Flood Outlines 2023, *Find open data - data.gov.uk*, www.data.gov.uk.

Rosen, S 1974, 'Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition', *Journal of Political Economy*, vol. 82, no. 1, pp. 34–55.

Shr, Y-H (Jimmy) & Zipp, KY 2019, 'The Aftermath of Flood Zone Remapping: The Asymmetric Impact of Flood Maps on Housing Prices', *Land Economics*, vol. 95, no. 2, pp. 174–192.

Sibley, A 2010, 'Analysis of extreme rainfall and flooding in Cumbria 18-20 November 2009', *Weather*, vol. 65, no. 11, pp. 287–292.

Skantz, T & Strickland, T 1987, ‘House Prices and a Flood Event: An Empirical Investigation of Market Efficiency’, *Journal of Real Estate Research*, vol. 2, no. 2, pp. 75–83.

Spencer, P, Faulkner, D, Perkins, I, Lindsay, D, Dixon, G, Parkes, M, Lowe, A, Asadullah, A, Hearn, K, Gaffney, L, Parkes, A & James, R 2017, ‘The floods of December 2015 in northern England: description of the events and possible implications for flood hydrology in the UK’, *Hydrology Research*, vol. 49, no. 2, pp. 568–596.

Speyrer, J & Ragas, WadeR 1991, ‘Housing prices and flood risk: An examination using spline regression’, *The Journal of Real Estate Finance and Economics*, vol. 4, no. 4.

Stamou, M, Mimis, A & Rovolis, A 2017, ‘House price determinants in Athens: a spatial econometric approach’, *Journal of Property Research*, vol. 34, no. 4, pp. 269–284.

Stewart, EJ, Morris, DG, Jones, DA & Spencer, PS 2010, ‘Extreme rainfall in Cumbria, November 2009 – an assessment of storm rarity’, *Role of hydrology in managing consequences of a changing global environment*.

Surminski, S & Eldridge, J 2015, ‘Flood insurance in England - an assessment of the current and newly proposed insurance scheme in the context of rising flood risk’, *Journal of Flood Risk Management*, vol. 10, no. 4, pp. 415–435.

UK House Price Index 2022, Data.gov.uk, viewed 27 December 2023, https://landregistry.data.gov.uk/app/ukhpi/browse?from=2022-08-01&location=http%3A%2F%2Flandregistry.data.gov.uk%2Fid%2Fregion%2Fcumberland&to=2023-08-01&lang=en.

Votsis, A & Perrels, A 2015, ‘Housing Prices and the Public Disclosure of Flood Risk: A Difference-in-Differences Analysis in Finland’, *The Journal of Real Estate Finance and Economics*, vol. 53, no. 4, pp. 450–471.

Wedawatta, G, Ingirige, B & Proverbs, D 2013, ‘Small businesses and flood impacts: the case of the 2009 flood event in Cockermouth’, *Journal of Flood Risk Management*, vol. 7, no. 1, pp. 42–53.

Whyte, I 2009, ‘Floods and their Impact on the Landscape: Flood Histories from Cumbria’, *Landscapes*, vol. 10, no. 1, pp. 61–76.

Zhang, L 2016, ‘Flood hazards impact on neighborhood house prices: A spatial quantile regression analysis’, *Regional Science and Urban Economics*, vol. 60, pp. 12–19.

7. APPENDIX: NOTATION GLOSSARY

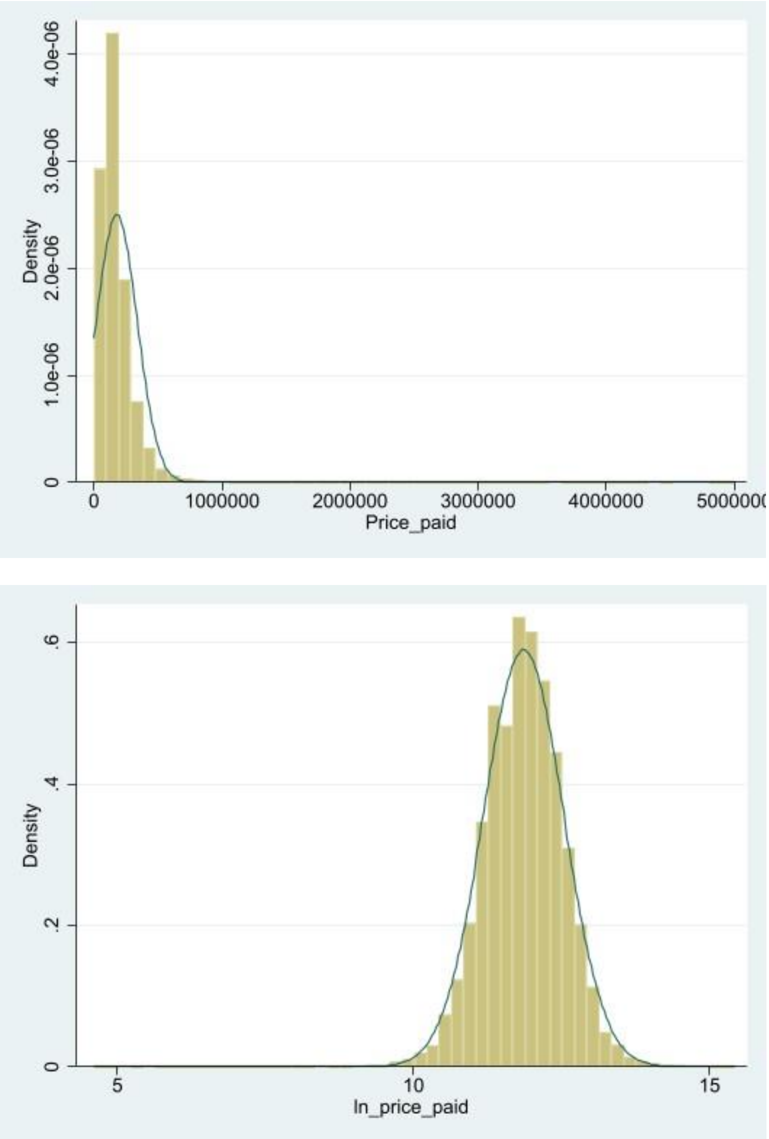
Empirical Model	
$Ln(P_{i,t})$	Natural logarithm of transaction prices
I	Transaction i = 1, 2, …, N
T	Time t = 1, 2, …, N
Cons	Constant to be estimated
β_i	Parameters to be estimated
Flood area 2005	Dummy variable, which equals 1 if property sold is located within the area affected by the flood of 2005
Flood area 2009	Dummy variable, which equals 1 if property sold is located within the area affected by the flood of 2009
Flood area2015	Dummy variable, which equals 1 if property sold is located within the area affected by the flood of 20015
Flood area 2005excl.	Dummy variable, which equals 1 if property sold is located within the area affected by the flood of 2005, excluding the capital city
Flood area 2015excl.	Dummy variable, which equals 1 if property sold is located within the area affected by the flood of 20015, excluding the capital city
Flood area 2005incl.	Dummy variable, which equals 1 if property sold is located within the area affected by the flood of 2005, including the capital city
Flood area 2015incl.	Dummy variable, which equals 1 if property sold is located within the area affected by the flood of 20015, including the capital city
Estate Type	Dummy variable, which equals 1 if property sold is freehold and equals 0 if property sold is leasehold
New Build	Dummy variable, which equal 1 if property sold has been newly built
Property Type	Categorical Variable, describing whether the property sold is detached, semi-detached, flat/maisonettes, terraced or other
District	Categorical variable, describing the district where the property sold is located
PostCode	Categorical variable, describing the postcode where the property sold is located
Sales Year	Categorical variable, describing the transaction year
Sales Month	Categorical variable, describing the transaction month

8. FIGURES & TABLES

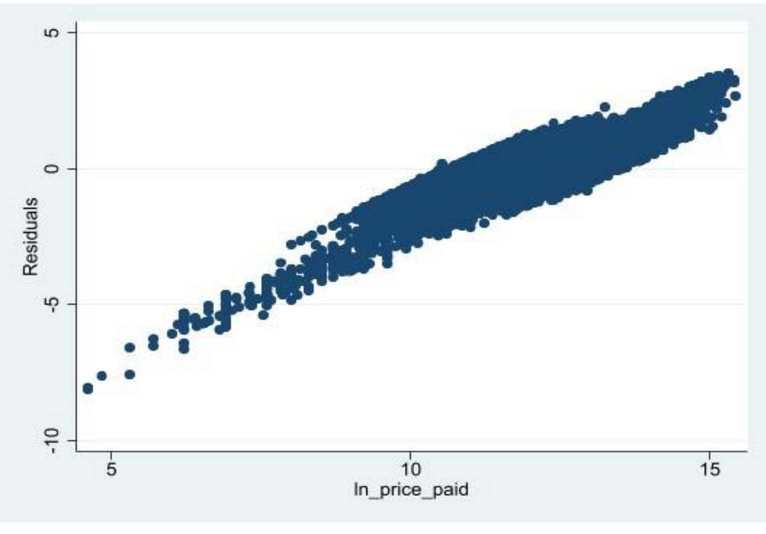
Table 12 Correlation Matrix

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Flood2005	1.000									
(2) Flood2009	-0.011	1.000								
(3) Flood2015	-0.021	-0.015	1.000							
(4) Estate Type	-0.038	-0.072	-0.073	1.000						
(5) Newbuild	-0.010	-0.007	-0.022	-0.071	1.000					
(6) Proper Type	0.033	0.003	0.020	0.175	-0.159	1.000				
(7) Sales Year	-0.002	0.004	-0.003	0.023	0.011	-0.039	1.000			
(8) Sales Month	0.005	-0.004	-0.000	0.003	0.010	-0.012	-0.006	1.000		
(9) Postcode	-0.121	0.004	-0.011	0.007	-0.090	0.017	-0.017	-0.008	1.000	
(10) District	-0.060	0.011	0.029	-0.072	-0.031	-0.129	0.014	-0.007	0.524	1.000

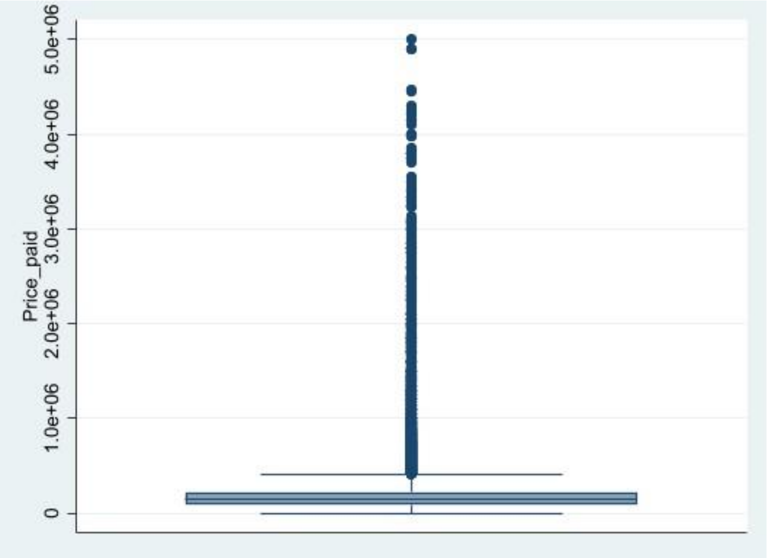
Graph 9: Histogram of log transaction prices indicating issues with Heteroscedasticity



Graph 10: Scatterplot of residuals



Graph 11: Outliers of dependent variable



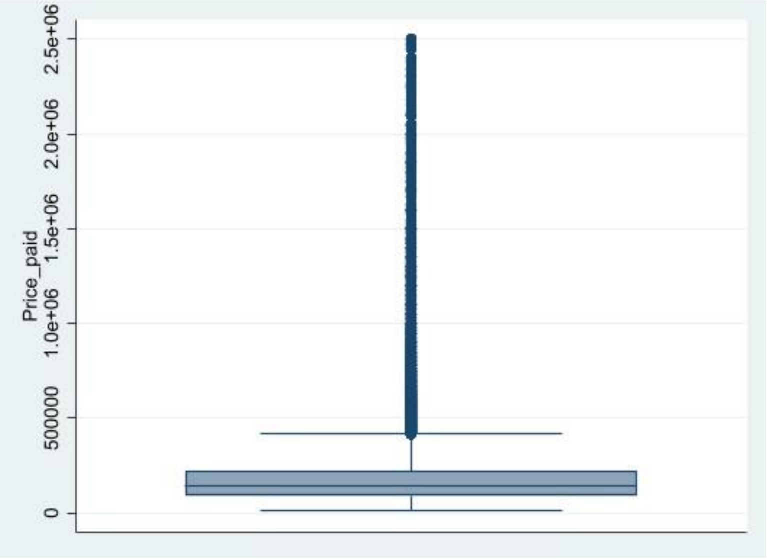


Table 14: Variance Inflation Factor (Base model)

	VIF	1/VIF
Flood area 2005excl.	1.009	.991
Flood area 2009	1.01	.99
Flood area 2015excl.	1.014	.986
Estate Type	2.111	.421
Newbuild	1.031	.969
Property Type (Detached)		
Flat/Maisonettes	2.222	.37
Other	1.143	.875
Semi-Detached	1.575	.635
Terraced	1.726	.579
Sales Year	1.037	.965
Sales Month	1	1
Postcode	1.69	.592
District	1.663	.601
Mean VIF	1.40	.

Note: This table denotes the variance inflation factor per independent variable. The variable Detached is used as reference variable and therefore is not included. A VIF value of 1 indicates that two variables are not correlated, a value between 1 and 5 indicates a moderate correlation and a value above 5 indicates a high correlation.

Table 21: Heterogeneity for Property Type (Primary Model)

Variables	Pooled	Detached	Semi-Detached	Flat/Maisonettes	Terraced	Other
Flood area 2005	-0.0888***	-0.418	0.123***	0.00884	0.0359***	-0.222
	(0.000)	(0.089)	(0.000)	(0.673)	(0.001)	(0.135)
Flood area 2009	-0.165***	-0.0251	-0.0214	0.110***	-0.121***	-0.637***
	(0.000)	(0.381)	(0.457)	(0.000)	(0.000)	(0.000)
Flood area 2015	-0.0834***	-0.0354*	0.0306*	0.0349*	0.0310***	-0.306**
	(0.000)	(0.024)	(0.015)	(0.016)	(0.000)	(0.002)
Estate Type (1=Freeh.)	0.421***	0.269***	0.177***	0.102***	0.0739***	0.583***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Newbuild (1=Yes)	0.308***	0.0452***	0.115***	0.324***	0.206***	-0.120
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.441)
Time Effects	Yes	Yes	Yes	Yes	Yes	Yes
Location Effects	Yes	Yes	Yes	Yes	Yes	Yes
_cons	11.08***	11.58***	11.17***	11.08***	11.12***	5.088***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	184048	43715	51535	15458	67998	5342
R^2	0.360	0.369	0.409	0.431	0.519	0.104
Adj. R^2	0.360	0.368	0.408	0.428	0.518	0.089
RSS	53900.95	6111.89	7146.80	3051.78	10899.91	10134.34

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: This table presents the regression results for each property type individually. The dependent variable is the ln (transaction price). Time effects are sales month and sales year, and location effects control for postcode and district. Parentheses include the p-values.

Table 22: Heterogeneity for Price Segments (Primary Model)

Variables	Pooled	Low-end properties	Middle-end properties	High-end properties
Flood area 2005	0.0453***	0.0399***	0.0248***	-0.0111
	(0.000)	(0.000)	(0.000)	(0.649)
Flood area 2009	-0.0528***	-0.0994***	0.0185*	-0.0656***
	(0.000)	(0.000)	(0.021)	(0.000)
Flood area 2015	0.0245***	0.0225**	0.00739	-0.0299*
	(0.000)	(0.004)	(0.077)	(0.013)
Estate Type (1=Freeh.)	0.158***	0.0418***	0.0467***	0.0291*
	(0.000)	(0.000)	(0.000)	(0.044)
Newbuild (1=Yes)	0.154***	0.166***	0.0376***	-0.0669***
	(0.000)	(0.000)	(0.000)	(0.000)
Property Type (Base: Detached)				
Flat/Maisonettes	-0.760***	-0.468***	0.138*	-0.180***
	(0.000)	(0.000)	(0.029)	(0.000)
Other	-0.431***	-1.118***	-0.0449***	0.405***
	(0.000)	(0.000)	(0.000)	(0.000)
Semi-Detached	-0.467***	-0.180***	0.141***	-0.119***
	(0.000)	(0.000)	(0.000)	(0.000)
Terraced	-0.759***	-0.405***	-0.154***	-0.150***
	(0.000)	(0.000)	(0.000)	(0.000)
Time Effects	Yes	Yes	Yes	Yes
Location Effects	Yes	Yes	Yes	Yes
_cons	11.65***	11.36***	12.17***	12.71***
	(0.000)	(0.000)	(0.000)	(0.000)
N	184048	98748	64076	21224
R^2	0.531	0.305	0.179	0.236
Adj. R^2	0.531	0.304	0.178	0.232
RSS	39046.90	14294.93	1994.35	1737.73

*$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: This table presents the regression results for each price segment individually. The dependent variable is the ln (transaction price). Time effects are sales month and sales year, and location effects control for postcode and district. Parentheses include the p-values.

Stata Do-File

```
Cleaned

**# Bookmark

*CODE USED FOR INVESTIGATING THE PRIMARY MODEL OF MY ANALYSIS

clear

cd "D:\RUG\Thesis Final Data\Cumbria\Property Transaction Dataset"
D:\RUG\Thesis Final Data\Cumbria\Property Transaction Dataset

import delimited "D:\RUG\Thesis Final Data\Cumbria\Property Transaction Dataset\Final.csv"

*In order to look for missing value
mdesc

*In order to do the Diagnostics
hettest
regcheck
predict e, res
twoway scatter ln_price_paid e
*or
predict residuals, r
scatter residuals ln_price_paid residuals

*The code for descriptive statistics
sum var1 var2 ...

*The code for a correlation matrix table
corr var1 var2 var3 ...

*To check if you have a normal distribution you create a histogram and observe it
hist ln_price_paid
hist price_paid

*This code is for generating the natural logarithm of our dependent variable
gen ln_price_paid = log(price_paid)

*Check for Outliers
graph box ln_price_paid
graph box price_paid
drop if price_paid > 2500000
drop if price_paid < 10000
```

```
*Encoding variables so I can run it in stata
encode property_type, gen (propertytype)
encode postcode, gen (postcode_)
encode district, gen (district_)

*CREATING DUMMIES

gen estatetype = (estate_type == "F")

gen newbuild = ( new_build == "Y")

*A flood dummy variable that is equal to 1 if the property/transaction is affected by each respective flood, and 0 otherwise

gen Flood2005 = (join_count == 1) & (name == "2005 Flood")

gen Flood2009 = (join_count == 1) & (name == "2009 Flood")

gen Flood2015 = (join_count == 1) & (name == "2015 Flood")


*Regressing variables
reg ln_price_paid Flood2005 Flood2009 Flood2015 estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_

*Regression with robust standard errors
reg ln_price_paid Flood2005 Flood2009 Flood2015 estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_ , r

*To control for VIF you have first to run a regression and then enter the code
estat vif or vif

*create multiple regression table
*First run each model and then save each model
reg ln_price_paid Flood2005 Flood2009 Flood2015,r
eststo model1
reg ln_price_paid Flood2005 Flood2009 Flood2015 estatetype newbuild i.propertytype ,r
eststo model2
reg ln_price_paid Flood2005 Flood2009 Flood2015 estatetype newbuild i.propertytype i.year i.month, r
eststo model3
reg ln_price_paid Flood2005 Flood2009 Flood2015 estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_ ,r
eststo model4
```

```
*To export the table I can Use:
esttab using regtab.doc, r2 ar2 p
*or
outreg2 using myfile.doc, word


***Chow Test for Property types
gen detchaded_dummy = ( property_type == "D")
gen semi_detached_dummy = ( property_type == "S")
gen flat_dummy = ( property_type == "F")
gen terraced_dummy = ( property_type == "T")
gen other_dummy = ( property_type == "O")

eststo clear
*Pooled
eststo: reg ln_price_paid Flood2005 Flood2009 Flood2015 estatetype newbuild i.year i.month i.postcode_ i.district_
*Regressions for each Type of Property
eststo: reg ln_price_paid Flood2005 Flood2009 Flood2015 estatetype newbuild i.year i.month i.postcode_ i.district_ if detchaded_dummy == 1
eststo: reg ln_price_paid Flood2005 Flood2009 Flood2015 estatetype newbuild i.year i.month i.postcode_ i.district_ if semi_detached_dummy == 1
eststo: reg ln_price_paid Flood2005 Flood2009 Flood2015 estatetype newbuild i.year i.month i.postcode_ i.district_ if flat_dummy == 1
eststo: reg ln_price_paid Flood2005 Flood2009 Flood2015 estatetype newbuild i.year i.month i.postcode_ i.district_ if terraced_dummy == 1
eststo: reg ln_price_paid Flood2005 Flood2009 Flood2015 estatetype newbuild i.year i.month i.postcode_ i.district_ if other_dummy == 1

esttab using ChowTestPrimary.rtf, r2 ar2 p

***CHOW TEST FOR PRICE SEGMENT

gen price_segment = cond(price_paid <= 150000, "Low-End", cond(price_paid <= 300000, "Middle-End", "High-End"))
tabulate price_segment
gen lowend_dummy = (price_segment == "Low-End")
gen middleend_dummy = (price_segment == "Middle-End")
gen highend_dummy = (price_segment == "High-End")

eststo clear
*Pooled
eststo: reg ln_price_paid Flood2005 Flood2009 Flood2015 estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_
*Regressions for each Type of Property
eststo: reg ln_price_paid Flood2005 Flood2009 Flood2015 estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_ if lowend_dummy == 1
eststo: reg ln_price_paid Flood2005 Flood2009 Flood2015 estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_ if middleend_dummy == 1
eststo: reg ln_price_paid Flood2005 Flood2009 Flood2015 estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_ if highend_dummy == 1

esttab using PrimaryChowTest.rtf, r2 ar2 p
```

```
****EXCLUDING CARLISLE BASE MODEL ANALYSIS
clear

cd "D:\RUG\Thesis Final Data\Cumbria\Property Transaction Dataset"
D:\RUG\Thesis Final Data\Cumbria\Property Transaction Dataset

import delimited "D:\RUG\Thesis Final Data\Cumbria\Property Transaction Dataset\ExcludingCarlisle.csv"

*In order to look for missing value
mdesc

*In order to do the Diagnostics
hettest
regcheck
predict e, res
twoway scatter lnsaleprice e
*or
predict residuals, r
scatter residuals ln_price_paid residuals

*The code for descriptive statistics
sum var1 var2 ...

*The code for a correlation matrix table
corr var1 var2 var3

*To check if you have a normal distribution you create a histogram and observe it
hist ln_price_paid
hist price_paid

*This code is for generating the natural logarithm of our dependent variable
gen ln_price_paid = log(price_paid)

*Check for Outliers
graph box ln_price_paid
graph box price_paid
drop if price_paid > 2500000
drop if price_paid < 10000
```

```
*Encoding variables so I can run it in stata
encode property_type, gen (propertytype)
encode postcode, gen (postcode_)
encode district, gen (district_)

*CREATING DUMMIES

gen estatetype = (estate_type == "F")

gen newbuild = ( new_build == "Y")

gen Flood2005excl = (join_count == 1) & (name == "2005 Flood")

gen Flood2009 = (join_count == 1) & (name == "2009 Flood")

gen Flood2015excl = (join_count == 1) & (name == "2015 Flood")

*Regressing variables
reg ln_price_paid Flood2005excl Flood2009 Flood2015excl estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_

*Regression with robust standard errors
reg ln_price_paid Flood2005excl Flood2009 Flood2015excl estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_ , r

*To control for VIF you have first to run a regression and then just click the code
estat vif or vif

*create multiple regression table
*First run each model and then save each model
reg ln_price_paid Flood2005excl Flood2009 Flood2015excl,r
eststo model1
reg ln_price_paid Flood2005excl Flood2009 Flood2015excl estatetype newbuild i.propertytype ,r
eststo model2
reg ln_price_paid Flood2005excl Flood2009 Flood2015excl estatetype newbuild i.propertytype i.year i.month, r
eststo model3
reg ln_price_paid Flood2005excl Flood2009 Flood2015excl estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_ ,r
eststo model4

*To export the table I can Use:
esttab using regtab1.doc, r2 ar2 p
*or
outreg2 using myfile1.doc, word
```

```
***Chow Test for Property types
gen detchaded_dummy = ( property_type == "D")
gen semi_detached_dummy = ( property_type == "S")
gen flat_dummy = ( property_type == "F")
gen terraced_dummy = ( property_type == "T")
gen other_dummy = ( property_type == "O")

eststo clear
*Pooled
eststo: reg ln_price_paid Flood2005excl Flood2009 Flood2015excl estatetype newbuild i.year i.month i.postcode_ i.district_
*Regressions for each Type of Property
eststo: reg ln_price_paid Flood2005excl Flood2009 Flood2015excl estatetype newbuild i.year i.month i.postcode_ i.district_ if detchaded_dummy == 1
eststo: reg ln_price_paid Flood2005excl Flood2009 Flood2015excl estatetype newbuild i.year i.month i.postcode_ i.district_ if semi_detached_dummy == 1
eststo: reg ln_price_paid Flood2005excl Flood2009 Flood2015excl estatetype newbuild i.year i.month i.postcode_ i.district_ if flat_dummy == 1
eststo: reg ln_price_paid Flood2005excl Flood2009 Flood2015excl estatetype newbuild i.year i.month i.postcode_ i.district_ if terraced_dummy == 1
eststo: reg ln_price_paid Flood2005excl Flood2009 Flood2015excl estatetype newbuild i.year i.month i.postcode_ i.district_ if other_dummy == 1

esttab using ChowTestBase.rtf, r2 ar2 p


*CHOW TEST FOR PRICE SEGMENT

gen price_segment = cond(price_paid <= 150000, "Low-End", cond(price_paid <= 300000, "Middle-End", "High-End"))
tabulate price_segment
gen lowend_dummy = (price_segment == "Low-End")
gen middleend_dummy = (price_segment == "Middle-End")
gen highend_dummy = (price_segment == "High-End")

eststo clear
*Pooled
eststo: reg ln_price_paid Flood2005excl Flood2009 Flood2015excl estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_
*Regressions for each Type of Property
eststo: reg ln_price_paid Flood2005excl Flood2009 Flood2015excl estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_ if lowend_dummy == 1
eststo: reg ln_price_paid Flood2005excl Flood2009 Flood2015excl estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_ if middleend_dummy == 1
eststo: reg ln_price_paid Flood2005excl Flood2009 Flood2015excl estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_ if highend_dummy == 1

esttab using BaseChowTest.rtf, r2 ar2 p
```

```
****INCLUDING CARLISLE THIRD MODEL ANALYSIS

clear

cd "D:\RUG\Thesis Final Data\Cumbria\Property Transaction Dataset"
D:\RUG\Thesis Final Data\Cumbria\Property Transaction Dataset

import delimited "D:\RUG\Thesis Final Data\Cumbria\Property Transaction Dataset\Final.csv"

*In order to look for missing value
mdesc

*In order to do the Diagnostics
hettest
regcheck
predict e, res
twoway scatter ln_price_paid e
*or
predict residuals, r
scatter residuals ln_price_paid residuals

*The code for descriptive statistics
sum var1 var2 ...

*The code for a correlation matrix table
corr var1 var2 var3

*To check if you have a normal distribution you create a histogram and observe it
hist ln_price_paid
hist price_paid

*This code is for generating the natural logarithm of our dependent variable
gen ln_price_paid = log(price_paid)

*Check for Outliers
graph box ln_price_paid
graph box price_paid
drop if price_paid > 2500000
drop if price_paid < 10000
```

```
*Encoding variables so I can run it in stata
encode property_type, gen (propertytype)
encode postcode, gen (postcode_)
encode district, gen (district_)

*CREATING DUMMIES

gen estatetype = (estate_type == "F")

gen newbuild = ( new_build == "Y")

*A flood dummy variable that is equal to 1 if the property/transaction is affected by each respective flood, and 0 otherwise

gen Flood2005 = (join_count == 1) & (name == "2005 Flood")

gen Flood2015 = (join_count == 1) & (name == "2015 Flood")


*Regressing variables
reg ln_price_paid Flood2005 Flood2015 estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_

*Regression with robust standard errors
reg ln_price_paid Flood2005 Flood2015 estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_ , r

*To control for VIF you have first to run a regression and then just click the code
estat vif or vif

*create multiple regression table
*First run each model and then save each model
reg ln_price_paid Flood2005 Flood2015,r
eststo model1
reg ln_price_paid Flood2005 Flood2015 estatetype newbuild i.propertytype ,r
eststo model2
reg ln_price_paid Flood2005 Flood2015 estatetype newbuild i.propertytype i.year i.month, r
eststo model3
reg ln_price_paid Flood2005 Flood2015 estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_ ,r
eststo model4

*To export the table I can Use:
esttab using regtab2.doc, r2 ar2 p
*or
outreg2 using myfile2.doc, word
```

```
***Chow Test for Property types
gen detchaded_dummy = ( property_type == "D")
gen semi_detached_dummy = ( property_type == "S")
gen flat_dummy = ( property_type == "F")
gen terraced_dummy = ( property_type == "T")
gen other_dummy = ( property_type == "O")

eststo clear
*Pooled
eststo: reg ln_price_paid Flood2005 Flood2015 estatetype newbuild i.year i.month i.postcode_ i.district_
*Regressions for each Type of Property
eststo: reg ln_price_paid Flood2005 Flood2015 estatetype newbuild i.year i.month i.postcode_ i.district_ if detchaded_dummy == 1
eststo: reg ln_price_paid Flood2005 Flood2015 estatetype newbuild i.year i.month i.postcode_ i.district_ if semi_detached_dummy == 1
eststo: reg ln_price_paid Flood2005 Flood2015 estatetype newbuild i.year i.month i.postcode_ i.district_ if flat_dummy == 1
eststo: reg ln_price_paid Flood2005 Flood2015 estatetype newbuild i.year i.month i.postcode_ i.district_ if terraced_dummy == 1
eststo: reg ln_price_paid Flood2005 Flood2015 estatetype newbuild i.year i.month i.postcode_ i.district_ if other_dummy == 1

esttab using ChowTestThird.rtf, r2 ar2 p


***CHOW TEST FOR PRICE SEGMENT

gen price_segment = cond(price_paid <= 150000, "Low-End", cond(price_paid <= 300000, "Middle-End", "High-End"))
tabulate price_segment
gen lowend_dummy = (price_segment == "Low-End")
gen middleend_dummy = (price_segment == "Middle-End")
gen highend_dummy = (price_segment == "High-End")

eststo clear
*Pooled
eststo: reg ln_price_paid Flood2005 Flood2015 estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_
*Regressions for each Type of Property
eststo: reg ln_price_paid Flood2005 Flood2015 estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_ if lowend_dummy == 1
eststo: reg ln_price_paid Flood2005 Flood2015 estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_ if middleend_dummy == 1
eststo: reg ln_price_paid Flood2005 Flood2015 estatetype newbuild i.propertytype i.year i.month i.postcode_ i.district_ if highend_dummy == 1

esttab using ThirdChowTest.rtf, r2 ar2 p
```