The Effect of Seismic Hazard Risk on Property Values: Evidence from a Regression Discontinuity

by

Wantong Wang

A Thesis Submitted in Partial Fulfillment of the Requirements for the
Degree of Bachelor of Arts, Honours
in the Department of Economics
University of Victoria
April 2017

Supervised by Dr.Martin Farnham

for

Dr. Chris Auld, Honours co-advisor

Dr. Merwan Engineer, Honours co-advisor

Special Thanks to
Daniel Brendle-Moczuk
Jun Duan

Abstract

The government of British Columbia published the Relative Earthquake Hazard Map of Greater Victoria in 2000, affecting homebuyers' perception of earthquake risk and, possibly, their willingness to pay for properties adjacent to a boundary between two earthquake risk zones. I use Regression Discontinuity (RD) design to estimate the effect of earthquake risk information on house prices and find little evidence of discontinuities in sales prices of houses at the risk boundaries in District of Saanich in Greater Victoria, BC. My results are consistent with there being no effect of the publication of the Hazard Map on sales price.

Table of Contents

1. Introduction

In this paper, I use Regression Discontinuity (RD) design to evaluate how seismic hazard affects property values using data from Greater Victoria (BC). Specifically, I estimate whether there is a discontinuity in sales price of houses at the boundary between zones with different earthquake risk levels and whether consumers respond rationally to the publication of earthquake hazard maps. Many properties around the world lie in seismically active areas such as California, Japan, and the west coast of Canada. Therefore, this is a question of interest to property owners, insurers, and governments. Governments often summarize seismic risk through maps that make information available to home buyers. However, this information is usually imperfectly summarized. For example, earthquake risk zone maps usually display average risks across an area rather than the precise risk at each location. Therefore, it would be useful to estimate whether such publication of information distorts real estate markets.

Earthquake risk can vary dramatically over small distances due to local variation in soil geology. The ability of a structure to weather an earthquake with minimal damage is much higher on rocky ground than on deep soil. Therefore, while different properties across an earthquake-active area are equally likely to experience the same large earthquake, there is significant local variation in the likely damage that would result from that earthquake. In earthquake-prone areas, homeowners can buy insurance against earthquake losses. However, due to the enormous potential insured loss caused by major earthquakes, insurance companies in high-risk areas usually do not provide full-coverage earthquake insurance, or only provide it with very high premiums. Even if full insurance was available, one would expect the premiums to be higher in less geologically stable areas due to the larger expected losses in those areas. Some combination of higher expected losses or higher earthquake premiums in less stable parts of region would be expected to be capitalized into the relative sales prices of properties.

I use property transactions and earthquake risk data for Greater Victoria, BC to estimate the effect of seismic risk on sales prices because Greater Victoria is threatened by potentially significant earthquakes and associated property losses. There is a 33% chance of a major earthquake (magnitude 6.5 and above on the Richter Scale) on Vancouver Island within the next 50 years, and there is a 10% chance the earthquake will be the Big One (magnitude 9.0 or higher), according to Earthquake Canada seismologist, Alison Bird. According to the Insurance Bureau of Canada, the estimated expected damage from a catastrophic earthquake in British Columbia is approximately 75 billion dollars. In 2000, the government of British Columbia issued the Relative Earthquake Hazard Map of Greater Victoria (the Hazard Map) online and in local newspapers. According to the Hazard Map, within a small neighbourhood, earthquake risk levels can vary from low to high, over six different risk zones (Figure 1.1). The Hazard Map not only enables insurees to evaluate the potential earthquake losses of their properties but also enables insurers to price the earthquake insurance premiums based on potential insured losses, affecting home buyers' willingness to buy and earthquake insurance premium. The publication of the Hazard Map allows me to identify the different earthquake risks for properties in Greater Victoria and the distance of the properties to the boundaries between different earthquake risk zones. With these variables, I can evaluate how earthquake risk is capitalized into property prices in Greater Victoria. Specifically, I use RD design to estimate whether higher earthquake risk is negatively capitalized into sales prices of homes. In addition, I use RD design to estimate whether the ordinal and discrete earthquake zone rating information displayed on the Hazard Map causes a discontinuity in the sales price of houses at earthquake zone boundaries. Since the real seismic risk is likely to be continuous across risk boundaries and since the Hazard Map only displays ordinal and discrete ratings, this also amounts to a test of whether consumers respond rationally to the publication of information about seismic risks. Finally, using a parametric RD hedonic pricing model, I estimate whether sales prices fall as consumers move deeper into a relatively higher earthquake risk zone.

Hedonic regression is commonly used to estimate the effect of amenities on real estate values (Smith and Huang, 1995). Using a hedonic pricing model, Clarke (2013) runs regressions of property values in Greater Victoria on risk characteristics and other land and structure characteristics, and find that on average properties in high-risk zones are worth less than those in relatively low-risk zones. However, due to unobservable heterogeneity and omission of relevant variables, the estimates of hedonic pricing models may be biased. For example, if the ground type near the coast is rock, then expensive houses with ocean views would tend to be located in safe zones. If more of the inland ground type is soil, then relatively cheaper houses with less spectacular views will tend to be located in relatively risky zones. In this example, failure to control for the view of the home (unobservable in most datasets, including Clarke, 2013) could lead to OLS estimates of the effect of earthquake risk on property prices being biased.

RD design is a relatively credible and transparent approach for estimating treatment effects (Lee and Lemieux 2008). Hidano et al. (2015) use two-dimensional RD design to evaluate how the property market in Tokyo responded to earthquake risk information. They find that the perceived seismic hazard rating lowers the value of residential properties in riskier zones. Using local-polynomial RD design (Lee and Lemieux 2010), I include only properties that lie within a certain bandwidth on each side of an earthquake risk zone boundary to deal with the potential bias in hedonic models. The assignments of earthquake zones are determined by the level of risk that an area will face when an earthquake is shaking and based on geographical knowledge; therefore, the treatment assignments are independent of locations of the houses and can be viewed as local randomization, subject to certain caveats. As a result, I can use local-polynomial RD design to control for unobservable heterogeneity of properties and reduce potential bias without including other property-related characteristics. Instead of estimating the effect of earthquake risk on property values on average (as done by Clarke, 2013), I use parametric RD design to estimate whether there is a discontinuity in sales price at an earthquake zone boundary and how properties' distance to a zone boundary affect their sales price.

This is one of the only papers to estimate the effect of seismic risk on property values using RD design. It is the first time that the relationship between property values and earthquake risk has been estimated by RD design with Greater Victoria data. Actual earthquake risk levels are likely to be continuous across space, but the earthquake risk zones displayed on the Hazard Map are discrete and vary within small geographical areas. Consequently, for two adjacent properties in different risk zones, naive consumers without proper geographical knowledge (or sophisticated consumers who expect other consumers to be naïve) should be expected to value the property in the low-risk zone more than its neighbour in the relatively high-risk zone, which may cause a discontinuity in sales price at the risk zone boundaries. If consumers are rational and understand that real earthquake risk is unlikely to jump discretely across risk boundaries, the sales prices are expected to be lower for properties that are far away from the boundaries and in the centre of a high-risk zone. In this paper, taking all of my estimates together, I find very little evidence of discontinuities in sales prices at the earthquake zone boundaries and of a negative relationship between sales price and a property's distance to the boundaries. Also, my results do not support that the publication of the Hazard Map affects consumers' willingness to pay for housing.

2. Data

In this paper, I construct a unique dataset by combining property transaction data from Landcor Data Corporation and the shapefiles of the Relative Earthquake Hazard Map of Greater Victoria. The data from Landcor Data Corporation include the sales price,sales date, street address, Parcel Identifier (PID), property size, the number of bedrooms, the number of bathrooms, year built and foundation type. I use single-family dwellings traded between 1997 and 2003 in District of Saanich in Greater Victoria, BC.

The Relative Earthquake Hazard Map of Greater Victoria, published online and on the front page of the local newspaper in 2000, is a composite map of three maps of Greater Victoria: the Relative Liquefaction Hazard Map, the Relative Amplification of Ground Motion Hazard Map, and the Seismic Slope Stability Map. The Relative Liquefaction Hazard Map measures the susceptibility of soil to behave like a liquid when an earthquake is shaking.

The Relative Amplification of Ground Motion Hazard Map shows areas in which ground motion will be enhanced during an earthquake. The Seismic Slope Stability Map measures the susceptibility to slope failures for a site during an earthquake. The Hazard Map classifies all areas in Greater Victoria as zone 1 to 6 based on the composition of the three earthquake hazards mentioned above (class 1=low hazard , Class 2=low to moderate hazard, Class 3=low to high hazard, Class 4=moderate hazard, Class 5=moderate to high hazard, Class 6=high hazard).

In this paper, I use Geographic Information System (GIS) software to process my data. First, I generate three earthquake-related variables for each property in the dataset based on the earthquake information and shapefiles of the Hazard Map. These three variables are 1) the hazard zone that each property is located in, 2) the nearest different hazard zone (other than the property's own hazard zone) that the property is close to, and 3) the distance of each property to the nearest different hazard zone. Second, I use the PID of each property to exclude from my sample properties that straddle two or more hazard zones, since they could bias my estimate of boundary effects downward. Third, I use PID's of the properties to present property location parcels with earthquake hazard zone layers on the Hazard Map and show the distributions of the single-family dwellings in Saanich (Figure1.1 and Figure 1.2; .Figure 1.2 is an enlarged version of Figure 1.1). Each polygon represents a property from the dataset.

Figure1.1

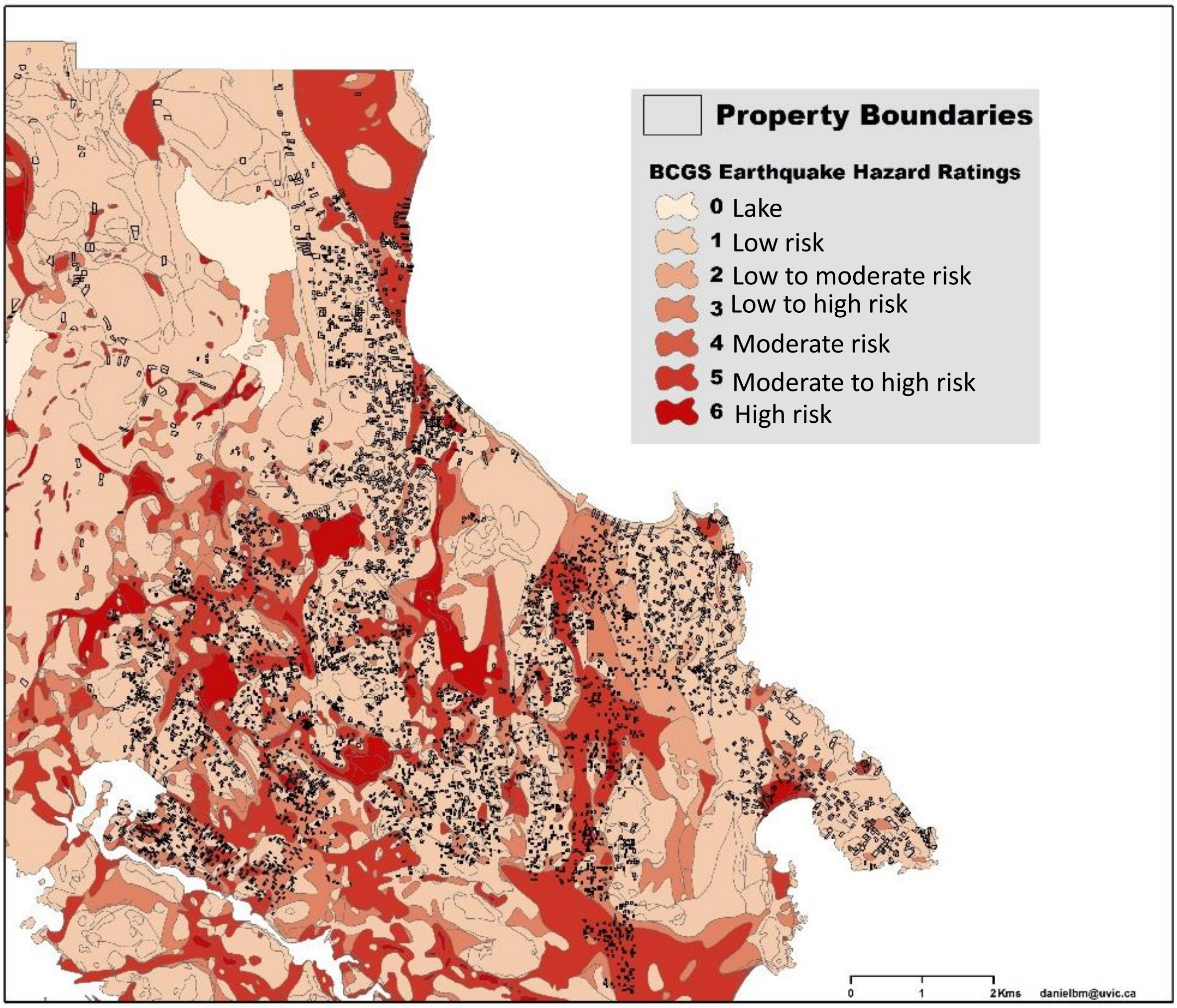
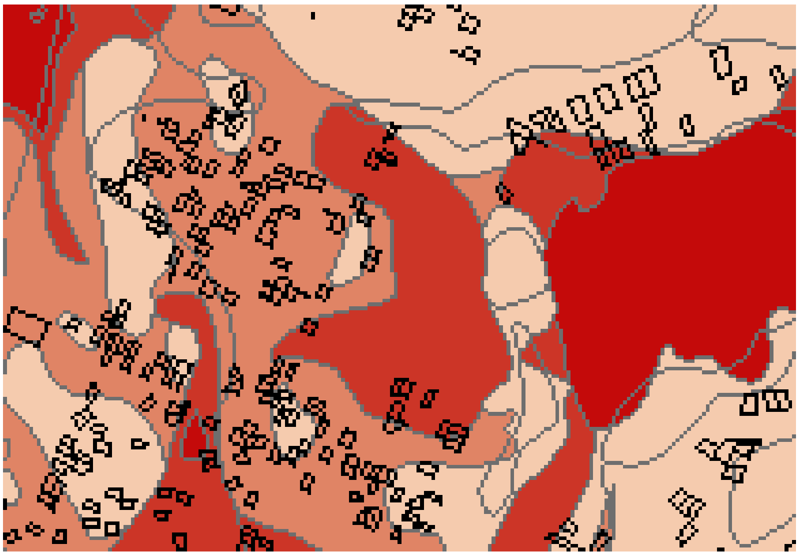


Figure 1.2



Most observations Saanich are located in hazard zones 1, 3 and 5. In total, there are 3560 properties, with 2197 properties in zone 1, 729 properties in zone 3 and 729 in zone 5. Because there are very few observations in 2, 4 and 6, I only focus on analyzing properties in hazard zones 1,3 and 5. Tables 1.1, 1.2, and 1.3 present the summary statistics of the sales price and other property characteristics for zones 1, 3 and 5, respectively. The average sales prices are \$276,793, \$233,021and \$222,063 for hazard zones 1, 3 and 5, respectively. The average land sizes of properties in zone 3 and 5 are similar, but the average land size of properties in zone 1 is 3,000 square feet larger than in zone 3 and 5. The average number of bedrooms (approx. 3.3) and bathrooms (approx. 2.1) do not vary substantially among the three hazard zones. Since there are four foundation types and properties were built in eight different decades in the dataset, I introduce dummy variables for each of them. For example, the dummy variable, foundation basement, equals one if the foundation type of the property is a basement, and equals zero otherwise. If a property was built in the 1950s, then the dummy variable “built in 1950s” equals one, and equals zero otherwise. Therefore, the mean of a dummy variable is the proportion of the properties that have the specific characteristic indicated by the dummy variable. For example, from the summary statistics for hazard zone 1, the mean of “built in the 1980s” is 0.23, revealing that 23% of properties in my sample were built in 1980s.

Table 1.1 Statistics Summary of Hazard Zone 1 (low risk)

Variable	Mean	Std. Dev.
sale price	276,793.60	121,830.20
property size (sq.ft)	10,349.10	6,727.24
number of bedrooms	3.37	0.94
number of bathrooms	2.38	0.92
foundation basement	0.37	0.48
foundation crawl	0.33	0.47
foundation slab	0.24	0.43
found. partial basement	0.06	0.24
built in 1930s	0.06	0.23
built in 1940s	0.09	0.29
built in 1950s	0.18	0.38
build in 1960s	0.11	0.31
built in 1970s	0.13	0.34
built in 1980s	0.23	0.42
built in 1990s	0.13	0.34
built after 2000	0.06	0.24

Table 1.2 Statistics Summary of Hazard Zone3 (low to high risk)

Variable	Mean	Std. Dev.
sale price	233,021.00	73,778.66
property size(sq.ft)	7,827.82	2,649.59
number of bedrooms	3.21	0.91
number of bathrooms	2.08	0.93
foundation basement	0.42	0.49
foundation crawl	0.32	0.47
foundation slab	0.17	0.37
found. partial basement	0.10	0.29
built in 1930s	0.06	0.24
built in 1940s	0.16	0.36
built in 1950s	0.22	0.42
build in 1960s	0.11	0.31
built in 1970s	0.14	0.35
built in 1980s	0.20	0.40
built in 1990s	0.06	0.23
built after 2000	0.06	0.23

Table 1.3 Statistics Summary of Hazard Zone 5 (moderate to high risk)

Variable	Mean	Std. Dev.
sale price	222,063.40	82,035.36
property size(sq.ft)	7,704.38	3,389.12
number of bedrooms	3.28	0.96
number of bathrooms	2.03	0.86
foundation types	1.80	0.89
built year	6.14	1.90
foundation basement	0.46	0.50
foundation crawl	0.35	0.48
foundation slab	0.13	0.34
found. partial basement	0.06	0.23
built in 1930s	0.04	0.20
built in 1940s	0.12	0.33
built in 1950s	0.27	0.44
build in 1960s	0.17	0.38
built in 1970s	0.15	0.36
built in 1980s	0.12	0.33
built in 1990s	0.03	0.18
built after 2000	0.08	0.28

3. Empirical Strategy

RD design was first introduced by Thistlethwaite and Campbell (1960). It can be used to analyze the treatment effect of a program when the treatment is assigned to candidates based on whether the value of some rating variable exceeds or falls below a certain cut-off point (Lee and Lemieux, 2010). For example, RD design can be used to estimate the effect of receiving a scholarship on students' post-scholarship grades . Suppose when students' grades (the rating variable) are above a certain threshold (cut-off point), they are given the scholarship (treatment). One can use RD design to estimate the difference in post-scholarship grades (outcome) of students just above and just below the qualifying threshold to see the effect of the scholarship on students' performance. In my case, the rating variable is a property's distance to the nearest earthquake zone boundary. This distance approaches zero for properties right next to risk boundaries. Some properties are located on one side of the boundary in the relatively low-risk zone, while properties on the other side of the boundary in the relatively high-risk zone. The comparison of sales prices for properties on either side of risk boundaries forms the basis of my RD design. I estimate whether the discrete change in earthquake risk ratings across these risk boundaries cause the sales price (outcome) of the properties to jump.

One of the advantages of RD design is its transparency. In this paper, before undertaking any formal statistical methods, I use a graphical presentation if my RD design to informally test whether there is evidence of a discontinuity in sales price at the risk boundaries. I can use this graphical presentation to show the unconditional relationship a between distance and sales prices. I divide my dataset into three subsamples based on which risk boundaries are being considered: subsample 1-3, subsample 3-5, and subsample 1-5. Take subsample 1-3 as an example: this subsample includes the properties that lie in risk zone 1 (3) and are close to risk zone 3 (1). I construct graphs for each subsample. To estimate the relationship between sales price and distance, I define the distance of properties in the relative low-risk zone to be negative and those in the high-risk zone to be positive. In this way, a lower distance value corresponds to lower risk. If one assumes that average seismic risk declines moving further from a risk-zone boundary into the lower risk zone, or alternatively that seismic risk increases moving further from a risk-zone boundary into the higher risk zone, then given the way I define the distance variable, one should expect a negative relationship between distance and sales price.

Since my sample size is large and the plot of individual properties is noisy, I use smoothed plots for the graphical presentations (Lee and Lemieux, 2010). First, I divide the distance to the adjacent hazard zone of the properties into equal-sized intervals (bins). Second, I calculate the mean of the sale price and the median of the distance and count the number of observations in each bin. Third, I create weighted scatter plots in Stata, with the average sale price for each bin on the vertical axis and the mid-point of the distance to the adjacent hazard zone on the horizontal axis; the size of bubbles in scatter plots reflects the number of observations in each bin. Figures 2.1, 2.2 and 2.3 are the scatter plots for subsamples 1-3, 3-5 and 1-5 respectively with a bin size of 10 metres. The graphs do not show evidence of discontinuity of sales price at the cut-off point, so it is unlikely for statistical estimation to have statistically significant discontinuity results. Except for subsample 3-5, sales price and distance have a negative relationship in the graphs. The downward sloping relationship between sales price and distance to the boundary means that when a property is located in a relatively low-risk zone, the closer it is to the boundary, the lower its sales price is. When a property is located in a relatively high-risk zone, the closer it is to the boundary, the higher its sales price is. However, without statistical analysis, I cannot say whether this relationship is statistically significant.

Figure 2.1. Subsample 1-3, Sales Sales Price vs. Distance to Risk Boundary

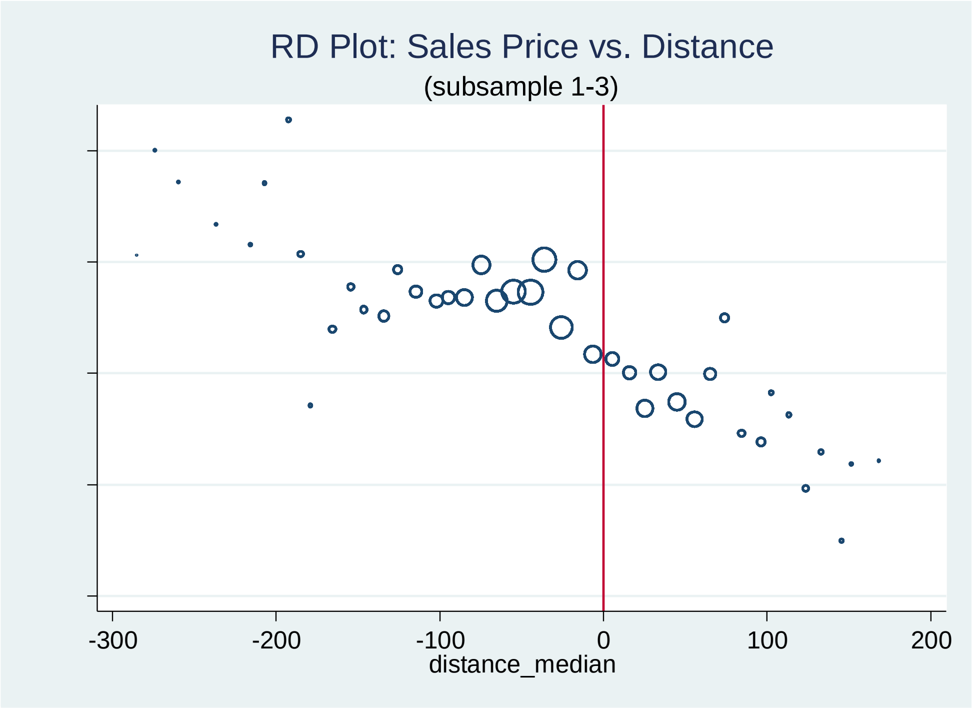


Figure 2.2. Subsample 3-5, Sales Price vs. Distance to Risk Boundary

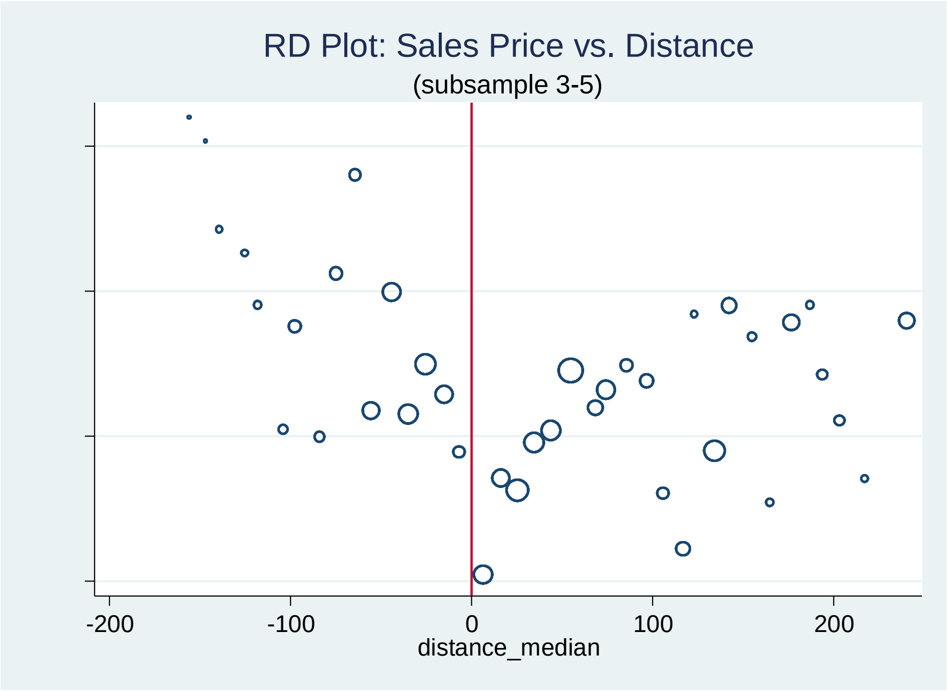
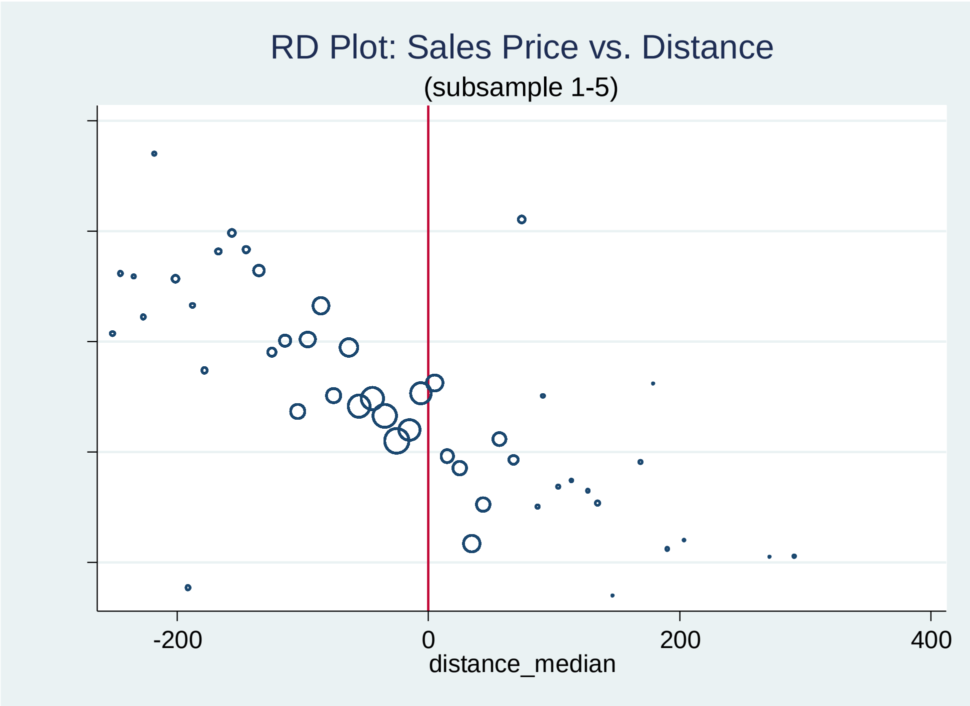


Figure 2.3. Subsample 1-5, Sales Price vs. Distance to Risk Boundary



3.2. *Parametric RD Design*

The following models allow me to analyze the discontinuity in sales price at the cut-off point and relationship between sales price and distance.

Model (1.1): Linear $\ln(P_{i,j}) = \alpha_j + \theta_j \cdot D_{i,j} + \beta_j \cdot d_{i,j} + \varepsilon_{i,j}$

Model (2.1): Linear interaction $\ln(P_{i,j}) = \alpha_j + \theta_j \cdot D_{i,j} + \beta_j \cdot d_{i,j} + \gamma_j \cdot d_{i,j} \cdot D_{i,j} + \varepsilon_{i,j}$

Model (3.1): Quadratic $\ln(P_{i,j}) = \alpha_j + \theta_j \cdot D_{i,j} + \beta_j \cdot d_{i,j} + \varphi_j \cdot d_{i,j}^2 + \varepsilon_{i,j}$

Model (4.1): Quadratic interaction

$$\ln(P_{i,j}) = \alpha_j + \theta_j \cdot D_{i,j} + \beta_j \cdot d_{i,j} + \varphi_j \cdot d_{i,j}^2 + \gamma_j \cdot d_{i,j} \cdot D_{i,j} + \mu_j \cdot d_{i,j}^2 \cdot D_{i,j} + \varepsilon_{i,j}$$

I can estimate these models separately for the three subsamples. In other words, j can equal 13, 35 and 15 for subsamples 1-3, 3-5 and 1-5, respectively. Take subsample 1-3 as an example. $\ln P_{i,13}$ is the natural logarithm of the CPI-adjusted sales price of property i in subsample 1-3. $d_{i,13}$ is property i's distance to the boundary. It is negative if the property i is in risk zone 1 and positive otherwise. When $d_{i,13}$ equals zero, the risk zone boundary occurs. $D_{i,13}$ equals one if property i is in risk zone 1 and zero if it is in risk zone 3. θ_{13} is the estimator of the treatment effect on the sales price for subsample 1-3. In models 2, 4 and 6, I interact distance $d_{i,13}$ with the risk zone dummy variable $D_{i,13}$, allowing the slope of the regression line on each side of the cut-off point to be different. I do not include such interactions in models 1, 3 and 5, thus constraining the slope of the regression line on each side of the cut-off point to be the same. While one could estimate the models above separately on each subsample, a more efficient use of the data pools all three subsamples together. To include more information and reduce the standard error of my estimation, I construct and estimate variants of the following pooled regression model:

Model (5.1)

$$\ln(P_{i,j}) = Z_{13} \cdot [\theta_{13} \cdot D_{i,13} + f(d_{i,13})] + Z_{35} \cdot [\theta_{35} \cdot D_{i,35} + f(d_{i,35})] + Z_{15} \cdot [\theta_{15} \cdot D_{i,15} + f(d_{i,15})] + \varepsilon_{i,j}$$

All variables are defined as before. $f(d_{i,j})$ is the relationship between the running variable $d_{i,j}$ and the outcome variable $\ln P_{i,j}$. It follows the six specifications mentioned above. Z_{13}, Z_{35}, and Z_{15} are risk zone pair dummy variables. For example, Z_{13} equals 1, Z_{35} equals 0, and Z_{15} equals 0 if the property is in subsample 1-3. Therefore, $Z_{13} * \theta_{13}$, $Z_{35} * \theta_{35}$ and $Z_{15} * \theta_{15}$

are the boundary treatment effects for subsamples 1-3, 3-5 and 1-5, respectively.

3.3 *Non-Parametric RD Design*

There may be some other variables that are correlated with the risk zone variable and influence the sales price, which may cause a bias in the parametric RD design. The assignments of earthquake zones are based on geographical knowledge rather than locations of the houses and can be viewed as local randomization, subject to certain caveats. To reduce bias, I run kernel-based local polynomial regressions (Calonico, Cattaneo, and Titiunik, 2014) separately for observations in the two bins adjacent to the cut-point (bandwidth) to estimate the boundary treatment effect. However, the precision of the estimates is decreased due to smaller sample size. As a result, the non-parametric strategy can only be a complement but not a substitute for the parametric strategy. My local linear regression and local quadratic regression are generated with the Stata comment *rdrobust* introduced by Calonico et al. (2014). The cross-validation method that selects the optimal bandwidth is also included in the robust comment.

3.4 *Effect of Publication of the Hazard Map on Sales Price*

The analysis above considers the effect of earthquake risk information on the level of sales prices. One might also be interested in how the release of earthquake risk information changes sales prices. In this subsection, I describe my analysis of the effect of releasing earthquake maps on property values in Greater Victoria.

Taking subsample 1-3 as an example, to estimate the effect of the publication of the

Hazard Map on sales price, I inlcude post publication dummy $post_{i,13}$ and interact it with *distance*, risk zone dummy regress $D_{i,13}$, and the interaction between *distance* and $D_{i,13}$ on the right hand side of models (1.1) to (4.1), constucting models(1.2) to (4.2). The post publication dummy $post_{i,13}$ equals one if property i was sold in 2001-2003; it equals zero if property i was sold in 1997-1999 (the Hazard Map was published in 2000). With the earthquake information displayed on the Hazard Map, if consumers are naive and believe the earthquake risk jumps across boundaries given the information provided by the Hazard Map, then the coefficients of the two interactions are expected to be negative. $X_{i,13}$ is a vector containing other property-related characteristics such as property size, the number of bedrooms, the number of bathrooms, build year and foundation type.

Model (1.2): Linear

$$\ln(P_{i,j}) = \alpha_j + \theta_j \cdot D_{i,j} + \beta_{0,j} \cdot d_{i,j} + \delta_j \cdot post_{i,j} + \rho_j \cdot D_{i,j} \cdot post_{i,j} + \omega_{i,j} \cdot d_{i,j} \cdot post_{i,j} + X_{i,j} \cdot \tau_j + \varepsilon_{i,j}$$

Model (2.2): Linear interaction

$$\ln\left(P_{i,j}\right) = \alpha_j + \theta_j \cdot D_{i,j} + \beta_{0,j} \cdot d_{i,j} + \gamma_j \cdot d_{i,j} \cdot D_{i,j} + \delta_j \cdot post_{i,j} + \rho_j \cdot D_{i,j} \cdot post_{i,j} + \omega_j \cdot d_{i,j} \cdot post_{i,j}$$

$$+\beta_{1,j} \cdot post_{i,j} \cdot d_{i,j} \cdot D_{i,j} + X_{i,j} \cdot \tau_j + \varepsilon_{i,j}$$

Model (3.2): Quadratic

$$\ln(P_{i,j}) = \mathrm{\alpha_j} + \mathrm{\theta_j} \cdot \mathrm{D_{i,j}} + \mathrm{\beta_{0,j}} \cdot \mathrm{d_{i,j}} + \mathrm{\varphi_j} \cdot \mathrm{d_{i,j}^2} + \delta_j \cdot post_{i,j} + \rho_j \cdot D_{i,j} \cdot post_{i,j} + \omega_{i,j} \cdot d_{i,j} \cdot post_{i,j}$$

$$+\beta_{1,j} \cdot d_{i,j}^2 \cdot post_{i,j} + X_{i,j} \cdot \tau_j + \mathrm{\varepsilon_{i,j}}$$

Model (4.2): Quadratic interaction

$$\ln(P_{i,j}) = \alpha_j + \theta_j \cdot D_{i,j} + \beta_{0,j} \cdot d_{i,j} + \varphi_j \cdot d_{i,j}^2 + \gamma_j \cdot d_{i,j} \cdot D_{i,j} + \mu_j \cdot d_{i,j}^2 \cdot D_{i,j} + \delta_j \cdot post_{i,j} + \rho_j \cdot D_{i,j} \cdot post_{i,j}$$

$$+\omega_{i,j} \cdot d_{i,j} \cdot post_{i,j} + \beta_{1,j} \cdot d_{i,j}^2 \cdot post_{i,j} + \beta_{2,j} \cdot d_{i,j} \cdot D_{i,j} \cdot post_{i,j} + \beta_{3,j} \cdot d_{i,j}^2 \cdot D_{i,j} \cdot post_{i,j} + X_{i,j} \cdot \tau_j + \varepsilon_{i,j}$$

As I demonstarate in section 3.2, to include more information and reduce the standard error of my estimation, a more efficient use of the data pools all three subsamples together. I construct and estimate variants of the following pooled regression model:

Model (5.2)

$$\ln(P_{i,j}) = Z_{13} \cdot f\left(d_{i,13}, D_{i,13}, post_{i,13}\right) + Z_{35}. f\left(d_{i,35}, D_{i,35}, post_{i,35}\right) + Z_{15}. f\left(d_{i,15}, D_{i,15}, post_{i,15}\right) + \varepsilon_{i,j}$$

f(.) can be the four specifications (1.2) to (4.2) . *Z13*, *Z35*, and *Z15* are risk zone dummies defined in the last section.

4. Results

4.1 *Parametric RD Design Estimation*

The estimates of the model (5.1) in Section 3.2 using the four different specifications are given in Table 2. Columns 2 to 5 are estimates of the models (4.1) to (4.4) respectively. The coefficient estimate of the variable (higher risk)*Z_{13}, *(higher risk)*$*Z_{35}$ and *(higher risk)*Z15* measures boundary treatment effects for subsamples 1-3, 3-5 and 1-5 respectively. Collectively, the boundary treatment effects are not statistically significant for three subsamples, indicating there is no discontinuity in sales price at the zone boundary. Except for subsample 3-5, sales price negatively depends on the distance to the earthquake zone boundary. For subsample 1-3, when a property is 100 metres (the unit of distance to the boundaries is metre/100) closer to zone 3, its sales price decrease by approximately 8%. For subsample 1-5, when a property is 100 metres closer to zone 5, its sales price decrease by approximately 10%. These two results are consistent with consumers being sophisticated and and knowing that the earthquake risk changes continuously across the risk boundary.

Table 2. Parametric RD Design Regression Results
Dependent variable: sales price
(without controls)

	Model(1.1)	Model(2.1)	Model(3.1)	Model (4.1)
N	3555	3555	3555	3555
constant*Z13	12.44***	12.45***	12.44***	12.46***
	(0.01)	(0.02)	(0.02)	(0.02)
(higher risk)*Z13	-0.06**	-0.04	-0.04	-0.06
	(0.03)	(0.03)	(0.03)	(0.04)
dist*Z13	-0.08***	-0.06***	-0.10***	-0.04
	(0.02)	(0.02)	(0.03)	(0.06)
constant*Z35	12.37***	12.32***	12.34***	12.33***
	(0.03)	(0.04)	(0.04)	(0.07)
(higher risk)*Z35	-0.08*	-0.05	-0.05	-0.06
	(0.05)	(0.05)	(0.06)	(0.08)
dist*Z35	0.02	-0.06	-0.01	-0.02
	(0.02)	(0.07)	(0.04)	(0.26)
constant*Z15	12.43***	12.41***	12.43***	12.41***
	(0.02)	(0.02)	(0.02)	(0.03)
(higher risk)*Z15	-0.02	-0.05	-0.04	-0.06
	(0.03)	(0.04)	(0.04)	(0.05)
dist*Z15	-0.11***	-0.13***	-0.08***	-0.14**
other controls	no	no	no	no
	(0.02)	(0.02)	(0.03)	(0.07)
inter*Z13		-0.07		-0.04
		(0.05)		(0.15)
inter*Z35		0.09		0.07
		(0.08)		(0.27)
interZ15		0.00**		0
		0.00		0.00
dist^2*Z13			-0.02	0.01

	(0.01)	(0.03)
dist^2*Z35	0.01	0.03
	(0.02)	(0.19)
dist^2*Z15	0.02*	-0.01
	(0.01)	(0.03)
inter^2*Z13		-0.05
		(0.10)
inter^2*Z35		-0.03
		(0.19)
inter^2*Z15		-0.02
		(0.06)

Z13,Z35 and Z15: *risk zone dummies:Z13=1, Z35=Z15=0 for subsample 1-3*
higher risk*: equals to 1 for properties in high-risk zone;0 otherwise*
dist: *distance (metre/100)*
inter: *dist*(higher risk)*
inter^2: *(dist^2)*(higher risk)*
other controls*: he number of bedrooms, the number of bathrooms, the foundation type dummies (basement, crawl, slab and partial basement)and built year dummies (in 1930s to after 2000)*

* p<0.10, ** p<0.05, *** p<0.01

To increase the completeness and reduce sample variability of my models, I include non-affected covariates, property size, the number of bedrooms, the number of bathrooms, the foundation type dummies (basement, crawl, slab and partial basement) and built year dummies (in the 1930s to after 2000) in the four models. To avoid collinearity problem, I omit the first dummy in each dummy groups (basement foundation and built year the 1930s). Also, to present the estimates in the tables better, I scale up the coefficient estimates of variable distance (metres) and property size(sq.ft) by dividing them by 100 and 1000 respectively. The estimates for the model (5.1) are in Table 3. Columns 2 to 5 display the estimates for models (1.1) to (4.1).

Collectively, with the inclusion of covariates, the boundary treatment effects for three subsamples are still not statistically significant, which is consistent with consumers knowing earthquake risk is unlikely to change discretely across boundaries. However, I find a mixed evidence of a negative relationship between sales price and distance because some of the coefficient estimates of distance are statistically significant and some are not. The statistically significant estimates are consistent with there being a negative relationship between sales price and distance, while the statistically insignificant estimates are consistent with there being no relationship between sales price and distance.

Table 3. Parametric RD Design Regression Results
Dependent variable: sales price
(with controls)

	model(1.1)	model(2.1)	model(3.1)	model(4.1)
N	3555	3555	3555	3555
constant*Z13	11.889***	11.888***	11.890***	11.896***
	(0.03)	(0.03)	(0.03)	(0.03)
(higher risk)*Z13	-0.021	-0.023	-0.025	-0.059
	(0.02)	(0.03)	(0.03)	(0.04)
dist*Z13	-0.046***	-0.048***	-0.041*	-0.015
	(0.02)	(0.02)	(0.02)	(0.05)
constant*Z35	11.874***	11.863***	11.878***	11.803***
	(0.04)	(0.05)	(0.04)	(0.07)
(higher risk)*Z35	-0.080**	-0.07	-0.084*	-0.03
	(0.04)	(0.05)	(0.05)	(0.07)
dist*Z35	0.025	0.001	0.029	-0.243
	(0.02)	(0.06)	(0.04)	(0.22)
constant*Z15	11.892***	11.885***	11.894***	11.879***
	(0.03)	(0.03)	(0.03)	(0.04)
(higher risk)*Z15	-0.006	-0.022	-0.016	-0.038
	(0.03)	(0.03)	(0.03)	(0.04)
dist*Z15	-0.041**	-0.052***	-0.028	-0.067
	(0.02)	(0.02)	(0.02)	(0.06)
property size	0.006***	0.006***	0.006***	0.006***
	(0.00)	(0.00)	(0.00)	(0.00)
bedroom	0.014**	0.014**	0.014**	0.014**
	(0.01)	(0.01)	(0.01)	(0.01)
bathroom	0.141***	0.141***	0.141***	0.141***
	(0.01)	(0.01)	(0.01)	(0.01)
	(0.03)	(0.03)	(0.03)	(0.03)
other controls	yes	yes	yes	yes
itnZ13		0.008		0.095

	(0.04)		(0.13)
itnZ35	0.026		0.315
	(0.06)		(0.23)
itnZ15	0.001		0.002
	0.00		(0.00)
disqrZ13		0.004	0.018
		(0.01)	(0.03)
disqrZ35		-0.002	-0.182
		(0.01)	(0.16)
disqrZ15		0.01	-0.008
		(0.01)	(0.03)
itnsqrZ13			-0.109
			(0.09)
itnsqrZ35			0.164
			(0.16)
itnsqrZ15			-0.034
			(0.05)

Z13,Z35 and Z15: *risk zone dummies:Z13=1, Z35=Z15=0 for subsample 1-3*
higher risk*: equals to 1 for properties in the high-risk zone;0 otherwise*
dist: *distance (metre/100)*
inter: *dist*(higher risk)*
inter^2 *(dist^2)*(higher risk)*
other controls: other controls*: foundation type dummies (basement, crawl, slab and partial basement) and built year dummies (in the 1930s to after 2000)*

* p<0.10, ** p<0.05, *** p<0.01

4.2 *Non-Parametric RD Design Estimation*

My non-parametric RD design estimates are based on observations that lie within 169 metres, 153 metres, and 169 metres of a risk boundary for subsamples 1-3, 3-5 and 1-5 respectively. This approach, in principle, controls for unobservable heterogeneity that may contribute to bias in my parametric estimates above. The boundary treatment effects are still not statistically significant for three subsamples, which is consistent with my result that consumers are rational and know earthquake risk vary continuously across boundaries.

Table 4. Local Polynomial Estimates (Non-parametric RD)

subsample 1-3				
Bandwidth 1.693 (metre/100)				
Method	Coef.	Std. Err.	z	P>|z|
(1-3 local linear)Higher risk	-0.046	0.033	-1.397	0.163
(1-3 local quadratic) Higher risk	-0.026	0.05	0.526	0.599
subsample 3-5				
Bandwidth 1.533 (metre/100)				
(3-5 local linear) Higher risk	-0.065	0.049	-1.33	0.183
(3-5 local linear) Higher risk	-0.058	0.067	-0.864	0.388
subsample 1-5				
Bandwidth 1.693 (metre/100)				
(3-5 local linear) Higher risk	-0.031	0.086	-0.359	0.72
(3-5 local linear) Higher risk	-0.019	0.098	-0.192	0.847

4.3 *Effect of Publication of the Hazard Map on Sales Price*

Interacting a post map dummy variable (indicating that local home buyers have access to seismic risk information in the form of the Hazard Maps) with the risk zone variable and distance allows me to estimate how the publication of the hazard map affects sales price. The estimates for the model (5.2) are in Table 5. Columns 2 to 5 display the estimates of models (1.2) to (4.2). None of the coefficient estimates of interactions between *distance* and *post map*, and interactions between *higher risk* and *post map* are statistically significant, indicating the publication of the Hazard Map does not affect consumers' willingness to pay is consistent with Clarke's results (2013). Perhaps, consumers already knew the earthquake risk information before the Hazard Map was published, and the Hazard Map, therefore, did not provide any new information on earthquake risk for consumers. Another possible explanation for the insignificant results may be that consumers do not care about the difference in earthquake risk when they buy or sell properties. Consumers not caring about the difference in earthquake risk may also be the reason of the insignificant coefficient estimates of distance when I interact distance and risk zone dummy with post map variable. The boundary treatment effects for three subsamples are still not statisitcally significant, supporting the result that there is no discontinuity in sales price at the risk zone boundary.

Table 5. Effect of Publication of the Hazard Map on Sales Price				
	model(1.2)	model(2.2)	model(3.2)	mode(4.2)
N	3555	3555	3555	3555
constant*Z13	11.855***	11.857***	11.856***	11.865***
	(0.03)	(0.03)	(0.03)	(0.04)
(higher risk)*Z13	-0.034	-0.032	-0.038	-0.07
	(0.03)	(0.04)	(0.04)	(0.05)
dist*Z13	-0.043*	-0.041	-0.039	-0.007
	(0.02)	(0.03)	(0.03)	(0.05)
(post map)*Z13	0.067***	0.063**	0.066***	0.064**
	(0.03)	(0.03)	(0.03)	(0.03)
(higher risk)*(post map)*Z13	0.018	0.012	0.019	0.011
	(0.05)	(0.05)	(0.05)	(0.07)
dist*(post map)*Z13	-0.002	-0.007	-0.002	-0.006
	(0.03)	(0.04)	(0.03)	(0.04)
constant*Z35	11.859***	11.875***	11.862***	11.818***
	(0.04)	(0.06)	(0.05)	(0.08)
(higher risk)*Z35	-0.101*	-0.113*	-0.104*	-0.051
	(0.05)	(0.06)	(0.06)	(0.09)
dist*Z35	0.03	0.057	0.032	-0.172
	(0.03)	(0.08)	(0.04)	(0.23)
(post map)*Z35	0.025	-0.034	0.025	-0.031
	(0.05)	(0.08)	(0.05)	(0.08)
(higher risk)*(post map)*Z35	0.047	0.093	0.048	0.045
	(0.08)	(0.09)	(0.08)	(0.10)
dist*(post map)*Z35	-0.009	-0.126	-0.01	-0.117
	(0.04)	(0.12)	(0.04)	(0.12)
constant*Z15	11.860***	11.855***	11.862***	11.848***
	(0.04)	(0.04)	(0.04)	(0.04)
(higher risk)*Z15	-0.013	-0.024	-0.024	-0.022
	(0.04)	(0.04)	(0.04)	(0.06)
dist*Z15	-0.045*	-0.053*	-0.032	-0.068
	(0.03)	(0.03)	(0.03)	(0.06)
(post map)*Z15	0.056*	0.052	0.055*	0.053*
	(0.03)	(0.03)	(0.03)	(0.03)
(higher risk)*(post map)*Z15	0.018	0.01	0.02	-0.022
	(0.06)	(0.06)	(0.06)	(0.08)
dist*(post map)*Z15	0.002	-0.005	0.001	-0.004
	(0.04)	(0.04)	(0.04)	(0.04)
(higher risk)*dist*Z13		-0.01		0.084
		(0.06)		(0.18)
(higher risk)*dist*post*Z13		0.026		0.017
		(0.08)		(0.23)

(higher risk)*dist*Z35	-0.03		0.179
	(0.09)		(0.25)
(higher risk)*dist*post*Z35	0.13		0.243
	(0.13)		(0.18)
(higher risk)*dist*Z15	0		0.001
	(0.00)		(0.00)
(higher risk)*dist*post*Z15	0		0.002
	(0.00)		(0.00)
(dist^2)*Z13		0.003	0.02
		(0.01)	(0.03)
(dist^2)*(post map)*Z13		0	0
		(.)	(.)
(dist^2)*Z35		-0.001	-0.167
		(0.01)	(0.16)
(dist^2)*(post map)*Z35		0	0
		(.)	(.)
(dist^2)*Z15		0.01	-0.007
		(0.01)	(0.03)
(dist^2)*(post map)*Z15		0	0
		(.)	(.)
(dis^2)*(higher risk)*Z13			-0.12
			(0.13)
(dis^2)*(higher risk)*post*Z13			0.008
			(0.17)
(dis^2)*(higher risk)*Z13			0.175
			(0.16)
(dis^2)*(higher risk)*post*Z13			-0.051
			(0.05)
(dis^2)*(higher risk)*Z13			0.002
			(0.07)
(dis^2)*(higher risk)*post*Z13			-0.07
			(0.09)

5. Conclusion

In this paper, using parametric and non-parametric RD design, I find that given the discrete earthquake risk ratings in the Hazard Map, the sales prices of properties in District of Saanich do not have a discontinuity at the earthquake zone boundary. When I regress CPI-adjusted sales price only on distance, risk zone dummies and the interaction between these two variables, I find that sales price negatively depends on the distance of properties. Take subsample 1-3 as an example. When a property is 100 metres closer to zone 3 (the relative high-risk zone), the sales price of this property decreases by 8%. However, when I include other characteristics of the property (property size, the number of bedrooms, post map dummy etc.), the negative relationship between sales price and distance disappear. Moreover, I do not find evidence of effect of the publication of the Hazard Map on the sales price. Rational consumers are expected to respond to new information. However, in my case, consumers have no reaction to the Hazard Map, suggesting either they might already know the information before the publication or they do not care about the difference in earthquake risk when they trade houses.

Using RD design, I estimate how sales price of homes change at the earthquake zone boundary and how it changes with properties' distance to the nearest seismic risk boundary. However, since I include all the observations that are close to or far away from the boundary, the unobservable heterogeneity problem could exist in the parametric RD design. One possible extension of this paper is to include a larger dataset with properties in risk zone 1 to 6 rather than only 1, 3 and 5. More recent data could be used to estimate better the effect of the publication of the Hazard Map on the sales price since not everyone had a computer back in 2003 and it usually takes some time for consumers to process the new information.

References

Angrist, J. D., & Pischke, J. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton [NJ]; Oxford [UK]: Princeton University Press.

Bayer, P., Ferreira, F., & McMillan, R. (2007). A Unified Framework for Measuring Preferences for Schools and Neighborhoods. Journal of Political Economy, 115(4), 588-638. doi:10.1086/522381

Bin, O., Kruse, J. B., & Landry, C. E. (2008). Flood Hazards, Insurance Rates, and Amenities: Evidence from the Coastal Housing Market. Journal of Risk and Insurance, 75(1), 63-82.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust Data-Driven Inference in the Regression-Discontinuity Design. *Stata Journal*, *14*(4), 909-946.

CBC News (2016). The 'Big One' near Vancouver you may not know about. Retrieved from http://www.cbc.ca/news/canada/british-columbia/fault-lines-big-one-1.3795416

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

Greenstone, M., & Gallagher, J. (2008). Does Hazardous Waste Matter? Evidence from the Housing Market and the Superfund Program. The Quarterly Journal of Economics, 123(3), 951-1003. doi:10.1162/qjec.2008.123.3.951

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.

Imbens, G. W., & Lemieux, T. (2008). Regression Discontinuity Designs: A guide to Practice. Journal of Econometrics, 142(2), 615-635.

Insurance Bureau of Canada (n.d.). Earthquake Insurance. Retrieved from http://www.ibc.ca/ns/home/types-of-coverage/optional-coverage/earthquake-insurance

Jacob, R., Zhu, P., Somers, M. A., & Bloom, H. (2012). A Practical Guide to Regression Discontinuity. MDRC.

Lee, D. S., & Lemieuxa, T. (2010). Regression Discontinuity Designs in Economics. Journal of Economic Literature, 48(2), 281-355.

Smith, V. K., & Huang, J. C. (1995). Can Markets Value Air Quality? A Meta-Analysis of Hedonic Property Value Models. *Journal of Political Economy*, *103*(1), 209-227.

Today's Earthquake (n.d.). Recent Earthquakes near Vancouver Island, Canada. Retrieved from http://earthquaketrack.com/r/vancouver-island-canada/recent

Van der Klaauw, W. (2002). Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression–Discontinuity Approach. International Economic Review, 43(4), 1249-1287.