PROXIMITY TO SEISMIC ACTIVITY AND ITS EFFECT ON PROPERTY VALUES: A CASE STUDY OF VANCOUVER, BC

Master RES Thesis

Olaf Welboren

S5542928

November 2024

COLOFON

Title	PROXIMITY TO SEISMIC ACTIVITY AND ITS EFFECT ON PROPERTY VALUES: A CASE STUDY OF VANCOUVER, BC
Version	Final Version
Author	Olaf Welboren
Supervisor	Dr. C. O'Driscoll
Second assessor	Dr. Niels Kuiper
E-mail	o.welboren@student.rug.nl
Date	November 7, 2024

For Emma

ABSTRACT

This thesis investigates the impact of seismic activity on property values in Vancouver, a city on the coast of the Pacific Northwest, characterized by both high earthquake risks and an expensive real estate market. Previous research on the topic suggests a negative effect of earthquake activity and risk awareness in the region on property values, although in some areas this effect can be negligible when risk is disclosed at the time of purchase of a property. Utilizing a Fixed Effects Linear Regression model and Panel data, this study investigates whether the proximity to earthquakes significantly influences property values. The results of this study show a significant effect of earthquake distance to Vancouver property prices between 2013 and 2023. As distance to an earthquake increases, so does the property value. This effect is in line with the theory, however, the effect measured with the available data and the model used is minimal. This impact may, on the one hand, be negligible because the population is unaware of the risks, or, the earthquake risk may not be priced into the real estate values because the risk is not large enough.

TABLE OF CONTENTS

1 INTRODUCTION

1.1 MOTIVATION

Along the West Coast of the United States and the southern West Coast of Canada you will find the Cascadia Subduction Zone, also known as the "CSZ" (Natural Resources Canada, 2011). This region is prone to seismic activity with hundreds of earthquakes annually. Most earthquakes in this region go undetected by human senses but can nonetheless inflict significant structural damages to real estate (Clarke, 2013; Natural Resources Canada, 2011). Besides the constant risk of smaller sized earthquakes, there is also a prevalent threat of a devastating megathrust earthquake. This high magnitude ($ML \geq 9$)[1] earthquake occurs as a result of the friction that has been built up by the Juan de Fuca plate being forced beneath the North American plate (Onur & Seemann, 2004). The megathrust earthquake, commonly referred to as "the Big One," last occurred in 1700 and has an average interval of 500-600 years (Government of Canada, 2024). Predictions for Vancouver to be struck by the CSZ megathrust earthquake in the next 30 years range from 12% to 35% (Onur & Seemann, 2004). An earthquake with a magnitude 9 or higher will have catastrophic effects on the metropolitan region of Vancouver (Nemetz & Dushnisky, 1994; Schulz, 2015). Early research on megathrust earthquakes in the Vancouver peninsula by Nemetz & Duschnisky (1994) predicted a potential loss of urban structures up to 93%. Additionally, road-, electricity-, communications- and other civil infrastructure will be damaged, which will pose a threat to living conditions in the area.

Even though the Pacific Northwest, including metro Vancouver, is rich in seismic activity the land and real estate markets in Vancouver, Seattle, and Victoria are exceedingly competitive (Clarke, 2013; Gold, 2017). Mortgage and rent payments take up a large portion of the expenditures of Vancouver residents, with an average of 61% of the net income spent (Carey, 2024). As of 2022, Vancouver ranks as the third most expensive city to live in North America (Shepert, 2022). The societal relevance and motivation of this thesis lies in trying to understand if there are any significant effects of the seismic events around

[1] *ML* = magnitude on Richter's scale, on surface level.

Vancouver on its property values. A property is in most cases the biggest financial investment any individual will make in their lifetime. Therefore, understanding how natural disasters like earthquakes, among other things, influence the cost and value of this investment is important for all current and future property owners, especially in the Vancouver region.

1.2 ACADEMIC RELEVANCE

Earlier literature on the CSZ and the seismic activity in Vancouver mainly focuses on the geography of the earthquakes and the physical impacts it will have on the region, such as landslides and tsunami's (Cherniawsky, Titov, Wang, & Li, 2007; Molnar, Assaf, Sirohey, & Adhikari, 2020). This literature, primarily situated in the realm of physical geography and geology, highlights how the physical geography of the Pacific Northwest is characterized by rocky terrain, mountains and great variation in elevation. It is therefore vulnerable to shifts in the terrain and landslides if a high magnitude earthquake occurs (Onur & Seemann, 2004). This risk, in combination with the characteristics of the land, logically imposes a reflection of risk in the land values, but this remains a relatively underexplored topic within existing real estate and urban economics literature. (For exceptions, see (Cheung, Wetherell, & Whitaker, 2018; Ikefuji, Laeven, Magnus, & Yue, 2022; Porter et al., 2004))

Nemetz & Dushnisky (1994), investigate potential losses in housing stock (among other losses) associated with a magnitude 9 earthquake in Vancouver. They conclude that between 28 and 93% of the housing supply could be lost in the case of a megathrust earthquake (Nemetz & Dushnisky, 1994). Yet, little is known about the effects of earthquakes on the real estate market in Vancouver. Research studying regions outside metro Vancouver indicate that earthquakes, and the risks of earthquakes, result in a negative discounting of property values (Keskin, Dunning, & Watkins, 2017; Thayer, Murdoch, & Singh, 1993).

Contrastingly, according to Kelly et al. (2020), this negative discounting of property values was countered by an increase in property insurance take-up rate. After the BC government announced that it will not cover any natural hazard damage claims, the amount spent on property insurance by BC residents increased. Kelly's (2020) theory suggests that earthquake risk is initially perceived as a threat, leading to a discount in property values as buyers and sellers adjust for the potential future costs and associated losses. As property owners try to mitigate the risk by taking up insurance, this reduces the perceived severity of the risk as the threat of earthquake damage becomes more manageable from a financial perspective (MacDonald, Murdoch, & White, 1987). This counteracts the negative discounting by stabilizing or even increasing property values. Exploring the extent to which these mechanisms influence property markets constitutes a meaningful contribution to the existing literature on this topic.

The literature also shows that the increased awareness of the risks of earthquakes among developers in Metro Vancouver has resulted in more anticipated urban development (Chang, Yip, & Tse, 2019). Chang et al. (2019) found that the increased exposure to natural hazard risk, including earthquakes, resulted in more sustainable development projects to reduce potential financial losses and casualties. An important influence in anticipated development are the geological surveys that were conducted in southwest British Columbia in the 1960s that mapped earthquake risks. These seismic hazard models were adopted in the National Building Code of Canada (Onur & Seemann, 2004). In combination with major improvements in earthquake engineering during the 1970s, almost all newly build properties after 1980 were structurally enforced to withstand smaller to medium sized earthquakes (Chang, Gregorian, Pathman, Yumagulova, & Tse, 2012; Onur & Seemann, 2004). Moreover, none of the scientific research on property values and earthquakes separated the effect on building value from those on land value, they only considered the overall property value.

Thus, the literature on the topic of earthquake risk and properties indicates that risk and risk-awareness are considered or are in someway included in property values. Although earlier research has been conducted on the possible consequences of earthquakes in the Vancouver area, the question on whether this is reflected in property values, and if so to what extent it is reflected, remains unclear.

1.3 Research problem statement

Although some research has been done on earthquake risk and the built environment in the Pacific Northwest, most of this focusses on the effect on the physical and social environment. Existing literature has seldom studied the relationship between real estate markets and earthquake risk in the Vancouver area. In fact, the most recent study on this topic has been written 30 years ago by Ventura & Schuster (1994), where they project the effect of a magnitude 7.0 earthquake on building structures and the economy in the Vancouver area (Ventura & Schuster, 1994).

This thesis tries to contribute through broadening the debate on earthquake risks and property values by providing a new geographical context for this research and to address a 30-year gap in the literature, using more recent data in the analysis.

This thesis will analyze the effects of distance to earthquakes and property values in Vancouver. Therefore, the main research question is:

To what extent is the distance to earthquakes related to property values in Vancouver?

To answer this question, it is broken down into three sub questions. First focusing on the distance to earthquakes, followed by analyzing if the effect differs between building age, and finally, touching on the ratio of building value to land value. The three sub questions that will be used to conclude an answer to the main research question are as follows:

- Is there a significant relation between earthquake location and property values?
- Does the relation of earthquake distance and property values differ between properties built before and after 1980?
- Is there a significant relation between earthquake distance and change in the ratio of building value to land value?

By answering these questions, I will be able to formulate an answer to the main research question. The aim of this research is to contribute to existing research, while also providing new insights on a subject that can have a large societal impact.

1.4 Outline

This thesis is structured in 5 sections. The first section is an introduction, followed by a comprehensive literature review in Section 2, where the theoretical framework is laid out on which the hypotheses of the thesis are built. Section 3 describes the approach of the data collection and selection, explains the descriptive statistics of the dataset, and sets out the assumptions and methodology of the Fixed Effects model. Section 4 includes the results and interpretation of the output. Finally, Section 5 concludes this study by answering the main research question and discussing any limitations of the research findings.

2 THEORY & HYPOTHESES

2.1 THE CANADIAN AND VANCOUVER PROPERTY MARKET

Over the last 10 years, after the 2007-2008 financial/real estate crisis, there has been a large influx of foreign capital investments into Canadian real estate markets (Rherrad, Bago, & Mokengoy, 2021). Unlike the U.S housing market, which was characterized by relatively lax mortgage regulations combined with a high degree of speculation, the Canadian mortgage market maintained a more conservative approach (Macgee, 2009, 2010). Mortgages in the US between 2000-2008 were often accessible with minimal scrutiny, as credit checks were light, and mortgage brokers were incentivized to maximize mortgage sales. In contrast, the Canadian mortgage market was more conservative, credit evaluations were more comprehensive, and interest rates were less flexible. This led to a significantly lower rate of foreclosures in Canada compared to the high foreclosure rates observed in the U.S (Macgee, 2009, 2010).

As a result, the Canadian real estate market recovered to pre-crash levels in about two years, whereas the US market took about a decade to recover to pre-crash values (Schembri, 2015). This is one of the main causes of the relatively high cost of housing in Canadian cities. While the market crash in the United States resulted in a major market correction that cancelled out a significant portion of the market speculation that caused the high property values, the Canadian market saw only a minor correction and property prices continued to rise (Schembri, 2015). The quick rebound of the market and the ongoing price increases have been driven by factors such as strong demand, limited housing supply, and ongoing market speculation. This contrasts with the U.S. market, where the correction significantly reduced speculative forces on property values (Schembri, 2015). The two global cities in Canada, Toronto, and Vancouver in particular, have seen a rapid incline of property values in the most recent years (Rherrad et al., 2021). From 2021 to 2022 average property prices increased 55% in the Vancouver area (Rherrad et al., 2021). This increase in property values does not correspond with a parallel increase in the average salary for Canadian households, as economic theory might suggest (Schembri, 2015). Canadian household income increased by 37% over the last decade, while property prices more than doubled (The

Measure of a Plan, 2024). It has become clear that most Canadians cannot afford purchasing a house (The Measure of a Plan, 2024). Over the last decades, real estate investment trusts and foreign investors, particularly from China, have taken over a large part of the real estate market by investing in residential properties to collect rents (Ley, 2018). This large entry of capital by these wealthy players has reduced the competitive power of individuals and households that are on the market for buying property.

2.2 EARTHQUAKE EFFECTS ON PROPERTY

Earthquakes can be classified as a natural hazard. Other natural hazards include, but are not limited to, floods, hurricanes, wildfires, tsunamis and extreme heat. The risk for disasters and (economic) losses caused by natural hazards tend to increase with population growth (Chang et al., 2012). Earthquakes occur randomly, and because they are very hard to predict, the threat of an earthquake can be considered a constant risk. Risk is often compensated by means of a 'premium'[2] (MacDonald et al., 1987). In the case of land and real estate properties many owners do not want to carry the risk associated with such an event occurring and therefore buy natural hazard or earthquake insurance (Kelly, Bowen, & McGillivray, 2020).

Buying earthquake insurance and decreasing property values go hand in hand. For one, buying insurance is a signal by buyers that they believe that the risk of an event, in this case a natural hazard, is substantial enough to purchase an insurance that will cover any potential losses (Brookshire, A, Tschirhart, & Schulze, 1985). The principle of risk aversion applies in this case. An asset that involves a higher level of risk is less attractive and its price will decrease (Brookshire et al., 1985). Through this mechanism of market sentiments, the perception of uncertainty grows and so does the take up of insurance (Akerlof, 1970). Because of the take up of insurance, inherently insurers push the value of the property down, since they hold the majority of the risk – as opposed to the property owners.

Additionally, natural hazard or earthquake insurance will take on a substantial portion of the (monthly) payment for the future property owner. This causes a budget constraint. If the insurance take-up rate

[2] A risk premium refers to the additional costs of an investment to compensate for risks associated by investing in a particular asset, compared to a risk-free investment such as government bonds.

increases when awareness rises, there are less funds available for buying the actual property, and the overall value of properties in the area will decrease (Kelly et al., 2020). Kelly et. al (2020) concludes that Canadians spend relatively more on earthquake insurance compared to Americans, partly because when Canadian property owners are told to buy insurance, they are more willing to purchase it, as opposed to US property owners (Kelly et al., 2020). By this logic, property values should be more prone to change in Canada after an earthquake has occurred, especially as the magnitude increases.

An examination of the US Geological Survey (2024) found that an earthquake with a magnitude lower than 2 on the Richter's scale cannot be felt by humans and does not inflict any damage on structures. Thousands of earthquakes with a magnitude higher than 2 occur in Southwestern BC and can be felt by humans (Government of Canada, 2024). The observations of earthquakes by people can influence awareness of earthquakes, and based on the earlier discussed theory, influence the property market. Earthquakes above a magnitude of 3 can cause structural damage to buildings, especially older buildings (U.S. Geological Survey, n.d.).

Contrary, a study by Palm (1981) concludes that when earthquake risks were disclosed to property buyers, it, in fact, has a negligible effect on the purchasing price. This is supported by a study by Hidano et.al (2015) where it was concluded that seismic hazard risk did not significantly affect real estate prices of newly constructed apartment buildings. Newly constructed buildings have updated building regulations compared to older constructions that consider possible earthquake or ground shaking (Chang et al., 2012, 2019; Onur & Seemann, 2004).

2.3 Building Values and Land Values

The value of a property is the total worth of a piece of real estate, combining two components: land value and building value (or improvement value). The land value refers to the value of the parcel or the plot of land itself. The building or improvement value refers to the added worth of any buildings or structures on top of the land. In this thesis, I will distinguish land value as the value of the plot of land and building value as the value of the structures, with property value being the combination of both.

The value of land is not only influenced by the price landowners are willing to spend on it, Ricardo's rent theory explains that scarcity of land, in combination with a strategic location and high productivity rate leads to an increase in rent, or land value (Lackman, 1976). As population increases, so does the demand for space and consequently land to build on. This is supported by Henry George's (1879) theory that claims, economic growth will create more capital and attract people. The combination of economic and population growth drives up land values (George, 1879). Therefore, as long as the economy keeps growing, so will land values.

Land values in practice rarely decrease in value, this cannot be said about the value of buildings (George, 1879; Lackman, 1976). A building is more prone to damage than land is, as it is a structure that is in constant maintenance to preserve its quality. Research done on property values before and after damage-inflicting earthquakes saw a difference in price shifts between the value of buildings and land values (Brookshire et al., 1985; Nakagawa et al., 2007; Yamamura, 2016). Building values seemed to be affected more by earthquakes than land values, and furthermore saw a large decrease in value, while land values stayed relatively constant (Nakagawa et al., 2007; Yamamura, 2016).

Thus, the theory suggests that although there is a theoretical role for earthquake risk in influencing property valuations, its practical significance is unclear. I will now begin to explore this problem empirically in subsequent sections.

2.4 HYPOTHESES

Based on the theory discussed in 1.2, 2.1 and 2.2 we can express hypotheses on the outcomes of the research sub-questions. Research sub-question 1 concerns the distance between real estate properties and earthquake locations. From the existing literature we can suppose that the effect of earthquakes decay over distance. This means that properties closer to an earthquakes' location will experience more ground movement than properties located farther away. Hence, locations close to earthquakes should have more price effects relative to properties farther away. The null – and alternative hypothesis for research sub-question 1, based on the literature review is formulated as follows:

i. There is no significant positive relation between the distance to earthquake locations and the value of a real estate property in Vancouver.
ii. There is a significant positive relation between the distance to earthquake locations and the value of a real estate property in Vancouver.

Research sub-question 2 dives deeper into the characteristics of the property and its construction quality. The stricter building codes created and implemented in the 1970s in British Columbia would have resulted in properties built with higher construction standards and earthquake resilience. Properties built before 1980 should in theory be exposed to a greater level of risk of earthquake damage than properties built in or after 1980 and should therefore be more impacted by the price effects of distance to earthquakes. The hypothesis for the second research sub-question considering the construction year of the properties can be expressed as:

i. There is no significant difference of the effect of distance to earthquake locations on property values between properties in Vancouver built before 1980 and in or after 1980.
ii. There is a significant difference of the effect of distance to earthquake locations on property values between properties in Vancouver built before 1980 and in or after 1980.

Research sub-question 3 is complementary to the first sub-question but seeks to isolate the specific effect of earthquakes on building value relative to the inherent value of the land it is located on. Theory suggests that land values are largely related to location and are less prone to earthquake damage, while buildings are more prone to earthquakes as it can inflict structural damage which can greatly impact the quality and value of the building. Considering this, the hypothesis for the third and last sub-question is as follows:

i. There is no significant relationship between the distance to an earthquake and a change in the ratio of building value to land value of properties in Vancouver.
ii. There is a significant relationship between the distance to an earthquake and a change in the ratio of building value to land value of properties in Vancouver.

These hypotheses will be tested using Fixed Effects models and a Chow test. In the concluding chapter the results of these models will be used to reject and accept the above hypotheses and formulate an answer to the main research question.

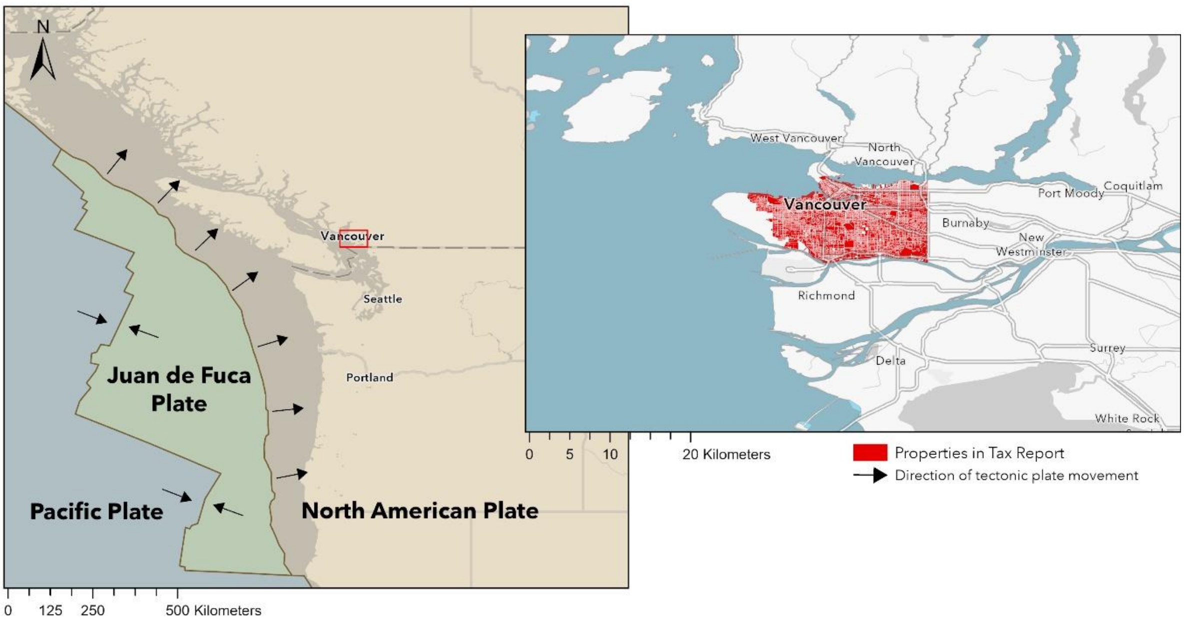
3 DATA & METHODS

3.1 STUDY CONTEXT

The study area concerns the city of Vancouver. Consistent open data covering the whole metro region of Vancouver (an urban region spanning 12 districts) is limited, hence this analysis only focuses on the city of Vancouver.

Vancouver is in the Southwest of British Columbia, about 25 km north of the US-Canada border, and about 175 km to Seattle. The city of Vancouver lies in the middle of the metro region Vancouver. The districts directly surrounding it are, West Vancouver, North Vancouver, Burnaby, and Richmond. It is located on the Burrard Peninsula. The area is surrounded by the English Bay and Vancouver Harbour on the North side and on the South size by the Fraser River. Figure 1 shows the location of the research area in the Pacific Northwest.

Figure 1. Map of tectonic plates in the Pacific Northwest, and Vancouver property parcels used in this analysis



The area of interest has a total of 100,021 property parcels that span an area of ~ 86 km². The district is characterized primarily by urban parcels. The most recent census conducted in 2021 concluded that a population of 662,248 people reside in the city of Vancouver (Government of Canada, 2022). This makes it the most populated city in the province of British Columbia and the most densely populated city in Canada. Surrounded by mountainous areas North and East however, Vancouver is relatively flat.

The region is characterized by a moderate oceanic climate, with the highest amount of rainfall per annum in Canada. Besides the rainfall the region is also characterized by a high number of earthquakes. Figure 1 shows the movement of the tectonic plates in the Pacific Northwest. The Juan de Fuca oceanic plate subducts underneath the lighter North American Plate, creating seismic activity. If we consider all recorded earthquakes in the region South-West British Columbia, over the years 2020 to 2023 a total of 2245 earthquakes were recorded (with differences in magnitude) (Government of Canada, 2023). In this analysis only earthquakes in a range of 75 kilometer from the city of Vancouver that can inflict structural damage to buildings are considered (ML > 2). Figure 2. shows the spatial distribution of these potentially damage inflicting earthquakes that occurred between 2013 and 2023. Between 2013 and 2023, 49 earthquakes were recorded from which 45 were of a magnitude between 2 and 3 on Richter's scale, 3 earthquakes were of a magnitude between 3 and 4, and 1 earthquake was recorded with a magnitude higher than 4 at 4.7. Table 1 shows the distribution of earthquakes per year.

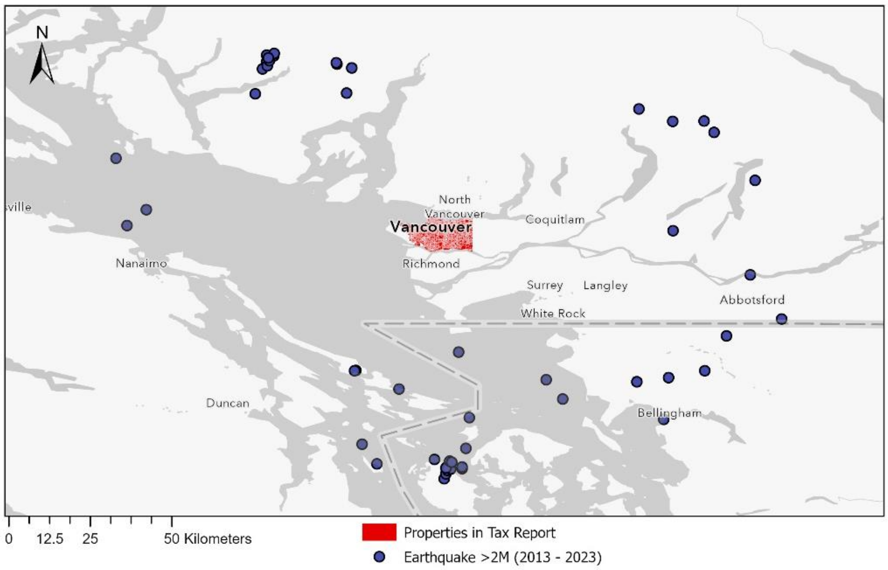
Table 1. Earthquake records per year by magnitude within a 75 radius of downtown Vancouver

MAGNITUDE	2023	2022	2021	2020	2019	2018	2017	2016	2015	2014	2013	Total
ML 2 to 3	4	5	4	3	3	2	6	7	5	4	2	45
ML 3 to 4			1					1	1			3
ML > 4									1			1
Total	4	5	4	3	3	2	6	7	5	4	2	**49**

Notes: ML corresponds to Magnitude on Richter's magnitude scale; records represent magnitude on surface level

In the period during which data was collected, no severe earthquakes (i.e., those with a magnitude of 5 or larger) have occurred in a range of 75 km from downtown Vancouver. There are only three years that have 'heavier' earthquakes recorded, 2015, 2016, and 2021. At the end of 2015 and the beginning of 2016 two earthquakes with a magnitude higher than 3 occurred in a span of 4 months. The start of 2015 saw an earthquake of a magnitude of 4.7, causing items to fall off walls and out of cupboards. However, no severe damage was reported (Government of Canada, 2023). In the years 2016 and 2017 most earthquakes were recorded, the years directly after were the least active years. This highlights not only the unpredictability of how often these earthquakes occur but also the volatility in their severity. This furthermore adds to the uncertainty surrounding future earthquakes and increases the sensitivity of risk assessments for earthquake damage. This sensitivity should, based on the theory, be reflected in the pricing of real estate properties. Table 6. in the Appendix shows a more detailed overview of the earthquakes used in this analysis.

Figure 2. Earthquake locations with ML >= 2 between 2013 and 2023



3.2 DATA COLLECTION

For this analysis, property tax information and assessed property values were collected at parcel level across the City of Vancouver area. Data on the number of earthquakes that occurred, alongside the location of the epicentre and magnitude, was also used. The data was collected for each year between 2013 and 2023 from 10 different datasets.

Property parcel information was collected from the Open Data Portal of the City of Vancouver. This portal has open access data on a wide range of locational data in the city of Vancouver. The dataset 'Property parcel polygons' contains all property parcels in the Vancouver municipal area (COV Open Data Portal, 2023a). There are a total of 100,021 unique parcels in the dataset. Each observation has data on location, size, shape, site id, and tax coordinate which is the first 8 digits of the BC assessment folio number. BC Assessment is the government body that collects and stores data on all properties in British Columbia, including assessed building value, assessed land value and tax assessments. By importing the Property parcel dataset into a Geographic Information System (GIS) it is possible to calculate the size of each property in m^2. All parcels that have an area of less than 50 m^2 are parcels with no building, therefore these parcels are excluded from the data (this deletes 177 parcels, leaving 99,874 for further analysis). The parcels used in this analysis are visualized in Figure 2.

The names, size, and location of the 22 Vancouver areas are found in the "Local area boundary" dataset in the Open Data Portal of the City of Vancouver (COV Open Data Portal, 2024). By geoprocessing each parcel polygon to the local area boundary, it is made possible to calculate the distance from each parcel to the downtown area.

Property value data was collected from the dataset "Property tax report" in the Open Data Portal of the city of Vancouver (COV Open Data Portal, 2023b). This dataset includes data of tax reports of each property per year, including land and improvement values. The data is collected by BC Assessment and is exported for each year between 2013 and 2023. Multiple tax reports can be filed in a year for an individual property, hence the number of observations in this dataset do not correspond to the property parcel data. The dataset contains 193,392 entries for the year 2013. The number of tax reports filed increases every year by about 2,500. The number of tax reports filed in 2023 count to 222,197 tax reports. For each tax report the dataset has information on Property ID, Land coordinate (which is the first 8 digits of BC Assessment's folio number), zoning district, zoning classification, street name and postal code, year of construction, current land value, tax assessment year, land value in previous year, year built, current improvement value and previous improvement value. The improvement value is the value of the additional improvements made to the plot of land; this is in almost all cases a building. In this analysis I use the variable improvement value as the building value. By combining the improvement value with the land value, it is possible to create a new variable property value.

The tax report dataset is imported to R where data cleaning is done (see Appendix 7.3.2 for full code). Missing values are removed, and outliers are excluded by winsorizing at 1%. By removing missing values, a significant portion of the residential properties were excluded, as most of the entries have missing values as a result of privacy standards. Residential properties are still included in the analysis as the category *Comprehensive Development* describes mixed-use parcels, where residential properties stand on the same parcel as commercial real estate. In cases with multiple tax reports entered in a single year, only the most recent entry is kept.[3] The data transformed from a long format to wide format with parcels in rows and variables classified by year. By exporting the dataset to a GIS software, it is possible to calculate the property value of the parcel per meter.

The earthquake data is collected from the Earthquake Database hosted by the Government of Canada. By filtering on year, magnitude of at least 2 ML, and within a span of 75 kilometers from downtown Vancouver, I find the number of potentially damage inflicting earthquakes in the area. The year, magnitude and location of each filtered earthquake exported to a GIS software and placed in the right location on the map, then, by using a Python script the distance to each earthquake is measured for each parcel. The Python script can be found in the Appendix.

After initial data collection and combining in ArcGIS, the dataset is exported to Stata for final data preparation and to run the empirical model. The data variables are renamed, and the variables are converted to fit the panel data format. The final data frame used consists of 76,605 parcels, with 50 variables.

3.3 DESCRIPTIVE ANALYSIS

In the next section a short summary and interpretation is given about the data that was collected and used in the analysis. It should be noted that some variables have less observations than the majority. In the case of 'Distance to water body' this is caused by the entrance of 0 as a missing value in the case of parcels that are directly on the shore. This is dealt with accordingly in Stata. For the variables of 'Building year' and 'Average income' there are missing values for the year 2013 that did not have any

[3] The dataset 'tax report' includes all tax reports in the City of Vancouver, tax reports under the same site id can be entered multiple times in a year when in that year the zoning classification changes, a development on the property has finished, or the property changed ownership (City of Vancouver, 2024).

data entries. The decision was made to still use the observations as these variables are only control variables.

Table 2. Summary of statistics

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
Average distance to earthquakes per year (m)	76,605	62,733	5,567	52,456	75,004
Property price per meter ($CAD)	76,605	10,935	20,782	1.006	952,712
Log of property price per meter	76,605	8.425	1.588	0.00643	13.77
Size (m²)	76,605	2,220	9,202	52.18	547,743
Box Cox of Size (m²)	76,605	2.061	0.0484	1.823	2.172
Ratio	76,605	1.16	60.47	0	9,080
Distance downtown (m)	76,605	2,536	2,399	0	9,330
Distance water body (m)	76,605	986.5	797.3	0	3,481
Average income	69,650	49,982	14,415	31,534	118,668
Average age	76,605	41.69	2.407	38.3	47.3
Building year	69,650	1964	30.67	1800	2023
Number of earthquakes	76,605	4.8	1.833	2	7
Consumer Price Index (CPI)	76,605	2.726	1.634	0.600	6.700

Notes: Ratio considers the ratio between building value and land value of a property parcel; Average distance to earthquakes per year considers the average distance to all earthquakes in that year

Table 2. presents a detailed statistical overview of the variables related used in the analysis on property prices, geographic characteristics, and social and economic statistics of the City of Vancouver. The sample size after data cleaning is 76,605 observations. The table provides the number of observations (N) per variable, their mean, standard deviation (sd), minimum (min), and maximum (max) values. The average distance to earthquakes per year is approximately 52,5 kilometers with a standard deviation of 5,567 meters, indicating that the distances are relatively concentrated around the mean, which can be explained by the fact that the properties are all relatively close by each other.

The property price per meter in Canadian dollars (CAD) shows a large variability, reflecting large differences in property prices, which range from as low as 1,006 CAD to a high of 952,712 CAD per square meter. This is in part due to the size of the parcel and its location; smaller plots usually have a higher price per meter than larger plots of land. The logarithm of property price per meter has a mean of 8,425 and is a more normalized distribution of property prices considering the log transformation. Property sizes also vary widely, with an average of 2,220 square meters but a standard deviation of 9,202 square meters, indicating that while many properties are around the mean size, some are significantly larger.

Moreover, Table 2. includes variables such as the distance of a property to downtown, and to the closest water bodies, with averages of 2,536 meters and 986.5 meters, respectively. The ratio variable, which represents the ratio between the lot value and the property structure value, has an average of 1.16, indicating that, the structure that is built on the lot is 1.16 times the value of the lot. It has a large standard deviation of 60.47, implying a large variability, with a maximum where the structure is 9,080 times the value of the lot.

Demographic factors per district, such as average income and age of residents, are also presented. The means of these factors at 49,982 CAD and 41.69 years, respectively, highlight a reasonably wealthy and middle-aged population considering income over time. The average building year is 1964, implying that most properties were built in the mid-20th century, with some as old as the 1800s. The table also notes the number of earthquakes and the Consumer Price Index (CPI) per year, which have means of 4.8 and 2.726, respectively, providing a broader context of environmental and economic conditions affecting the properties.

3.4 METHODOLOGY

This research aims to find a significant relationship between distance to earthquakes and property values in Vancouver over time. To determine a relationship, I utilize the collected dataset in a Fixed Effect Log-Linear Regression model using panel data. With this model, a linear relation between the dependent and independent variables can be analyzed over a period of time, using recurring observations (Allison, 2009). Panel data and Fixed Effects (FE) models offer several advantages over cross-sectional or time-series data alone. By tracking the same units over time, panel data can control for unobserved variables that vary across entities but are constant over time, reducing omitted variable bias (Allison, 2009). Similarly, Panel data allows the study of dynamic behaviors, such as how past outcomes affect current outcomes (e.g., lagged effects). In practice, FE models are most often used to control for all time-invariant characteristics (both observed and unobserved) of the individuals, groups, or entities in the panel data. This helps address omitted variable bias by accounting for these unchanging factors, and thereby also help to address endogeneity concerns (Allison, 2009). In this analysis the Fixed Effects model is used to control for meso- and macroeconomic conditions (e.g., Inflation rates, average household incomes, population numbers). Distance to downtown is included in the Fixed Effects models as, according to Ricardian theory, distance to downtown influences the price of the land. Distance to water bodies is included in the model to account for the effects of other natural hazards like tsunamis and floods.

This thesis aims to find the relation between earthquake distance and property values. The left-hand variable in the model equation should define an indication of property value. To achieve this, I use the logarithm of the variable property price per meter (ppm_{it}). The descriptive analysis concluded that the variable *ppm* was not normally distributed (see Figure 3. And 4. In Appendix 7.1.2 for distributions), hence we take the logarithm *ppm* as our dependent variable in our first model that will find an answer to the first research sub-question: *Is there a significant relation between earthquake distance and change in property values?* To answer this question, we use a Fixed Effects model using Year Fixed Effects and District Fixed effects, with the average distance to an earthquake in year *t* as the independent variable of interest, and property price per meter as the dependent variable. Control variables include property and location characteristics. As property prices change over time, we want to measure the differences between the years. Time and District Fixed Effects were included to capture locational effects, and account for trends over time. The FE equation for the model is as follows:

$$\ln(ppm_{it}) = \alpha + \beta_1\left(\frac{1}{Nt-1}\sum_{j=1}^{Nt-1} d_{ijt-1}\right) + \beta_2 sqm\ boxcox_i + \beta_3 building\ year_i + \beta_4 distance\ downtown_i + \beta_5 distance\ water_i + \beta_6 zoning\ classification_i + \nu_{t,i} + \epsilon_{it} \quad (1)$$

Where $\ln(ppm_{it})$ is the natural log of property price per meter in \$CAD for property *i* in year *t*, the constant is captured in α and the error term in ϵ, $\beta_1(\frac{1}{Nt-1}\sum_{j=1}^{Nt-1} d_{ijt-1})$ represents the estimator β for the average distance of the number of earthquakes *N* in year *t* – 1 from property *i* to earthquake *j*, [4] $sqm\ boxcox_i$ is the Box-Cox transformed size of property *i*, $building\ year_i$ is the categorical variable of the year that property *i* was built with groups built before and after 1980, $distance\ downtown_i$ is the distance in meters from property *i* to downtown Vancouver, $distance\ water_i$ is the distance in meters from property *i* to the body of water surrounding Vancouver, $zoning\ classification_i$ is the zoning classification of property *i*,, and $\nu_{t,i}$ represents the year-specific Fixed Effects and district-specific Fixed Effects capturing unobserved heterogeneity across years.

After data cleaning, the variable sqm (size of property in m^2) was not normally distributed. The standard transformations by taking the natural logarithm, winsorizing, and taking the square root of the variable all did not improve the distribution of the variable. Therefore, the variable is transformed with the optimal Box-Cox transformation as follows.

$$sqm_{boxcox_i} = \frac{sqm_i^{\lambda} - 1}{\lambda} \text{ where } \lambda = -0.4594055 \quad (2)$$

[4] I used a lagged variable for average distance to earthquake because housing price data is only available on an annual basis. As a result, any change in property values due to seismic events can only be observed in the next annual tax report, reflecting the delayed effect on housing prices.

This transformation improved the normality of the variable ‘size in square meter’ but still does not show a perfect normal distribution. A graph of the distribution of the variable can be found in Appendix 7.1.2 (Figure 5. and 6.).

To answer the second research sub-question: *Does the relation of earthquake distance and property values differ between properties built before and after 1980?,* a comparison between property building year is conducted. A distinction is made between buildings constructed before and after 1980, as this marks the adoption of new building codes that significantly improved earthquake resilience. To determine whether the relationship of average distance of a property to earthquakes in a year and the value of a property differs significantly between properties built before and in or after 1980, I performed a Chow-test. The Chow-test measures whether the coefficients in the two restricted regression models (Model (2) and (3) in Table 4.) are equal, and therefore showing structural stability or are susceptible to change and show significant differences. The Chow-test calculates an F-statistic that measures the improvement in fit of the model by specifying in groups built before and in or after 1980. In general, a large F-statistic indicates a strong observed difference between the restricted models. The calculated F-statistic based on the pooled and restricted models is then referenced to the critical F-statistic at a 0.01 significance level. To calculate the F-statistic in a Chow test I used the following equation

$$F - statistic = \frac{RSS - (RSS_1 + RSS_2)}{RSS_1 + RSS_2} \times \frac{T - 2k}{k} \tag{3}$$

To answer the last research sub-question, *Is there a significant relation between earthquake distance and change in the ratio of building value to land value?,* I use a similar model as the first research question but instead of price per meter as the dependent variable I use the variable *$ratio_{it}$* as the dependent variable. This variable represents the ratio of the building value to the land value of the parcel. The regression formula is as follows

$$ratio_{it} = \alpha + \beta_1\left(\frac{1}{Nt-1}\sum_{j=1}^{Nt-1} d_{ijt-1}\right) + \beta_3 sqm_{boxcox_i} + \beta_4 building\ year_i + \beta_5 distance\ downtown_i + \beta_6 distance\ water_i + \beta_7 zoning\ classification_i + \nu_{t,i} + \epsilon_{it} \tag{4}$$

Where *ratio*$_{it}$ is the ratio of structural property to land value of property *i* in year *t*.

To test the assumptions of the regression model, the function regcheck in Stata could unfortunately not be run with the specific model. Ergo, for multicollinearity a VIF test was ran. The VIF test and correlation matrix did not indicate any multicollinearity between the variables of interest. In addition to this, to test for autocorrelation a Woolridge test was performed in combination with a correlation matrix. The Woolridge test for both the main model and the ratio model resulted in no first-order autocorrelation. The output of the correlation matrix and the VIF scores can be found in the Appendix 7.2.

4 RESULTS & DISCUSSION

4.1 RESULTS

After data collection and preparation, I implement the theoretical models into a regression model. For this, I used Stata and created the following output. In Table 3. the output of 4 models concerning formula (1) is visible. Model (1) illustrates a simple regression between the average distance to earthquakes in the previous year as the independent variable, and a log of price per meter as the dependent variable. This model suggests no significant relationship between the main variable of interest and the dependent variable. To create a more robust model, additional control variables were included. This effort produced Model (2). The results of this model suggest a highly significant positive relationship between the main variable of interest and the dependent variable. To better exploit the panel structure of the data, Random-Effects (3) and Fixed-Effects (4) specifications were employed. Model (3) did not find a significant relationship, while Model (4) did. To elaborate which offered a better fit for the data, a Hausman test was used. This test suggests that the Fixed-Effects model was the best fit for the specified regression. The main model for this analysis is therefore Model (4).

Table 3. Main model

VARIABLES	(1) OLS	(2) OLS Pooled model	(3) Random Effects	(4) Fixed Effects
Lag of average distance to earthquake (meters)	5.17e-08	7.16e-06***	-1.10e-05	2.08e-06*
	(9.78e-07)	(9.50e-07)	(1.67e-05)	(1.62e-06)
Lag of number of earthquakes in year *N*		-0.00436*		
		(0.00251)		
Boxcox transformation of size (m2)		-13.21***	-13.23***	-13.22***
		(0.109)	(0.0781)	(0.433)
Dummy variable of year built				
Built after 1980		-0.560***	-0.560***	-0.560***
		(0.01000)	(0.0127)	(0.0303)
Distance to downtown (meters)		-0.000116***	-0.000120***	-0.000117***
		(7.94e-06)	(3.99e-06)	(2.24e-05)
Distance to body of water (meters)		0.000217***	0.000216***	0.000216***
		(9.62e-06)	(4.57e-06)	(3.00e-05)

Dummy variables of zoning classification				
Comprehensive development		-0.477***	-0.477***	-0.477***
		(0.0155)	(0.0223)	(0.0529)
Historical area		-0.266***	-0.267***	-0.266***
		(0.0278)	(0.0173)	(0.0926)
Industrial		0.0798***	0.0808	0.0798*
		(0.0154)	(0.0507)	(0.0421)
Limited agriculture		-0.302***	-0.304***	-0.302*
		(0.0531)	(0.0422)	(0.178)
Constant	8.422***	35.37***	36.86***	35.97***
	(0.0624)	(0.241)	(1.144)	(0.897)
Observations	76,605	76,605	76,605	76,605
R-squared	0.000	0.385	0.372	0.403
Year FE			YES	YES
Census district FE			YES	YES
Number of years			11	11

Notes: Dependent variable is Log of price per meter of property parcel (in CAD$). Lag of number of earthquakes in year *N* is captured in Year Fixed Effects in (3) and (4). Boxcox transformation of size used the value -0.4594055 for transformation. Excluded dummy variable of year built is *Built before* 1980. Excluded dummy variable of zoning classification is *Commercial.* Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

The relation of the main variable of interest (Lag of average distance to earthquake) significant at the 10% confidence level. The positive relation suggests that a one-meter increase in the average distance of a property to an earthquake, results in a 0.00000208 % increase of price per meter. This amounts to 0.00208% increase per km. This would mean that on a property valued at $500,000 CAD, the price per meter would increase by $10.40 per 1 km increase of average distance to earthquakes. A one standard deviation of average distance to earthquakes at 5567 meter will result in a 0.0116% increase in price per meter. This would mean a $5789.70 CAD increase for a property valued at $500,000. This implies that proximity to earthquakes is perceived negatively in terms of property value. This is in line with the theory that states that properties closer to earthquakes have a higher risk of damage by earthquakes and are therefore valued lower.

Next, to answer the second research sub-question on the differences between houses built before and in or after the year 1980, I performed a Chow test. Table 4. includes three models, Model (1) represents the pooled main model namely the Fixed-Effects model (4) in Table 3. including all observations in the data, Model (2) is the first unrestricted model only including observations of properties built before 1980, and Model (3) is the second unrestricted model only including observations of properties built in or after 1980. The formula below shows the equation used for the performed Chow F-Test on the models in Table 4.:

$$\frac{115339.23 - (17266.05 + 89960.41)}{17266.05 + 89960.41} \times \frac{76605 - 2 * 30}{30} \quad (4)$$

This equation resulted in a F-statistic of 193.05. By referencing the F-critical value table at a 0.01 significance level with the degrees of freedom of the model at df(30, 76545) we find a F-critical value of 1.70. As 193.05 is greater than 1.70, the Chow F-test therefore suggests that the estimated coefficients of models 2 and 3 differ significantly between the groups built before and in or after 1980.

Table 4. Chow test models

VARIABLES	(1) Pooled model	(2) Built before 1980	(3) Built after 1980
Lag of average distance to earthquake (meters)	2.08e-06*	-1.41e-06	1.84e-07
	(1.62e-06)	(1.56e-06)	(2.30e-06)
Boxcox transformation of size (m2)	-13.22***	-5.494***	-16.63***
	(0.433)	(0.394)	(0.678)
Dummy variable of year built			
Built after 1980	-0.560***		
	(0.0303)		
Distance to downtown (meters)	-0.000117***	-0.000151***	-0.000142***
	(2.24e-05)	(1.92e-05)	(3.74e-05)
Distance to waterbody (meters)	0.000216***	0.000128***	0.000214***
	(3.00e-05)	(2.66e-05)	(4.63e-05)
Dummy variables of zoning classification			
Comprehensive development	-0.477***	-0.343***	-0.531***
	(0.0529)	(0.0569)	(0.0736)
Historical area	-0.266***	-0.492***	-0.418**
	(0.0926)	(0.0890)	(0.181)
Industrial	0.0798*	-0.177***	0.0495
	(0.0421)	(0.0347)	(0.0736)
Limited agriculture	-0.302*	-2.279***	0.391*
	(0.178)	(0.333)	(0.209)
Constant	35.97***	20.92***	42.58***
	(0.897)	(0.808)	(1.436)
Observations	76,605	36,540	40,065
R-squared	0.403	0.515	0.310
Year FE	YES	YES	YES
Census district FE	YES	YES	YES

Notes: Dependent variable is Log of price per meter of property parcel (in CAD$). Boxcox transformation of size used the value -0.4594055 for transformation. Excluded dummy variable of zoning classification is *Commercial.* Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

The relationship of the main variable of interest (Lag of average distance to an earthquake) is in both unrestricted models not significant. This implied that although the relation between the years and the price per meter is different, only the combination of the two categories has a large enough statistical power to capture the relationship.

To answer the third research sub-question, instead of the price per meter as the dependent variable, I take the ratio of building value and land value and perform the same models as in table 3, a simple regression, a Pooled model, a Random Effects model and Fixed Effects model. Table 5. displays the results of these tests. To elaborate on which model offered a better fit for the data, a Hausman test was used. This test suggests that the Fixed-Effects model was the best fit for the specified regression. The main model for the third research sub-question is therefore Ratio model (4).

Table 5. Ratio models

VARIABLES	(1) OLS	(2) OLS Pooled model	(3) Random effects	(4) Fixed effects
Lag of average distance to earthquake (meters)	4.00e-05	3.77e-05	3.77e-05	7.09e-05
	(3.72e-05)	(3.76e-05)	(4.16e-05)	(0.000175)
Boxcox transformation of size (m2)		15.40***	15.40***	15.42**
		(5.308)	(3.756)	(7.531)
Dummy variable of year built				
Built after 1980		1.220**	1.220**	1.220
		(0.485)	(0.475)	(1.036)
Distance downtown (meters)		0.000835**	0.000835***	0.000843
		(0.000385)	(0.000226)	(0.000831)
Distance to waterbody (meters)		0.00140***	0.00140***	0.00141
		(0.000466)	(0.000433)	(0.00182)
Dummy variables of zoning classification				
Comprehensive development		-0.825	-0.825*	-0.826
		(0.753)	(0.499)	(1.181)
Historical area		-2.245*	-2.245**	-2.243
		(1.347)	(0.986)	(1.454)
Industrial		1.035	1.035*	1.034
		(0.748)	(0.547)	(1.792)
Limited agriculture		9.192***	9.192***	9.196
		(2.575)	(0.817)	(5.929)
Constant	-1.383	-41.12***	-41.12***	-43.29*
	(2.375)	(11.42)	(10.51)	(24.40)
Observations	76,605	76,605	76,605	76,605
R-squared	0.000	0.002		0.002
Number of year			11	

Notes: Dependent variable is ratio of building value to land value. Boxcox transformation of size used the value -0.4594055 for transformation. Excluded dummy variable of year built is *Built before* 1980. Excluded dummy variable of zoning classification is *Commercial.* Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

When looking at the output of both models, in which the ratio of building value and land value is exchanged for price per meter as the dependent variable, I conclude that the relation is not significant. The variance of ratio cannot significantly be linked to the average distance to earthquakes. This also must be concluded for all other control variables except for the size of the property, which is positively significant. Possible reasons for this insignificance can be that the specified model is not optimal to measure the relationship, or the relation is not linear.

4.2 DISCUSSION

The outcomes from the three models show interesting results. In the first Fixed Effects model, there is a small positive significant relationship. This suggests that properties located further away from earthquakes experience an increased price per meter. The results supports the assumption that properties closer to earthquakes would have lower property values due to the increased risk of damage. However, the results show a significant relation at a 10% confidence interval, highlighting that the variable of interest should be interpreted with caution. The analysis found an outcome, whereby the null hypothesis and for research sub-question 1 is rejected. The answer for research sub-question 1 is: *There is a significant positive relation between the distance to earthquake locations and the value of a real estate property in Vancouver.*

The Chow test considered if the relationship between earthquake distance and property value differed for houses built before and in or after 1980. The Chow test found that properties built before 1980 showed a significantly different relationship between distance and property value compared to those built in or after 1980. This difference might be because of changes in building standards, suggesting that older properties are more sensitive to earthquake damage. The analysis found a significant difference between the two groups. The null hypothesis for research sub-question 2 is therefore rejected and the alternative hypothesis is accepted; namely *There is a significance difference of the effect of distance to earthquake locations on property values between properties in Vancouver built before 1980 and in or after 1980*.

The ratio model did not find a significant effect, indicating that the ratio of building value to land value is not significantly influenced by earthquake proximity. The lack of significance in this model can be caused by several factors. First, the model used might not be optimal for capturing the effect. Second, the relationship may not be a linear relationship. Or finally, there is no effect. So, the null hypothesis cannot be rejected for the last research sub-question and the concluding answer for this question is: *There is no significant relationship between the distance to an earthquake and a change in the ratio of building value to land value of properties in Vancouver.*

Although the model found a significant positive relationship the effect is considerably small. This can have multiple causes. One possible explanation for this could be that buyers and real estate valuers, may not be informed or educated well enough about the risks of earthquakes. This lack of information could lead to an underestimation of risks and therefore do not, or only slightly incorporate in into the property value. On the other hand, the risk of earthquakes may already be priced into the market due to insurance take-up. Another explanation is that the effect of earthquake proximity on property values might be small because the earthquakes measured were not big enough to influence property prices significantly. The magnitudes of the earthquakes between 2013-2023 may not have been strong enough, or the observed earthquakes are located too far away and lost too much energy over distance to have a more influential relation. Lastly, the measure of average distance to earthquake might not be the ideal measurement to assess the relation between distance and property value. A property that was very close to one earthquake but far away to another can have the same measurement of distance as a property that has not been close to any earthquake. This can distort the outcome of the results.

5 CONCLUSION & LIMITATIONS

5.1 CONCLUSION

This thesis aimed to investigate the relationship between proximity to earthquakes and property values in the city of Vancouver, using a Fixed Effect Log-Linear Regression model with panel data. By analyzing a dataset of property tax reports over time, the study controlled for several time-varying factors including inflation rates and average household incomes. This is done to isolate the effect of earthquake distance on property prices. The primary research question asked:

To what extent is the distance to earthquakes related to property values in Vancouver?

This thesis also looked at whether this relationship varies between properties built before and in or after 1980, as well as if the effect differs between building value and land value.

The findings revealed a significant outcome, namely that properties located farther from earthquakes experienced a slight increase in price per square meter. This supports the theory that closer proximity to earthquake zones would reduce property values due to higher risk (Hidano et al., 2015; Porter et al., 2004). Additionally, this study did find a significant difference between buildings built before and after 1980, but neither of these groups has a significant relationship with property values when isolated. Furthermore, the analysis on the relationship between distance to earthquakes and the ratio between building value and land value did not find a significant relationship. Indicating that seismic activity does not have a stronger relationship with building value than land value.

The implications of these findings can be significant for the residents of Vancouver, professionals in real estate, and policymakers. For the residents of Vancouver, the results highlight the need for better public awareness and increased understanding of earthquake risks. This is could also become applicable for regions beyond the Vancouver metropolitan area with a history of seismic activity. Real estate professionals, including appraisers and developers, may need to reconsider how they assess property values in relation to earthquake proximity. Finally, policymakers should potentially also consider these findings when developing zoning regulations and building codes, by making sure that they reflect on the actual risk of earthquakes and to encourage more durable construction practices.

In conclusion, while the relationship between earthquake distance and property values in Vancouver is statistically significant, it is complex and influenced by a range of factors beyond simple proximity.

5.2 LIMITATIONS AND FUTURE RESEARCH

This study, while providing valuable insights into the relationship between earthquake proximity and property values in Vancouver, has notable limitations that must be acknowledged.

One limitation is the under representation of residential properties after data cleaning. The data cleaning step where I excluded missing values, deleted a large part of the residential properties. Residential properties are still taken into account in the analysis but is underrepresented. This limits the generalizability of the findings, as residential properties represent a large segment of the real estate market. The analysis, therefore, primarily reflects the dynamics of commercial or non-residential properties, which may behave differently from residential properties in response to earthquake risks.

Additionally, in must be acknowledged that the data used is a combination of multiple sets of secondary data. Since the data is secondary, I cannot guarantee its accuracy or quality, the datasets used might be incomplete or have a bias.

Furthermore, this analysis has not taken into account the differences in magnitudes of earthquakes. By only using earthquakes above a certain threshold (>2 ML) that can inflict damage, I have tried to incorporate some factor of magnitude into the analysis, but this does not measure the full effect of earthquake risk. It must be acknowledged that the magnitude of earthquakes differs and significantly depreciates over distance. The risk-perception and damage infliction are therefore different for each earthquake. The relationship between earthquake magnitude and property prices is a variable that I did not adopt in this study, this can introduce bias into the data as each earthquake is inherently weighted identically, regardless of magnitude. Incorporating the magnitude of earthquakes is a highly relevant topic for future research.

Future research should also address other limitations by including residential properties into the analysis in order to provide a more comprehensive understanding of how earthquake proximity affects different types of real estate. Additionally, studying larger datasets or extending the analysis to Metro Vancouver and regions with more frequent severe earthquakes could help determine whether the measured effect size in this thesis is context-specific or a broader trend. Finally, future research could investigate the role of public awareness and risk perception of earthquakes in explaining property values and find a more in-depth understanding of real estate markets in seismically active regions.

6 REFERENCES

Akerlof, G. A. (1970). The Market for “Lemons”: Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, *84*(3), 488–500. https://doi.org/10.2307/1879431

Allison, P. D. (2009). *Fixed Effects Regression Models*. SAGE Publications.

Brookshire, D. S., A, T. M., Tschirhart, J., & Schulze, W. D. (1985). A Test of the Expected Utility Model: Evidence from Earthquake Risks. *Journal of Political Economy*, *93*(2), 369–389.

Carey, C. (2024, April 12). Vancouver renters spend 60% of monthly income on housing. Retrieved August 28, 2024, from CityNews Vancouver website: https://vancouver.citynews.ca/2024/04/12/metro-vancouver-renters-60-income/

Chang, S. E., Gregorian, M., Pathman, K., Yumagulova, L., & Tse, W. (2012). Urban Growth and Long-Term Changes in Natural Hazard Risk. *Environment and Planning A: Economy and Space*, *44*(4), 989–1008. https://doi.org/10.1068/a43614

Chang, S. E., Yip, J. Z. K., & Tse, W. (2019). Effects of urban development on future multi-hazard risk: The case of Vancouver, Canada. *Natural Hazards*, *98*(1), 251–265. https://doi.org/10.1007/s11069-018-3510-x

Cherniawsky, J., Titov, V., Wang, K., & Li, J.-Y. (2007). Numerical Simulations of Tsunami Waves and Currents for Southern Vancouver Island from a Cascadia Megathrust Earthquake. *Pure and Applied Geophysics*, *164*, 465–492. https://doi.org/10.1007/s00024-006-0169-0

Cheung, R., Wetherell, D., & Whitaker, S. (2018). Induced earthquakes and housing markets: Evidence from Oklahoma. *Regional Science and Urban Economics*, *69*, 153–166. https://doi.org/10.1016/j.regsciurbeco.2018.01.004

Clarke, D. M. (2013). *Earthquake Risk and Housing Prices: A Hedonic Analysis of the Victoria Real Estate Market*. University of Victoria, Victoria, Canada.

COV Open Data Portal. (2023a). Property parcel polygons. Retrieved May 31, 2024, from https://opendata.vancouver.ca/explore/dataset/property-parcel-polygons/export/?disjunctive.streetname&sort=-civic_number&location=12,49.25681,-123.12645

COV Open Data Portal. (2023b). Property tax report. Retrieved December 9, 2023, from https://opendata.vancouver.ca/explore/dataset/property-tax-report/

COV Open Data Portal. (2024). Local area boundary. Retrieved May 31, 2024, from https://opendata.vancouver.ca/explore/dataset/local-area-boundary/export/?disjunctive.name

George, H. (1879). Progress and Poverty. Retrieved October 13, 2024, from OLL Liberty Fund website: https://oll.libertyfund.org/titles/george-progress-and-poverty

Gold, K. (2017, March 24). As Seattle market mirrors Vancouver's, housing experts search for answers. *The Globe and Mail*. Retrieved from https://www.theglobeandmail.com/real-estate/vancouver/as-seattle-market-mirrors-vancouvers-housing-experts-search-for-answers/article34418725/

Government of Canada. (2023). Earthquake Reports for 2022 [Database]. Retrieved December 9, 2023, from https://earthquakescanada.nrcan.gc.ca/recent/2022/index-en.php#wb-auto-5

Government of Canada. (2024). Questions and Answers on Megathrust Earthquakes. Retrieved December 9, 2023, from Earthquakes Canada website: https://earthquakescanada.nrcan.gc.ca/zones/cascadia/qa-en.php

Government of Canada, S. C. (2022, February 9). Profile table, Census Profile, 2021 Census of Population—Vancouver, City (CY) [Census subdivision], British Columbia. Retrieved May 31, 2024, from https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/index.cfm?Lang=E

Hidano, N., Hoshino, T., & Sugiura, A. (2015). The effect of seismic hazard risk information on property prices: Evidence from a spatial regression discontinuity design. *Regional Science and Urban Economics*, *53*, 113–122. https://doi.org/10.1016/j.regsciurbeco.2015.05.005

Ikefuji, M., Laeven, R. J. A., Magnus, J. R., & Yue, Y. (2022). Earthquake Risk Embedded in Property Prices: Evidence From Five Japanese Cities. *Journal of the American Statistical Association*, *117*(537), 82–93. https://doi.org/10.1080/01621459.2021.1928512

Kelly, M., Bowen, S. G., & McGillivray, R. G. (2020). The Earthquake Insurance Protection Gap: A Tale of Two Countries. *Journal of Insurance Regulation*, 1–37. https://doi.org/10.52227/23843.2021

Keskin, B., Dunning, R., & Watkins, C. (2017). Modelling the impact of earthquake activity on real estate values: A multi-level approach. *Journal of European Real Estate Research*, *10*(1), 73–90. https://doi.org/10.1108/JERER-03-2016-0014

Lackman, C. L. (1976). The Classical Base of Modern Rent Theory. *The American Journal of Economics and Sociology*, *35*(3), 287–300.

Ley, D. (2018). Global China and the making of Vancouver's residential property market. In *The Globalisation of Real Estate*. Routledge.

MacDonald, D. N., Murdoch, J. C., & White, H. L. (1987). Uncertain Hazards, Insurance, and Consumer Choice: Evidence from Housing Markets. *Land Economics*, *63*(4), 361–371. https://doi.org/10.2307/3146293

Macgee, J. (2009). Why didn't Canada's housing market go bust? *Economic Commentary*.

Macgee, J. (2010). *Not Here? Housing Market Policy and the Risk of a Housing Bust*.

Molnar, S., Assaf, J., Sirohey, A., & Adhikari, S. R. (2020). Overview of local site effects and seismic microzonation mapping in Metropolitan Vancouver, British Columbia, Canada. *Engineering Geology*, *270*, 105568. https://doi.org/10.1016/j.enggeo.2020.105568

Nakagawa, M., Saito, M., Yamaga, H., 中川雅之, 齊藤誠, & 山鹿久木. (2007). Earthquake Risks and Land Prices: Evidence from the Tokyo Metropolitan Area. *Japanese Economic Review*, *60*. https://doi.org/10.1111/j.1468-5876.2008.00438.x

Natural Resources Canada. (2011). *Earthquakes in southwestern British Columbia*. Geofacts.

Nemetz, P. N., & Dushnisky, K. (1994). Estimating Potential Capital Losses from Large Earthquakes. *Urban Studies*, *31*(1), 99–121. https://doi.org/10.1080/00420989420080061

Onur, T., & Seemann, M. R. (2004). *Probabilities of significant earthquake shaking in communities across British Columbia: Implications for emergency management*. World Conference on Earthquake Engineering. https://doi.org/10.4095/215329

Palm, R. I. (1981). PUBLIC RESPONSE TO EARTHQUAKE HAZARD INFORMATION*. *Annals of the Association of American Geographers*, *71*(3), 389–399. https://doi.org/10.1111/j.1467-8306.1981.tb01364.x

Porter, K. A., Beck, J. L., Shaikhutdinov, R. V., Au, S. K., Mizukoshi, K., Miyamura, M., … Masuda, M. (2004). Effect of Seismic Risk on Lifetime Property Value. *Earthquake Spectra*, *20*(4), 1211–1237. https://doi.org/10.1193/1.1810536

Rherrad, I., Bago, J.-L., & Mokengoy, M. (2021). Real estate bubbles and contagion: New empirical evidence from Canada. *New Zealand Economic Papers*, *55*(1), 38–51. https://doi.org/10.1080/00779954.2020.1791940

Schembri, L. (2015). *The Long-Term Evolution of House Prices: An International Perspective*. Bank of Canada.

Schulz, K. (2015, July 13). The Earthquake That Will Devastate the Pacific Northwest. *The New Yorker*. Retrieved from https://www.newyorker.com/magazine/2015/07/20/the-really-big-one

Shepert, E. (2022, June 17). Vancouver ranks 3rd most expensive city in North America. Retrieved December 9, 2023, from Vancouver Is Awesome website: https://www.vancouverisawesome.com/local-news/vancouver-ranks-3rd-most-expensive-city-in-north-america-5490661

Thayer, M., Murdoch, J., & Singh, H. (1993). The Impact of Natural Hazards on Housing Values: The Loma Prieta Earthquake. *Real Estate Economics*, *21*, 167–184. https://doi.org/10.1111/1540-6229.00606

The Measure of a Plan. (2024, April 27). Tracking Canadian Housing Market Affordability (1999 to 2024). Retrieved August 28, 2024, from The Measure of a Plan website: https://themeasureofaplan.com/canadian-housing-affordability/

U.S. Geological Survey. (n.d.). Earthquake Magnitude, Energy Release, and Shaking Intensity. Retrieved October 13, 2024, from https://www.usgs.gov/programs/earthquake-hazards/earthquake-magnitude-energy-release-and-shaking-intensity

Ventura, C. E., & Schuster, N. D. (1994). Seismic Risk and Hazard Reduction in Vancouver, British Columbia. In B. E. Tucker, M. Erdik, & C. N. Hwang (Eds.), *Issues in Urban Earthquake Risk* (pp. 221–236). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-015-8338-1_14

Yamamura, E. (2016). Natural disasters and social capital formation: The impact of the Great Hanshin-Awaji earthquake. *Papers in Regional Science*, *95*, S143–S165. https://doi.org/10.1111/pirs.12121

7 APPENDIX

7.1 FIGURES & TABLES

7.1.1 Recorded earthquakes

Table 6. Recorded earthquakes 2013 – 2023 in span of 75 in Vancouver

Date/Time (UTC)	Latitude (°)	Longitude (°)	Depth (km)	Magnitude	Description
27/07/2023 15:58	48.794	-122.775	22.1	2.4 ML	35 km SE of Tsawwassen, BC
03/03/2023 3:53	48.602	-123.052	18.2	2.2 ML	26 km E of Sidney, BC
01/03/2023 11:35	48.607	-123.052	15.6	2.5 ML	26 km E of Sidney, BC
02/01/2023 18:51	49.529	-122.357	10.0	2.1 ML	48 km NW of Chilliwack, BC
05/12/2022 11:42	49.316	-123.923	63.0	2.4 ML	15 km N of Nanaimo, BC
08/09/2022 6:43	49.559	-122.471	2.7	2.1 ML	51 km E of Squamish, BC
19/08/2022 21:18	48.658	-123.042	22.5	2.1 ML	26 km E of Sidney, BC
05/04/2022 0:16	48.923	-123.062	20.4	2.5 ML	10 km SE of Tsawwassen, BC
12/02/2022 9:09	49.635	-123.623	6.7	2.2 ML	19 km NE of Sechelt, BC
22/12/2021 2:21	48.873	-123.345	22.8	2.3 ML	21 km SW of Tsawwassen, BC
17/12/2021 12:13	48.872	-123.349	22.7	3.8 ML	21 km SW of Tsawwassen, BC
19/10/2021 14:56	49.704	-123.603	1.5	2.6 ML	27 km NE of Sechelt, BC
20/07/2021 10:49	48.853	-122.483	31.4	2.3 ML	28 km SW of Abbotsford, BC
16/06/2021 16:25	49.258	-122.470	2.5	2.1 ML	21 km NE of Langley, BC
24/03/2021 19:57	48.669	-123.328	20.9	2.4 ML	5 km E of Sidney, BC
07/12/2020 17:11	48.872	-122.383	27.2	2.0 ML	23 km SW of Abbotsford, BC
15/05/2020 8:15	48.574	-123.102	15.2	2.8 ML	23 km E of Sidney, BC
08/01/2020 17:37	49.740	-123.571	5.2	2.4 ML	31 km W of Squamish, BC
31/12/2019 17:13	48.588	-123.097	13.8	2.2 ML	24 km E of Sidney, BC
05/04/2019 20:04	48.597	-123.092	15.7	2.0 ML	23 km E of Sidney, BC
24/03/2019 1:52	48.968	-122.323	9.7	2.0 ML	12 km SW of Abbotsford, BC
08/06/2018 17:09	48.822	-123.226	60.8	2.8 ML	21 km SSW of Tsawwassen, BC
04/02/2018 6:57	48.628	-123.128	12.2	2.8 ML	20 km E of Sidney, BC
18/10/2017 17:59	49.707	-123.356	0.0	2.0 ML	13 km W of Squamish, BC

06/02/2017 3:22	48.842	-122.570	2.4	2.1 ML	29 km SSE of Langley, BC
01/02/2017 20:34	48.601	-123.084	14.4	2.5 ML	23 km E of Sidney, BC
30/01/2017 11:03	48.606	-123.096	16.6	2.7 ML	22 km E of Sidney, BC
30/01/2017 2:27	48.604	-123.097	16.4	2.4 ML	22 km E of Sidney, BC
24/01/2017 13:44	48.743	-123.032	18.4	2.1 ML	28 km ENE of Sidney, BC
22/10/2016 15:26	49.594	-122.564	0.0	2.4 ML	43 km NNW of Mission, BC
11/10/2016 8:49	49.637	-123.371	13.1	2.3 ML	16 km WSW of Squamish, BC
09/10/2016 0:47	48.738	-122.496	19.3	2.1 ML	38 km SW of Abbotsford, BC
17/07/2016 0:42	49.560	-122.385	1.0	2.9 ML	36 km N of Mission, BC
08/05/2016 4:41	49.397	-122.244	1.0	2.3 ML	19 km NNE of Mission, BC
03/05/2016 16:13	49.724	-123.591	0.7	2.4 ML	30 km NNE of Sechelt, BC
05/04/2016 14:45	49.742	-123.591	3.0	2.0 ML	30 km W of Squamish, BC
04/04/2016 18:36	49.712	-123.589	5.0	3.1 ML	29 km NNE of Sechelt, BC
30/12/2015 7:39	48.615	-123.287	57.5	4.7 Mw	9 km ESE of Sidney, BC
21/09/2015 7:35	49.747	-123.570	1.0	2.0 ML	29 km W of Squamish, BC
29/08/2015 11:39	49.136	-122.257	4.3	2.1 ML	9 km NNE of Abbotsford, BC
27/08/2015 2:20	49.272	-123.976	62.0	2.1 ML	12 km NNW of Nanaimo, BC
05/07/2015 7:45	49.015	-122.171	7.7	2.7 ML	9 km ESE of Abbotsford, BC
23/02/2015 18:21	49.727	-123.584	3.5	2.0 ML	30 km W of Squamish, BC
15/02/2015 4:12	49.734	-123.586	3.0	3.4 ML	30 km W of Squamish, BC
11/12/2014 14:53	49.458	-124.007	65.8	2.8 ML	18 km W of Sechelt, BC
24/10/2014 10:02	48.847	-122.820	19.2	2.1 ML	28 km ESE of Tsawwassen, BC
19/02/2014 7:23	49.717	-123.397	1.0	2.3 ML	16 km W of Squamish, BC
18/02/2014 23:54	49.721	-123.399	1.0	2.3 ML	16 km W of Squamish, BC
21/06/2013 5:18	48.623	-123.086	13.5	2.2 ML	23 km E of Sidney, BC
21/06/2013 4:35	48.620	-123.081	15.1	2.4 ML	23 km E of Sidney, BC

Source: Government of Canada (2024)

7.1.2 Distribution histograms

Distribution of price per meter and natural logarithm of price per meter

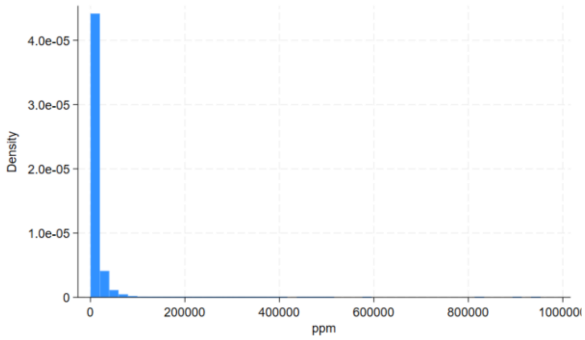


Figure 3. Distribution of price per meter

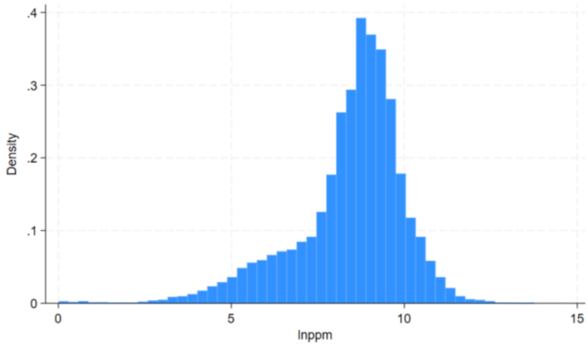


Figure 4. Distribution of the natural logarithm of price per meter

Distribution of size in m² and Box Cox distribution of size in m²

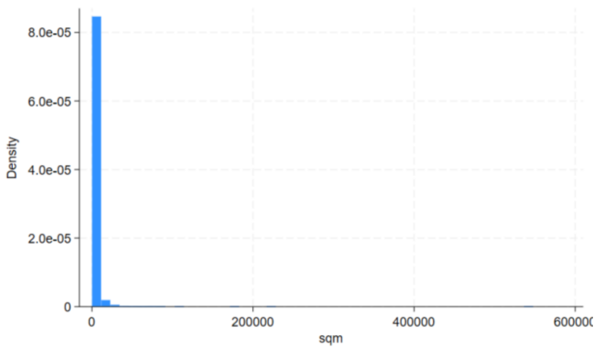


Figure 5. Distribution of size in m²

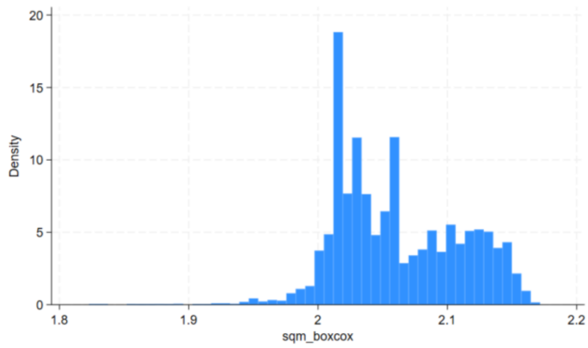


Figure 6. Distribution of the Box Cox distribution of size in m²

7.2 Assumptions

7.2.1 Correlation matrix

	lnppm	lagged..	sqm_bo~x	year_b~t	dist_d~n	dist_w~r	zoning~m
lnppm	1.0000						
lagged_avg~q	0.0002	1.0000					
sqm_boxcox	-0.5182	-0.0280	1.0000				
year_built~t	-0.3750	-0.0197	0.3795	1.0000			
dist_downt~n	-0.2693	-0.1291	0.1681	0.1369	1.0000		
dist_water	-0.0150	-0.0685	-0.0776	-0.0324	0.4880	1.0000	
zoning_cla~m	-0.1021	0.0089	0.1704	0.0199	-0.0530	-0.3719	1.0000

7.2.2 VIF output

Variable	VIF	1/VIF
lagged_avg~q	1.02	0.983185
sqm_boxcox	1.31	0.760502
2.year_bui~t	1.21	0.825961
dist_downt~n	1.56	0.642540
dist_water	1.61	0.622656
zoning_cla~m		
2	1.41	0.710454
3	1.15	0.869118
4	1.41	0.711251
5	1.15	0.868860
Mean VIF	1.31	

7.3 DATA MANIPULATION STEPS

7.3.1 Step by step description

This is a short description of the data manipulation steps used for this thesis, with the corresponding R and Stata code.

1. Get property tax data from:
2. https://opendata.vancouver.ca/explore/dataset/property-tax-report/information/
3. Get property parcel polygons in geojson. file from: https://opendata.vancouver.ca/explore/dataset/property-parcel-polygons/export/?disjunctive.streetname&sort=-civic_number&location=12,49.25681,-123.12645
4. Load parcels to ArcGIS Pro with geoprocessing tool "JSON to Feature"
5. Calculate m2 of each polygon with geoprocessing tool "Calculate Geometry Attributes"
6. Delete all polygons that have m2 < 50 (this deletes 177 polygons – 99847 left). This is because they are not parcels but small areas that do not have a construction built in that area.
7. Create OBJECTID variable
8. Export to R with option Export table
9. Run R file (see R code below).
10. Add CSV file from R to map and use join by objectid with parcels to place on map
11. Get local areas from https://opendata.vancouver.ca/explore/dataset/local-area-boundary/export/?disjunctive.name
12. Join local areas with parcels on map with spatial join
13. Place earthquake data on map by longitude and latitude
14. Save property parcels as shapefile and add to map again, otherwise next step is not possible
15. Run the Python file below in ArcGIS Pro to calculate and create new columns for distances to earthquakes
16. Use the Geoprocessing tool Near to calculate distances to downtown
17. Export data as CSV, and run the Stata file (see below)

7.3.2 R code

```
library(tidyverse)
library(tidyr)
library(fuzzyjoin)

# load Vancouver tax report dataset and the properties dataset
taxreport20_24 <- read.csv("../../data/property_tax_report.csv", sep = ";", header = T)
taxreport16_19 <- read.csv("../../data/property_tax_report16_19.csv", sep = ";", header = T)
taxreport11_15 <- read.csv("../../data/property_tax_report11_15.csv", sep = ";", header = T)
properties <- read.csv("../../data/property_parcels.csv", sep = ",", header = T)

# rename coordinates
taxreport20_24 <- rename(taxreport20_24, "coordinate" = "LAND_COORDINATE")
taxreport16_19 <- rename(taxreport16_19, "coordinate" = "LAND_COORDINATE")
taxreport11_15 <- rename(taxreport11_15, "coordinate" = "LAND_COORDINATE")
properties <- rename(properties, "coordinate" = "tax_coord")

# exclude unused variables from table
taxreport20_24 <- taxreport20_24 %>%
  select(PID, coordinate, ZONING_CLASSIFICATION, CURRENT_LAND_VALUE,
CURRENT_IMPROVEMENT_VALUE,
         PREVIOUS_LAND_VALUE, PREVIOUS_IMPROVEMENT_VALUE, YEAR_BUILT, REPORT_YEAR)
taxreport16_19 <- taxreport16_19 %>%
  select(PID, coordinate, ZONING_CLASSIFICATION, CURRENT_LAND_VALUE,
CURRENT_IMPROVEMENT_VALUE,
         PREVIOUS_LAND_VALUE, PREVIOUS_IMPROVEMENT_VALUE, YEAR_BUILT, REPORT_YEAR)
taxreport11_15 <- taxreport11_15 %>%
  select(PID, coordinate, ZONING_CLASSIFICATION, CURRENT_LAND_VALUE,
CURRENT_IMPROVEMENT_VALUE,
         PREVIOUS_LAND_VALUE, PREVIOUS_IMPROVEMENT_VALUE, YEAR_BUILT, REPORT_YEAR)
properties <- properties %>%
  select(OBJECTID, size_m2, coordinate, geo_point_2d)

# join tables
tax_rep11_24 <- rbind(taxreport11_15, taxreport16_19, taxreport20_24)

# drop rows with missing values
tax_rep11_24_clean <- tax_rep11_24[complete.cases(tax_rep11_24[, c("CURRENT_LAND_VALUE",

"CURRENT_IMPROVEMENT_VALUE",
                                                                    "PREVIOUS_LAND_VALUE",

"PREVIOUS_IMPROVEMENT_VALUE")]), ]

# create the new variables for total value and ratio
```

```
tax_rep11_24_clean$tot_val_curr <- tax_rep11_24_clean$CURRENT_LAND_VALUE +
tax_rep11_24_clean$CURRENT_IMPROVEMENT_VALUE

tax_rep11_24_clean$tot_val_prev <- tax_rep11_24_clean$PREVIOUS_LAND_VALUE +
tax_rep11_24_clean$PREVIOUS_IMPROVEMENT_VALUE

tax_rep11_24_clean$ratio_curr <- tax_rep11_24_clean$CURRENT_IMPROVEMENT_VALUE /
tax_rep11_24_clean$CURRENT_LAND_VALUE

tax_rep11_24_clean$ratio_prev <- tax_rep11_24_clean$PREVIOUS_IMPROVEMENT_VALUE /
tax_rep11_24_clean$PREVIOUS_LAND_VALUE


#rename dataframe

tax_rep14_24 <- tax_rep11_24_clean


# remove duplicates

tax_rep14_24_unique <- tax_rep14_24 %>%

  distinct(PID, coordinate, YEAR_BUILT, REPORT_YEAR, .keep_all = TRUE)


# go from long to wide format

tr_wide <- tax_rep14_24_unique %>%

  pivot_wider(

    names_from = REPORT_YEAR,

    values_from = c(CURRENT_LAND_VALUE, CURRENT_IMPROVEMENT_VALUE,

                    PREVIOUS_LAND_VALUE, PREVIOUS_IMPROVEMENT_VALUE,

                    tot_val_curr, tot_val_prev,

                    ratio_curr, ratio_prev),

    names_glue = "{.value}{REPORT_YEAR}"

  )


# rearrange columns

tr_wide <- tr_wide %>%

  select(coordinate, ZONING_CLASSIFICATION, YEAR_BUILT,

         sort(tidyselect::peek_vars())[-(1:3)])


# remove missing values

tr_wide2 <- tr_wide %>%

  drop_na()


# join tables, and keep only unique values of coordinate

properties_tax_all <- merge(properties, tr_wide, by = "coordinate") %>%

  distinct(coordinate, .keep_all = TRUE)


# export to csv

write.csv(properties_tax_all, "../../data/properties_tax_all2.csv", row.names = FALSE)
```

7.3.3 Python code

```
import arcpy

# set workspace
arcpy.env.workspace =
r"C:\Users\Olaf\Documents\RES\Thesis\thesis_data\gis\thesis1606\thesis1606.gdb"

# define input layers
property_parcels = "propertyparcels"
earthquakes = "earthquakes"

# define utm zone 10N spatial reference
utm_zone_10n = arcpy.SpatialReference(32610)

# convert eq points to list with year and objectid
earthquake_points = []
with arcpy.da.SearchCursor(earthquakes, ["SHAPE@", "OBJECTID", "Date_Time"]) as cursor:
    for row in cursor:
        eq_geom = row[0].projectAs(utm_zone_10n)  # eq point to utm
        eq_id = row[1]
        eq_year = row[2].year  # extract year
        earthquake_points.append((eq_geom, eq_id, eq_year))

# add field to property parcels dataframe for each eq distance with year in name
for idx, eq in enumerate(earthquake_points):
    field_name = f"dist_eq_{idx + 1}_{eq[2]}"
    arcpy.AddField_management(property_parcels, field_name, "DOUBLE")

# create list for update cursor
field_names = ["SHAPE@"] + [f"dist_eq_{idx + 1}_{eq[2]}" for idx, eq in
enumerate(earthquake_points)]

# calculate distance to each eq for each property parcel
with arcpy.da.UpdateCursor(property_parcels, field_names) as ucursor:
    for urow in ucursor:
        parcel_geom = urow[0].projectAs(utm_zone_10n)
        for idx, eq in enumerate(earthquake_points):
            distance = parcel_geom.distanceTo(eq[0])
            urow[idx + 1] = round(distance, 1)
        ucursor.updateRow(urow)
```

7.3.4 Stata code

```
version 18

clear

set more off

capture log close


* Fill in own working directory

cd "C:\Users\Olaf\Documents\RES\Thesis\thesis_data\code\stata"


import delimited "final2_prop_data2106.csv", delimiter(",")


* List of all ratio variables to be converted

local ratio_vars "ratio_curr2014 ratio_curr2015 ratio_curr2016 ratio_curr2017 ratio_curr2018
ratio_curr2019 ratio_curr2020 ratio_curr2021 ratio_curr2022 ratio_curr2023 ratio_curr2024
ratio_prev2014 ratio_prev2015 ratio_prev2016 ratio_prev2017 ratio_prev2018 ratio_prev2019
ratio_prev2020 ratio_prev2021 ratio_prev2022 ratio_prev2023 ratio_prev2024"


* Remove any non-numeric characters (such as commas) and replace "NA" with empty string

foreach var of varlist `ratio_vars' {

    replace `var' = subinstr(`var', "NA", "", .)

    replace `var' = subinstr(`var', ",", "", .)

    replace `var' = subinstr(`var', " ", "", .)

}


* Convert ratio variables to numeric

foreach var of varlist `ratio_vars' {

    destring `var', replace force

}


* Convert remaining variables to numeric

local other_vars "current_improvement_value2016 current_improvement_value2017
current_improvement_value2018 current_improvement_value2019 current_improvement_value2020
current_improvement_value2021 current_improvement_value2022 current_improvement_value2023
current_improvement_value2024 current_land_value2014 current_land_value2015
current_land_value2016 current_land_value2017 current_land_value2018 current_land_value2019
current_land_value2020 current_land_value2021 current_land_value2022 current_land_value2023
current_land_value2024 pid previous_improvement_value2014 previous_improvement_value2015
previous_improvement_value2016 previous_improvement_value2017 previous_improvement_value2018
previous_improvement_value2019 previous_improvement_value2020 previous_improvement_value2021
previous_improvement_value2022 previous_improvement_value2023 previous_improvement_value2024
previous_land_value2014 previous_land_value2015 previous_land_value2016
previous_land_value2017 previous_land_value2018 previous_land_value2019
previous_land_value2020 previous_land_value2021 previous_land_value2022
previous_land_value2023 previous_land_value2024 tot_val_curr2014 tot_val_curr2015
tot_val_curr2016 tot_val_curr2017 tot_val_curr2018 tot_val_curr2019 tot_val_curr2020
tot_val_curr2021 tot_val_curr2022 tot_val_curr2023 tot_val_curr2024 tot_val_prev2014
tot_val_prev2015 tot_val_prev2016 tot_val_prev2017 tot_val_prev2018 tot_val_prev2019
tot_val_prev2020 tot_val_prev2021 tot_val_prev2022 tot_val_prev2023 tot_val_prev2024"


foreach var of varlist `other_vars' {

    replace `var' = subinstr(`var', "NA", "", .)

    destring `var', replace

}
```

```
*drop missing values (-86,149) 7006 left

drop if missing(current_improvement_value2016, current_improvement_value2017,
current_improvement_value2018, current_improvement_value2019, current_improvement_value2020,
current_improvement_value2021, current_improvement_value2022, current_improvement_value2023,
current_improvement_value2024, current_land_value2014, current_land_value2015,
current_land_value2016, current_land_value2017, current_land_value2018,
current_land_value2019, current_land_value2020, current_land_value2021,
current_land_value2022, current_land_value2023, current_land_value2024, pid,
previous_improvement_value2014, previous_improvement_value2015,
previous_improvement_value2016, previous_improvement_value2017,
previous_improvement_value2018, previous_improvement_value2019,
previous_improvement_value2020, previous_improvement_value2021,
previous_improvement_value2022, previous_improvement_value2023,
previous_improvement_value2024, previous_land_value2014, previous_land_value2015,
previous_land_value2016, previous_land_value2017, previous_land_value2018,
previous_land_value2019, previous_land_value2020, previous_land_value2021,
previous_land_value2022, previous_land_value2023, previous_land_value2024, tot_val_curr2014,
tot_val_curr2015, tot_val_curr2016, tot_val_curr2017, tot_val_curr2018, tot_val_curr2019,
tot_val_curr2020, tot_val_curr2021, tot_val_curr2022, tot_val_curr2023, tot_val_curr2024,
tot_val_prev2014, tot_val_prev2015, tot_val_prev2016, tot_val_prev2017, tot_val_prev2018,
tot_val_prev2019, tot_val_prev2020, tot_val_prev2021, tot_val_prev2022, tot_val_prev2023,
tot_val_prev2024)


* Rename variables

rename name district

rename size_m2 sqm

rename zoning_classification zoning_class

rename av avg_age


* List of all dist_eq variables

local eq_vars "dist_eq_1_2023 dist_eq_2_2023 dist_eq_3_2023 dist_eq_4_2023 dist_eq_5_2022
dist_eq_6_2022 dist_eq_7_2022 dist_eq_8_2022 dist_eq_9_2022 dist_eq_10_2021 dist_eq_11_2021
dist_eq_12_2021 dist_eq_13_2021 dist_eq_14_2021 dist_eq_15_2021 dist_eq_16_2020
dist_eq_17_2020 dist_eq_18_2020 dist_eq_19_2019 dist_eq_20_2019 dist_eq_21_2019
dist_eq_22_2018 dist_eq_23_2018 dist_eq_24_2017 dist_eq_25_2017 dist_eq_26_2017
dist_eq_27_2017 dist_eq_28_2017 dist_eq_29_2017 dist_eq_30_2016 dist_eq_31_2016
dist_eq_32_2016 dist_eq_33_2016 dist_eq_34_2016 dist_eq_35_2016 dist_eq_36_2016
dist_eq_37_2016 dist_eq_38_2015 dist_eq_39_2015 dist_eq_40_2015 dist_eq_41_2015
dist_eq_42_2015 dist_eq_43_2015 dist_eq_44_2015 dist_eq_45_2014 dist_eq_46_2014
dist_eq_47_2014 dist_eq_48_2014 dist_eq_49_2013 dist_eq_50_2013"


*get shortest distance

egen closest_dist_eq = rowmin(`eq_vars')


* Get unique years from variable names

local years "2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023"


* Loop over each year and sum the corresponding variables and count the non-missing values

foreach year of local years {

    * Create a list of variables for the current year

    local year_vars

    foreach var of local eq_vars {

        if strpos("`var'", "`year'") {

            local year_vars `year_vars' `var'

        }

    }


    * Sum the variables for the current year and count the non-missing values

    generate tot_dist_eq_`year' = 0
```

```
    generate num_earthquakes_`year' = 0
    foreach var of local year_vars {
        replace tot_dist_eq_`year' = tot_dist_eq_`year' + `var'
        replace num_earthquakes_`year' = num_earthquakes_`year' + !missing(`var')
    }
}

* Convert year_built to numeric and handle "NA" values
gen year_built_num = .
replace year_built_num = real(year_built) if year_built != "NA"

* Create year built categories
gen year_built_cat = .
replace year_built_cat = 1 if year_built_num <= 1969
replace year_built_cat = 2 if year_built_num >= 1970
drop year_built

* Remove unused variables
drop pid civic_number streetname tax_coord site_id shape_length shape_area name_1
shape_length_1 shape_area_1 join_count target_fid near_fid dist_eq_1_2023 dist_eq_2_2023
dist_eq_3_2023 dist_eq_4_2023 dist_eq_5_2022 dist_eq_6_2022 dist_eq_7_2022 dist_eq_8_2022
dist_eq_9_2022 dist_eq_10_2021 dist_eq_11_2021 dist_eq_12_2021 dist_eq_13_2021 dist_eq_14_2021
dist_eq_15_2021 dist_eq_16_2020 dist_eq_17_2020 dist_eq_18_2020 dist_eq_19_2019
dist_eq_20_2019 dist_eq_21_2019 dist_eq_22_2018 dist_eq_23_2018 dist_eq_24_2017
dist_eq_25_2017 dist_eq_26_2017 dist_eq_27_2017 dist_eq_28_2017 dist_eq_29_2017
dist_eq_30_2016 dist_eq_31_2016 dist_eq_32_2016 dist_eq_33_2016 dist_eq_34_2016
dist_eq_35_2016 dist_eq_36_2016 dist_eq_37_2016 dist_eq_39_2015 dist_eq_40_2015
dist_eq_41_2015 dist_eq_42_2015 dist_eq_43_2015 dist_eq_44_2015 dist_eq_45_2014
dist_eq_46_2014 dist_eq_47_2014 dist_eq_48_2014 dist_eq_49_2013 dist_eq_50_2013

* Reshape to long format
reshape long current_improvement_value current_land_value previous_improvement_value
previous_land_value ratio_curr ratio_prev tot_val_curr tot_val_prev tot_dist_eq_
num_earthquakes_, i(objectid district dist_downtown dist_water dist_eq_38_2015 coordinate sqm
zoning_class year_built_cat) j(year)

*gen ppm
gen ppm = tot_val_curr / sqm
gen prev_ppm = tot_val_prev / sqm

*Get average distance to earthquakes per property
gen avg_dist_eq = tot_dist_eq_ / num_earthquakes_

* Get interest rates per year
gen cpi = .
replace cpi = 3.7 if year == 2024
replace cpi = 4.3 if year == 2023
replace cpi = 6.7 if year == 2022
replace cpi = 2.7 if year == 2021
replace cpi = 0.6 if year == 2020
replace cpi = 2.4 if year == 2019
```

```
replace cpi = 2.9 if year == 2018
replace cpi = 2.2 if year == 2017
replace cpi = 2.2 if year == 2016
replace cpi = 1.2 if year == 2015
replace cpi = 1.1 if year == 2014
replace cpi = 0.2 if year == 2013

*Set time
xtset objectid year

*Sort, and get the lagged distance and future sqrt root
sort objectid year
by objectid: gen lagged_avg_dist_eq = L.avg_dist_eq
by objectid: gen lagged2_avg_dist_eq = L2.avg_dist_eq
by objectid: gen lagged_number_of_eqs = L.num_earthquakes_
by objectid: gen lagged2_number_of_eqs = L2.num_earthquakes_
by objectid: gen lagged_cpi = L.cpi
by objectid: gen lagged2_cpi = L2.cpi
by objectid: gen lagged_dist_eq38 = L.dist_eq_38_2015

* Remove outliers
drop if ppm > 1000000
drop if ppm < 1

drop if missing(ratio_curr)
drop if missing(ratio_prev)

*As sqm is not normally distributed, and both transforming it with log() and sqrt() do not
make it normally distributed, I use boxcox to find the best transformation
boxcox sqm
gen sqm_boxcox = (sqm^(-0.4594055 ) - 1) / -0.4594055
gen log_sqm = log(sqm)

*Transform ppm
gen lnppm = log(ppm)
gen lagged_lnppm = L.lnppm

gen boxcox_av_dist = (avg_dist_eq^(0.8230472)-1)/0.8230472

encode district, gen(district_num)
encode zoning_class, gen(zoning_class_num)
```

```
****************************************************************

*   Thesis models

****************************************************************

*Linear regression model

reg lnppm lagged_avg_dist_eq

****************************************************************

*Pooled model

reg lnppm lagged_avg_dist_eq lagged_number_of_eqs sqm_boxcox i.year_built_cat dist_downtown
dist_water i.zoning_class_code i.district_code avg_age avg_inc lagged_cpi

estimates store pooled_ols

****************************************************************

*Main model

*FE model

reghdfe lnppm lagged_avg_dist_eq lagged_number_of_eqs sqm_boxcox i.year_built_cat
dist_downtown dist_water i.zoning_class_code i.district_code avg_age avg_inc, absorb(year)
vce(cluster objectid)


*RE model

xtset year

xtreg lnppm lagged_avg_dist_eq lagged_number_of_eqs sqm_boxcox i.year_built_cat dist_downtown
dist_water i.zoning_class_num i.district_num avg_age avg_inc, re robust

****************************************************************

*Chow test

*Pooled model

reghdfe lnppm lagged_avg_dist_eq lagged_number_of_eqs sqm_boxcox i.year_built_cat
dist_downtown dist_water i.zoning_class_num i.district_num avg_age avg_inc, absorb(year)
vce(cluster objectid)


*Before 1980

*FE model

reghdfe lnppm lagged_avg_dist_eq lagged_number_of_eqs sqm_boxcox dist_downtown dist_water
i.zoning_class_num i.district_num avg_age avg_inc if year_built_cat == 1, absorb(year)
vce(cluster objectid)


*After 1980

reghdfe lnppm lagged_avg_dist_eq lagged_number_of_eqs sqm_boxcox dist_downtown dist_water
i.zoning_class_num i.district_num avg_age avg_inc if year_built_cat == 2, absorb(year)
vce(cluster objectid)

*****************************************************************

*Ratio model

*FE model

reghdfe ratio_curr lagged_avg_dist_eq lagged_number_of_eqs sqm_boxcox i.year_built_cat
dist_downtown dist_water ratio_prev i.zoning_class_num i.district_num avg_age avg_inc,
absorb(year) vce(cluster objectid)


*RE model

xtset year

xtreg ratio_curr lagged_avg_dist_eq lagged_number_of_eqs sqm_boxcox i.year_built_cat
dist_downtown dist_water ratio_prev i.zoning_class_num i.district_num avg_age avg_inc, re
robust

*****************************************************************
```

```
*****************************************************************
*    Assumptions
*****************************************************************
* Conduct the Hausman test
xtset year
xtreg lnppm lagged_avg_dist_eq sqm_boxcox i.year_built_cat dist_downtown dist_water
i.zoning_class_num i.district_num, fe
estimates store fixed_effects


xtset year
xtreg lnppm lagged_avg_dist_eq sqm_boxcox i.year_built_cat dist_downtown dist_water
i.zoning_class_num i.district_num, re
estimates store random_effects


hausman fixed_effects random_effects
*******************************************************************
* VIF
reg lnppm lagged_avg_dist_eq sqm_boxcox i.year_built_cat dist_downtown dist_water
i.zoning_class_num
vif
*******************************************************************
*Woolridge test for autocorrelation in panel data
xtset objectid year
xtserial lnppm lagged_avg_dist_eq lagged_number_of_eqs sqm_boxcox dist_downtown dist_water
avg_age avg_inc


xtserial ratio_curr lagged_avg_dist_eq lagged_number_of_eqs sqm_boxcox dist_downtown
dist_water avg_age avg_inc


*******************************************************************
*Correlation matrix
corr lnppm lagged_avg_dist_eq sqm_boxcox year_built_cat dist_downtown dist_water
zoning_class_num
*******************************************************************




****************************************************************
*     Tables
****************************************************************
*Table 1: Regression output


*Linear regression model
reg lnppm lagged_avg_dist_eq
outreg2 using t69.doc, replace ctitle(OLS)
```

```
*Pooled model

reg lnppm lagged_avg_dist_eq sqm_boxcox i.year_built_cat dist_downtown dist_water
i.zoning_class_num i.district_num avg_age avg_inc lagged_cpi

outreg2 using t69.doc, append ctitle(OLS Pooled model)


*RE model

xtset year

xtreg lnppm lagged_avg_dist_eq sqm_boxcox i.year_built_cat dist_downtown dist_water
i.zoning_class_num i.district_num avg_age avg_inc, re robust

outreg2 using t69.doc, append ctitle(Random Effects) addtext(Year FE, YES, Census district FE,
YES)


*FE model

reghdfe lnppm lagged_avg_dist_eq sqm_boxcox i.year_built_cat dist_downtown dist_water
i.zoning_class_num i.district_num avg_age avg_inc, absorb(year) vce(cluster objectid)

outreg2 using t69.doc, append ctitle(Fixed Effects) addtext(Year FE, YES, Census district FE,
YES)

*****************************************************************

*Table 2: Chow test

*Chow test

*Pooled model

reghdfe lnppm lagged_avg_dist_eq sqm_boxcox i.year_built_cat dist_downtown dist_water
i.zoning_class_num i.district_num, absorb(year) vce(cluster objectid)

outreg2 using t667.doc, append ctitle(Pooled model)


*Before 1980

*FE model

reghdfe lnppm lagged_avg_dist_eq sqm_boxcox dist_downtown dist_water i.zoning_class_num
i.district_num if year_built_cat == 1, absorb(year) vce(cluster objectid)

outreg2 using t667.doc, append ctitle(Built before 1980)


*After 1980

reghdfe lnppm lagged_avg_dist_eq sqm_boxcox dist_downtown dist_water i.zoning_class_num
i.district_num if year_built_cat == 2, absorb(year) vce(cluster objectid)

outreg2 using t667.doc, append ctitle(Built after 1980)


********************************************************************

*Table 3: Ratio model

*Linear regression model

reg ratio_curr lagged_avg_dist_eq

outreg2 using t3099.doc, replace ctitle(OLS)


*Pooled model

reg ratio_curr lagged_avg_dist_eq sqm_boxcox i.year_built_cat dist_downtown dist_water
i.zoning_class_num i.district_num avg_age avg_inc

outreg2 using t3099.doc, append ctitle(OLS Pooled model)


*RE model
```

```
xtset year

xtreg ratio_curr lagged_avg_dist_eq sqm_boxcox i.year_built_cat dist_downtown dist_water
i.zoning_class_num i.district_num avg_age avg_inc, re robust

outreg2 using t3099.doc, append ctitle(Random effects)


*FE model

reghdfe ratio_curr lagged_avg_dist_eq sqm_boxcox i.year_built_cat dist_downtown dist_water
i.zoning_class_num i.district_num avg_age avg_inc, absorb(year) vce(cluster objectid)

outreg2 using t3099.doc, append ctitle(Fixed effects)


*********************************************************************

*Table 4: Descriptives

*Variables

outreg2 using t40.doc, replace sum(log) keep(avg_dist_eq ppm lnppm sqm sqm_boxcox ratio_curr
dist_downtown dist_water avg_inc avg_age year_built_num  num_earthquakes_ cpi zoning_class_num
i.district_num)
```