Robin Antonsen & Stian Lien

Quick Clay Landslides & Housing Prices

The effects of quick clay landslide risk and landslides occurrences on housing prices in the event location, closely located, and remotely located municipalities

Master's thesis spring 2022
Oslo Business School
Oslo Metropolitan University
MSc in economics and business administration

Abstract

In this thesis, we employ a difference-in-differences approach to estimate the price effect of quick clay landslide risk on dwellings and the price effect of quick clay landslide occurrences on both dwellings at risk and not at risk of quick clay landslides. The price effects are estimated by including a dummy variable for being at risk of quick clay landslides and controlling for housing characteristics, unobserved time effects, and location effects. The estimations are based upon transactional data for dwellings within municipals that are either the place of the event, nearby municipalities, or remote municipalities. This thesis has included three such events: the quick clay landslide in Alta, 2020; Lyngen, 2010; and Gjerdrum, 2020. There are indications and evidence of price discounts for dwellings at risk. However, we also find contradictive results, giving us reasons for questioning the validity of the results. For the prices of dwellings after a quick clay landslide, we find similar results. However, the findings for the Gjerdrum case clearly indicates an additional negative price effect on dwellings at risk after landslide events compared to dwellings not at risk, which may be due to a salience effect of such events.

Preface

Writing this thesis during the last semester of our studies has been both challenging and time-demanding, to say the least. It has pushed us to apply valuable knowledge acquired both recently and throughout our studies. All this hard work has now accumulated in what is our greatest academic achievement, and we are happy to say we have grown due to this experience.

Although the research was executed and presented by us, we must highlight the excellent supervision provided by our supervisors at Housing Lab Oslo Metropolitan University, Ph.D. candidates Andreas Eidspjeld Eriksen and Nini Barth. Your creative and professional input, availability, and honest feedback have pushed us to be more efficient during this entire process. It has been invaluable, and we greatly appreciate it.

Aside from our supervisors, we would like to thank the following people.
First, a big thank you goes to the remaining people at Housing Lab for going to great lengths to assist us. Without your help and resources, we would not have achieved these results.

Second, Andreas Jensen and Eiendomsverdi AS provided us with the necessary data. Without it, this thesis would not be feasible to write, and we genuinely hope you find the results interesting. Your help has been highly valued.
Third, thank you to our partners for their support and understanding throughout the countless hours of writing.

Last, we would like to thank the Faculty of Social Sciences at OsloMet for believing in this project and their support through a scholarship. We hope to meet the hopes and expectations of everyone involved.

The errors that may occur within this thesis are our responsibility.

Table of content

List of Figures

List of Maps[1]

List of Models

List of Tables[2]

[1] This list excludes maps included in appendices. See appendices A
[2] This list excludes tables included in appendices. See appendices B

1. Introduction

There is no denying that the economic consequences a household is potentially exposed to may be vast in the event of a quick clay landslide; from here on interchangeable with the abbreviation QCL. According to Flavin & Yamashita (2002), dwellings are considered to count for the most significant fraction of the household`s assets. In addition to housing being one, if not the most important investment most households make in their life, Norway has a relatively high amount of dwellings built on areas with quick clay according to Mordt et al. (2021), attributing some exogenous risk. In the Norwegian housing market, there is a tendency toward a preference for households owning their dwelling rather than renting (Langberg, 2016). In 2016, 84.4% of Norwegians lived in households where they owned their dwelling compared to countries such as Switzerland with 44.5%, Germany with 52.5%, and Austria with 57.2% of their population owning their own home. Since 2016 the tendency has become weaker, and in 2021 the reported percentage was 7.4% (SSB, 2022).

Moreover, pre-existing papers studying the effect of a natural disaster on housing prices indicate (a) a lasting, (b) a short-term drop-in, or (c) no significant effect on housing prices regionally after such events as landslides, and floods, when housing areas are exposed to damages. In the pre-existing papers, there are different results in how the prices are affected by the event, but we expect there may be a long-term consequence for housing prices for all dwellings built on quick clay, not only those in the proximity of the event. However, it is essential to notice that though we do expect similar results, pre-existing papers primarily consist of analyses on housing prices when exposed to exogenous shocks in the forms of other natural disasters, such as forest fires, earthquakes, and floods. There is far less publicized research conducted for landslides, hereunder QCLs. One possible explanation for changes in price in the event of a natural disaster been revealed by Garnache (2020) and Naoi et al. (2009), a paper we will discuss in detail later, where they point out *salience* as a plausible explanation of why the prices changes after such events. Salience can, according to Taylor and Thompson (1982, p.175), be defined as "…the phenomenon that when one's attention is differentially directed to one portion on the environment rather than to others, the information contained in that portion will receive disproportionate weighing in subsequent judgments". From here, we can interpret salience regarding our thesis as a noteworthy and visible change in the perception of the risk the households are exposed to. Another way we can look at it is if the attitude towards the risk of damages to housing if the event was to occur is being updated as they are reminded what the risk entails.

In this thesis, we conduct a difference-in-differences (DD) approach like the designs employed by the likes of Kiel and Matheson (2018) and Kim et al. (2017) to assess: (a) whether there is a price effect for dwellings at risk of a QCL, (b) whether there is a price effect for dwellings at risk of a QCL after the occurrence of a QCL, and (c) whether there is a price effect for dwellings not at risk of a QCL after the occurrence of a QCL. The two prior research papers, among others, will be discussed in detail in section 2 of this thesis. For the remainder of the thesis, we will focus on three QCLs, selected from Norway`s long history of QCLs. Each case will be treated separately and give us a more detailed indication of how the housing market responds to such events. An introduction to each of the landslides can be found in section 1.2.

This thesis contributes to the literature in two ways. First and foremost, with essential insights into the effects of being at risk and how a QCL affect Norwegian homeowners' assets, measured in the price of dwellings. We think the insights from this thesis may be of great use to inform entities such as homeowners, policymakers, and insurance companies about different mechanisms regarding housing prices and the perception of being at risk of a QCL. This additional information may contribute to home buyers' decisions regarding purchasing and the timing of entering the housing market, and the areas they may choose to purchase a home.

When a dwelling is up for sale, the seller is, by law, obligated to inform the buyers about all the information that may be considered significant knowledge about the property in question (Fæste, 2017). In addition to the obligation to inform, there are publicly available maps from NVE over risk areas in Norway. Regardless of this, we think that households, specifically private buyers in the housing market, do not tend to focus on whether their present or future home is at risk of a QCL but rather on more visible and tangible attributes. Therefore, we hope this thesis will contribute to making households more aware of the risk that comes with such housing attributes as we study. Regarding the policymakers, we think the thesis can contribute to the discussion of whether there is a need for more strict and better implementation of such policies within areas consisting of quick clay sediments. According to Botzen et al. (2019), such policies may also help mitigate the impacts of such events on both the economy and society. Secondly, this thesis is based upon a larger dataset than most pre-existing papers on similar topics.

Probably one of the most important sources of bias when estimating price changes in the housing market is that there are price effects based on where the dwellings are located. An extreme example of this is rapidly climbing and expensive dwellings in cities like Oslo and relatively more inexpensive dwellings in rural areas. This challenge is dealt with by controlling for location-specific effects by including a dummy variable for postcodes.

As the three cases are analysed separately, we also present the results of these analyses separately in section 6. For this reason, there are mixed results across the three cases, which might be somewhat surprising, as we expect the markets to behave similarly. The most significant differences in results are between the case of Gjerdrum, which shows expected results of there being a price discount for dwellings at risk of QCLs up to approximately 76,000 NOK and indicates that these prices drop further after a QCL. Somewhat surprisingly, dwellings not at risk also see a drop in prices after a QCL. Contrary to this, results indicate that dwellings at risk included in the Alta and Lyngen cases experience a price premium of approximately 70,000 and 300,000 NOK, respectively. We find it reasonable to believe that there are amenity effects that create bias in these estimates. Furthermore, we present results of a drop in sales volume per quarter for dwellings at risk for all three cases, which curbs the results of the price effects.

After the QCLs in Lyngen, Alta, and Gjerdrum, policies for building on sediments with quick clay have been up for debate by both local and national newspapers. There are many reasons, whereas we believe that the seriousness of the landslide in Gjerdrum plays a considerable role. More of this will be covered in section 1.2. It is also important to point out that not all areas are mapped, so there might be larger areas with sediments such as quick clay that we are not familiar with, hence areas where we do not know whether the policies should be implemented.

In the event of a landslide, the households may have secured their monetary value by getting their assets replaced by the insurance company. In addition to the insurance companies offering their products to the households, they also hold an important humanitarian role when such events occur. For example, the Norwegian insurance company Fremtind played a considerable role in the community experiencing loss in the aftermath of the landslide in Gjerdrum (Fremtind, 2020). They sent out representatives to help the people affected by the landslides find a place to stay after being evacuated, report the damages to the insurance company, and guide them during a difficult time.

We need to point out that the impact on households' economies of such events goes beyond the economic framework addressed in this thesis. In addition, it is vital to remember and consider that the people affected by such events may experience losses beyond their economies. Worst-case scenarios, such as the QCL in Gjerdrum, show the extent of the importance of the subject, especially for the likes of the policymakers.

1.1. A general introduction to quick clay

According to the Norwegian Geotechnical Institute, also known as NGI, quick clay is a specific type of clay mainly found in Norway and Sweden. Even though quick clay is also found in other regions, such as Finland, Canada, Russia, and Alaska, it is not as common as in the two Scandinavian countries (NGI, n.d.a). The report publicized after the examinations of the landslide in Gjerdrum also revealed that the risk of quick clay in general in Norway is much more present than what we, as a society, should accept (*NOU 2022:3*, 2022). The name, "Quick clay", derives from the clay's trait, being that the clay collapses and starts flowing like a liquid when overloaded, both naturally and by man-made actions (Spjeldnæs, 2021).
The clay itself is found below marine level, which is defined by "the present elevation of where the sea level was at end of the last ice age" (NGI, n.d.a). Today the marine level is measured at about 220 meters above the current sea level. When the ice melted, tiny clay particles flowed with the melted water and sedimented in the marine environment and got mixed in with saltwater along the seashore. As the salt water mixed with the clay and the other sediments, the salt made the clay particles form a highly unstable structure, later known as quick clay. The process of how quick clay

came to be formed can be part of the explanation of why quick clay is primarily found in countries in or close to Scandinavia, as Scandinavian countries were covered in an ice layer that was about 3000 meters, about 20 000 years ago.

After the more significant QCL in Rissa in Trøndelag, demanding one life and creating considerable material damage in 1978, NVE started with national mapping to identify areas that could be more prone to more significant QCLs. NVE wanted to map these areas to prioritize better how to secure areas for erosion and reduce the risk of a landslide being triggered (NVE, 2021). In addition to better prioritizing and reduction in risk, the municipalities and other entities will benefit from having a better tool at their hands when planning for building dwellings and other infrastructural construction. The mapping consists of the degree of danger, measured in the probability of a landslide, consequences of a landslide, and degree of risk of a QCL, which is measured by the two latter means of measure (NGI, n.d.b).

At first, the methods developed were adapted to identify the more significant and most dense masses of marine clays, primarily in Trøndelag and the eastern regions of Norway, the regions being most prone to QCLs at the time (NVE, n.d.). Even though the purpose of these maps was to identify the areas where there is a more significant risk of a larger QCL to occur, there are still areas below marine level that have not yet been mapped. Therefore, the Norwegian Water Resource and Energy Directorate advise people to be cautious when below marine level. There is also important to remember that these mappings of areas with a risk of quick-clay landslides do not exclude the outlying areas.

1.2. A general introduction to each case

For this thesis, we have chosen three cases of QCLs. When selecting the cases for our analysis, we studied the list of QCLs in Norway found on NGI's website (NGI, n.d. The criteria were that the landslide had a significant size, that at least one dwelling was affected by the masses, and that the landslide happened within the last 15 years.

The first QCL we include in this thesis is the large QCL by the coast at Solhov in Lyngseidet, Norway, on the 3rd of September in 2010. The landslide did not result in fatalities, but the landslides' immense forces took two houses. There was immediately initiated a rescue operation, where one 60-year-old man was rescued through the loft window at sea. The man did not get severe injuries. In addition, people living in neighbouring houses were evacuated (Tiltnes, 2010). After the landslide, it was concluded that man-made actions were the cause of the landslide to occur (Matre et al., 2010). Reportedly, at least 1000m3 masses of rock were dumped on a new industrial property by the seashore without permission from the municipal. The entities involved disagreed on whether the permission was given orally, but such permission was never given as a formal approval. In addition to the new industrial area, there had also been the construction of a new nursing home, road work, and leakage from water soil in the nearby area. Even though all these events were part of the explanation for why the landslide occurred, the dumping of the rock by the seashore is, according to the reports by NVE, Jernbaneverket, & Statens vegvesen, the triggering cause (NVE et al., 2013, p. 15). The event and circumstances were covered both by national and local media after the event.

The second landslide is the large QCL at Kråknes in Alta, Norway the third of June 2020. Fortunately, the landslide did not result in any fatalities, but the landslides' immense forces took eight buildings. The landslide was about 650 meters and 40 meters in height (NVE, 04.06.2020). After the landslide in Alta, the examination of the events conducted by the Norwegian Water Resources and Energy Directorate, more commonly known as NVE, found that man-made actions were one of the main reasons for the landslide to occur (NVE, 2021). In 2015, one of the residences conducted construction for their cabin and dumped 80 truckloads of rock on their property. The dumping, in addition to several days of intense melting of snow in the area are said to be the reasons for the landslide to occur. The event and circumstances were covered both by national and local media after the event.

The third and final landslide included in this thesis is the massive QCL in Ask in Gjerdrum, Norway on the 30th of December in 2020. Large areas in Ask were evacuated instantly, and over 1600 people were affected by the landslide in total, eighter by the landslide itself or through the evacuation. A thorough rescue operation was initiated in a short amount of time, but unfortunately, as many as ten people died. The QCL in Gjerdrum is one of the worst cases throughout history. In February

2022, Gjerdrum municipal was charged by the police for not following the Act relating to protection against natural damage, stating that the Norwegian municipalities are obligated to take precautions against natural perils (Gjerdrum Kommune, 2022). In this case, the police claim that the municipality did not have systems for receiving alerts, seeing a connection between the many alerts received, and that they did not initiate measures to avoid erosion in the nearby water stream. The erosion in Tistilbekken is said to be the causative reason for the quick clay not being able to hold its solid state (Gjerdrum Kommune, 2022). In addition to being one of the most catastrophic QCLs throughout the country's history and has been a highly publicized event for the last one and a half years.

The circumstances for why the three landslides occurred when they did may imply that the policies for building on sediments such as quick clay are not well constructed or implemented. Also, the municipalities' tendency to presumably not act by the law is alarming.

The rest of the thesis is structured as follows: section 2 will be a review of similar, pre-existing research. Section 3 will discuss the econometric strategies, theoretical background, and foundations for the DD design. Section 4 will describe the initial data, how the data was trimmed, and then a final description of the trimmed data. Section 5 presents the empirical conditions for answering the research question, while section 6 presents the results from our DD approach. The results are to be found in table 6.1-6.9. Also, we will discuss the results and the possible mechanisms that potentially can explain the estimates in this section, while we will conclude the conducted research in section 7. The thesis's maps, models, figures, and tables follow a section-based structure. For example, each map, model, figure, and figure in section 3 will be named with the numeric 3.X.
For maps in the Appendices, see appendices A. Maps. For tables in the appendices, see appendices B. General summary statistics.

2. Literature Review

Because quick clay is not the most common sediment of foundation measured on a world basis, the research on housing prices and quick clay is scarce, if not more or less non-existing, especially not regarding the Norwegian housing market. For the literature review, we have therefore chosen studies with similar traits, hence studies on changes in housing prices before and after a natural disaster. The review of the pre-existing literature follows a chronological order from 2009 to 2017.

The first paper we want to highlight was written by Naoi, Seko & Sumita (2009), who studied the effect on housing prices of earthquake risk before and after massive earthquakes. Their approach utilised a DD approach using an earthquake risk probability variable that interacted with a post-earthquake dummy variable in a hedonic price regression model. The post-quake dummy has the following specification "...equals one if an earthquake event occurred in the previous year and zero otherwise." (Naoi et al., 2009, p. 665). The earthquake risk variable contains the probability of an earthquake occurrence "... with a given seismic intensity at a fairly disaggregated geographical level (1km x 1km grid cells)." (Naoi et al., 2009, p.661). When conducting their study, they chose not to include all the data for earthquake occurrence probability by dropping all risk measures for earthquakes with a ground motion less than seismic intensity 6. By annualising the provided 30-year occurrence probability and aggregating the data to construct city-level probabilities, the authors argue that this result in an earthquake risk variable approaching heterogeneity (Naoi et al., 2009, p. 661). Also, the models include time dummy variables to control for "...unobserved time-varying effects that change year by year such as housing market changes after the earthquake." (Naoi et al., 2009, p. 665).

In total they estimated four baseline models, which did not include the distinguishment between observations before and after the earthquake. Out of these four, there were two models attributed to renter holds, and two models for homeowners, from which there is one OLS model for each, and an additional OLS model where they control for respondent fixed effects. The effects that are controlled for are characteristics of the survey respondents that yielded the dataset, "...and dummy variables for housing types, prefectures, and city-sizes..." (Naoi et al., 2009, p. 665). The OLS model for renter households shows that there are results that

indicate “...that earthquake probability has a significant negative effect on housing rents” with -10,49% (Naoi et al., 2009, p. 665). In contrast, the FE model for renter households indicates negative results that are not statistically significant. For homeowners, the OLS model indicates adverse effects on the price of the dwellings, but only the FE model shows a negative coefficient that is statistically significant with -9,6% (Naoi et al., 2009, p. 665).
Following, the regression analysis of the estimated hedonic regression model with the DD approach shows significantly negative results on housing prices attributed to the interacted term in the model. “This suggests that massive quakes in neighbouring cities/towns changed the perception of earthquake risk for renter households and homeowners.” (Naoi et al., 2009, p. 665). Overall, the average rent prices were reduced by 16%, with a fall of 13% in housing values. Also, there was a 0.2% increase in the annual earthquake probability in the post-quake period. The authors argue that the most plausible interpretation of the results “... is that households are initially unaware of, or at least, underestimate the earthquake risk in the pre-quake period.”. Following this initial interpretation, renter households and homeowners dramatically adjust their perception of earthquake risk after a massive earthquake in neighbouring cities/towns. This interpretation is made because of the combination of the negative coefficient of the interaction term (DD) and the indication of a non-significant negative coefficient of the earthquake risk variable. (Naoi et al., 2009, 665-666).

The second paper was written by Atreya, Ferreira & Kriesel (2013), who used a spatial DD model to estimate the effect of flood(s) on prices for houses. Similar spatial DD approaches have also been utilised by Bin and Landry (2012) and Kousky (2010), among others, to investigate the “information effects of a natural disaster” (Atreya, et al., 2013, p. 582). The idea of using this method is to determine the effect of the flood on property prices. Two dummy variables were incorporated into their model to measure flood risk: one for the 100-year floodplain and one for the 500-year floodplain. This risk is captured by floodplain location, and thus, the treatment group are properties that fall within the floodplains, and the control group is the properties outside the floodplains. The definition of treatment and control group found in this paper is quite similar to how we define the treatment and control groups for this thesis, presented in detail in section 5.1. Their model uses a dummy variable to indicate whether a sale occurred after the flood. In this case, the flood occurred in 1994, after the Flint River overran, which caused a major flood in southwestern Georgia. Using this setup, the researchers create interaction terms between the risk variables and the after-flood dummy to examine the effect of the flood on property prices in the area studied. The spatial element for this model was introduced by accounting for "...spatial dependence among neighbouring properties via a combination of spatial lagging of the dependent variable and correcting for autocorrelation in the error term." (Atreya et al., 2013, p. 578). We resign from explaining this further, as this is something we will not incorporate in our thesis. The model was applied to a dataset of property sales data in Doughtery County from 1985 to 2004 (Atreya et al., 2013, p. 585-586). The results show a weak significant finding of a 9% price discount on the properties within the 100-year floodplain before the 1994 flood, which was not the case for the properties within the 500-year floodplain. Based on their findings, they argue that the buyers of the properties within the 500-year floodplain were not aware of the flood risk; thus, the prices were not discounted. For a model with a linear time decay function right after the flood, the properties in the 100-year floodplain experience a 32% discount in price. This discount is calculated from a 9% baseline discount estimate from property within a floodplain, which means the flood attributes a 23% discount. (Atreya et al., 2013, p. 589). The flood risk time decay function shows statistically significant results for the discount decaying rapidly. After four years, the price discount vanishes before it turns positive nine years after the flood (Atreya et al., 2013, p. 593).

Third, we have chosen a paper written by Kiel and Matheson (2018), where we are introduced to a hedonic pricing model with an incorporated DD approach for dwellings in a multitude of risk levels for forest fires. Their paper conducts a study of the Fourmile-Lefthand Canyon Forest fire, spreading September 2010, on housing prices in vulnerable neighbouring areas that were not directly affected by the fire. They approach their research using a DD method, hence examining housing prices across a treatment and a control group before and after a given event. This method is employed by including dummy variables for the four risk levels and a dummy that is equal to one if the observation was made after the forest fire. Their research is based on a trimmed data set containing 9377 single-family housing transactions between January 2009 and April 2012 in Boulder County, where all homes were bought from the Warren Group. Their level of risk variable has four values: “very high”, “high”, “medium”, and “low”.

Kiel & Matheson (2018) created a general model using a hedonic pricing approach for their estimations. The general model was later used to specify two different models, which will be discussed later in this thesis. This model consisted of variables for housing characteristics such as the age of the dwelling, which includes age2, the number of bathrooms and bedrooms, lot acreage, and the square footage of the home’s living area. Their model also includes neighbourhood characteristics based on where the dwelling is located with included variables such as the percentage of the population living in what is defined as poverty, the composition of the population and a dummy variable for what city in which the dwelling is located. The specification of the risk of exposure to a forest fire for the given dwelling was denoted using dummy variables for low, medium, high, or very high risk. They also included a dummy with a similar specification to those presented earlier – the dummy containing the value 1 when the transaction is registered after the forest fire, but 0 if not.

In their model, Kiel & Matheson (2018) controlled for the level of risk within each area. We found this practically impossible for our thesis due to the time constraints as it would take an ambitious amount of work. There are feasible ways of doing this by plotting all the coordinates from the dataset, cross-referencing these with the NVE Quick clay map, and categorising each observation across five different levels of risk. In addition to the time constraint, this could also lead to us ending up with a smaller sub dataset when conducting the DD approach. Further, they use the general model described above to create two different models, where model 1 groups together all dwellings with a risk dummy that is equal to one if the risk level for the given dwelling is above low; hence in the categories “very high”, “high”, and “medium”, and zero otherwise. The results of model 1 suggest a fall in prices for dwellings in risky areas. Dwellings in risky areas seems to sell for 5.6% less than Dwellings in non-risky areas; else held constant. However, the statistics are not statistically significant. If taking the negative interaction term between risk and after into account, the results show that dwellings in risky areas sold for around another 5% less after the fire. In similarity to the latter discussed coefficient, this did not yield a statistically significant result, but it still does suggest a change in the perception of risk after the event.

In model 2, the risk measures were included by making dummy variables for each level of risk instead, meaning they in had the dummy variables: “very high risk”, “high risk”, “medium risk”, and “low risk”, where the low-risk variable served as the omitted variable in the regression model. This model found that dwellings in the very high-risk area show a statistically significant impact on the sales of dwellings, estimating that the dwellings in this area sold for a 17.5% higher price than dwellings in the low-risk area. On average, the dwellings in the medium risk area also show a statistically significant change, as they sell for 7.9% less than the dwellings in low-risk areas. Also, the dwellings in high-risk areas are likely to see a reduction in prices by 8.7% compared to those in low-risk areas. However, the results are not statistically significant when controlling for other aspects of the dwelling and neighbourhood effects. The conclusion is that there is no impact on prices by being in the high-risk area. After the fire, dwellings in the very high-risk area saw a reduction in sales prices of 21.9% compared to those in the low-risk area. This measure is statistically significant. However, this seems a little suspicious as only 0.36% (~34 homes) have this specification.

One issue that may arise when not having an extensive number of observations is that the estimations are statistically significant, while they may still be biased. Having few observations when estimating coefficients for an Ordinary Least Square model yields higher uncertainty as less observations are included when the coefficients are estimated (Tufte, 2020). Because the sample size for the data used in our thesis is bigger than the sample size for similar, pre-existing research, we do not deem this is a likely problem for us. Regarding the significant change in sale price at 21.9% for the very high-risk areas compared to the low-risk areas, they had 17 observations before and 17 observations after the event. Meanwhile, the estimators indicate a significant drop in prices following the event in interest. Neither of the other tests performed across the risk classifications yielded any results indicating a significant difference in housing prices. We deem it strange that the only significant result is yielded from the smallest sample within their data.

From their study, we want to point out that the dwellings included in their data sample were not affected directly by the fire but rather dwellings located in the fire's proximity. This is important for us because we deem dwellings at risk of a QCL as directly affected by the event, as the dwellings taken by the landslides cannot be sold, hence not part of the transactional data. A crucial error within this paper is Kiel and Matheson's (2018) omitting of the assessment of whether the data included in the treatment and control group fulfilled the requirements for the parallel trend's assumption, also known as the common trends' assumption. Not having included an assessment for this assumption leaves us with many questions regarding their work, as this is the most central assumption and prerequisite to employing a DD approach. Did the two groups follow a similar trend before the fires, or did one group's trend over time deviate from the other in the first place? If the trends were to deviate, the estimators would no longer measure the price effects of the event. We will discuss the theoretical background of this assumption further in detail in section 3.1.1

The fourth and final paper in our literature review is written by Kim, Park, Yoon, and Cho (2017), who conducted a study on housing prices utilising a similar approach to the previously presented literature. Their paper aimed to determine whether the landslide in Woomyeon National Park (WNP) in Seoul, South Korea, influenced housing prices for dwellings located within 1 kilometre from the park. The study found that the dwellings located close to the national park, defined by 1 km, fell by 11.3% after the landslide in July 2011 (Kim et al., 2017)

Their model estimated the logarithmic transformation of the transaction price, adjusting for trends with the housing market index for Seocho-gu, one of the 25 districts that constitute south Korea's capital Seoul, with monthly time variables. For estimating the logarithmic transformation of the price, they included a variable for housing characteristics such as usage area, first floor, age, and squared age. Their model also included location characteristics such as distance to the closest school and subway, a dummy variable representing each month, a dummy variable for the dwelling's distance from WNP, which measures both hazard and amenity effects of the WNP, and time dummies for measuring short-term and long-term effects of the landslide in interaction with the dummy for distance. The short-term effects are measured with a variable that contains the value 1 if the transaction is registered between August 2011 and July 2012 and 0 if not. The long-term measures were interpreted with the dummy containing the value 1 if the transaction is registered after August 2012 and 0 if not.

We found certain treats to their study quite interesting when reading their work. First off, their study only included 5,758 housing transactions from 2008 to 2014 and for dwellings within 1km of the WNP. Another interesting detail is that their study contained data fulfilling the assumption of homogeneity within the type of dwelling included. For fulfilling the assumption, the dataset only contained transactions across 212 apartment complexes, hence high-rise condominiums, and multifamily housing with five stories or more. For our thesis, we will control for the housing type, hence fulfilling the same assumption. Their study also found the parallel trends assumption to be fulfilled for their chosen treatment and control group, indicating that the effect of different time trends is minimal.

3. Theory

This section will dig further into the different theories and frameworks that we will use to build our models and estimate how a QCL affects housing prices. First, we will introduce you to the econometric theory for the DD method and what assumptions need to be fulfilled. Next, we will look at the theory for hedonic pricing and how this can be useful for this thesis. We will also investigate market efficiency and discuss how this is relevant to the housing market.

3.1. Difference-in-Differences

A DD model is an approach to fixed effects estimation, which lets us investigate the regressor or variable of interest when this varies at an aggregate or group level (Angrist & Pischke, 2009, p. 227). This method is applied to sets of groups, where certain groups are exposed to a treatment of interest and other groups are not, making this the perfect suit for our thesis (Angrist & Krueger, 1999, p. 21). For example, a treatment may be a landslide impacting a particular region in a country, such as a municipality, or smaller areas within a municipality.

Running a DD model is feasible when the treatment is as-if randomly assigned, conditional on some observed control variables, such as the size of an apartment (Stock & Watson, 2020, p.492). If we are to conduct a DD model, we can do this by using regression estimation. This is accomplished by using two dummy variables in the regression for the treatment and control groups, where the dummy takes the value of 1 for the treatment group and 0 for the control group. The second dummy to be included in the model is a time dummy that takes the value of 1 if the observation takes place after the event we are interested in, in our case, the three landslides. This variable will have a value of 0 if the observation is registered before the treatment. (Wooldridge, 2018, p. 436) apply this method in a regression equation that we have altered to match better for this thesis. To do so more intuitively, let us say that we are to make a simple DD model for a landslide in Gjerdrum. We can then specify model 3.1

Model 3.1: A simple DD model for a landslide in Gjerdrum municipal

$$P_{Gjerdrum} = \beta_0 + \delta_0 after_{Gjerdrum} + \beta_1 LR_{Gjerdrum} + \delta_1 after_{Gjerdrum} \cdot LR_{Gjerdrum} + \epsilon_{Gjerdrum}$$

Using the reasoning of Wooldridge (2018, p. 436), we can say that $LR_{Gjerdrum}$ is the dummy variable denoting the treatment group, being sold dwellings in Gjerdrum at risk of QCLs. $after_{Gjerdrum}$ is the time dummy, which equals one if the observation is registered after the event, in this case the landslide in Gjerdrum, which is the time of treatment. Following this model, the interaction term $\delta_1 after_{Gjerdrum} \cdot LR_{Gjerdrum}$ denotes the observations from Gjerdrum built on quick clay sold after the landslide.

Taking advantage of a regression approach will, according to Angrist and Pischke (2009, p.233-234), allow us to estimate the DD estimates and standard errors, as well as it will make it easy for us to include several control groups, such as municipals and additional periods. This will be especially useful in our model as we want to include several control variables. Using this method, we can then analyse the data by comparing differences by a regression that includes the treatment indicator and control variables, such as housing characteristics like usage area, building age, etc. The differences estimator is then According to Stock & Watson (2020, p.476-477) “the difference in the sample averages for the treatment and control groups”. These differences can be computed by regressing the outcome variable, $P_{Gjerdrum}$ on the binary treatment dummy $LR_{Gjerdrum}$. Further, by including control variables that help explain the variation in $P_{Gjerdrum}$ the standard error of the regression is reduced, allowing us to obtain more precise estimations. To include these control variables, they must be pre-treatment individual characteristics and the treatment dummy, $LR_{Gjerdrum}$, must be randomly assigned, so that the error term, $\epsilon_{Gjerdrum}$, satisfies the conditional mean independence condition.

By using the DD approach, we can estimate the effect of changes in, for example, the economic environment or changes in government policy. Such changes in the economic environment may be external or exogenous events such as landslides. The practical usefulness has made the method a widespread approach within econometric research in the last fifty years or so (Angrist & Krueger, 1999, p. 21). Thus, we can estimate the treatment effect, precisely the effect of a QCL on dwelling prices. Similar cases have been presented in various studies, like the papers included in the literature review. Here, the researchers analysed data on dwellings from natural experiments caused by exogenous events; hence they conducted what is known as a natural experiment (Rosenzweig & Wolpin, 2000, p. 828). If such an event were to occur, two groups would be exposed: a control group and a treatment group. Thus, we have one group that has been impacted by the event, referred to as the treatment group, and another group that the event has not impacted, that we will refer to as the control group. The two groups have not been randomly chosen for our thesis, as we stated must be the case earlier. Therefore, there may be systematic differences between the two groups, which must be controlled for later. To control for these systematic differences, we include data from before and after the exogenous event; hence we have four groups to consider: the control group before and after the event, and the treatment group before and after the event (Kiel & Matheson, 2018, p.5).

To better explain how the logic of the DD approach works, we can utilize model 3.1 and create the following example. Suppose there are two groups and two time periods. In the first period, neither group is exposed to a treatment. In the second period, however, one group is exposed to a treatment, whereas the other group is not (Schwerdt & Woessman, 2020, p.13).

Let us say we have transactional data on housing prices from two municipalities over a two-year period, where the first observation is registered at the beginning of 2020, and the last observation is registered on the last day in 2021. The period averages can then be arranged quarterly and monthly accordingly.
Let us say a QCL occurred in one of these municipalities 1st of January in year two. We now have two periods which in this case yields $P_{Gjerdrum}2020, and\ P_{Gjerdrum}2021$ where the first denotes the price for observations recorded before the landslide, and the second denotes the price for the period after the landslide occurred. At the same time, we have the treatment group, with observations within the municipality where the QCL occurred, and the control group for observations within the municipal where there was no landslide. Alternatively, we can define the treatment group as sold dwellings built on or at risk of being affected by the landslide within both municipalities and the control group as sold dwellings that are considered not to have any risk of the landslide being attributed to them within both municipalities. From here, we can use DD estimation to estimate the different effects of the

landslide on housing prices for the two groups, using the general regression equation from Model 3.1.

In model 3.1, $P_{Gjerdrum}$ is the outcome variable of interest, and δ_1 is the DD estimator, which measures the effect of the exogenous event (Wooldridge, 2018, p. 434).

When conducting the DD approach, we want to estimate the parameter δ_1, which in the standard approach can be done in two ways: firstly, we can compute the differences in averages between the treatment and control groups in each time-period, and then difference the results over time. Alternatively, we can compute the averages over time for each treatment and control group and then difference these changes (Wooldridge, pg. 434, 2018). The two alternatives to conducting the standard approach of the DD estimator are illustrated in table 13.3, found in Wooldridge (2018, p.435)

For the first alternative, we compute the differences in averages between the treatment and control groups in each period, and then differencing the results over time, hence resulting in the model presented below.

Model 3.2: Averages between treatment and control group for each period

$$\hat{\delta}_1 = \left(\bar{P}_{Gjerdrum\_2021,T} - \bar{P}_{Gjerdrum\_2021,C}\right) - \left(\bar{P}_{Gjerdrum\_2020,T} - \bar{P}_{Gjerdrum\_2020,C}\right)$$

Where $\bar{P}_{Gjerdrum\_2021,T}$ and $\bar{P}_{Gjerdrum\_2021,C}$ is the average of the outcome variable of interest in 2021, for the treatment (T) and control group (C), and $\bar{P}_{Gjerdrum\_2020,T}$ and $\bar{P}_{Gjerdrum\_2020,C}$ is the average of the outcome variable of interest in 2020, for the treatment (T) and control group (C).

By simply rearranging the equation above, we can get the following equation found in model 3.3, which by construction, yields two different interpretations of the DD estimator, with the same estimate of $\hat{\delta}_1$.

Model 3.3: Averages between periods between the treatment and control group

$$\hat{\delta}_1 = (\bar{P}_{Gjerdrum\_2021,T} - \bar{P}_{Gjerdrum\_2020,T}) - (\bar{P}_{Gjerdrum\_2021,C} - \bar{P}_{Gjerdrum\_2020,C})$$

3.1.1. Parallel Trends Assumption

There may be an issue with fulfilling the *parallel trends assumption* regarding a two-group, two-period DD setup. This assumption says that any trends in the outcome variable, $P_{Gjerdrum}$, will trend at the same rate in the absence of the event of our interest. A threat to the identification strategy used when conducting a DD will arise by violating this assumption. The reason for this is that the DD estimate is simply the difference in the estimated trends for the treatment and control groups. Of course, there may be fundamental differences between the control and treatment groups because of differences in areas, housing charachteristics, etc., within the municipals. If the assumption holds, we can interpret the difference in trends between the groups as the actual difference induced by the treatment. To investigate whether the parallel trends assumption holds, one would need data on several periods (Angrist & Pischke, 2009, p. 231). As multiple periods of data are needed, there are multiple points in time with data that we can draw a fitted line between, which deems the trend line. To check whether this assumption holds for our data before conducting the DD approach, we can plot the linear group trends together and see if the trend lines are parallel before the given event. If this is the case, the assumption holds. In addition, such plots can also be used to visualize the price changes, both before and after the event, by looking at how the trend lines for the treatment and control group shift. This approach to assessing whether the treatment and control groups follow a parallel trend over time will be conducted in section 5.2.

3.2. Hedonic Pricing

It is natural to include control variables in the DD regression approach to include variables on housing attributes, akin to the essence of a hedonic pricing approach, which is commonly used in combination with a DD model.

When it comes to assigning prices to or valuating different goods, such as housing, there are many ways to go. One of the advanced methods to evaluate or appraise real estate presented by Pagourtzi et al. (2003) is hedonic pricing. Early contributions to the theory of hedonic prices come from Sherwin Rosen (1974). He

defines hedonic prices as “the implicit prices of attributes”. Rosen also goes further into detail by saying that hedonic prices “are revealed to economic agents from observed prices of differentiated products and the specific amounts of characteristics associated with them” (Rosen, 1974, p.34).
By utilizing a hedonic pricing approach, one can decompose attributes, such as the number of rooms, square feet, age of the housing, and lot size, to better analyse housing prices by examining the effect of these attributes on the price of the housings. Typical attributes for housings are variables such as type of housing, the number of rooms, size, building materials, etc. Using an approach to a hedonic pricing model to estimate housing prices, we can analyse attributes which does not have an observable market price by itself (Pagourtzi et al., 2003, p.395-396).

According to Selim (2009, p. 2844), most studies conducted on the pricing of the housing market are based on a multiple regression analysis. The definition of a multiple regression analysis is simplified by Braut and Dahlum (2021) as a statistical method of describing the coherence between one or more independent variables and a dependent variable (Braut & Dahlum, 2021) and is, therefore, an appropriate method to use if the case is a straightforward estimation between the price of the dwelling in question and its various characteristics. Though this seems like an appropriate method for the purpose of this thesis, the method have been subject to criticism for potential problems because of the model’s assumptions and estimation of supply and demand, market segmentation, choice of independent variables, and so on (Selim, 2009, p.2844).

Following the likes of Chau and Chin (2002), there are several key assumptions of applying hedonic pricing theory to the housing market. First, one assumes the housing product is homogeneous, which seems a little unnatural given the heterogeneous nature of the product. In terms of attributes, housing can be differentiated by considerations such as locational, structural or neighbourhood factors, in line with what we saw in the pre-existing papers. In addition to these attributes, the housing can be differentiated by other criteria such as type of housing (Chau & Chin, 2002, p.6). Even though this assumption seems unnatural at first, we approach the issue of variety in the housing product by introducing control variables within our model for housing characteristics, such as housing age, living area, plot area, etc. For more details on included variables for our models, see section *5.3*. According to Tufte (2020), the issue is that heteroscedasticity may lead to a reportedly lower standard errors, the reason being the underestimation of the uncertainty of the estimators. Regarding heteroscedasticity there is also important to consider the different degrees of seriousness of the problem. For example, specification error, such as omitting less relevant exploratory variables or interactions between independent variables, is less of a problem compared to omitting highly relevant exploratory variables, include outlier or use the wrong functional form. We also want to point out that conducting a regression analysis of any kind, such as DD approach used in this thesis, would not be ideal if the housing product had perfect homogeneity. Regarding the DD approaches conducted in this thesis, there are certain things we have not been able to control for. Examples of such variables can be the standard of the dwelling, and more specific location characteristics as what view the dwelling have, and whether the dwelling is located nearby schools, malls, etc.

Secondly, the housing market is assumed to be operated as a perfectly competitive market, where there are numerous buyers and sellers. This assumption is according to Chau & Chin (2002, p.6) justified as no individual buyer or seller can affect the price of the properties.

A third assumption for the hedonic pricing model for the housing market is the assumption of complete freedom in entering or exiting the market for both buyers and sellers. There are no such things as legislation, requirements, or restrictions for the housing market, except for each individual’s budget constraints that set an upper limit for what housings can be bought. It is also assumed that the buyers and sellers have perfect information about the housing product and the pricing (Chau & Chin, 2002, p.6). When we look at this assumption given this thesis research question, it is fair to assume so. By doing some research on the mapping of quick clay in Norway, we can say that information on whether the dwelling in question is built on a quick clay is publicly available. Also, as mentioned earlier, sellers are obligated to inform of such information by law. Although the information is available, it does not necessarily mean that the buyers are searching for this information, hence, they may not be aware of such risk.

Finally, it is assumed that for the hedonic price model to work, the market must be in equilibrium, and that there are no interrelationships between the implicit prices of the attributes of the housing product (Chau & Chin, 2002, p.6).

3.3. Market Efficiency

In a perfect world, we would expect the housing market always to reflect all the available information in the market. This means that the dwellings would always be priced in the market at the point where the market deems it fair, given the information about the dwelling in question. Such information might be housing characteristics, such as size, the number of rooms, integrated air conditioning or not, if there is a garage attached to the house, etc. However, all information also includes things that might appear less apparent, obvious, or salient, such as the fact that a building might be built on a foundation consisting of quick clay. This inherently faces the area that the building is built on with a degree of risk; will there be a landslide, and the scope of the damages the landslide will cause. If this information and all other information in the market are reflected in the housing market, it means that the housing market is efficient. Given findings in pre-existing papers indicating that the housing market is inefficient, we find it reasonable to expect that the risk factor of QCLs is not fully reflected in the selling prices of the homes. Thus, we do expect that there will be an effect on prices for these dwellings following a QCL.

Market efficiency was first widely popularized by Fama (1970), whom he presents that the market efficiency hypothesis "…states that the prices in the market should always fully reflect available information…" (Fama, 1970, p. 383). If a market would behave in such a way, it would, for many theorists, be considered a theoretically "perfect world"-scenario upon which to work upon. If a market is efficient, the prices year-to-year should follow a random walk pattern. If this is the case, the price development will show that the price changes will not be followed by a price change in the same direction in the subsequent year. Alternatively, this shows there is no time structure in the development of housing prices. Although this hypothesis is mainly considered when talking about financial markets, you can also consider the efficiency of other markets, such as the housing market.

In the case of housing market efficiency, Case and Shiller (1989, p.125), among others, hypothesized that the market for single-family homes is inefficient. They approached this by looking at dwellings sold twice that also had no obvious quality change and had conventional mortgages applied for purchases. Their data consisted of sales between 1970 and 1986 in Atlanta, Chicago, Dallas, and San Francisco/Oakland. In total, this added up to 39,210 observations used in the paper.

Their estimations were accomplished by undertaking a three-step weighted least squares procedure, first consisting of constructing the Weighted Repeated Sales index. Then, the price changes for each quarter in the index were then tested to see whether there was a random walk pattern. (Case & Shiller, 1989, p.125-128). This was accomplished by producing "…estimates and standard errors for an index of housing prices by regressing, using ordinary least squares, the change in log price of each dwelling on a set of dummy variables, one dummy for each time period in the sample except for the first." (Case & Shiller, 1989, p.126). Through their research they found that prior housing prices tend to predict changes in pricing in the same direction for the following period. However, they were not able to definitively prove whether the housing market is efficient or not (Case & Shiller, 1989, p.135).

Since there has been further research on the issue, Pollkowski and Ray (1997) pointed out that recent papers indicating that housing markets are not informationally efficient. They point out three prior research projects that all point out this fact: (a) Rayburn, Devaney, and Evans (1987), (b) Case and Shiller (1989), and (c) Guntermann and Norrbin (1991). These results indicate that housing prices do not always reflect all publicly available information; hence the housing market is deemed inefficient. Pollkowski and Ray's contribution to the research extends the subject of research by investigating the diffusion of price changes. More specifically, they investigated "…whether housing prices in one location or for one type of housing can be predicted not only by its own history but also by housing price changes in other locations or for other housing types." (Pollowski & Ray, 1997, p.107).

Considering this market efficiency issue for the housing market in Norway, Larsen and Weum (2018) presented their paper regarding this topic. The study was conducted by replicating Case and Shiller's efficiency test, which is described earlier in this section, on housing data in Oslo, Norway. The method is applied to 9,513 pairs of dwellings sold by the OBOS between the third quarter of 1991 and the fourth quarter of 2002; that was sold two times or less, without reporting errors that are bigger than 20 square meters, and without implausible transaction prices. Their study found characteristics indicating that the Norwegian housing market is inefficient. For this thesis, an efficient market would mean that the housing prices will not be affected by events such as QCLs. As this thesis aims to measure salience, hence whether the perception of risk is updates when a landslide occurs, we would have nothing to update in an efficient market. As the household withholds all information regarding the housing product, the event will not be as much of a surprise as if the market is deemed inefficient.

4. Introduction of the data

To answer the research question of this thesis and underlying research questions, we first and foremost need the right set of data. The data presented in this section were collected by Eiendomsverdi AS, a Norwegian limited liability company developing and supplying the Norwegian residential real estate market with information tools and systems for estimating market value on properties, both on individual units and for portfolios (Eiendomsverdi, 2022). Eiendomsverdi will from hereon be referred to as EV. The initial data consist of Norwegian housing transactional data. Even though we got the data through EV, the data is collected from selected other data suppliers such as Finn.no, Eiendom Norge, and several real estate agencies in Norway. Therefore, the data collected for this thesis is not supplied from the primary source but part of a more extensive collection of data to which we were given access to parts of this data.

The data contains observations of sold dwellings within two different periods. The sample data for the period 2007 to 2021 for the municipalities found in the more northern region of Norway, being part of the two QCL cases Alta and Lyngseidet. For the municipalities in the more southern regions of Norway, in this thesis, part of the data for the QCL in Gjerdrum, we have data from 2015 to 2021. The first observation in the initial dataset is registered as sold 01.01.2007, whilst the last observation is registered as sold 31.12.2021. As the thesis is studying the price effects of a QCL, the study is dependent upon a time variable, in which the variable sales date will serve the purpose of. All the tidying of the dataset has been conducted in RStudio[3], while the DD estimations have been conducted in Stata[4].

The initial dataset contained 163,713 observations across 21 variables. The data included the transaction-related variables sales date and sales price including common debt. In addition, the data contains variables that are related to the property being sold. These variables are housing id, name of the municipality, postcode, type of housing, size of the usage area, measured in square meters, built year, and type of ownership.

[3] For the tidying and trimming of the data in RStudio we used the packages: tidyverse and readxl

[4] STATA v.16.

For this thesis, EV also provided a variable for each of the two risk types: Risk of QCL and risk of general landslide. The two variables will be used in this thesis to differentiate between sales of dwellings at risk of a QCL and dwellings not at risk of a QCL. This will be discussed more in detail in section 4.1. These are dummy variables and are thus not structured to interpret different levels of risk in the same way as the risk areas that we can locate in the risk area map from NVE, which is defined by the degree of risk on a scale from one to six. In addition to variables that will be included in the models created in section 5.3, the dataset included variables for the longitude and latitude for the sold units. This will be at much help when we create maps for visualization of the data.

For Alta and Lyngen, we have 42,666 observations in total within the timeframe 2007 and 2021, whilst we, for the Gjerdrum case, have 121,047 observations within the timeframe 2015 to 2021.

There were collected data on a total of 43 municipals, which are grouped into three categories, depending on the case in question. For Alta and Lyngen, there are 15 municipalities included in each dataset, while there are 21 municipalities in the data given for the Gjerdrum case. At the beginning of writing this thesis, we initially wanted to conduct a spatial DD, which will be discussed further in section 7.1. The municipalities were therefor chosen based on both location; hence if the municipality is located remote or nearby the municipality where the landslides occurred, and whether the municipals had any registered risk areas.

4.1. Deleting and creating new variables

When initiating the data cleaning, we decided to drop some variables from the dataset as these would not contribute to the regression model or the DD estimations. In addition to variables not contributing, there is also a need to include and create certain variables that are not part of the initial dataset. This section will give a quick overview of the deleted and created variables and why these variables are vital for the further process.

First off, the variable floor from the initial dataset is not clearly defined. For apartments, the floor variable represents which floor the apartment is located, but for apartments that consist of more than one floor, we can think that the apartment would either be registered as floor 1 or 2, 3 or 4 etc. For other types of housings, we think that the variable might measure how many floors there are. For example, for a two-floored house, the value would be equal to 2. This would be problematic if we included this variable in our dataset as the variable's values are not correctly registered and specified relative across the different housing types, hence being ambiguous. We, therefore, decided to not use this as a variable in the estimations of the DD estimator.

When all variables deemed not valuable have been deleted, we created a dummy variable based on the municipalities where the QCLs occurred, closely located municipalities, and remote municipalities. This will be very useful when creating the maps as we want to differentiate for more detailed maps for each group.

Next, we created dummies for each owner type, housing type, and for the two types of risk: QCL and other landslides. For the remainder of this thesis, we assume that all sold dwellings at risk of both quick clay and other landslides in the dataset are dwellings at risk of QCLs. We do this because we, for one, have a low number of observations consisting of dwellings at risk of a QCLs. Second, we do know that more minor landslides, could potentially be registered as other types of landslides, such as landslides composed of deposits (A. F. Berg, personal communication, February 18, 2022). Also, in addition to the errors in registration of type of landslide, we also think the two potential events will deem a similar response regarding the housing prices due to their similar exogenous traits. Further mentions of QCLs are therefore interchangeable with landslides.

For the construction year, we decided to split the dummies into four. The first dummy is assigned the value 1 if the dwelling was built before 1950, whereas the second dummy includes housing built between 1950 and 1979. The third dummy denotes housing built between 1980 and 1999, whereas the fourth and last dummy denotes all dwellings built After 1999. The creation of these dummies is based on Anundsen and Røed Larsen`s (2018) interpretation of how to deal with the uncertainty regarding registration errors for the dwelling's construction year. These dummies also distinct between dwellings with different characteristics in terms of housing age and dwellings built with different standards. Next, we created dummies where the value is 1 if the date of the registered sales is set after each of the cases selected for this thesis. The three dummies After Gjerdrum, After Alta and After Lyngen will be helpful as this distinct between sales registered before and after each of the cases, making it easy to compare trends before and after the landslides. We also created a new variable returning the price per square meter. This will allow us to eliminate transactions where the price per square meter is unusually high later in the tidying of the data.

Furthermore, it is reasonable to control for unobserved time-effects, such as seasonality, which is overcome by creating a variable for each quarter and year, based on the date variable.

4.2. Method of trimming

Upon receiving the data from EV, we can see that the data is not registered without flaws that can be problematic and thus need to be dealt with before conducting estimations of the DD estimators in STATA.

For a starter, the dataset contains observations with randomly missing values in one or more of the variables included. In addition to some missing values within certain observations, the dataset also contains observations where there are registered values that we consider beyond what is plausible and logical.

4.2.1. Removing Observations

In most cases, when given a dataset for research purposes, there are values and observations we deem as problematic. The steps conducted and the number of observations removed for each step are summed up in Table 4.1.
In the first step, within the process of removing observations, we removed all observations that do not have any information; hence outputted as NA within the dataset. We removed these values for the following variables: sales price, usage area, construction year, type of ownership and type of housing.

During our checks of the unique values within each variable, we found that the dataset included the four owner types: contractual leasehold flats (CLF), limited liability housing company flats (LLHCF), detached houses, semi-detached houses, apartments, and townhouses. Of these types of ownerships, contractual leasehold flats and limited liability housing company flats are not part of what we consider to be a typical transaction in the housing market. We the therefore chose to remove these observations from the dataset.

4.2.2. Dealing with problematic values

As part of the tidying, we first investigated whether there was a need for repairing randomly missing values within any observations. As we started this process, we found that none of the repeated Housing IDs had omitted values that other observations with the same Housing ID included.

In our dataset, we did have some registered values that we deem problematic. For usage area we found observations in our data returning the value zero. Having zero-values can be a problem as it is non-intuitive that a dwelling does not have any usage area. We also removed all observations where the variable construction Year was equal to zero, as the value zero implies that the dwelling was built 2022 years ago, which we deem highly improbable. We deem these data as wrongly registered, in the same way of observations returning the value NA, in sales price, usage area, and postcode as these would influence the results presented in section 6 of this thesis.

Table 4.1[5]: Summary of each step in the tidying of the dataset.

Step	Description of step taken	N of observations	N removed observations	Removed in %
0	Initial dataset	163,713	0	0%
1	NA's selling price	162,893	820	0.501%
2	NA's postcode	163,890	3	0.002%
3	NA's size	160,566	2,324	1.427%
4	NA's year of construction	160,135	431	0.269%
5	Size = 0	159,865	270	0.169%
6	Year of construction = 0	159,850	15	0.094%
7	CLF & LLHCF	159,061	789	0.494%
Sum				2.842%

4.3. Creating separate datasets for each case

Before we can load our dataset into Stata and conduct a regression based on our data, we split the data into three different datasets. The datasets are divided based on the three QCLs.

Within these datasets, we must further trim the data. The data still contains some extreme values, even though these are deemed as reliable registered data. We looked further into the potential outliers to approach a more normal distribution. For dealing with these observations, we decided to use upper and lower quantile values as a measure to cut the two-sided tales in the data distributions. During this step, we tested different percentages for the quantiles to ensure that the data within the variables approached normal distribution. After a closer study of the data distributions before and after the cut-offs, we ended up with a cut-off of the lower 1% and the upper 1% for the variable's sales price, usage area, and price per square meter. We trimmed all the variables within the same trimming to keep as many observations as possible while still deleting the most extreme and outlying

[5] Table 4.1 sums up in what way the trimming that was completed affected the dataset regarding the number of observations. This trimming was executed for the complete dataset and therefore there's no information on how these affected the each of the datasets presented in section 4.3. In this table, CLF is an abbreviated term for Contractual Leasehold Flat, while LLHCF is an abbreviated term for Limited Liability Housing Company Flat.

observations. In Table 4.2, there is a summary for each dataset following the trimming.

Table 4.2[6]: Summary before and after trimming with quantiles.

Case	Obs. before casewise trimming	Obs. after casewise trimming	Change in %	Obs. before in %	Obs. after in %
Initial dataset	159,061	159,061	0%	100%	100%
Gjerdrum	119,384	113,325	5.1%	75%	71%
Alta	16,766	15,870	5.3%	11%	10%
Lyngen	33,279	31,750	4.6%	21%	20%

4.3.1. Description of the final data - Gjerdrum

Before trimming for each case, there was 119,384 observations in the dataset regarding the Gjerdrum case, while a trimming of 1% in each of the variables listed led to a total of 113,325 observations in the final dataset This is equivalent to a cut-off by approximately 5.07%.

The dataset consists of 95,761 observations before the landslide, and 17,564 observations after the landslide in 2020. In total there are 4,738 observations, yielding that 4.18% of dwellings within the dataset are at risk of a QCL. Of these 3,999 are registered before the landslide, and 739 is registered after the landslide. The observations that are not at risk of a QCL are distributed with 91,762 before the landslide and 16,825 after the landslide. For the Gjerdrum dataset, none of the sold dwellings were registered as dwellings at risk of other landslides. Summary statistics for some of the control variables in the dataset are presented in Table 4.3. For a more detailed overview of the distribution of sold dwellings across the 22 municipals, please see appendices B.1.A – B.1.D.

[6] Obs is an abbreviation for observations, Obs before in % returns the share of observations before the trimming percentagewise of the initial dataset, and Obs after in % returns the share of observations after the trimming percentagewise of the initial dataset.
The casewise trimming was conducted by deleting all observations within the lower 1% and the upper 1% for the variable's sales price, usage area, and price per square meter. The same limit was utilised for all three cases.

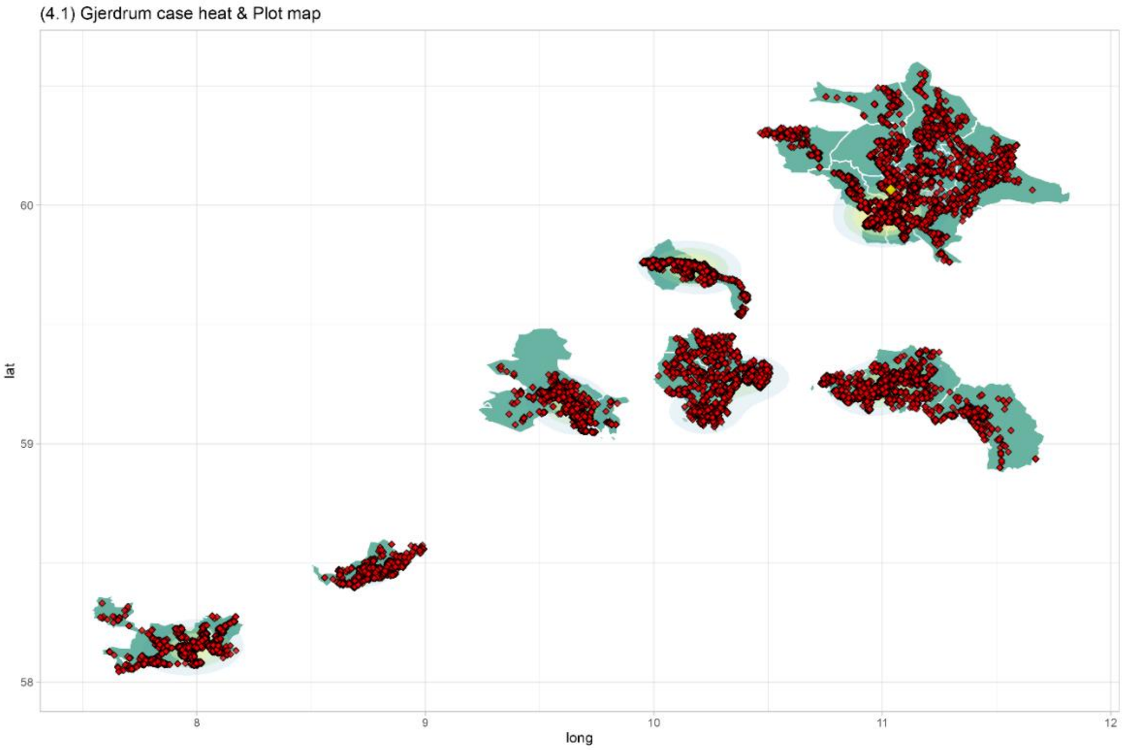
Table 4.3: Summary statistics for critical variables in Gjerdrum dataset.

VARIABLES	N	mean	sd	min	max
Selling price	113,325	3,205,284	1,391,789	1,007,723	8,900,000
Size in square meters	113,325	119	61	31.00	334
Year of construction	113,325	1976	32	1599	2022
Selling price per square meter	113,325	30,438	12,691	9,671	73,577

Next, we want to introduce the first map within this thesis. We have created maps to visualize the spread of the sold dwellings. In the case of Map 4.1, we have plotted all sold dwellings for all municipals in the Gjerdrum case. For this and the remainder of the maps in this thesis, including those in the appendix, the mapped area consists of municipals included in the case, or the specified area(s) of interest. For a more detailed map over Gjerdrum case, see Map 4.4 or map A.1.A – A.1.G in the appendices. The map consists of plotting of the sold dwellings in each dataset represented in red, and mapping of the risk areas in grey. In addition, the map also includes a heatmap, which visualize the probability of an observation to be in within the certain area, hence, the darker the colour, the bigger probability there is. Lastly, the place of event for each case is marked with yellow on the map.

For a more detailed view of the data using ggplot2, we have decided to split the municipals within each case into the four groups: municipal of the event, municipals in proximity of the municipal of the event, remote located municipals, and all municipals within the cases. Within the Gjerdrum case there are 21 municipals, as mentioned earlier, which is included in the maps that are presented in this thesis, excluding the maps in the appendix. The list of included municipalities and the number of observations for each of them for the Gjerdrum case cases can be found in appendix B.1.A These municipalities are grouped by the number of sold dwellings, whether they are at risk or not, and sold dwellings before or after the landslide event in Gjerdrum. Corresponding tables for the Alta case can be found in appendix B.2.A and B.2.D, and for the Lyngen case refer to appendix B.3.A and B.3.D.

Map 4.1[7]: Spread of sold dwellings for Gjerdrum case



4.3.2. Description of the final data - Alta

Before trimming for each case, there was 16,766 observations in the dataset regarding the Alta case, while a trimming of 1% in each of the variables listed led to a total of 15,870 observations in the final dataset. This is equivalent to a cut-off by approximately 4.84%.

[7] This map visualises the spread of the sold dwellings for all 21 municipals in the Gjerdrum case. For a list over the municipals see appendix B.1.A For this and the remainder of the maps in this thesis, including the appendix, the mapped area consists of municipals included in the case, or the area of interest. For a more detailed map see map 4.4 or see map A.1.A – A.1.G in the appendices. Sold dwellings are plotted in red. For all maps in section 4 of this thesis, including the interactive map, the risk zones only include the areas at risk of the risk quick clay landslide, as defined by NVE, and not by the assumption for this thesis. Risk areas for other landslides are not included as the maps for these were separated for each scenario, hence it would be too time consuming to plot. The heatmap is created by using density.2D. The heatmap therefor returns the probability of an observation to be in within the certain area, hence, the darker the colour, the bigger probability there is. The place of event is marked with yellow on the map.

The dataset consists of 13,700 observations before the landslide, and 2,170 observations after the landslide in 2020. In total there are 1,039 observations of dwellings at risk of a QCL, yielding that 6.55% of the dwellings within the dataset are at risk. Of these 906 are registered before the landslide, and 133 are registered after the landslide. The observations that are at risk of a QCL are distributed with 12,794 before the landslide and 2,037 after the landslide. Because we now include other landslides in what we define as quick clay for the remainder of this thesis, we want to point out that 728 observations were registered as sold dwellings at risk of QCL as the initial risk, while 311 observations were registered as sold dwellings at risk of other landslides. We present summary statistics for some of the control variables in the dataset in Table 4.4. For a more detailed overview of the distribution of sold dwellings across the 22 municipals, please see appendices B.2.A – B.2.D.

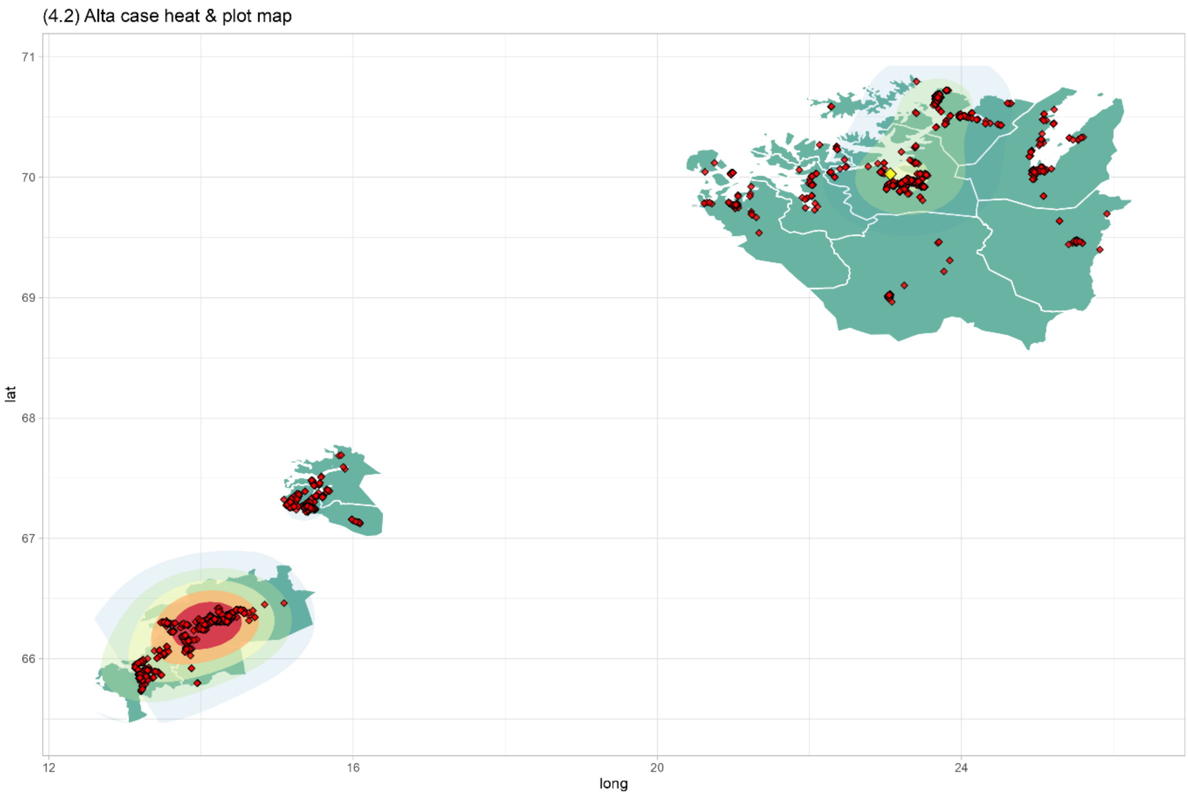
Table 4.4. Summary statistics for critical variables in Alta dataset

VARIABLES	N	mean	sd	min	max
Selling price	15,870	2,294,147	889,376	580,000	5,500,000
Size in square meters	15,870	133.00	59.79	38	323
Year of construction	15,870	1974	21.32	1,840	2021
Selling price per square meter	15,870	19,748	8,976	4,251	49,161

Next, we want to visualize the spread of the sold dwellings within the Alta case. In

Map 4.2, we have plotted all sold dwellings in all 15 municipals. This map gives the same visualization as discussed in Map 4.1: Spread of sold dwellings for Gjerdrum case

Map 4.2[8]: Spread of sold dwellings for Alta case.



4.3.3. Description of the final data – Lyngen

Before trimming for each case, there was 33,279 observations in the dataset regarding the Lyngen case, while a trimming of 1% in each of the variables listed led to a total of 31,717 observations left in the final dataset. This is equivalent to a cut-off by approximately 4.7%.

The dataset consists of 27,081 observations before the landslide, and 4,636 observations after the landslide in 2010. In total there are 6,539 observations with dwellings at risk of a QCL, yielding that 20.62% of the dwellings are at risk. Of these 742 are registered before the landslide and 5,797 are registered after the landslide. The observations that are not registered with a risk of quick clay are distributed with 3,894 before the landslide and 21,284 after the landslide.

[8] Map 4.2 visualises the spread of the sold dwellings for all 15 municipals in the Alta case. For a list of the municipals, see appendix B.2.A. The mapped area consists of municipals included in the case, or the area of interest. For a more detailed map for the Alta case, see map 4.4 or see map A.2.A – A.2.G in the appendices. Sold dwellings are plotted din red, while the risk areas for quick clay are defined in the same way as in footnote 7, are the areas with a light grey colour. The heatmap is created by using density.2D. The heatmap therefor returns the probability of an observation to be in within the certain area, hence, the darker the colour, the bigger probability there is. The place of event is marked with yellow on the map.

Because we now include other landslides in what we define as quick clay for the remainder of this thesis, we want to point out that 1,502 observations were registered as sold dwellings at risk of QCL as the initial risk, while 5,037 observations were registered as sold dwellings at risk of other landslides. We present summary statistics for some of the control variables in the dataset in Table 4.5. For a more detailed overview of the distribution of sold dwellings across the 22 municipals, please see appendices B.3.A – B.3.D.

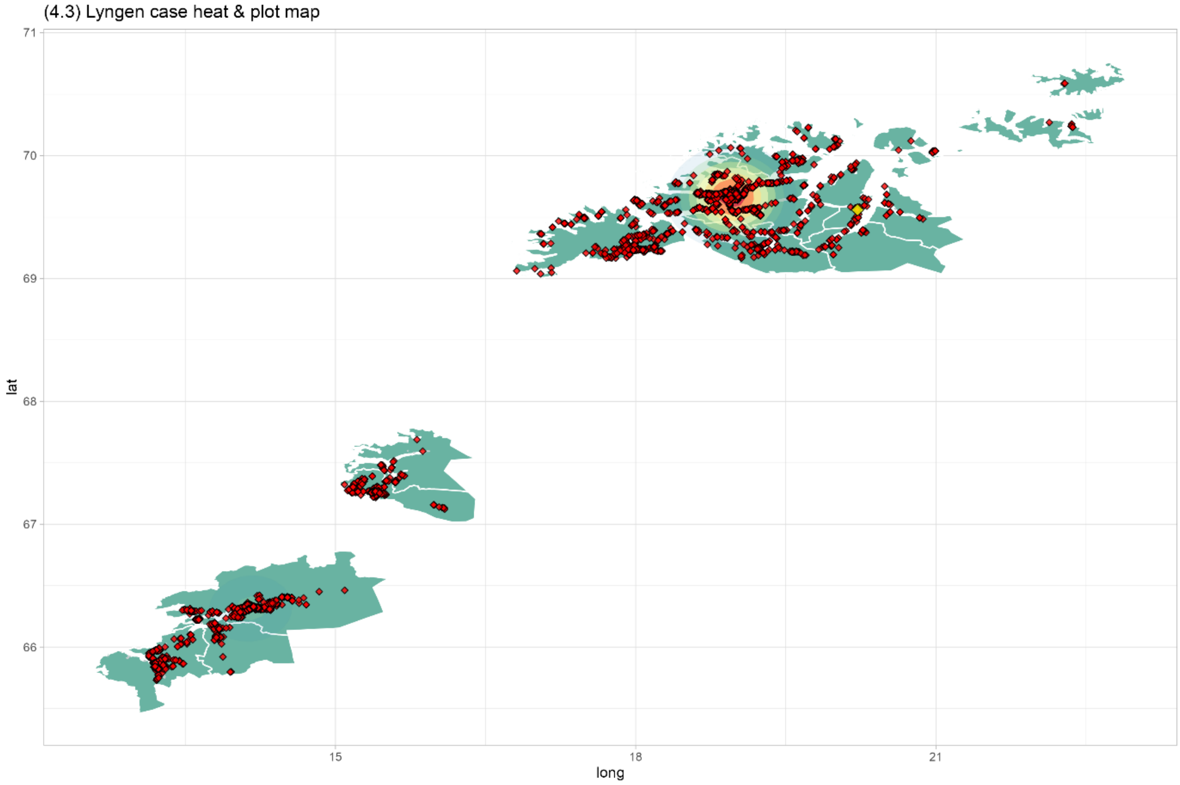
Table 4.5: Summary statistics for critical variables in Lyngen dataset

VARIABLES	N	mean	sd	min	max
Selling price	31,717	2,905,000	1,302,000	686,633	7,800,000
Size in square meters	31,717	116.4	61.34	31	320
Year of construction	31,717	1979	25.70	1697	2021
Price per square meter	31,717	29,900	14,734	4,660	71,429

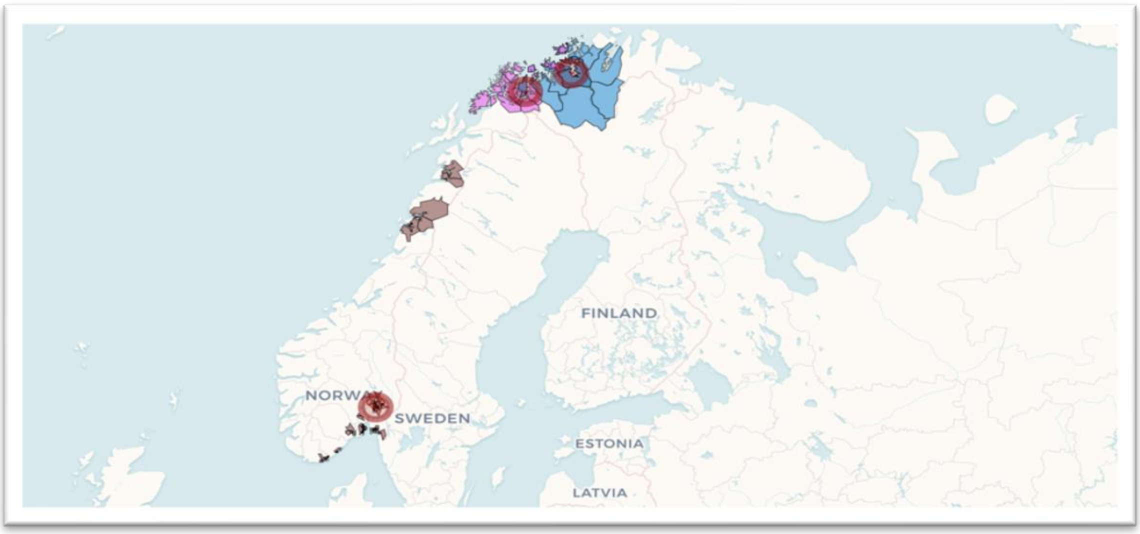
Next, we want to visualize the spread of the sold dwellings within the Lyngen case. In Map 4.3[9], we have plotted all sold dwellings in all 15 municipals. This map gives the same visualization as discussed in map 4.1.

[9] Map 4.3 visualises the spread of the sold dwellings for all municipals in the Lyngen case. For this and the remainder of the maps in this thesis, including the appendices, the mapped area consists of municipals included in the case, or the area of interest. For a more detailed map see map 4.4 or see map A.2.A – A.2.G in the appendices. Sold dwellings are plotted din red, while the risk areas for quick clay are defined in the same way as in footnote 7, are the areas with a light grey colour. The heatmap is created by using density.2D. The heatmap therefor returns the probability of an observation to be in within the certain area, hence, the darker the colour, the bigger probability there is. The place of event is marked with yellow on the map.

Map 4.3: Spread of sold dwellings for Lyngen case.



Map 4.4[10]: Click the picture for an Interactive map that visualises the dataset



[10] To enter the interactive map, please click the picture. The interactive map will open in a new tab in your browser, but it may take some time depending on the Rpubs server. The map includes features such as having three different types of map layers, as well as a search function for checking the risk of quick clay in any area of your interest. The map also gives options for visualising the spread of observations with a heatmap, a cluster for all sales, and clusters for sold dwellings within the treatment and control groups for each case. The cluster only returns a polygon in which the sold dwellings that are grouped can be located within rather than exposing their exact location The map is created from map files created and published publicly by NVE and GeoNorge.

"tomato" coloured area = Gjerdrum + closely located municipals in Gjerdrum case. Blue = Alta + closely located municipals in Alta case. Pink = Lyngen + closely located municipals in Lyngen case. Brown = municipals defined as remotely located for all cases. Municipals located below Gjerdrum on the map belong to the Gjerdrum case; meanwhile, those above belong to Alta & Lyngen.

5. Empirical conditions

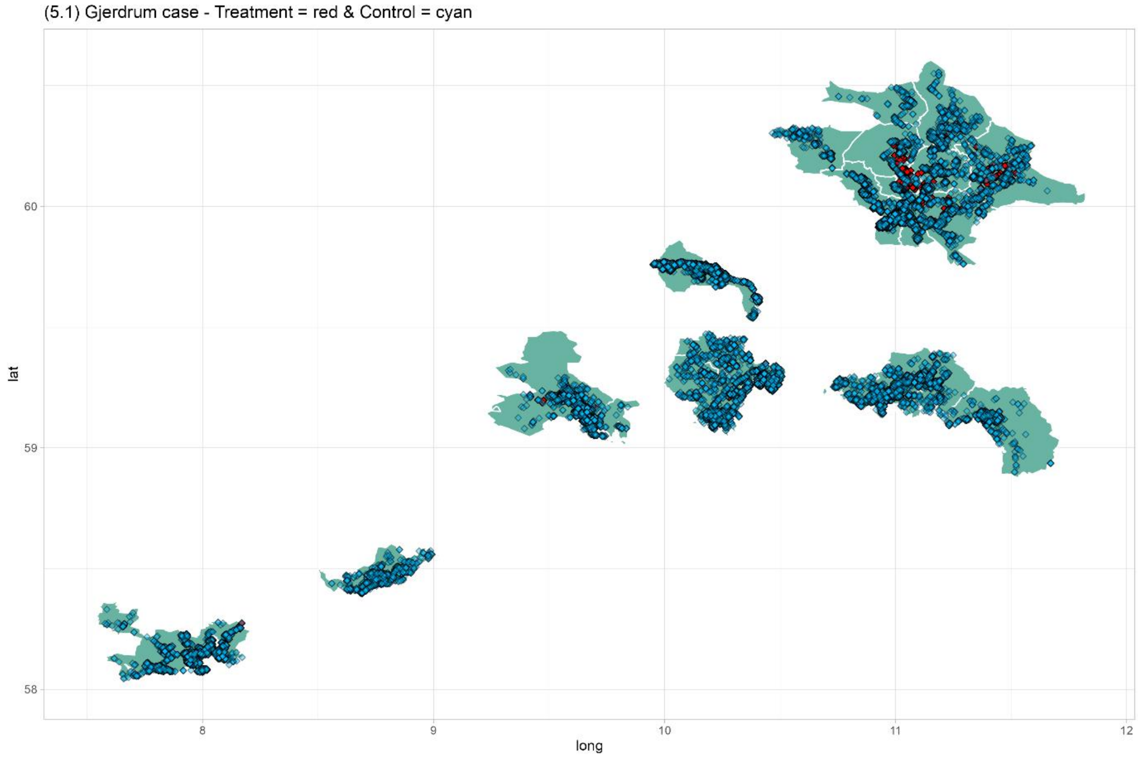
5.1. Treatment and control groups

When conducting a DD approach, we first and foremost need to define our treatment and control group. As mentioned in section 1, this thesis aims at assessing whether the landslides influence dwellings at, and dwellings not at risk of a QCL after the three selected QCLs. In addition to this, we attempt to assess if there is a general price effect for dwellings that are at risk of such events. To be able to assess what we sought out for, the dwellings built on quick clay, and therefore has the risk element of a QCL, are the group of dwellings that are affected by a landslide, i.e., the treatment group. Thus, the treatment group should intuitively be the sold dwellings at a risk of quick clay. The control group is then sold dwellings that are not at risk of a QCL. For example, we can estimate a model with the hedonic pricing aspect of dwellings in Gjerdrum. If the trend charts for the treatment group and the control group seem to follow a similar pattern over time, we can assume they also would follow each other into the future, given there is no treatment intervention, i.e., landslide. As mentioned in the theory section of this thesis, this is one of the most crucial assumptions and criteria to fulfil when conducting a DD.

When we test whether an event is affecting the housing prices, we need to be certain that these patterns would follow each other also in the future, given that the event we are interested in would not occur. As mentioned in section 4.1, we have chosen to treat the two types of risk that were reported in the initial dataset as one, hence the risk off QCL. As a result of this, dwellings at risk of a QCL and dwellings at risk of a general landslide are both included in what we define as our treatment group. Regarding this assumption, we want to point out that there were zero sold dwellings in the Gjerdrum case at risk of other types of landslides, hence for the Gjerdrum case, all sold dwellings included in the treatment group is at risk of a QCL regardless of this assumption.

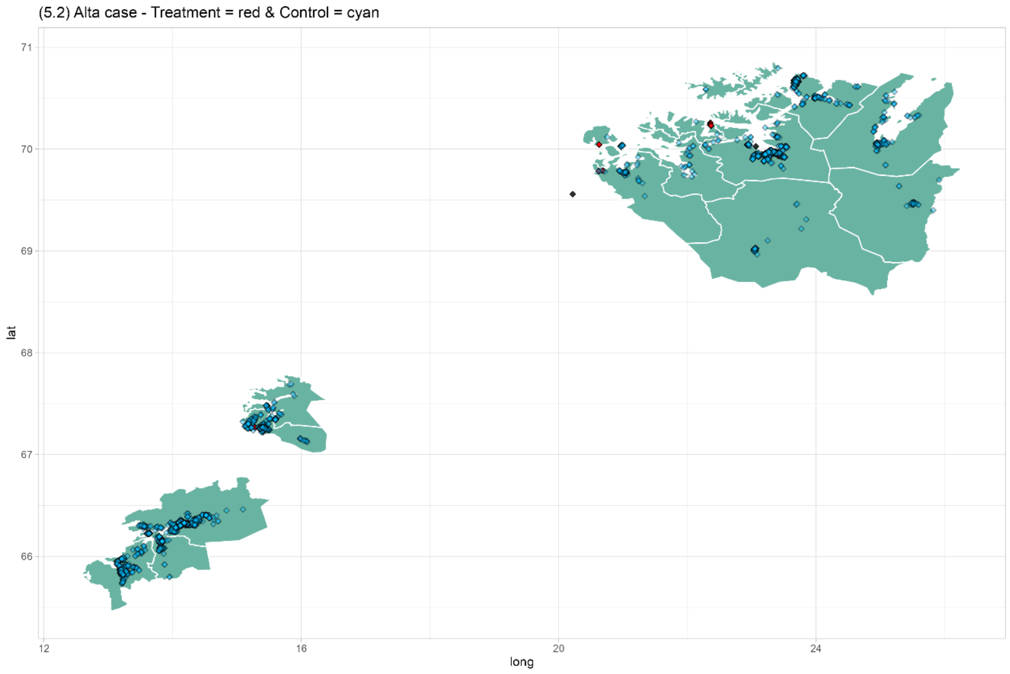
Because the chosen treatment group is defined by sold dwelling at risk of a QCL; hence the control group is sold dwellings not at a risk. The treatment and control groups therefore follow the same distribution as mentioned in section 4.3.1, 4.3.2, and 4.3.3. With this definition, Map 5.1 yields a visualisation of the spread of observations within the treatment and control group for the Gjerdrum case. The same goes for the treatment and control group in the Alta and Lyngen case; hence Map 5.2 and Map 5.3 yield the same visualisation. In addition, we refer to appendix A.1.F, A.1.G, A.2.F, A.2.G, A.3.F, and A.3.G for maps that visualises the spread of the sold dwellings, included in the treatment and control group, both before and after the event.

Map 5.1[11]: Spread of sold dwellings in treatment (red) and control groups (cyan) for Gjerdrum.

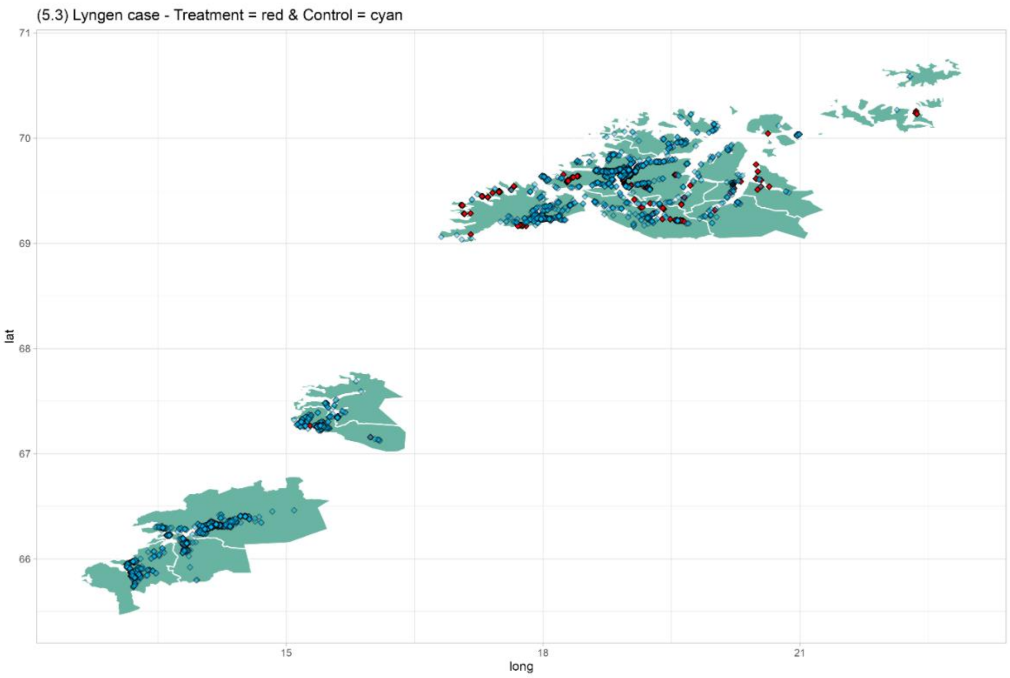


[11] Map 5.1 is a visualisation of dwellings included in our defined treatment and control groups for the Gjerdrum case, where the included municipals are visualised in the colour "green". Sold dwellings that are part of the treatment group is plotted in red, while the sold dwellings that are part of the control groups for Gjerdrum is plotted in cyan. The risk areas for quick clay risk areas are defined in the same way as in footnote 7, are the areas with a light grey colour. The place of event (landslide) is plotted in yellow. The distribution between the two can be found in section 4.3.1.

Map 5.2[12]: Spread of sold dwellings in treatment (red) and control groups (cyan) for Alta.



Map 5.3[13]. Spread of sold dwellings in treatment (red) and control groups (cyan) for Lyngen



[12] Map 5.2 is a visualisation of dwellings included in our defined treatment and control groups for the Alta case, where the included municipals are visualised in the colour "green". Sold dwellings that are part of the treatment group is plotted in red, while the sold dwellings that are part of the control groups for Gjerdrum is plotted in cyan. The risk areas for quick clay are defined in the same way as in footnote 7, are the areas with a light grey colour The place of event (landslide) is plotted in yellow. The distribution between the two can be found in section 4.3.2

[13] Map 5.3 is a visualisation of dwellings included in our defined treatment and control groups for the Lyngen case, where the included municipals are visualised in the colour "green". Sold dwellings that are part of the treatment group is plotted in red, while the sold dwellings that are part of the control groups for Gjerdrum is plotted in cyan. The risk areas for quick clay are defined in the same way as in footnote 7, are the areas with a light grey colour. The place of event (landslide) is plotted in yellow. The distribution between the two can be found in section 4.3.3.

5.2. Visual inspection of parallel trends

To assess whether the parallel trend assumptions hold or not, we can plot trend lines by predicting values of the selling prices, based on a hedonic price regression model. In the trend plots, depicted in figures 5.1.-5.3., we plot the observations on selling prices on the quarterly date. The selling prices of the dwellings for each case area will be depicted on the y-axis of the plots presented, while the time variable of choice is the quarterly dates attributed to the x-axis. Our main objective here is to determine whether the trends between the treatment and control groups before the treatment date are parallel or not. The linear trend lines are yielded by estimating the regression, by utilising Model 5.1, four times. By estimating the model four times, there are imposed different conditions on the model, such as if the sale of the dwelling is registered before or after the given landslide and whether the dwellings are at risk of QCLs, the two conditions combined yield four identical regression models with four different sets of conditions that draw the trend lines. The regression is thus estimated for (1) the control group before the given landslide, (2) the control group after the given landslide, (3) the treatment group before the given landslide, and (4) the treatment group after the given landslide.

Model 5.1: Estimating linear trend lines to check for parallel trends

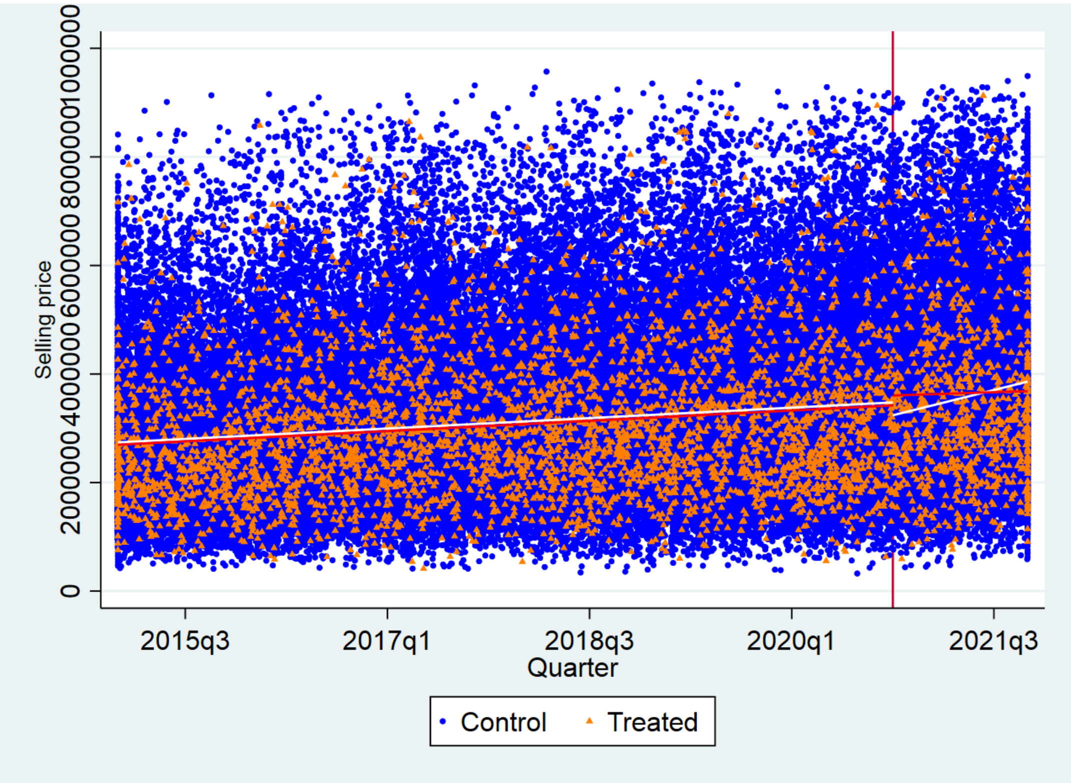
$$P_{it} = \beta_0 + \beta_1 size_i + \beta_2 between1950a1979_i + \beta_3 between1980a1999_i + \beta_4 after1999_i + \beta_5 dwellingtype_i + \beta_6 ownertype_i + \beta_7 qdate_i + \varepsilon_i$$

where P_{it} is the selling price of the dwelling of property i at time t, $size_i$ is the floor size in square meters of the dwelling, $between1950a1979_i$ is a dummy indicating if the dwelling was built between 1950 and 1979, $between1980a1999_i$ is a dummy indicating if the dwelling was built between 1980 and 1999, $after1999_i$ is a dummy indicating if the dwelling was built after 1999, $dwellingtype_i$ denoted what type of housing the dwelling is (detached house, apartment, townhouse or semi-detached house), $ownertype_i$ denotes whether the dwelling is owned through a housing association or is a freehold ownership, $qdate_i$, denotes the quarter and year the dwelling was sold.

Utilising model 5.1, we estimate values of the dependent variable, selling price, using the ordinary least squares (OLS) method and fitting a linear line through the observations of the selling price based on the time of sale using the quarterly dates. OLS fits the values of the dependent variable by minimizing the sum of squared residuals. The residuals are the difference between the actual values of the observations on the variables in question, and the predicted values of the variables.

The plot beneath shows some of the scattered values of selling prices for the dwellings in the Gjerdrum case area and period, where the orange triangles are the observations of the treatment group, and the blue dots are the observations of the control group. Not all observations are scattered as it would clutter the plot area by having many plotted observations on top of each other. To avoid such, we used the jitter function in STATA when generating the plot. This function "...adds spherical random noise to the data before plotting." (stata.com, n.d, p.16). Due to the approach of plotting observations this way, there will occur plotted observations for selling prices initially do not exist within the dataset. The vertical red line is drawn when the landslide occurred, hence the treatment date. This line is drawn to indicate the two time periods for each trend line, yielding Figure 5.1. The figure shows jittered values of sales price for each quarter. The values of the observations on sales price are presented on the y-axis, while each quarterly date is shown on the x-axis. The white line is the estimated trend line for the control group, while the red one is for the treatment group.

Figure 5.1[14]: Parallel trends plot Gjerdrum case.

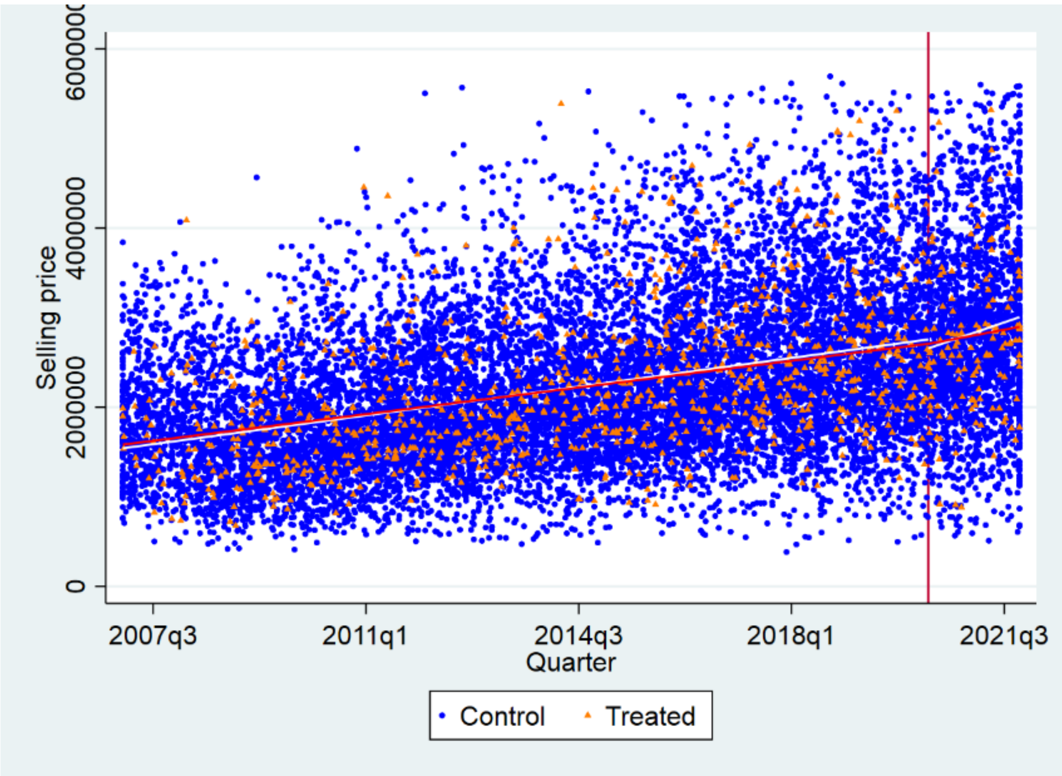


We can see that the trend lines for the selling prices of the dwellings before the treatment date are, in fact, parallel and close to coinciding. Thus, we conclude that the parallel trend assumption holds. Looking at the plot, we can see a positive trend, as expected under normal circumstances. Also, the prices on dwellings within the treatment group are lower than those in the control group before the treatment, in line with our expectations. After the treatment, we see a drop in prices for the control group before the trend relatively rapidly climbs compared to before the treatment. We see a slight jump in prices for the treatment group, but a positive slacker trend compared to before the treatment and compared to the control group after the treatment.

[14] For a more comprehensive overview of what municipals that are included and the distribution of sold dwellings for the treatment and control group, see appendix B.1.B. The landslide in Gjerdrum occurred 30th of December 2020, which is marked in the figure by xline.

Figure 5.2 yields the same plot, but for the data included in the Alta dataset. Like in Similar to Figure 5.1, Figure 5.2 shows jittered values of the sales price for each quarter, in which the values of sales price are on the y-axis and each quarterly date are on the x-axis. The estimated trend lines are white for the control group and red for the treatment group.

Figure 5.2[15]: Parallel trends plot for Alta case



Here, the trend lines before the treatment are again close to coinciding, but with the prices of the treatment group being marginally lower than those of the control group. Dwellings in both groups sees a growth in prices before the landslide.

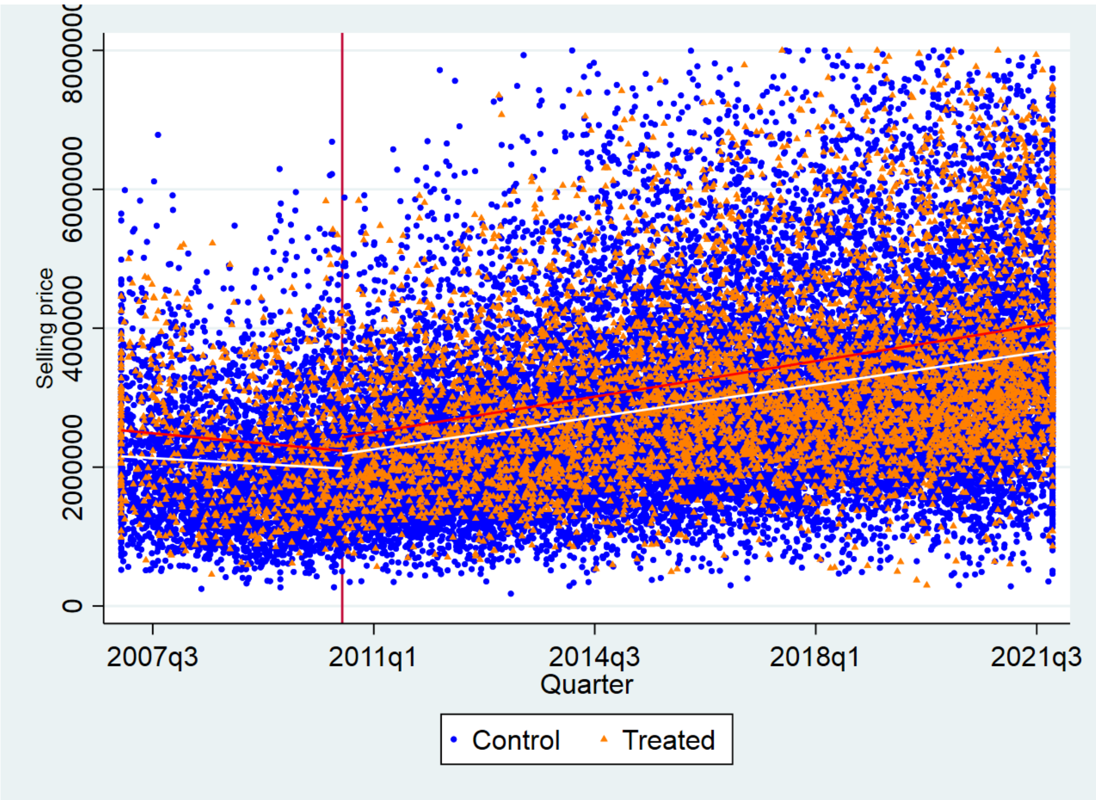
Again, the trend lines are parallel, and we conclude that the parallel trend assumption holds. After the landslide, however, the control group experiences a slight drop in prices, but with a more rapid growth rate, such that the prices exceed those of before the landslide. The treatment group's prices after the treatment continues to trend at approximately the level in which it did before the landslide.

.

[15] For a more comprehensive overview of what municipals that are included and the distribution of sold dwellings for the treatment and control group, see appendix B.2.B. The landslide in Alta occurred June third, 2020, which is marked in the figure by xline.

Again, Figure 5.3 yields the same plot as the previous parallel trend's plots, but for Lyngen case area over time. Once again, the plot area depicts jittered values of sales price for each quarter, whilst the values of sales price are given on the y-axis, and each quarterly date is shown on the x-axis. The white line is the trend line for the control group, while the red line is the trend line for the treatment group.

Figure 5.3[16]. Parallel trends plot for Lyngen case



For the Lyngen case area, we see a plot quite different from the previously presented plots for the Alta and Gjerdrum case areas. Here, the treatment group has higher prices than the control group, both before and after the treatment date. Also, the trends are interesting. For these dwellings, the trend of the prices is negative before the treatment date and are not entirely parallel to each other. Although there are differences in the trends, they are approximately parallel. We, therefore, choose not to reject the parallel trends assumption and will move on with our analysis in the results section without any further manipulation of the data. Curiously, the trend for both groups turn around after the treatment date and is thus positive. However,

[16] For a more comprehensive overview, see appendix B.3.B. The landslide in Alta occurred June third, 2020, which is marked in the figure by xline.

again, the trends are very close to perfectly parallel, with the treatment group having higher prices than the control group.

Interestingly, we have found different trends and reactions in prices after the treatment date between our three case areas. The most significant differences between the cases are found in the Lyngen case, compared to those of the Gjerdrum and Alta cases, even if the resulting trends of the latter cases are similar. Although these differences are interesting, we will not pursue further research into what causes these differences. Neither will we further investigate what causes the trends in the pricing of the dwellings. Through the visual analysis of the parallel trends, we have deemed that the parallel trends assumptions for all three cases holds, and thus, we can continue our analysis of the results without further implications.

5.3. Specification of the models

Our interest in this thesis is to estimate the effect on selling prices of being at risk of quick clay and its changes after a landslide in relative proximity. To approach this, we estimated 18 regression models, out of which six models are estimated for each case (Alta, Lyngen, Gjerdrum). Out of these six models, three models aim to estimate the baseline results for the effect on prices for dwellings at risk of a QCL.

Model 5.2: Estimating the baseline results

$$P_{it} = \beta_0 + \beta_1 LR_i + \beta_2 size_i + \beta_3 between1950a1979_i + \beta_4 between1980a1999_i + \beta_5 after1999_i + \beta_6 dwellingtype_i + \beta_7 ownertype_i + \varepsilon_i$$

Where we follow the same logic as with Model 5.1, but also includes the risk dummy, LR_{it} (Landslide risk) in the model. The risk dummy denotes whether the dwelling is at risk of a QCL. To control for time trends and seasonality, we include the variable $qdate_i$, which is treated as a dummy variable and indicates the quarter and year the dwelling was sold, yielding Model 5.3.

Model 5.3: Estimating the baseline results, controlling for time trends and seasonality

$$P_{it} = \beta_0 + \beta_1 LR_i + \beta_2 size_i + \beta_3 between1950a1979_i + \beta_4 between1980a1999_i + \beta_5 after1999_i + \beta_6 dwellingtype_i + \beta_7 ownertype_i + \beta_8 qdate_i + \varepsilon_i$$

Further, we extend the model by including the dummy $postcode_{it}$, which indicate which area the dwelling is located, based on the areas post code. By including the post code dummy, we control for location-specific effect in Model 5.4.

Model 5.4: Estimating the baseline results, controlling for time trends, seasonality, and location specific effects

$$P_{it} = \beta_0 + \beta_1 LR_i + \beta_2 size_i + \beta_3 between1950a1979_i + \beta_4 between1980a1999_i + \beta_5 after1999_i + \beta_6 dwellingtype_i + \beta_7 ownertype_i + \beta_8 qdate_i + \ \beta_9 postcode_i + \varepsilon_i$$

By also including the variable $after_{it}$, which is a dummy that equals one if the sale happened after the given landslide. Also, by including the interaction term between the landslide risk dummy, LR_{it}, the post-landslide dummy, $after_{it}$, we can estimate the DD effect, which measures the effect of the landslide on housing prices. Then, we estimate yet another three models for each case, in an identical fashion, for the exception of the inclusion of the post-landslide dummy and the interaction term, yielding in Model 5.5, 5.6, and 5.7

Model 5.5: DD estimations by including after dummy and its interaction with $\boldsymbol{LR_i}$

$$P_{it} = \beta_0 + \beta_1 LR_i + \delta_0 after_{it} + \delta_1 LR_i \cdot after_{it} + \beta_2 size_i + \beta_3 between1950a1979_i + \beta_4 between1980a1999_i + \beta_5 after1999_i + \beta_6 dwellingtype_i + \beta_7 ownertype_i + \varepsilon_i$$

In similarity with the inclusion of $qdate_i$to control for time trends and seasonality in model 5.3, we create model 5.6 by including $qdate_i$with model 5.5 as a base. We also create model 5.7 by including the $qdate_i$ and postcode dummy with model 5.5 as a base. By creating these models, we can control for time trends and seasonality within the DD models, in addition to control for location specific effects within model 5.7.

As stated earlier, these six models will be estimated for each of the three cases in section 6. For the last three models, the coefficient of interest is δ_0, and δ_1. δ_0 will tell the price effect on dwellings that are not at risk after the given landslide, and δ_1 tells the additional effect of the dwellings at risk after the given landslide

6. Results

In the following analyses, we will estimate and present the results of the above-specified models. These six models will be conducted for each of our three cases: Gjerdrum, Alta, and Lyngen. Therefore, we have chosen to present the results for each case in individual sections. The analysis will not exclusively include the models we want to estimate but also a separate regression, where we attempt to investigate the cause of the effects found. In this regression model, we run models identical to Model 5.5, but with one exception: here, we use the sum of total sales in each quarter to help determine whether the effects on price stem from changes in demand.

Furthermore, this will show whether there are demand effects that will curb the price effect attributed to the risk of QCLs. It is reasonable to believe through knowledge of the price mechanisms of supply and demand effects that the price effect of the risk will be curbed if there is a significant drop in sales, which is interchangeable with a drop in demand. These hypothetical drops in sales volume are probably attributable to the fact that the owners of the dwellings that wish to sell their houses will have challenges doing this because buyers are less willing to buy.

6.1. Gjerdrum

To start, we present the baseline results for Gjerdrum utilizing Model 5.2-5.4. We do not distinguish whether the sale occurred before or after the given landslide in these models. Our interpretation of the results will then focus on the effect of being at risk of a QCL on prices by considering both the effects of the risk before and after a given QCL.

The baseline results for Gjerdrum are summarized in Table 6.1. The coefficients of the regression results refer to the change in price measured in Norwegian kroner (NOK) following a one-unit increase of the given explanatory variable.

Table 6.1[17]: Gjerdrum results

VARIABLES	(5.2) Selling price	(5.3) Selling price	(5.4) Selling price
Landslide risk (LR)	-22,756	-24,351*	-76,336***
	(14,696)	(14,095)	(12,643)
Size in square meters	14,902***	14,918***	14,359***
	(96.10)	(92.90)	(72.91)
Between1950a1979	-124,808***	-122,119***	-75,013***
	(11,103)	(10,700)	(8,715)
Between1980a1999	288,315***	297,770***	319,056***
	(11,822)	(11,382)	(9,112)
After1999	819,157***	800,261***	904,803***
	(11,287)	(10,873)	(9,507)
Apartment	191,124***	181,068***	-337,201***
	(12,902)	(12,404)	(9,761)
Townhouse	247,655***	226,177***	-165,572***
	(13,792)	(13,046)	(8,814)
Semi-detached house	-88,492***	-94,777***	-289,134***
	(12,926)	(12,406)	(8,622)
Freehold ownership	158,734***	145,444***	94,582***
	(7,813)	(7,495)	(6,150)
Constant	960,979***	468,469***	1,632,000***
	(20,227)	(23,701)	(33,785)
Quarter dummy	No	Yes	Yes
Postcode dummy	No	No	Yes
Observations	113,325	113,325	113,325
Adjusted R-squared	0.437	0.480	0.739

Notes: Robust standard errors in parentheses. *** , **, and *, indicate that the estimated regression coefficient is significant at the 0.01, 0.05, and 0.10 levels, respectively.

We start off by considering the baseline results, which tell us about the effect on housing prices from the dwelling being built on quick clay. The results can be interpreted by considering the estimated regression coefficient β_1 on the risk-dummy LR_i (Landslide risk). This variable indicates whether the dwelling is built on quick clay and thus tells us the implicit price of there being a risk of a QCL for the dwelling.

[17] Table 6.1 utilizes model 5.2 – 5.4 for Gjerdrum.
The regression coefficients of qdate and postcode are omitted from all regression results as these are of no interest. However, these variables are important to control for effects such as time-trends, seasonality, and location-specific effects. The evaluation of the results of model 5.2-5.7, for all cases, will be centred around the variable Landslide Risk$_{it}$ (LR), as this is the variable of interest. LR$_{it}$, measures the effect on price for dwellings being at risk of a quick clay landslide.

In all three models, the coefficient of the treatment is negative, but the results are not significant for Model 5.2. Still, it gives us an indication of there being a price effect on dwellings from being built on quick clay. Considering the changes in coefficients and standard errors when controlling for unobserved time effects and location-specific effects, by including the dummy variables for quarters and postcode, respectively, it seems a more negative price effect on dwellings is attributed to the risk of QCL. For Model 5.2 and Model 5.3, we see statistically significant negative results on the treatment coefficient on a 1% significance level. This tells us that, on average, there is between a 24,351 and 76,336 NOK discount on price for dwellings at risk of a QCL. Alternatively, this is the price effect of a QCL risk. This change in the coefficient and significance level is induced to there being controlled for previously unobserved effects attributed to certain areas, given the postcode of the area. These unobserved effects would prior to controlling for them attribute their effect to all the areas included in the model, which means that there likely are areas that are impacted to a low degree which create bias in the results, as it attributes a lower degree of impact to all areas included.

Next, we extend the model by including the variable after, which indicates whether the sale occurred before or after the given QCL. This is done to determine the effect of a landslide on housing prices for dwellings built within or outside areas with quick clay. Finally, we consider the coefficient on the interaction term of our risk and treatment dummies to evaluate the effect on prices for dwellings built on quick clay after the given landslide.

The regression results are summarized in Table 6.2. The interpretation of the coefficients follows the same logic as earlier.

Table 6.2[18]: Gjerdrum DD

VARIABLES	(5.5) Selling price	(5.6) Selling price	(5.7) Selling price
Landslide risk (LR)	-20,131	-22,511	-72,504***
	(15,297)	(14,923)	(13,285)
After dummy	586,018***	-497,459***	-126,930*
	(9,282)	(161,753)	(72,546)
LR*after	-11,601	-12,278	-24,364
	(43,234)	(43,122)	(32,936)
Size in square meters	14,919***	14,918***	14,359***
	(94.32)	(92.90)	(72.91)
Between1950a1979	-116,957***	-122,148***	-75,043***
	(10,875)	(10,700)	(8,715)
Between1980a1999	301,183***	297,723***	319,022***
	(11,578)	(11,382)	(9,113)
After1999	818,785***	800,223***	904,754***
	(11,063)	(10,873)	(9,508)
Apartment	179,862***	181,101***	-337,208***
	(12,631)	(12,404)	(9,762)
Townhouse	231,214***	226,255***	-165,547***
	(13,387)	(13,044)	(8,814)
Semi-detached house	-93,563***	-94,740***	-289,128***
	(12,650)	(12,407)	(8,622)
Freehold ownership	151,610***	145,456***	94,577***
	(7,656)	(7,495)	(6,150)
Constant	874,681***	468,403***	1,632,000***
	(19,839)	(23,700)	(33,784)
Quarter dummy	No	Yes	Yes
Postcode dummy	No	No	Yes
Observations	113,325	113,325	113,325
Adjusted R-squared	0.460	0.480	0.739

Notes: Robust standard errors in parentheses. ***, **, and *, indicate that the estimated regression coefficient is significant at the 0.01, 0.05, and 0.10 levels, respectively.

The regression results on the after dummy is quite interesting. The estimated coefficient indicates the price effect on dwellings that were not built on quick clay after the Gjerdrum landslide. In the first model, while not controlling for time and location effects, there are significant results of the selling prices of these dwellings

[18] Table 6.2 utilizes model 5.5 – 5.7 for Gjerdrum.
The regression coefficients of qdate and postcode are omitted from all regression results as these are of no interest. However, are important to control for effects such as time-trends, seasonality, and location-specific effects.

being 586,018 NOK higher after the landslide. Because of the omitted control variables, these results may be biased, since there are probably unobserved time effects, such as price growth. Also, there may have been more sales occurring in more expensive areas, or there being a higher degree of sales in seasons that are more prone to higher demand and, thus, through market mechanisms, prices. Also, by omitting the quarterly variable, the estimation for sales price does not take the overall changes in price for the housing market into consideration.

By controlling for unobserved time effects in model 5.6, the coefficient on after changes drastically. This model yields significant results of the prices on dwellings at no risk for QCLs being 497,459 NOK cheaper after the landslide than before. Of course, the potential bias of more sales of dwellings in more expensive areas still exists. By controlling for this effect, we obtain a significant negative coefficient by including a dummy-coded variable for postcodes. Hence, there are clear indications of a drop in housing prices in nearby areas after QCLs.

The interaction term of the risk dummy and the after dummy indicates the average treatment effect or the difference in prices between the treatment and control groups after the landslide. All three models show a negative coefficient on this variable, which indicates that there may be an additional drop in prices for dwellings at risk of QCLs compared to those that do not face the risk. This effect may be attributed to the changes in risk perception, hence there is a salience effect from the event. Because of the severity of damages from the landslide, the latter may be a likely explanation to the effect.

To further investigate what contributes to this possible effect, we will estimate an identical regression model to Model 5.5 but with the total sales volume in each quarter as the dependent variable. This is to help determine whether the price change is due to changes in demand or a change in homebuyers' perception of risk. Also, as stated earlier in this section, will help determine whether the price effect of the QCL risk is curbed. These results are summarized in Table 6.3.

Table 6.3[19]: Volume effects investigation

VARIABLES	Volume per quarter
After dummy	-0.716
	(1.424)
LR*after	-371.8***
	(22.39)
Size in square meters	0.000769
	(0.00919)
Between1950a1979	-0.379
	(1.423)
Between1980a1999	-0.224
	(1.503)
After1999	0.623
	(1.562)
Apartment	-0.786
	(1.372)
Townhouse	-1.619
	(1.491)
Semi-detached house	0.00248
	(1.304)
Freehold ownership	-0.195
	(1.174)
Constant	3,524***
	(2.849)
Quarter dummy	Yes
Postcode dummy	Yes
Observations	113,325
Adjusted R-squared	0.986

Notes: Robust standard errors in parentheses. *** , **, and *, indicate that the estimated regression coefficient is significant at the 0.01, 0.05, and 0.10 levels, respectively.

[19] The regression coefficients of qdate and postcode are omitted from all regression results as these are of no interest. Also, LR is omitted, since it gives little to no explanatory value, as there are less dwellings at risk than those that are not will cause there to be naturally fewer sales of dwellings at risk, compared to those that are not. However, they are important to control for effects such as time-trends, seasonality, and location-specific effects.

The regression on total number of dwellings sold per quarter yields negative coefficients on *after,* and the interaction term *LR*after*. Although only the coefficient of the interaction term is statistically significant, there are indications that the number of dwellings sold after the landslide occurred for those dwellings that are not at risk is also affected to some degree, given the negative coefficient of *after.* It also seems that there are even fewer sales of dwellings that are at risk of QCLs even before the landslide. To conclude, there seem to be a drop in sales volume or demand for dwellings with the risk of QCLs after the landslide occurred. This means, that, as stated in the start of the results section, the reduction in sales volume for dwellings at risk will cause the coefficient on the interaction term in Model 5.5-5.7 to be curbed, which means that the price effect attributed to dwellings at risk after the given landslide might be less significant if sales volume were held constant.

6.2. Alta

We will continue our analysis by considering the regression results for the Alta case. This will be done similarly to what was done earlier for the Gjerdrum case. First, we consider the baseline results, summarized in Table 6.4.

All three models result in positive and statistically significant coefficients for the variable *LR*. So, contrary to our expectations, the prices of dwellings that are at risk do have evidence of being higher than those that are not at risk. It is challenging to intuitively explain what the cause of this may be, especially for the case of the model that has controlled for location effects by including the postcode dummy variable. Since we also controlled for the construction years for the dwellings, it is not reasonable to assume that the prices in areas with quick clay are higher because of newer homes. The discussion of this case continues further in section 6.4.

Table 6.4[20]: Alta results

VARIABLES	(5.2) Selling price	(5.3) Selling price	(5.4) Selling price
Landslide risk (LR)	61,653**	47,239**	74,663***
	(24,505)	(21,485)	(19,909)
Size in square meters	7,772***	8,022***	7,906***
	(166.2)	(146.5)	(115.2)
Between1950a1979	84,013***	94,061***	-21,798
	(27,954)	(24,873)	(19,041)
Between1980a1999	376,098***	376,240***	298,361***
	(29,033)	(25,700)	(20,230)
After1999	973,151***	910,184***	758,837***
	(31,467)	(27,654)	(22,654)
Apartment	7,699	72,547***	-276,460***
	(24,425)	(20,972)	(18,326)
Townhouse	47,072*	1,397	-166,948***
	(27,819)	(23,148)	(19,252)
Semi-detached house	-44,866**	-33,470*	-281,728***
	(21,700)	(18,082)	(15,468)
Freehold ownership	-47,063***	4,478	-71,527***
	(17,887)	(15,032)	(13,770)
Constant	983,382***	279,960***	120,157*
	(41,976)	(63,917)	(67,200)
Quarter dummy	No	Yes	Yes
Postcode dummy	No	No	Yes
Observations	15,870	15,870	15,870
Adjusted R-squared	0.325	0.513	0.704

Notes: Robust standard errors in parentheses. *** , **, and *, indicate that the estimated regression coefficient is significant at the 0.01, 0.05, and 0.10 levels, respectively.

Now we extend the models by including the variable after and the interaction term between after and treat to investigate the effects on prices for dwellings in Alta and surrounding municipalities after the QCL in Alta. The results of these estimated regression models are summarized in Table 6.5.

[20] . Table 6.4 utilizes model 5.2 – 5.4 for Alta.
The regression coefficients of qdate and postcode are omitted from all regression results as these are of no interest. However, they are important to control for effects such as time-trends, seasonality, and location-specific effects.

Table 6.5[21]: Alta results DD

VARIABLES	(5.5) Selling price	(5.6) Selling price	(5.7) Selling price
Landslide risk (LR)	73,894***	49,515**	67,141***
	(25,173)	(22,376)	(20,844)
After dummy	605,849***	-52,309	-93,454
	(18,644)	(81,496)	(61,238)
LR*after	-61,019	-16,111	60,169
	(73,995)	(73,020)	(48,631)
Size in square meters	7,822***	8,022***	7,906***
	(161.0)	(146.6)	(115.2)
Between1950a1979	76,793***	94,156***	-21,391
	(27,258)	(24,874)	(19,031)
Between1980a1999	366,875***	376,265***	298,703***
	(28,254)	(25,700)	(20,223)
After1999	947,892***	910,226***	758,949***
	(30,604)	(27,653)	(22,651)
Apartment	35,371	72,439***	-276,531***
	(23,503)	(20,976)	(18,329)
Townhouse	37,881	1,131	-167,188***
	(26,969)	(23,153)	(19,238)
Semi-detached house	-50,871**	-33,333*	-281,531***
	(20,563)	(18,084)	(15,471)
Freehold ownership	-32,195*	4,390	-71,401***
	(17,075)	(15,036)	(13,774)
Constant	884,372***	279,999***	119,947*
	(40,961)	(63,920)	(67,200)
Quarter dummy	No	Yes	Yes
Postcode dummy	No	No	Yes
Observations	15,870	15,870	15,870
Adjusted R-squared	0.379	0.513	0.704

Notes: Robust standard errors in parentheses. *** , **, and *, indicate that the estimated regression coefficient is significant at the 0.01, 0.05, and 0.10 levels, respectively.

We found that the results change dramatically between the three models for the Alta case. First, considering the coefficient for *after*, we see that Model 5.5 yields a statistically significant coefficient of 605,849, implying that prices of dwellings with no risk of QCLs increased by 605,849NOK after the landslide. After controlling for

[21] Table 6.5 utilizes models 5.5 – 5.7 for Alta.

The regression coefficients of qdate and postcode are omitted from all regression results as these are of no interest. However, are important to control for effects such as time-trends, seasonality, and location-specific effects.

unobserved time effects, the coefficient is now negative 52,309, although it is not statistically significant in model 5.6. After controlling for location effects, the after coefficient is -93,454 in model 5.7. Furthermore, the prices of dwellings at risk of QCLs have no significant changes after the landslide occurred, and the estimated regression coefficient on this variable is negative for Model 5.5.6 and 5.7, but positive for Model 5.5.

Similarly, to the results of the control group after the landslide, the coefficient of the interaction term change as we add additional controls. In Model 5.5, the coefficient is -61,019, which is the effect we expect to see. This price effect diminishes greatly by adding time effects but is still negative. Further adding location effects, the coefficient is no longer negative. Since these results are not statistically significant, it is unreasonable to conclude about the effects on prices for dwellings at risk of QCLs. We have no conclusive indications of what may cause these effects, but we found conclusive indications of higher prices for dwellings built on quick clay than those that are not. However, there are non-observed factors that the data does not capture, such as whether the dwellings at risk of QCLs are located in popular neighbourhoods, by a nature preserve, or if the dwellings' view is stunning or spectacular. Similar measures of such amenity effects are discussed by Naoi (2009). As stated earlier, we will delve deeper into these effects in section 6.4.

Again, we investigate the volume effects for the treatment group and the landslide by estimating the identical regression, with the total sales volume for each quarter as the dependent variable, instead of the selling prices, for which the results are summarized in Table 6.6.

Table 6.6[22]: Volume effects

VARIABLES	Volume per quarter
After dummy	1.528
	(1.202)
LR*after	-62.64***
	(3.243)
Size in square meters	0.00228
	(0.00282)
Between1950a1979	0.362
	(0.576)
Between1980a1999	0.207
	(0.588)
After1999	-0.0523
	(0.675)
Apartment	0.271
	(0.504)
Townhouse	0.461
	(0.424)
Semi-detached house	0.679
	(0.443)
Freehold ownership	-0.0970
	(0.419)
Constant	120.3***
	(1.943)
Quarter dummy	Yes
Postcode dummy	Yes
Observations	15,870
Adjusted R-squared	0.977

Notes: Robust standard errors in parentheses. *** , **, and *, indicate that the estimated regression coefficient is significant at the 0.01, 0.05, and 0.10 levels, respectively.

Similar to the case of Gjerdrum, we see that there are significant results indicating fewer sales by quarter for dwellings at risk of QCLs after the landslide. This is likely due to the supply for these dwellings being lower, as the seller are less willing to sell these dwellings at a loss, and therefore, rather waiting for the salience effects of the landslide to diminish. Alternatively, buyers are less willing to buy these dwellings, as

[22] The regression coefficients of qdate and postcode are omitted from all regression results as these are of no interest. Also, LR is omitted, since it gives little to no explanatory value, as there are less dwellings at risk than those that are not will cause there to be naturally fewer sales of dwellings at risk, compared to those that are not. However, they are important to control for effects such as time-trends, seasonality, and location-specific effects.

their perception of the risk attributed to these dwellings have changed. This means that dwellings at risk may not be sold at all in the time period included in the dataset. As stated earlier in section 6, this probably causes an effect that curbs the effects on prices for dwellings at risk of landslides after the landslide occurred.

6.3. Lyngen

Finally, we will analyse the results of the last case, which is the case of Lyngen and its nearby municipals. The baseline results are presented in Table 6.7.

Table 6.7[23]: Lyngen results

VARIABLES	(5.2) Selling price	(5.3) Selling price	(5.4) Selling price
Landslide risk (LR)	312,367***	253,215***	104,790***
	(15,630)	(13,581)	(26,740)
Size in square meters	12,998***	13,575***	13,267***
	(205.7)	(189.4)	(133.9)
Between1950a1979	-162,509***	-178,759***	588.2
	(25,227)	(22,484)	(15,968)
Between1980a1999	303,540***	296,091***	316,842***
	(25,918)	(22,973)	(16,787)
After1999	782,313***	639,833***	519,704***
	(25,816)	(22,620)	(17,401)
Apartment	739,716***	808,996***	-123,365***
	(27,870)	(25,097)	(19,464)
Townhouse	818,467***	788,488***	122,337***
	(36,919)	(32,442)	(22,333)
Semi-detached house	605,149***	646,122***	-85,339***
	(28,187)	(24,877)	(18,359)
Freehold ownership	354,277***	380,715***	122,390***
	(13,904)	(11,463)	(10,925)
Constant	372,027***	-522,379***	-1,153,000***
	(45,713)	(68,540)	(72,708)
Quarter dummy	No	Yes	Yes
Postcode dummy	No	No	Yes
Observations	31,717	31,717	31,717
Adjusted R-squared	0.290	0.453	0.745

Notes: Robust standard errors in parentheses. *** , **, and *, indicate that the estimated regression coefficient is significant at the 0.01, 0.05, and 0.10 levels, respectively.

[23] The regression coefficients of qdate and postcode are omitted from all regression results as these are of no interest. However, they are important to control for effects such as time-trends, seasonality, and location-specific effects. Table 6.7 utilizes model 5.2 – 5.4 for Lyngen.

The regression results for Lyngen are similar to the results we found for the Alta case earlier. The results indicate a statistically significant price premium for dwellings built on quick clay. By adding the same controls as for the two earlier cases, the total effect in NOK on the prices of the dwellings at risk of QCLs decreases. The biggest changes in the coefficient of *LR* occur when we add control for location effects by adding the variable *postcode.* Including postcode in our model induces a significantly lower estimated regression coefficient and increases the standard errors.

Following the same course of the analysis, we turn to look at the effect of the landslide on housing prices by introducing the *after* dummy. Recall that estimates the price effect on dwellings not at risk after the landslide. Also Again, recall that this dummy equal one if the sale occurred after the landslide and zero if before. Furthermore, the interaction term between the variables *after* and *LR* is added to examine the additional price effect on the treatment group, alternatively the dwellings built on quick clay, and therefore at the price effect of the risk of QCLs after the landslide, compared to those not at risk.

For Model 5.5, we find significant results of a drastic jump in prices for dwellings that are not at risk for QCLs after the landslide event in Lyngen. By adding controls for unobserved time effects, the after coefficient is no longer significant and changes from a positive value of approximately 1,000,000 in Model 5.5Model 5.4 to –99,058 in model 5.6. This indicates that the general price growth in the housing market and that some periods induce bias in the estimated coefficient because of sales of relatively expensive homes in specific periods. By further adding controls for location effects, the after coefficient is positive again, however still not statistically significant.

Table 6.8[24]: Lyngen results DD

VARIABLES	(5.5) Selling price	(5.6) Selling price	(5.7) Selling price
Landslide risk (LR)	296,164***	296,697***	55,931*
	(27,107)	(26,856)	(33,355)
After dummy	995,958***	-99,058	41,791
	(14,150)	(66,574)	(53,658)
LR*after	-10,541	-49,533	56,439**
	(31,518)	(30,601)	(24,773)
Size in square meters	13,255***	13,572***	13,271***
	(197.3)	(189.4)	(133.9)
Between1950a1979	-162,435***	-178,762***	1,085
	(23,867)	(22,494)	(15,964)
Between1980a1999	314,354***	295,948***	317,248***
	(24,470)	(22,980)	(16,786)
After1999	724,665***	640,209***	519,251***
	(24,354)	(22,626)	(17,408)
Apartment	758,219***	808,545***	-123,119***
	(26,575)	(25,100)	(19,465)
Townhouse	823,074***	787,865***	122,728***
	(34,930)	(32,443)	(22,331)
Semi-detached house	621,713***	645,930***	-85,328***
	(26,642)	(24,879)	(18,358)
Freehold ownership	388,584***	380,716***	122,005***
	(13,009)	(11,460)	(10,933)
Constant	-523,383***	-529,763***	-1.144e+06***
	(45,442)	(68,866)	(72,873)
Quarter dummy	No	Yes	Yes
Postcode dummy	No	No	Yes
Observations	31,717	31,717	31,717
Adjusted R-squared	0.361	0.453	0.745

Notes: Robust standard errors in parentheses. *** , **, and *, indicate that the estimated regression coefficient is significant at the 0.01, 0.05, and 0.10 levels, respectively.

Considering the case of dwellings that are at risk, the resulting regression analysis yields negative but non-significant estimators of the coefficient of the interaction term in model 5.5 and 5.6. Again, by adding controls for location effects, this coefficient turns positive and is also statistically significant, as we did not expect. However, this does not mean that a landslide makes dwellings at risk more expensive. Again, this

[24] . Table 6.8 utilizes model 5.5 – 5.7 for Lyngen.
The regression coefficients of qdate and postcode are omitted from all regression results as these are of no interest. However, they are important to control for effects such as time-trends, seasonality, and location-specific effects.

estimates result may arise because of curbed effects following changes in sales volume for such dwellings.

Again, estimating the effect on volume arising from the landslide, to determine whether there are curbing effects of the price changes. This is summarized in Table 6.9.

Table 6.9[25]: Volume effects

VARIABLES	Volume per quarter
After dummy	21.70***
	(2.115)
LR*after	-121.3***
	(2.274)
Size in square meters	-0.00247
	(0.00418)
Between1950a1979	0.498
	(0.591)
Between1980a1999	0.0705
	(0.635)
After1999	0.0695
	(0.662)
Apartment	0.950
	(0.692)
Townhouse	0.301
	(0.850)
Semi-detached house	0.641
	(0.649)
Freehold ownership	0.660
	(0.580)
Constant	218.2***
	(2.352)
Quarter dummy	Yes
Postcode dummy	Yes
Observations	31,717
Adjusted R-squared	0.972

Notes: Robust standard errors in parentheses. ***, **, and *, indicate that the estimated regression coefficient is significant at the 0.01, 0.05, and 0.10 levels, respectively.

[25] The regression coefficients of qdate and postcode are omitted from all regression results as these are of no interest. Also, LR is omitted, since it gives little to no explanatory value, as there are less dwellings at risk than those that are not will cause there to be naturally fewer sales of dwellings at risk, compared to those that are not. However, they are important to control for effects such as time-trends, seasonality, and location-specific effects.

Looking at the regression results for the volume effects on housing purchases, there are findings of significance of the changes in sales volume for dwellings at risk and dwellings not at risk after the landslide. There is an increase in purchases of dwellings that are not at risk after the landslide and a quite dramatic drop in sales volume of dwellings at risk. Intuitively, this might indicate that people who buy homes in these areas choose to purchase dwellings that are not at risk, compared to those at risk, at a higher rate than before the landslide. These effects, again, plausibly curb the implicit price of QCL risk.

6.4. Further discussion

In this section of the thesis, we focus on summarising the results from the three different cases presented in the previous subsections of the section while comparing these to the results of the included reference articles from prior research. Also, we will discuss some of the empirical considerations we have not addressed earlier. Probably the most important source of bias in our model is that we do not include a scale of risk but rather just a dummy variable for all risk levels of QCLs. We can expect the negative price effect to be more substantial in areas with a higher reported degree of risk than in areas with lower risk. This effect was illustrated in Kiel and Matheson (2018), where they first estimated a model where all levels of risk over low risk were grouped in a dummy variable, which is similar to what we did. They report that housing sold in risky areas was sold for approximately 5.6% less than the housing in areas with no risk before the forest fire occurred, and housing in risky areas sold for an additional 4.5% less than housing in non-risky areas after the fire. However, none of these results was statistically significant, which is also recognisable in this thesis. By estimating the price effect of the four different levels of risk, they found significant evidence that dwellings in very high-risk areas saw a 21.7% reduction in selling prices compared to dwellings in low-risk areas after the event. This clearly illustrates a major concern about statistical bias in the estimates for the price effect on dwellings in areas prone to QCLs compared to those that are not. Therefore, we highly recommend incorporating the six-scale (0-6) degree of QCL risk from NVE. Alternatively, one can address this issue by using distance from areas with quick clay, comparable to Kim, Park, Yoon, and Cho's (2017) approach.

Recall that the estimated regression results of the variables of interest, LR, after, and their interaction term have yielded significantly different results across the three cases, Gjerdrum, Alta and Lyngen, that we included in the thesis. These results were far from expected, as prior research has quite consistent results for there being drops in prices for dwellings at risk following such natural disasters as landslides and so on. Although, these may cast light upon the possible fact that there may be significant differences in the risk perception and willingness to take a risk for buyers in different regions. In Model 5.2-5.5 for the Gjerdrum case, we find indications and possibly evidence of lower prices, or a price discount, on dwellings that are at risk of a QCL compared to those that are not. This aligns with our expectations and makes intuitively sense that there should be lower prices for these dwellings because of the added risk of landslides, holding all else constant.

On the contrary, there is an indication and possible evidence of higher prices relative to dwellings not at risk for the dwellings at risk in the Alta and Lyngen case. Such estimates were found both before and after the given landslides for each of the three cases. This may indicate a higher degree of risk-averse behaviour from homebuyers in the Gjerdrum case than in Alta and Lyngen. Alternatively, recall that in earlier subsections of this section, we discussed that there might be some significant differences in the quality, and therefore prices, for dwellings at risk of landslides in the Alta and Lyngen regions that are not captured. Such unobserved effects are not captured in our regression models, which may create bias in the estimates. Intuitively, it does not make sense that home buyers want to pay a premium for a dwelling built on quick clay and, therefore, a higher degree of risk. What is sensible is the opposite of this case, which may indicate that there are some amenity effects like those in Kim et al. (2017). Suppose there is a price premium attributed to amenity effects. In that case, these are probably present because there are qualities such as living close to a nature park, nice views, etc. These effects might be captured by controlling for location-specific effects differently than we did. One can, for example, include additional controls for these effects, such as distance to the nearest bus stop, city size, and so on, such as in Naoi, Seko & Sumita (2009).
Of course, there are probably different solutions to capturing this effect, such as using dummies for postcodes with three digits, yielding a higher degree of generality. There are also additional controls for housing characteristics, such as the number of

bathrooms, whether there is a balcony, and so on. Again, these are unobserved effects in our models that probably cause bias in the results. Naoi, Seko & Sumita (2009) also included fixed effects for the owners and renters of the dwellings in their analysis. This was probably included because their data were gathered through surveys rather than registered data. However, they show evidence that there are changes in the results based on controlling for these effects, including age, education, marital status, etc. This may contribute to the results because there might be differences in the subjective risk assessment of each home buyer compared to the objective risk measures provided by entities such as in this thesis, NVE.
Another thing to consider that may induce bias in the estimated results is the presence of other, unobserved risks, such as the risk of flooding and forest fires. Contrary to prior research, we used a linear price scale instead of a logarithmic one. In hindsight, the interpretation of the regression coefficients may have been more intuitive if we had used a logarithmic transformation of the selling prices as the dependent variable compared to the linear one used in this thesis.

Further, we think that the analysis of the results will yield more convincing estimates when there has been a more extended period to gather data on housing sales from. The landslide happened recently in the Alta and Gjerdrum regions, so we only have data for approximately one and a half years for the Alta case and one year for the Gjerdrum case. This would also open the possibility of looking at short-term and long-term effects, such as in Kim et al. (2017).

One of the most challenging issues was the definitions of the treatment and control groups. Our analysis is still of value, but by adjusting what areas are included in the groups, such as having dwellings at risk only in Gjerdrum and comparing these to dwellings that are not at risk in the neighbouring municipals.
Contrary to our approach, one should consider omitting remote areas relative to the landslide event area or using only the remote areas in the control group. By omitting remote areas, one can further control for location effects by different means than using a postcode or municipal dummy. Nevertheless, this would only be appropriate if we sought to answer other research questions. Alternatively, the analysis can be altered to a spatial DD model, where you analyse the effects on price based on the centroid of the given landslide. This also seems like an intuitive solution to further

analysing the issue presented in this thesis. We found evidence that the prices for dwellings that are not at risk experience a reduction after landslides, which we saw for the cases Gjerdrum and Alta. This may be due to an association effect of such events, where home buyers change their perception of risk after a QCL occurs close by to housing that is not built on quick clay but close to areas of quick clay. As mentioned earlier, the idea was to extend our analysis by including a model for a spatial DD. However, given the time constraints on our project, we had to dismiss this idea. Using a spatial approach will give a more detailed insight into how the housing prices are affected by the landslides while considering what distance the dwelling is from the actual place of the event. One example of such a study is Atreya et al. (2013), discussed in the literature review in section 2.

Another potential issue is the chosen municipals and periods to include in each case. There might be some spill-over effects for the Alta and Lyngen cases because the areas are quite close, and the neighbouring areas we chose to include are, of course, also close in distance. We also chose to include some of the same municipals for each of these two cases, which might be an issue. For lists of municipalities for each case, see appendix B.1.A, B.2.A, and B.3.A.

Our analysis in section 6 included an investigation of the effects on sales volume for dwellings at risk and the effects on sales for dwellings before and after landslides. This was to help determine whether the effects on prices can be attributed to market effects, given the changes in sales volume, or to salience effects that change home buyers' perception of risk. Although this gave some indication of the issue, we must consider that these effects may be strongly correlated. If homebuyers become more aware of the risk of QCLs and change their risk perception, it is not given that they would want to buy dwellings at risk at a discounted price. Instead, the buyers, or at least a fraction of the buyers, would potentially not want to buy. Subsequentially, we can expect to see a reduction in sales volume. Therefore, we can hypothesize that the reduction in sales volume is somewhat attributed to the salience effect.

To further ensure our results are reliable and valid, we could have allocated more time to do robustness tests, but we have unfortunately not included this in our thesis. When all taken into consideration, several issues could have been dealt with throughout conducting this thesis. Fixing these issues could yield a more convincing result for later research.

7. Conclusion

This thesis investigates the price effect of risk for QCLs on dwellings and the price effect on dwellings at risk and not at risk after a landslide in relative proximity. This was approached by estimating regression models with a DD design, which was formulated by adding an interaction term between a dummy variable for dwellings at risk of QCLs and a dummy for whether the housing sale occurred before or after the given QCL event.

For the municipalities included in the Gjerdrum case we find evidence that the sold dwellings at risk are approximately 76,000 NOK cheaper than those not at risk. For a detailed list over the municipals included in the Gjerdrum case, see appendix B.2.A. Also, there are indications and evidence that the sold dwellings not at risk see a drop in prices of approximately between 126,000 and 497,000 NOK after the landslide in Gjerdrum occurred. This is probably due to association effects through the major news coverage on the QCL, which cause home buyers to have a lower willingness to pay for dwellings in these areas. Also, there are clear indications of further price drops for dwellings at risk after this landslide, attributed to salience effects, hence, changes in the perception of risk. If the market is efficient, this would not be the case since the price discount of the risk would always be reflected in the prices of dwellings at risk. Although the estimated results of the price changes for dwellings at risk of QCL are not statistically significant, we believe that this effect is curbed due to a drop in sales volume for these dwellings. Such effect indicates that the sellers are less willing to sell, given the loss they would face, or that buyers are less willing to move to these areas.

Unlike the Gjerdrum case, we find that the prices of dwellings at risk of QCL, for Alta and the other municipals included in the case, have a higher price than those not at risk, which is unexpected. For a detailed list over the municipals for the Alta case, see appendix B.2.A. There are findings from all three models that are significant for this case, where the estimated results are that dwellings at risk are between approximately 47,000 and 74,000 NOK, more expensive than those not at risk. However, this does not mean that buyers are more willing to pay a premium for the risk. Instead, this is probably due to some unobserved amenity effects, making home buyers willing to pay more to live in the areas facing risk. Such unobserved effects may be due to amenities such as a pleasant surrounding environment, great views, etc. Also, surprisingly, the resulting estimates of model 5.2 show indications of the prices of dwellings not at risk increasing by over 600,000 NOK after the landslides, which is unexpected, to say the least. However, we find indications that the prices drop following the QCL by adding controls. In model 5.3, the coefficient is -93,454. Although not significant, it indicates some association effects for homebuyers in these areas. The results for the dwellings at risk after the landslide are mixed between the models and have no significant results. Again, these results are probably of little value since there is a significant drop in the number of these dwellings following the landslide in Alta, which means that the price effect gets curbed.

In similarity with the Alta case, the analysis of Lyngen, in addition to its surrounding and remote areas, tells us that at-risk dwellings have higher prices than those not at risk. These results are significant in all three models, and the price premium of at-risk dwellings is found to be between approximately 104,000 and 312,000 NOK higher than those not at risk. Again, this is probably due to amenity effects like those we hypothesize driving the price premium for the Alta case, which is unobserved. We also find evidence that the prices of at-risk dwellings further increase by approximately 56,000 NOK after the landslide event in Lyngen, 2010. However, before adding controls for postcodes, there are indications of the opposite – that the prices drop for these dwellings. We find no convincing estimates of the price changes for dwellings that are not at risk after the landslide. However, in model 5.2, there is a significant increase in prices of approximately 996,000 NOK. This effect is likely attributed to significant increases in housing prices in areas such as Tromsø, which induces bias in the model's estimation. This is likely since the effect is no longer significant by adding controls for postcode and drops to an increase of approximate 40,000 NOK, which is also unexpected. This is in combination with the market effects contributing to the prices dropping for these dwellings, as the average sales volume drops significantly for a longer time (2010-2021), which makes the results even more unexpected. This may indicate that the amenity effects in the areas at risk for landslides are more substantial than the effects of the risk and market mechanisms.

Given the mixed results, we find it reasonable to think that changes in prices and, therefore, home buyers' risk perception is correlated to the severity of the damages from the QCLs or likely the degree of risk, which both are not observed in our models. However, we have found an effect on prices both for the dwellings being at risk of QCLs and prices of dwellings after QCLs occur.

7.1. Future research suggestions

As we have discussed in section 6.4, several elements could be and should be conducted further in future research to research the subject of risk of and actual occurrences of QCLs' effect on housing prices.

First, we highly recommend using a scale with multiple levels of risk rather than just a dummy variable indicating whether there is a risk of QCLs, which is in line with prior research, such as in Atreya et al. (2013), Kiel & Matheson (2018), and Kim et al. (2017). In their paper, Kiel and Matheson (2018) even illustrate that this approach yields significant effects on the estimated results for the risk levels. The reason is that there may be lower effects on price for lower degrees of risk and more significant effects from larger degrees of risk, which are lost and curbed when only using a dummy for one level of risk.

Secondly, it is fascinating to determine whether the possible price effects of landslide occurrences are short-term or last for a more extended period. This will cast light upon if there is a price effect after the damages are repaired and if risk perceptions change after a while, such that the salience effect dissipates. A similar approach was attempted by Kim et al. (2017) and will demand a more extended period of data after the landslide event. This means that the research should be conducted further into the future since the landslides in Alta and Gjerdrum occurred on June third, 2020, and December 30th, 2020, respectively. In this thesis

Furthermore, approaching the model estimations using White heteroskedasticity robust standard errors may not be optimal. A more reasonable approach to control for heteroskedasticity could be conducted by using clustered standard errors. One could, for example, cluster by grouping the treatment and control group, different types of housing, such as apartments or townhouses, or by clustering postcodes. Clustering the standard errors will, in contrary to white heteroskedasticity robust standard errors, take heteroskedasticity across the chosen group into consideration. Doing so would therefore lead to a more correct estimation of the standard errors. Finally, including additional control variables for location-specific effects to eliminate bias stemming from unobserved effects will plausibly yield more convincing estimates. Doing so, one could expect the results for at least the Alta and Lyngen cases to change drastically since we determined there might be some significant amenity effects that impose bias in our estimates. We believe controlling for population size, the sociodemographic composition of the neighbourhoods, distances to grocery stores, bus stops, nature parks, etc., might significantly contribute to more convincing estimates.

References

Angrist, J. D., & Krueger, A. B. (1999). Empirical strategies in labor economics. In *Handbook of labor economics* (Vol. 3, pp. 1277–1366). Elsevier.

Angrist, J. D., & Pischke, J.-S. (2008). Mostly Harmless Econometrics: An Empiricist's Companion. In *Mostly Harmless Econometrics*. Princeton University Press. https://doi.org/10.1515/9781400829828

Anundsen, A. K., & Røed Larsen, E. (2018). Testing for micro-efficiency in the housing market. *International Economic Review*, *59*(4), 2133–2162. https://doi.org/10.1111/iere.12332

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

Berg, A. F. (2022, February 18). *Registrering av kvikkleireskred* [Personal communication].

Bin, O., & Landry, C. E. (2012). *Changes in Implicit Flood Risk Premiums: Empirical Evidence from the Housing Market* (SSRN Scholarly Paper No. 1850671). Social Science Research Network. https://doi.org/10.2139/ssrn.1850671

Botzen, W. J. W., Deschenes, O., & Sanders, M. (2019). The Economic Impacts of Natural Disasters: A Review of Models and Empirical Studies. *Review of Environmental Economics and Policy*, *13*(2), 167–188. https://doi.org/10.1093/reep/rez004

Braut, G. S., & Dahlum, S. (2021). Regresjonsanalyse. In *Store norske leksikon*. http://snl.no/regresjonsanalyse

Case, K. E., & Shiller, R. J. (1989). The Efficiency of the Market for Single-Family Homes. *The American Economic Review*, *79*(1), 125–137.

Chau, K. W., & Chin, T. L. (2002). *A Critical Review of Literature on the Hedonic Price Model* (SSRN Scholarly Paper No. 2073594). Social Science Research Network. https://papers.ssrn.com/abstract=2073594

Eiendomsverdi AS. (n.d.). Retrieved 7 January 2022, from https://eiendomsverdi.no/

Fæste, M. (2017, August 23). *Selgers opplysningsplikt ved salg av bolig*. Huseierne. https://www.huseierne.no/boligbloggen/2017/opplysningsplikt-ved-salg-av-bolig/

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417.

Flavin, M., & Yamashita, T. (2002). Owner-occupied housing and the composition of the household portfolio. *American Economic Review*, *92*(1), 345–362.

Fremtind. (2020, December 31). *Skredet i Gjerdrum: Hjelp fra forsikringsselskapet*. https://kommunikasjon.ntb.no/pressemelding/skredet-i-gjerdrum-hjelp-fra-forsikringsselskapet?publisherId=17846913&releaseId=17898545

Garnache, C. (2020). *Risk Salience or Information Disclosure? The Effect of Climate-Related Risks on the Residential Real Estate Market* (SSRN Scholarly Paper No. 3398404). Social Science Research Network. https://doi.org/10.2139/ssrn.3398404

Gjerdrum Kommune. (n.d.). *Alt om leirskredet i Gjerdrum*. Retrieved 2 February 2022, from https://www.gjerdrum.kommune.no/virksomheter/alt-om-leirskredet/

Gjerdrum Kommune. (2022, February 2). *Politiet sikter Gjerdrum kommune*. https://www.gjerdrum.kommune.no/aktuelt/politiet-sikter-gjerdrum-kommune/

Kiel, K. A., & Matheson, V. A. (2018). The effect of natural disasters on housing prices: An examination of the Fourmile Canyon fire. *Journal of Forest Economics*, *33*, 1–7. https://doi.org/10.1016/j.jfe.2018.09.002

Kim, J., Park, J., Yoon, D. K., & Cho, G.-H. (2017). Amenity or hazard? The effects of landslide hazard on property value in Woomyeon Nature Park area, Korea. *Landscape and Urban Planning*, *157*, 523–531. https://doi.org/10.1016/j.landurbplan.2016.07.012

Kousky, C. (2010). Learning from Extreme Events: Risk Perceptions after the Flood. *Land Economics*, *86*(3), 395–422. https://doi.org/10.3368/le.86.3.395

Langberg, Ø. K. (2016, December 27). *I dette landet har folk høy utdanning og mye penger. Likevel leier de bolig hele livet.* https://www.aftenposten.no/verden/i/a9QAE/mens-nordmenn-er-besatt-av-aa-kjoepe-egen-bolig-er-det-helt-vanlig-aa-leie-i-resten-av-det-rike-europa

Matre, J., Framstad, A. P., Kippernes, G. A., & Skjeggestad, H. (2010, June 9). *Ulovlig byggearbeid kan ha forårsaket skredet i Lyngen* [Newspaper]. https://www.vg.no/i/zaVyb

Mordt, H., Mikaelsen, K. S., & Tufan, Ø. (2021, January 2). *Over 110.000 bor på kvikkleire i Norge*. NRK. https://www.nrk.no/osloogviken/over-110.000-bor-pa-kvikkleire-i-norge-1.15310161

Naoi, M., Seko, M., & Sumita, K. (2009). Earthquake risk and housing prices in Japan: Evidence before and after massive earthquakes. *Regional Science and Urban Economics*, *39*(6), 658–669. https://doi.org/10.1016/j.regsciurbeco.2009.08.002

NGI. (n.d.-a). *Kvikkleireskred i Norge*. Norges Geotekniske Institutt (NGI). Retrieved 4 February 2022, from https://www.ngi.no/Tjenester/Fagekspertise/Kvikkleireskred/Kvikkleireskred-i-Norge

NGI. (n.d.-b). *Quick clay landslides*. Norwegian Geotechnical Institute (NGI). Retrieved 24 January 2022, from https://www.ngi.no/eng/Services/Technical-expertise/Quick-clay-landslides

NGI, N. G. I. (n.d.-c). *What is quick clay?* Norwegian Geotechnical Institute (NGI). Retrieved 24 January 2022, from https://www.ngi.no/eng/Services/Technical-expertise/Quick-clay-landslides/What-is-quick-clay

NOU 2022: 3—På trygg grunn (NOU 2022:3; På trygg grunn, p. 270). (2022). Olje- og energidepartementet. https://www.regjeringen.no/no/dokumenter/nou-2022-3/id2905694/

NVE. (2020, June 4). *Kvikkleireskredet i Alta: NVE undersøker grunnforholdene*. https://www.nve.no/nytt-fra-nve/nyheter-skred-og-vassdrag/kvikkleireskredet-i-alta-nve-undersoker-grunnforholdene/

NVE. (2021a, March 16). *Lav skråningsstabilitet og fyllingsarbeider årsak til kvikkleireskred i Alta kommune*. https://www.nve.no/nytt-fra-nve/nyheter-skred-og-vassdrag/lav-skraningsstabilitet-og-fyllingsarbeider-arsak-til-kvikkleireskred-i-alta-kommune/

NVE. (2021b, May 1). *Kartlegging av fareområder*. https://www.nve.no/om-nve/spoer-nve/om-kvikkleire/kartlegging-av-fareomrader/

NVE, Jernbaneverket, & Statens Vegvesen. (2013). *Vurdering av kartleggingsgrunnlaget for kvikkleire i strandsonen*. https://asp.bibliotekservice.no/nve/title.aspx?tkey=21300

Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T., & French, N. (2003). Real estate appraisal: A review of valuation methods. *Journal of Property Investment & Finance*. https://doi.org/10.1108/14635780310483656

Røed Larsen, E., & Weum, S. (2008). Testing the efficiency of the Norwegian housing market. *Journal of Urban Economics*, *64*(2), 510–517.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 34–55.

Rosenzweig, M. R., & Wolpin, K. I. (2000). Natural 'Natural Experiments' in Economics. *Journal of Economic Literature*, *38*(4), 827–874. https://doi.org/10.1257/jel.38.4.827

Schwerdt, G., & Woessmann, L. (2020). Empirical methods in the economics of education. *The Economics of Education*, 3–20. https://doi.org/10.1016/B978-0-12-815391-8.00001-X

Selim, H. (2009). Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. *Expert Systems with Applications*, *36*(2), 2843–2852. https://doi.org/10.1016/j.eswa.2008.01.044

Shiller, R. J., & Pound, J. (1989). Survey evidence on diffusion of interest and information among investors. *Journal of Economic Behavior & Organization*, *12*(1), 47–66. https://doi.org/10.1016/0167-2681(89)90076-0

Spjeldnæs, N. (2021). Kvikkleire. In *Store norske leksikon*. http://snl.no/kvikkleire

SSB. (2022, March 22). *Boforhold, registerbasert*. SSB. https://www.ssb.no/bygg-bolig-og-eiendom/bolig-og-boforhold/statistikk/boforhold-registerbasert

stata.com. (n.d.). Graph twoway scatter—Twoway scatterplots. Retrieved 15 May 2022, from https://www.stata.com/manuals13/g-2graphtwowayscatter.pdf

Stock, J. H., & Watson, M. W. (2020). *Introduction to econometrics 4th ed* (4th ed.). Pearson Education Limited.

Taylor, S. E., & Thompson, S. C. (19820101). Stalking the elusive. *Psychological Review*, *89*(2), 155–181. https://doi.org/10.1037/0033-295X.89.2.155

Tiltnes, A. H. (2010, September 3). *Her gikk raset*. Nordlys. https://www.nordlys.no/1-79-5277258

Tufte, P. A. (2020, February 10). *Statistiske metoder -forelesning 02.10.2020*. Multippel lineaer regresjon: Fallgruver, Oslo.

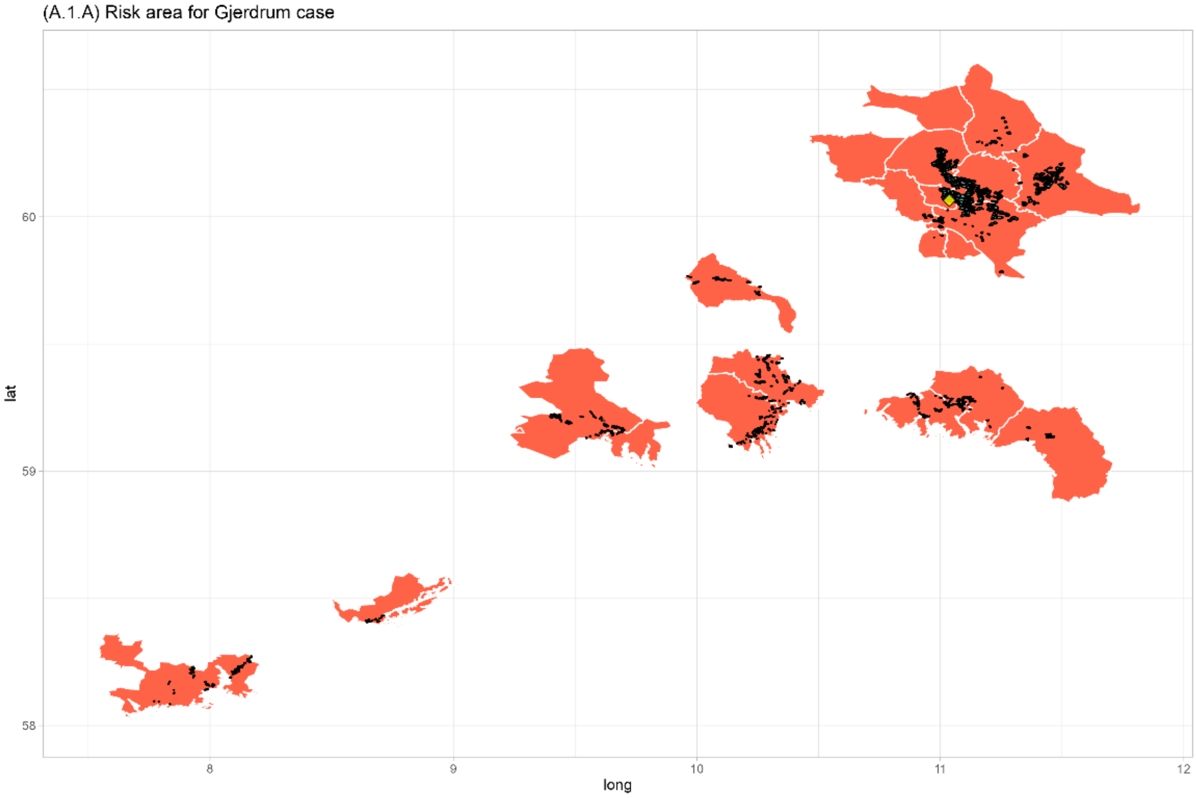
Wooldridge, J. M. (2018). *Introductory Econometrics: A Modern Approach* (7th ed.). Cengage Learning inc.

Appendices

A. Maps

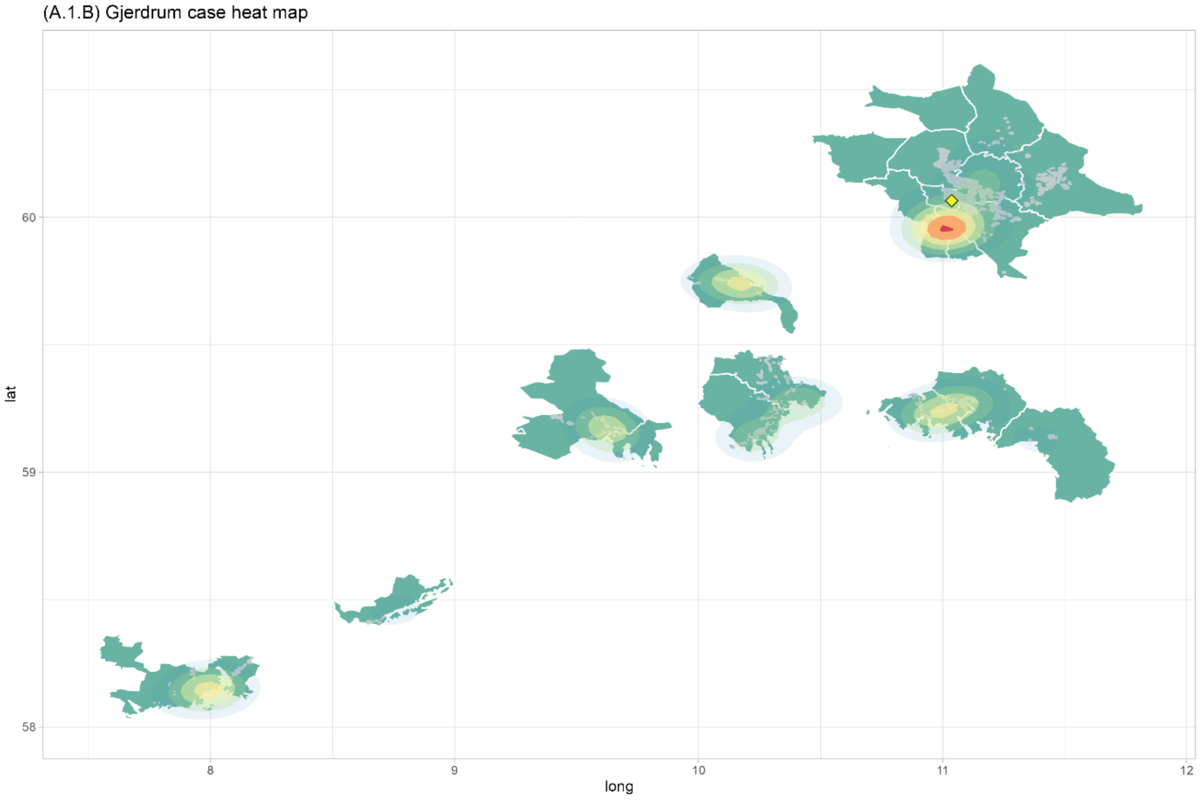
A.1. Gjerdrum case

A.1.A[26]. Risk area map for Gjerdrum case



[26] Appendix A.1.A visualises the risk areas for all included municipals for the Gjerdrum case in the colour "tomato", while the black areas on the map represents the risk areas for quick clay landslide for all municipals in the Gjerdrum case. For all maps in appendix A.1, and the interactive map, the risk zones only include the areas at risk of the risk quick clay landslide, as defined by NVE, and not by the assumption for this thesis. Risk areas for other landslides are not included as the maps for these were separated for each scenario, hence it would be too time consuming to plot. The maps are created based on map data from GeoNorge and NVE (2022).

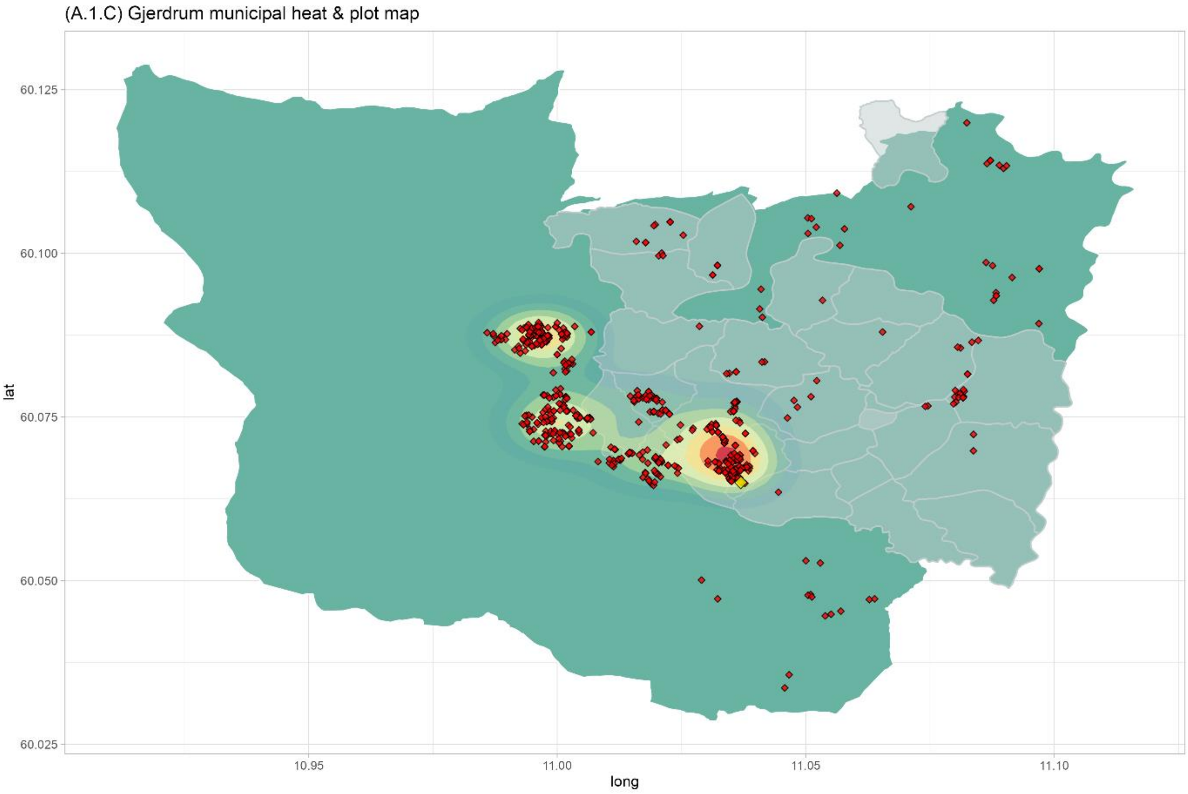
A.1.B[27] Heatmap for Gjerdrum case



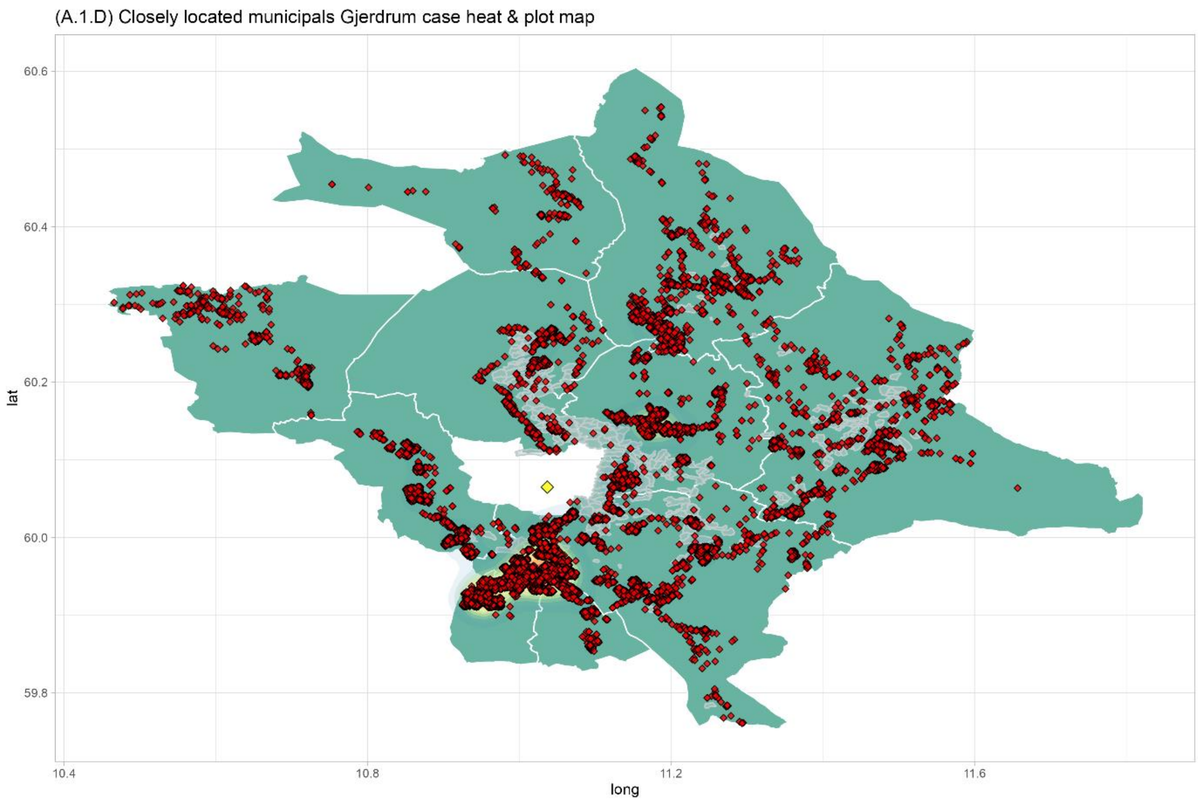
[27] Appendix A.1.B shows the included municipals for the Gjerdrum case in the colour "green". The heatmaps are based on a density probability, hence the darker the area, the more probable it is that a random observation is located within the filled area. The risk areas for quick clay, with the same perquisitions as mentioned in footnote 26, are the areas with a light grey colour. The place of event (landslide) is plotted in yellow, while the plotting of sold dwellings, for the remainder of the maps, are represented in red, given that no other explanation is given for the map in question. For a map over all sold dwellings in the Gjerdrum case, see map 4.1

For the remainder of the maps in appendix A.1, the municipals are separated into the three groups: 1) **Municipal of the event,** including the municipal Gjerdrum, 2) **closely located municipalities**, including the municipals Nittedal, Ullensaker, Nannestad, Lillestrøm, Lørenskog, Nes, Rælingen, Lunner, Hurdal, Eidsvoll, and 3) **Remote located municipalities**, including the municipals Kristiansand, Fredrikstad, Drammen, Sarpsborg, Skien, Arendal, Sandefjord, Tønsberg, Porsgrunn, Halden.

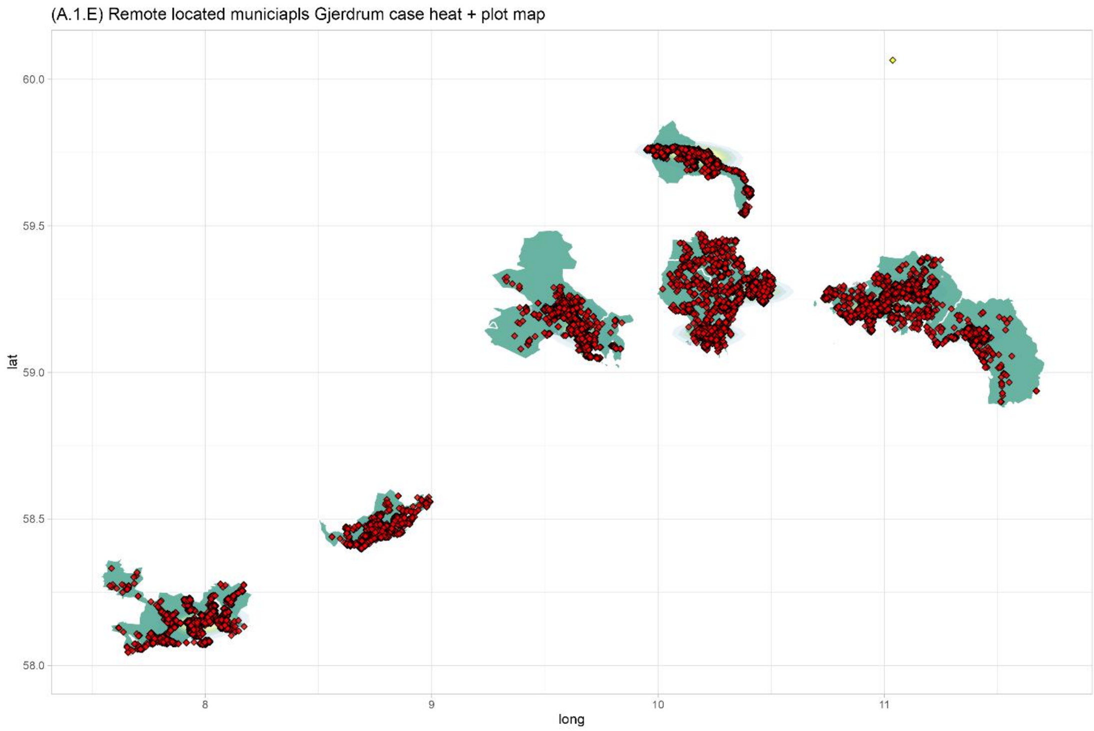
A.1.C Heatmap and spread of sold dwellings for Gjerdrum Municipal



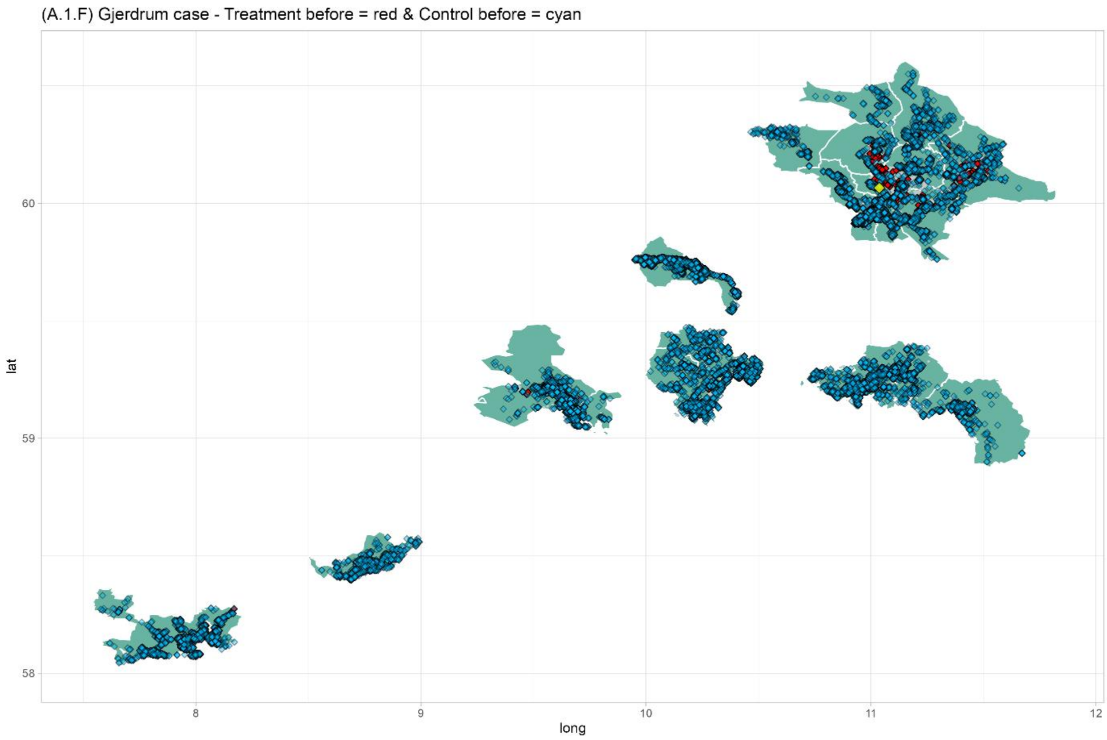
A.1.D Heatmap and spread for sold dwellings for closely located municipalities for Gjerdrum case



A.1.E Heatmap and spread of sold dwellings for remote located municipals for Gjerdrum case

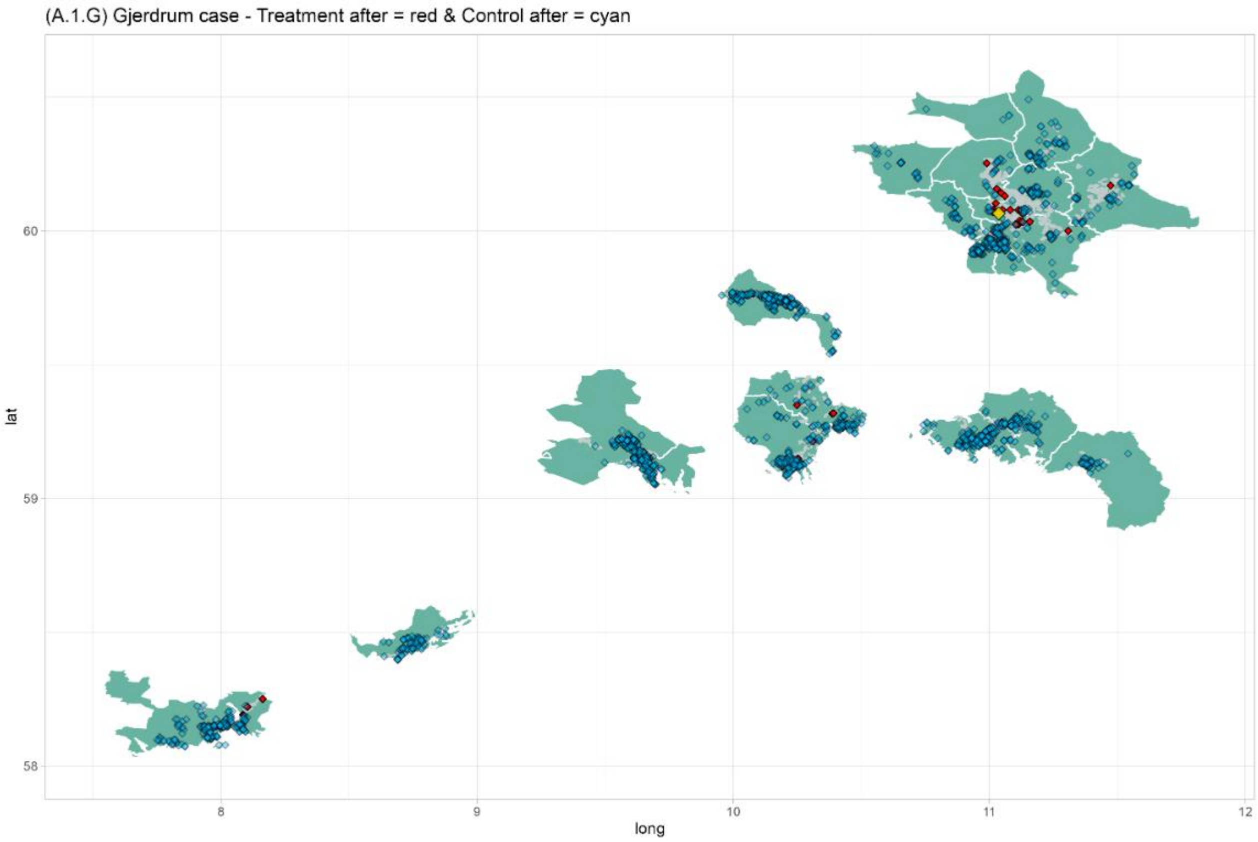


A.1.F[28] Spread of sold dwellings for treatment and control groups before the events.



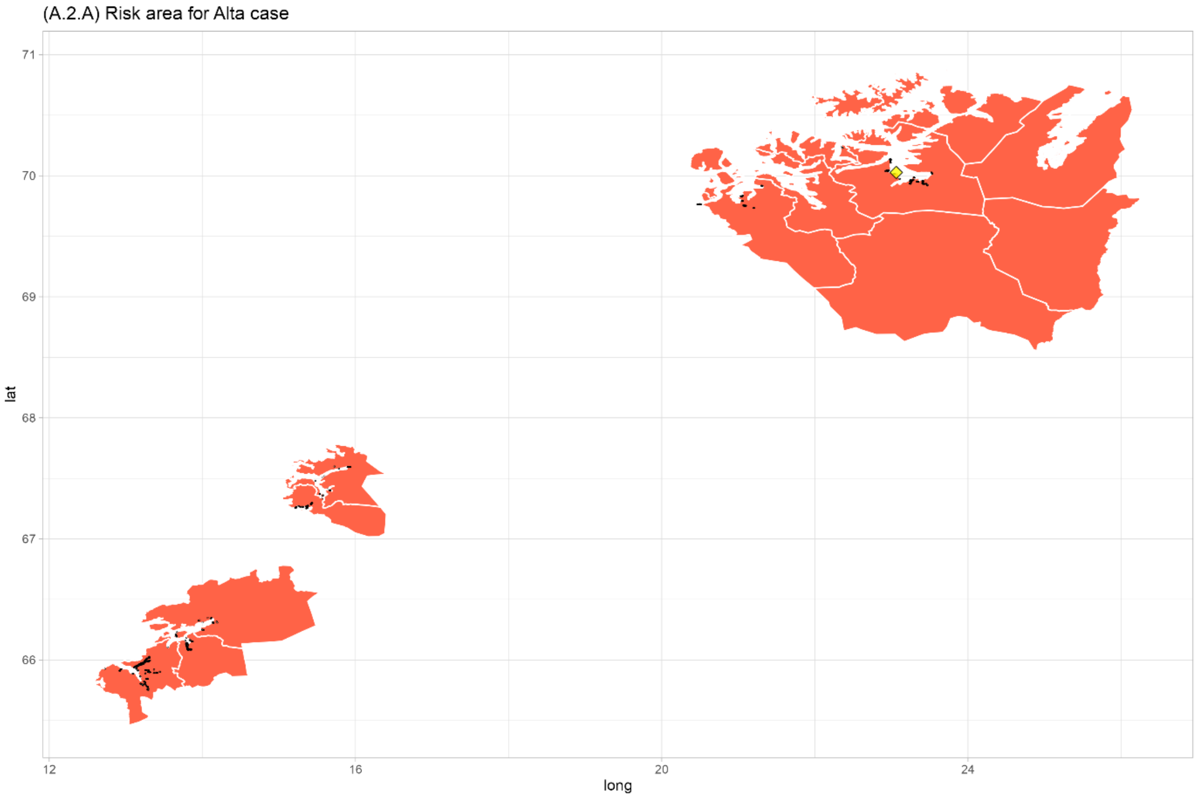
[28] This map follows the same logic as Map 5.1: Spread of sold dwellings in treatment (red) and control groups (cyan) for Gjerdrum., where the red plots are representing sold dwellings in treatment group for Gjerdrum, after the event and the cyan plots are representing sold dwellings in control group for Gjerdrum, after the event. The same logic goes for map A.1.G, where the plots represent sold dwellings before the event.

A.1.G: Spread of sold dwellings for treatment and control groups after the events.



A.2. Alta case

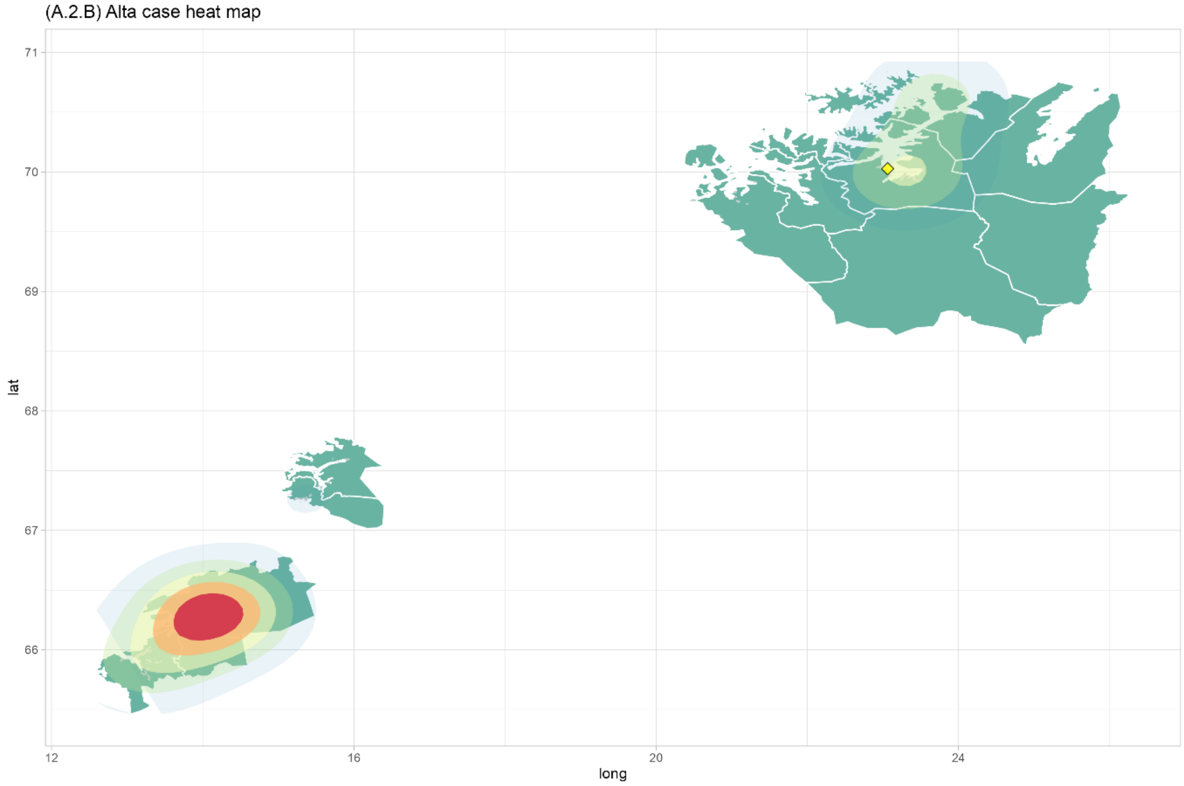
(A.2.A) Risk area map for Alta case[29]



29

Appendix A.2.A visualises all included municipals for the Alta case in the colour "tomato", while the black areas on the map represents the risk areas for quick clay landslide for all municipalities. For all maps in appendix A.1, and the interactive map, the risk zones only include the areas at risk of the risk quick clay landslide, as defined by NVE, and not by the assumption for this thesis. Risk areas for other landslides are not included as the maps for these were separated for each scenario, hence it would be too time consuming to plot. The maps are created based on map data from GeoNorge and NVE (2022).

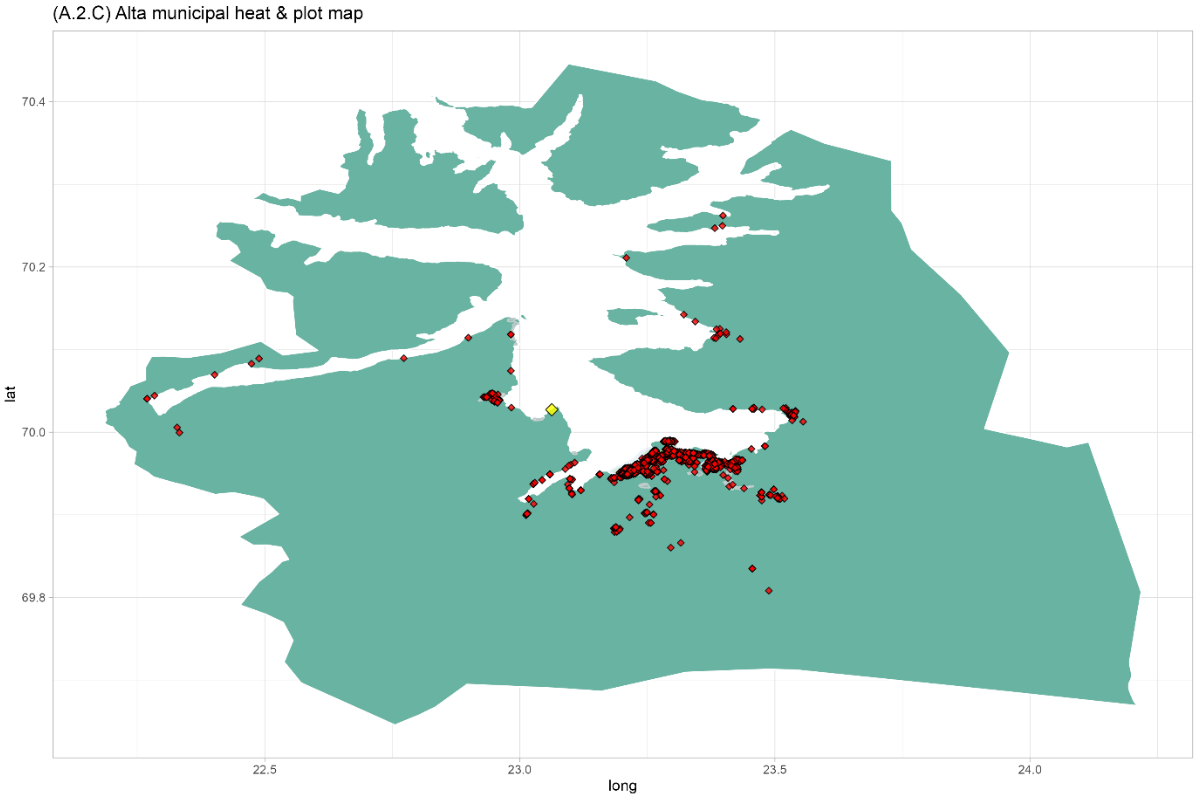
A.2.B[30]: Heatmap for Alta case



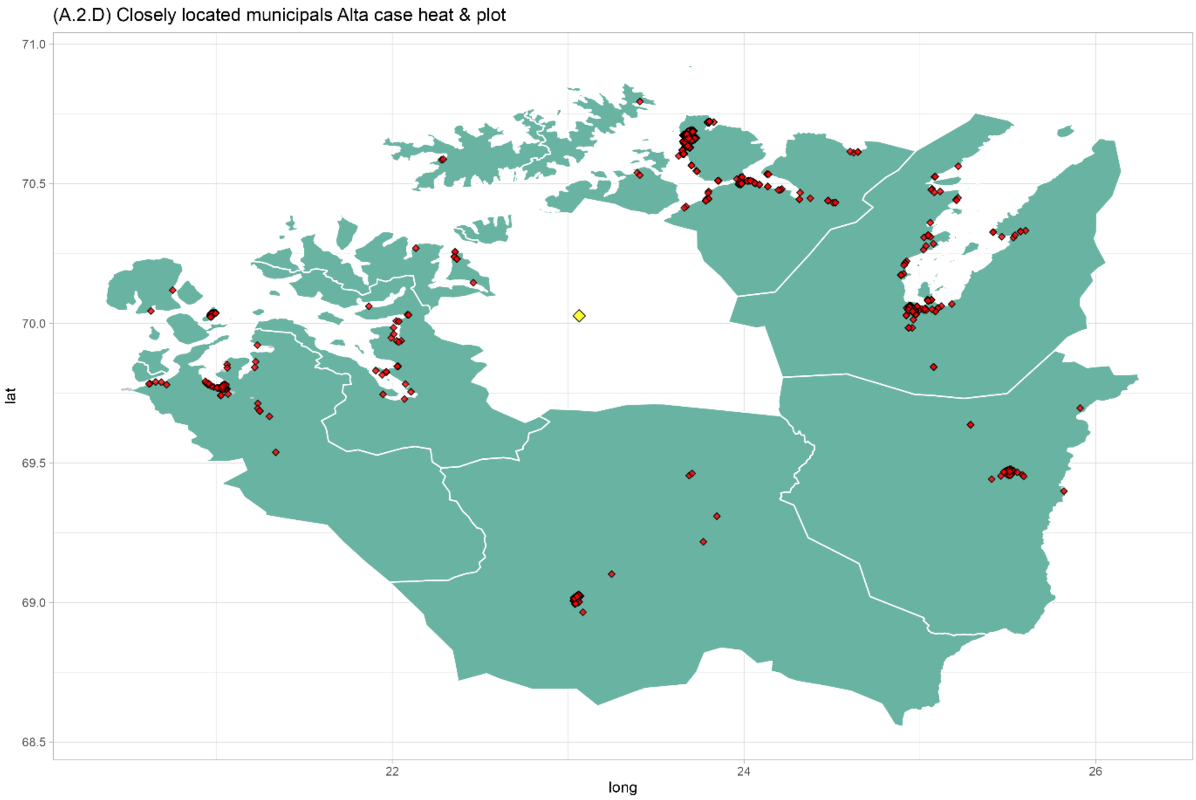
[30] Appendix A.2.B visualises all municipals in the Alta case in the colour “green”. The heatmaps are based on a density probability, hence the darker the area, the more probable it is that a random observation is located within the filled area. The risk areas for quick clay, with the same perquisitions as mentioned in footnote 29, are the areas with a light grey colour. The place of event (landslide) is plotted in yellow, while the plotting of sold dwellings, for the remainder of the maps, are represented in red, given that no other explanation is given for the map in question. For a map over all sold dwellings in the Alta case, see map 4.2

For the remainder of the maps in appendix A.2, the municipals are separated into the three groups: 1) Municipal of the event, including the municipal Alta, 2) closely located municipalities, including the municipals Porsanger, Hammerfest, Kvænangen, Kautokeino, Karasjok, Loppa, Nordreisa, Hasvik, Skjervøy, and 3) Remote located municipalities, including the municipals Vefsn, Hemnes, Rana, Fauske, Sørfold

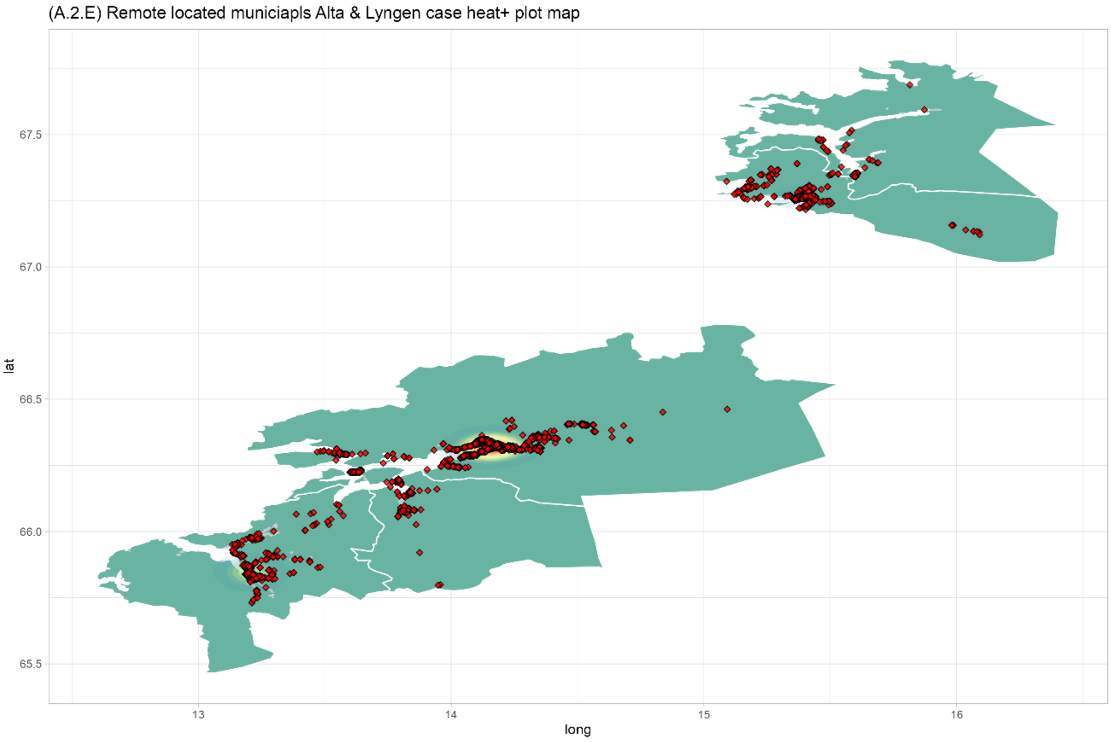
A.2.C: Heatmap and spread of sold dwellings for Alta Municipal



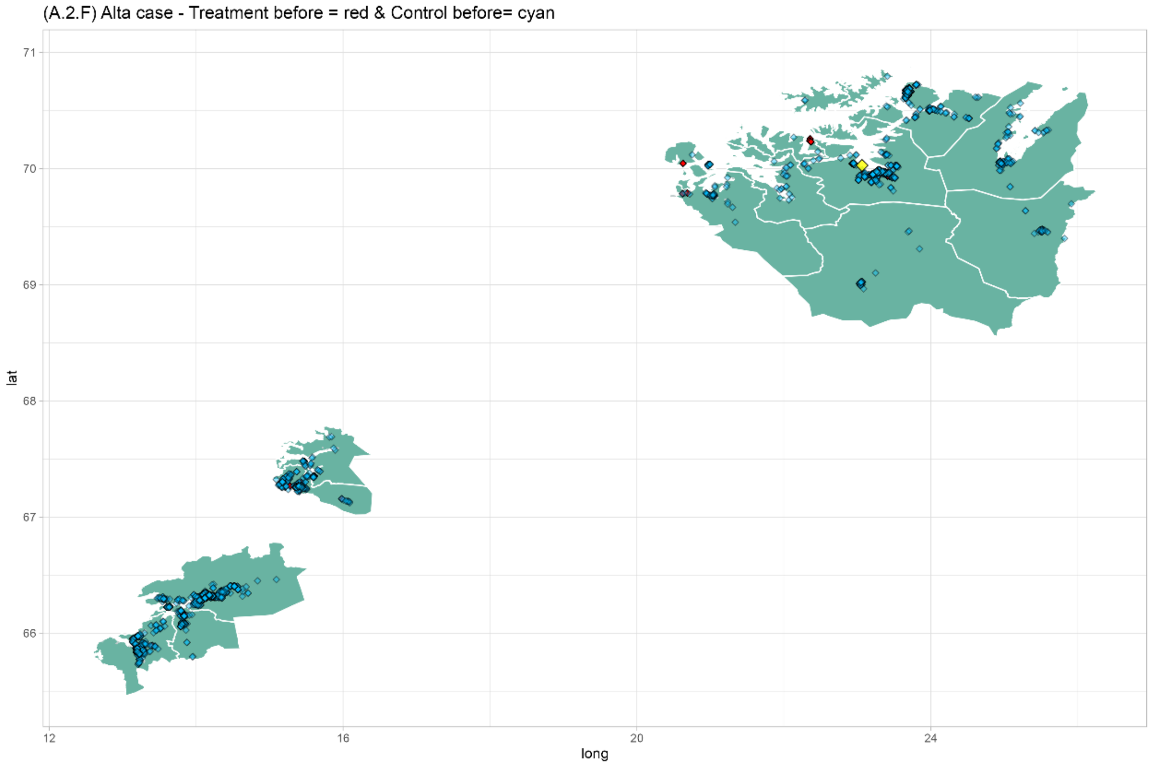
A.2.D: Heatmap and spread for sold dwellings for closely located municipalities for Alta case



A.2.E[31]: Heatmap and spread of sold dwellings for remote located municipals for Alta case

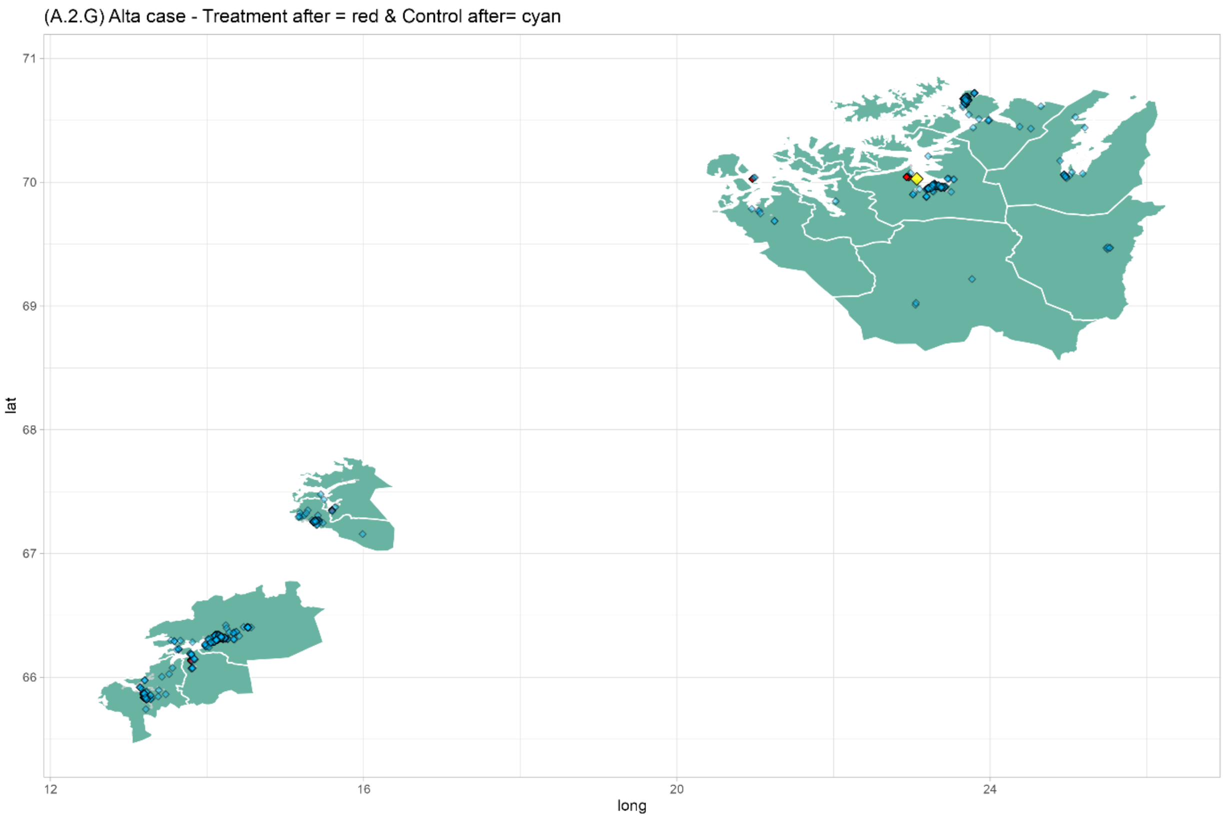


A.2.F[32]: Spread of sold dwellings for treatment and control groups before the event



[31] As the map would be distorted by marking the place of event on the maps for the remote located municipals the place of event is not plotted.
[32] This map follows the same logic as Map 5.2, where the red plots represent sold dwellings in treatment group for Alta, before the event, while the cyan plots are representing sold dwellings in control group for Alta, before the event.

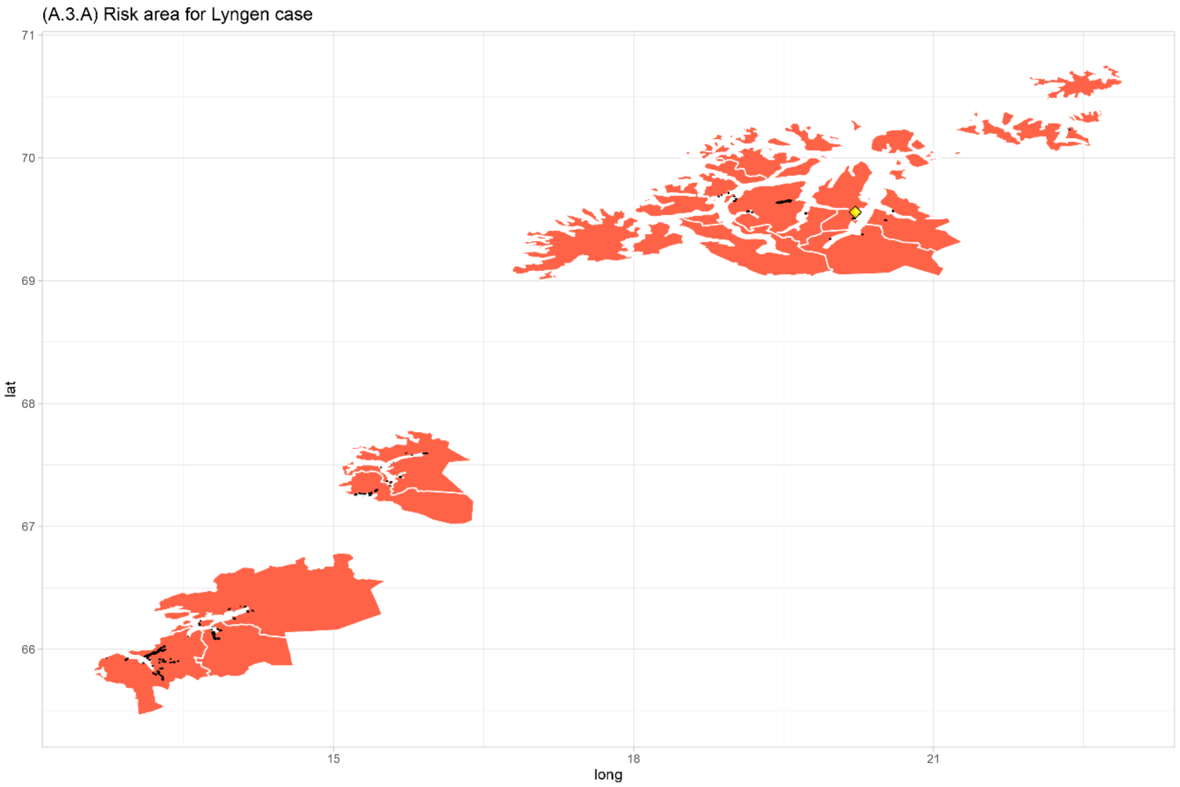
A.2.G[33]: Spread of sold dwellings for treatment and control groups after the event.



[33] This map follows the same logic as Map 5.2, where the red plots represent sold dwellings in treatment group for Alta, after the event, while the cyan plots are representing sold dwellings in control group for Alta, after the event.

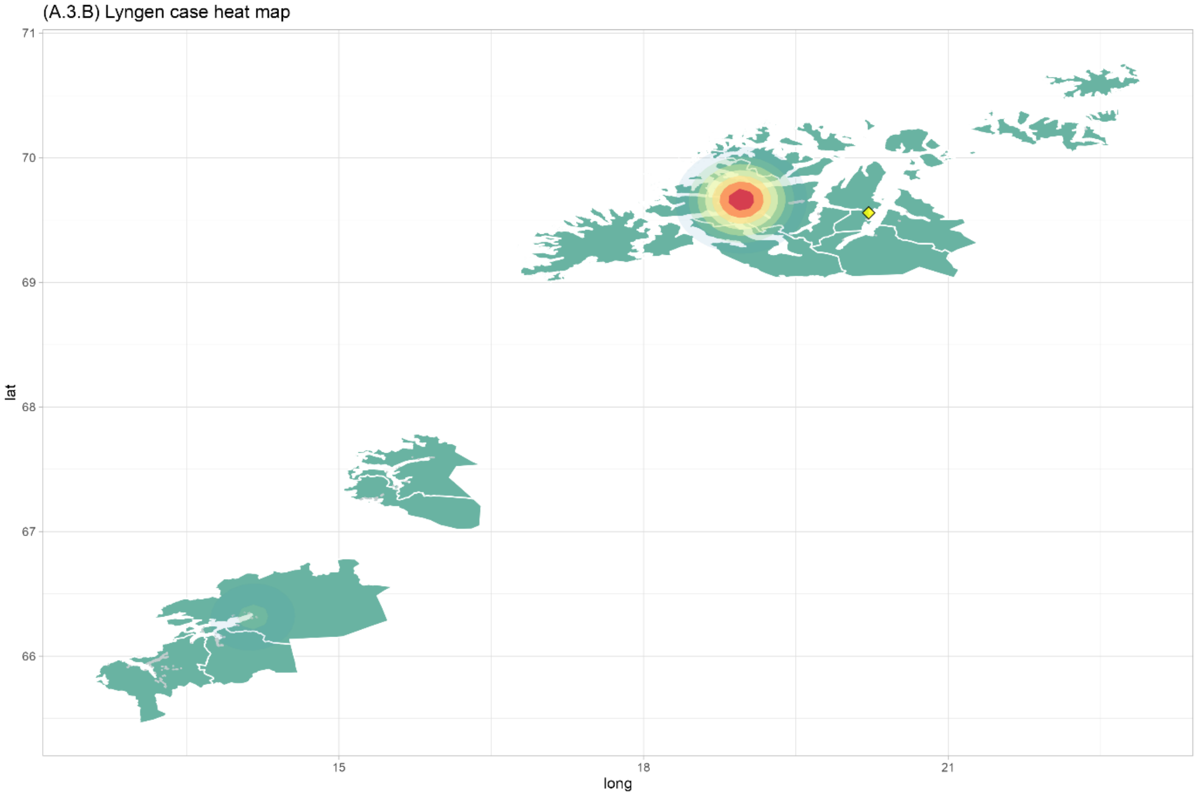
A.3. Lyngen case

A.3.A[34]: Risk area map for Lyngen case



[34] Map A.3.A visualises all the included municipals in the Lyngen case in the colour "tomato", while the black areas on the map represents the risk areas for quick clay landslide for all municipals in the Lyngen case. For all maps in appendix A.3, and the interactive map, the risk zones only include the areas at risk of the risk quick clay landslide, as defined by NVE, and not by the assumption for this thesis. Risk areas for other landslides are not included as the maps for these were separated for each scenario, hence it would be too time consuming to plot. The maps are created based on map data from GeoNorge and NVE (2022).

A.3.B[35]: Heatmap and spread of sold dwellings for Lyngen case

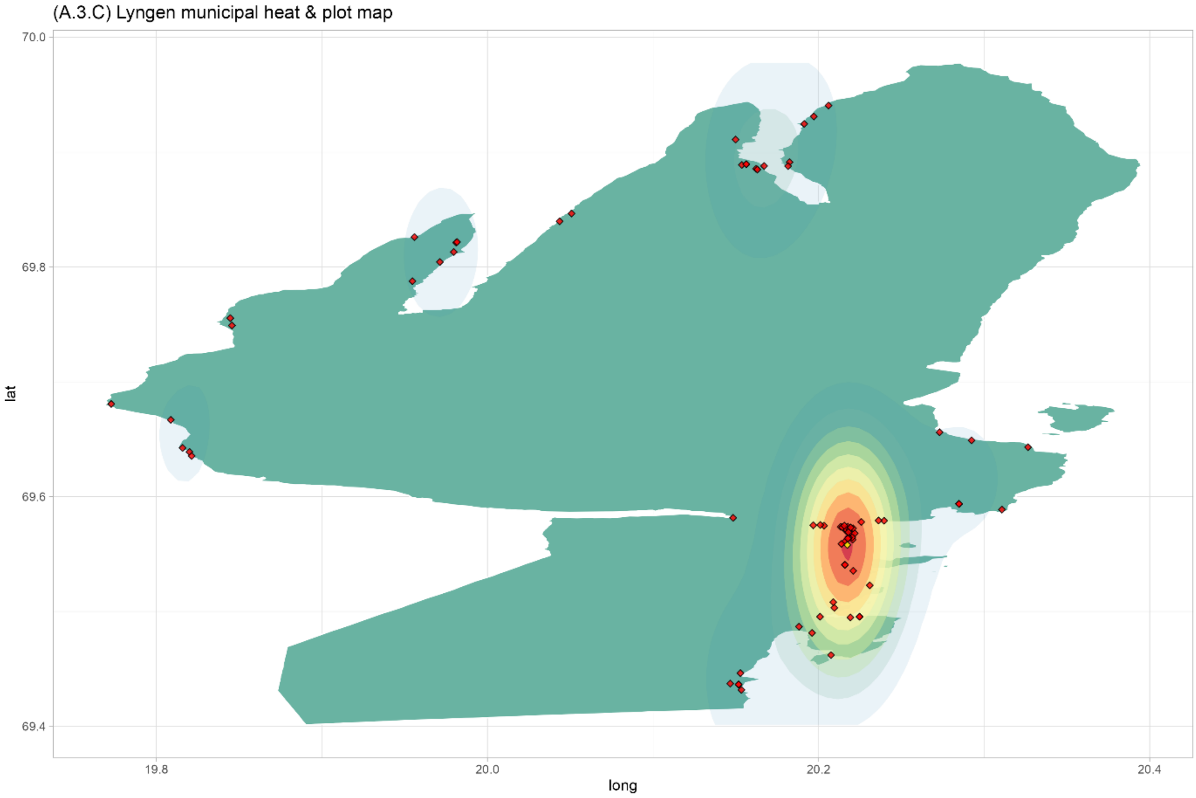


---

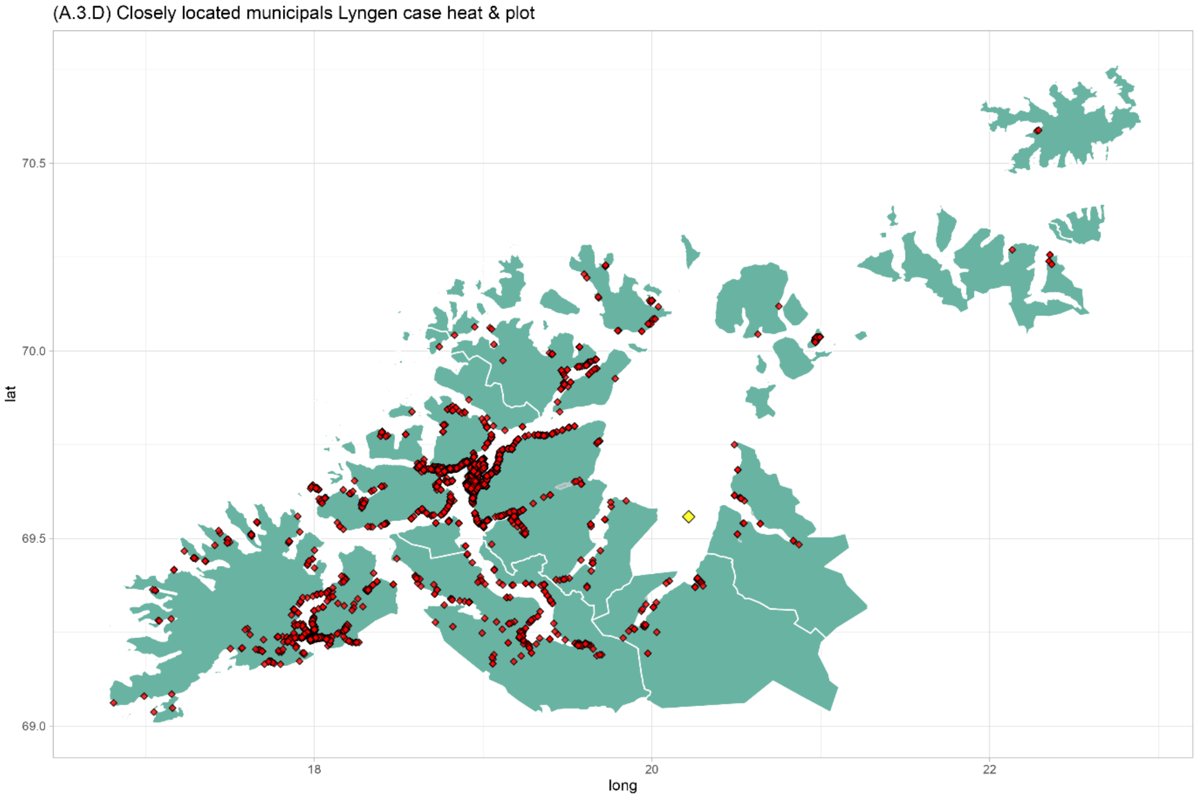
[35] The map shows the included municipals for the Lyngen case in the colour "green". The heatmaps are based on a density probability, hence the darker the area, the more probable it is that a random observation is located within the filled area. The risk areas for quick clay, with the same perquisitions as mentioned in footnote 34, are the areas with a light grey colour. The place of event (landslide) is plotted in yellow, while the plotting of sold dwellings, for the remainder of the maps, are represented in red, given no other explanations or notes say otherwise. For a map over all sold dwellings in the Lyngen case, see map 4.3

For the remainder of the maps in appendix A.3, the municipals are separated into the three groups: 1) **Municipal of the event,** including the municipal Lyngen, 2) **closely located municipalities**, including the municipals : Skjervøy, Kåfjord, Tromsø, Karlsøy, Storfjord, Loppa, Balsfjord, Hasvik, Senja, and 3) **Remote located municipalities**, including the municipals : Skjervøy, Kåfjord, Tromsø, Karlsøy, Storfjord, Loppa, Balsfjord, Hasvik, Senja

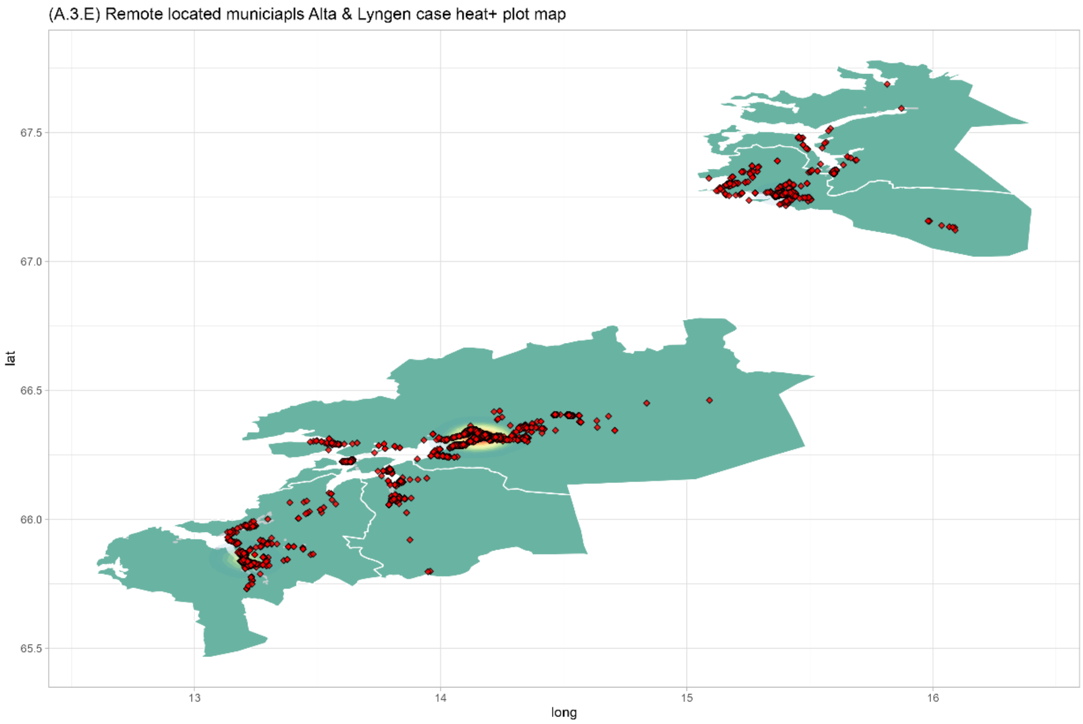
A.3.C: Heatmap and spread of sold dwellings for Lyngen Municipal



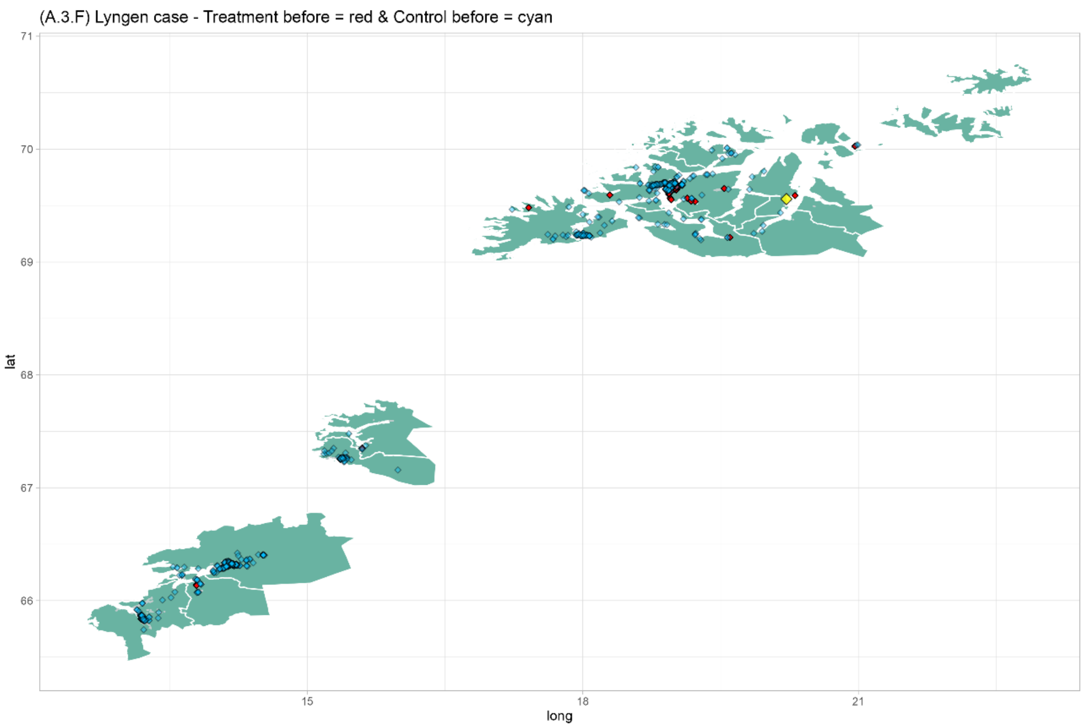
A.3.D: Heatmap and spread for sold dwellings for closely located municipalities for Alta case



A.3.E[36]: Heatmap and spread of sold dwellings for remote located municipals for Alta case

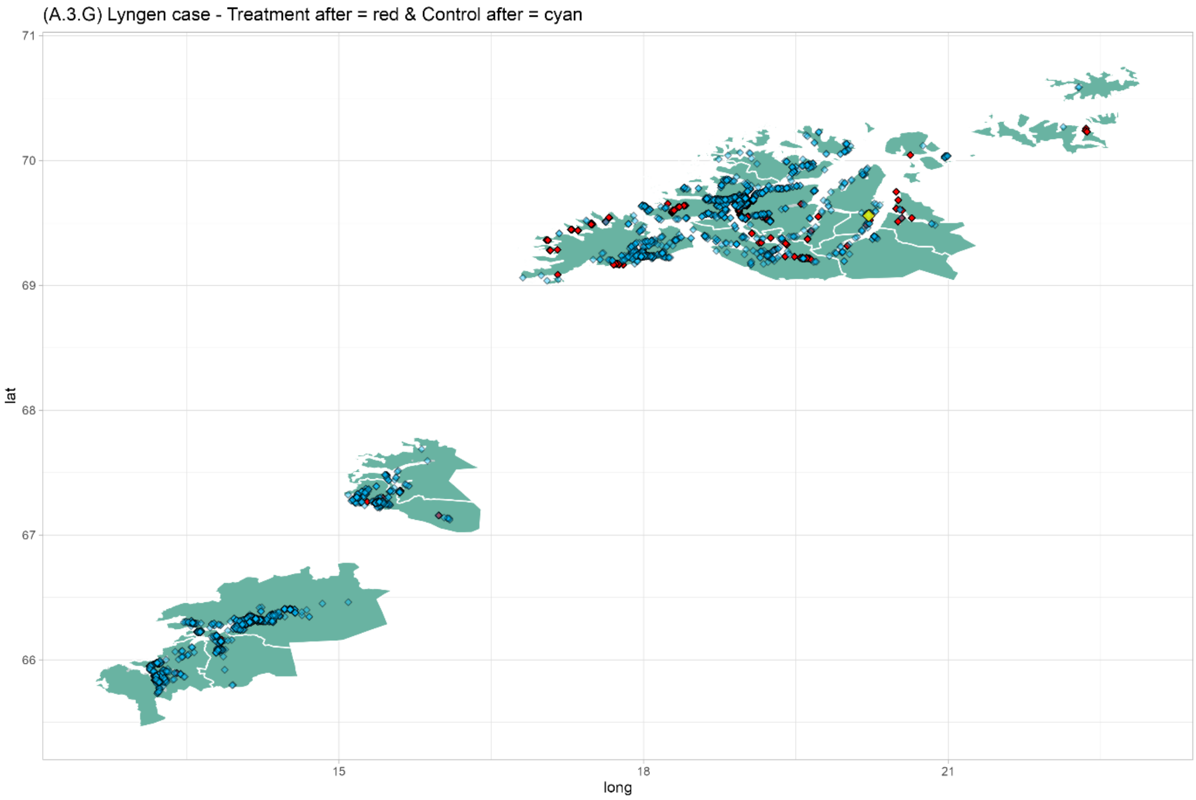


A.3.F[37]: Spread of sold dwellings for treatment and control groups before the event



[36] As the map would be distorted by marking the place of event on the maps for the remote located municipals the place of event is not plotted.
[37] This map follows the same logic as Map 5.3. The red plots represent sold dwellings in the treatment group for Lyngen before the event, and the cyan plots represent sold dwellings in the control group for Lyngen before the event.

A.3.G[38]: Spread of sold dwellings for treatment and control groups after the event



[38] This map follows the same logic as Map 5.3. The red plots represent sold dwellings in the treatment group for Lyngen after the event, and the cyan plots represent sold dwellings in the control group for Lyngen after the event.

B. General summary statistics

B.1. Gjerdrum

B.1.A: Table of observations for each municipal by after dummy for the Gjerdrum case

	After		
Municipal	0	1	Total
Arendal	3355	688	4043
Drammen	11631	2092	13723
Eidsvoll	2404	443	2847
Fredrikstad	7830	1472	9302
Gjerdrum	596	81	677
Halden	2817	509	3326
Hurdal	177	38	215
Kristiansand	11666	2284	13950
Lillestrøm	9085	1531	10616
Lunner	588	106	694
Lørenskog	4291	780	5071
Nannestad	1360	234	1594
Nes	2041	393	2434
Nittedal	2115	384	2499
Porsgrunn	4060	691	4751
Rælingen	2122	379	2501
Sandefjord	6819	1227	8046
Sarpsborg	5579	1096	6675
Skien	6099	1091	7190
Tønsberg	6702	1221	7923
Ullensaker	4424	824	5248
Total	95761	17564	113325

B.1.B: Table of observations for each municipal by risk dummy for the Gjerdrum case

	LR		
Municipal	0	1	Total
Arendal	4036	7	4043
Drammen	13069	654	13723
Eidsvoll	2844	3	2847
Fredrikstad	9008	294	9302
Gjerdrum	365	312	677
Halden	3326	0	3326
Hurdal	215	0	215
Kristiansand	13457	493	13950
Lillestrøm	10035	581	10616
Lunner	694	0	694
Lørenskog	5071	0	5071
Nannestad	1475	119	1594
Nes	2320	114	2434
Nittedal	2476	23	2499
Porsgrunn	4526	225	4751
Rælingen	2501	0	2501
Sandefjord	7554	492	8046
Sarpsborg	6427	248	6675
Skien	6605	585	7190
Tønsberg	7524	399	7923
Ullensaker	5059	189	5248
Total	108587	4738	113325

B.1.C: Table of observations per year by risk dummy for the Gjerdrum case

	LR			
Year	0	1	Total	Ratio treat/total
2015	14287	617	14904	0,041398282
2016	14308	627	14935	0,041981922
2017	14822	655	15477	0,042320863
2018	15463	656	16119	0,040697314
2019	15957	676	16633	0,040642097
2020	16941	768	17709	0,043367779
2021	16809	739	17548	0,042113061
Total	108587	4738	113325	

E.1.D: Table of observations for each type of dwelling for the Gjerdrum case

Housingtype	Freq.	Percent	Cum.
Detached house	38197	33,71	33,71
Apartment	54403	48,01	81,71
Townhouse	9479	8,36	90,09
Semi-detached house	11246	9,92	100
Total	113325	100	

B.2. Alta

B.2.A: Table of observations for each municipal by after dummy for the Alta case

	After		
Municipal	0	1	Total
Alta	2781	465	3246
Fauske	854	195	1049
Hammerfest	2124	289	2413
Hasvik	1	1	2
Hemnes	339	65	404
Karasjok	50	8	58
Kautokeino	49	13	62
Kvænangen	14	8	22
Loppa	5	3	8
Nordreisa	48	32	80
Porsanger	215	44	259
Rana	5089	670	5759
Skjervøy	13	15	28
Sørfold	74	29	103
Vefsn	2044	333	2377
Total	13700	2170	15870

B.2.B: Table of observations for each municipal by risk dummy for the Alta case

	LR		
Municipal	0	1	Total
Alta	3068	178	3246
Fauske	920	129	1049
Hammerfest	2329	84	2413
Hasvik	2	0	2
Hemnes	398	6	404
Karasjok	58	0	58
Kautokeino	62	0	62
Kvænangen	22	0	22
Loppa	2	6	8
Nordreisa	78	2	80
Porsanger	259	0	259
Rana	5355	404	5759
Skjervøy	24	4	28
Sørfold	65	38	103
Vefsn	2189	188	2377
Total	14831	1039	15870

B.2.C: Table of observations per year by risk dummy for the Alta case

	LR			
Year	0	1	Total	Ratio treat/total
2007	556	26	582	0,04467354
2008	685	39	724	0,053867403
2009	801	63	864	0,072916667
2010	922	59	981	0,060142712
2011	950	69	1019	0,067713445
2012	1000	58	1058	0,054820416
2013	887	66	953	0,069254984
2014	1028	61	1089	0,056014692
2015	1000	89	1089	0,081726354
2016	1162	84	1246	0,06741573
2017	1071	84	1155	0,072727273
2018	1178	89	1267	0,070244672
2019	1115	91	1206	0,075456053
2020	1253	84	1337	0,062827225
2021	1223	77	1300	0,059230769
Total	14831	1039	15870	

B.2.D: Table of observations for each type of dwelling for the Alta case

Housingtype	Freq.	Percent	Cum.
Detached house	7328	46,18	46,18
Apartment	6187	38,99	85,16
Townhouse	892	5,62	90,78
Semi-detached house	1463	9,22	100
Total	15870	100	

B.3. Lyngen

B.3.A: Table of observations for each municipal by after dummy for the Lyngen case

	After		
Municipal	0	1	Total
Balsfjord	31	259	290
Fauske	96	942	1038
Hasvik	0	2	2
Hemnes	45	341	386
Karlsøy	11	119	130
Kåfjord	1	19	20
Loppa	0	7	7
Lyngen	10	71	81
Rana	1082	4780	5862
Senja	157	928	1085
Skjervøy	4	23	27
Storfjord	4	37	41
Sørfold	5	92	97
Tromsø	2834	17388	20222
Vefsn	356	2073	2429
Total	4636	27081	31717

B.3.B: Table of observations for each municipal by risk dummy for the Lyngen case

	Risk		
Municipal	0	1	Total
Balsfjord	221	69	290
Fauske	910	128	1038
Hasvik	2	0	2
Hemnes	380	6	386
Karlsøy	130	0	130
Kåfjord	7	13	20
Loppa	1	6	7
Lyngen	75	6	81
Rana	5437	425	5862
Senja	1036	49	1085
Skjervøy	23	4	27
Storfjord	35	6	41
Sørfold	61	36	97
Tromsø	14619	5603	20222
Vefsn	2241	188	2429
Total	25178	6539	31717

B.3.C: Table of observations per year by risk dummy for the Lyngen case

	Treat			
Year	0	1	Total	Ratio Treatment/total
2007	841	177	1018	0,173870334
2008	881	176	1057	0,166508988
2009	1255	214	1469	0,145677332
2010	1405	286	1691	0,169130692
2011	1660	372	2032	0,183070866
2012	1773	363	2136	0,16994382
2013	1622	440	2062	0,213385063
2014	1827	438	2265	0,193377483
2015	1826	448	2274	0,197009675
2016	2000	572	2572	0,222395023
2017	1858	519	2377	0,218342448
2018	2016	601	2617	0,229652274
2019	1944	618	2562	0,241217799
2020	2140	680	2820	0,241134752
2021	2130	635	2765	0,22965642
Total	25178	6539	31717	0,20616704

B.3.D: Table of observations for each type of dwelling for the Lyngen case

Housingtype	Freq.	Percent	Cum.
Detached house	9847	46,18	31,05
Apartment	17073	38,99	84,88
Townhouse	1468	5,62	89,5
Semi-detached house	3329	9,22	100
Total	15870	100	

C. Description of visualization of data with maps

Even though the ambitions for the thesis were high, we did get coordinates as part of the data from EV. Even though the time constraint did not allow us to do a spatial DD, we still found it necessary and worthy to visualize the data through maps.
For this, we again used Rstudio[39]. The combination allowed us to create interactive and non-interactive maps with different features and use cases.

For the maps, we used two different websites. For maps regarding the risk areas concerning quick, we used NVE- kartdata. Getting the data directly from those who asses and is responsible for mapping the risk in Norway is undoubtedly a significant benefit as the data is reliable and up to date when downloading it. For mapping the different municipals that are part of each case, we downloaded a zip file including both municipals and counties. The file was downloaded from GeoNorge, and their catalogue consists of maps with different attributes. The map consists of municipals and counties cut according to the coastline and is based on municipals and counties from 2020. The files were downloaded in Geojson format using the following projection for all three maps: EUREF89 UTM sone 33, 2d.
For creating the maps, we first need to prepare the files so R can read the data within the file. First, we format the file into a utf8 format. These files are not initially formatted in utf8, a format that can read and write letters that are almost exclusively included in the Norwegian alphabet, such as Æ, Ø, and Å.

When the file is formatted, we can go on to the next step in the process. When dealing with what R defines as SpatialPolygonsDataFrame, we need to know what information the Geojson file withholds.
In step 4, we ask R to tell us what layers the dataset consists of. For example, the Geojson file with the data for quick clay risk zones in Norway has the layer "Skred_Kvikkleire", which we want to visualize. Moving on to step 5, we need to know what kind of geomType the data is formatted with. Looking back at the last step, we can find the layer type through the R-output. In all three files, we can see that we are dealing with wkbPolygon. Wkb stands for Well-known Binary, while polygons are the outlines that represent the visualization of the risk zones and

[39] For creating maps for the thesis, we used the packages ggplot2, ggmap, leaflet.extras, dplyr, maps, mapproj, RColorBrewer and utf8.

municipals. If the Layer type were any different, we would need to replace "wkbPolygon" with the given type. In step number 6, we produce a dataset based on the geomType where all the information we need is included. In this case, we delete the data that is not needed for this thesis. Next up, we must format the projection of the maps to make them able to read longitudes and latitudes for the region we want to visualize. We did this by using the commands proj4string and spTransform.

Next, we want to see what values the data consists of. For example, we can check unique values in the variables. In the mapping data for risk zones, we checked what municipals had a registered risk area. We found that the Municipal "Kåfjord", was named "Gáivuotna" in the data over risk zones. We then had to rename this value to match the data registered in the other datasets.

Next up, we had to make datasets that included a selection of municipals that match the selection we did with the transaction data from EV. This was reproduced for each case, for each composition of municipals within each case, that being the municipality where the landslide occurred, closely located or remote municipalities. Lastly, we did the same coding for the risk zones.
Now that we have the datasets, we need to create map functions in R, we create a new dataset by hand. The data frame used for the popup information in the interactive map was created using the "tibble" function. The data was then manually plotted and consisted of information such as where the landslide occurred, in what municipal, what year, and at what date the landslide found place. To place it on the map, we had to manually find the longitude and latitude of the centroid of the landslide. This took a decent amount of time as I had to manually check whether the coordinates found matched where the landslides found place and make sure it was approximately in the centre of the landslide. For the interactive map, we also wanted a summary of each case to pop up for the viewer when interacting with the markers for each case. This was produced by mutating a new variable named popup info into the data frame we created earlier.

Creating the interactive maps is, from this point, a straightforward process. Using leaflet, a web-based service, we can now create an interactive map with the features and information we want it to have. The interactive maps will be available through a link in map 4.4.

For maps included in the thesis and the appendix, we used the ggplot2 package to visualize the data. First, we must format the data so that the package can read the information. We do this by using the tidy function. We create new datasets with the same name as the earlier datasets in a different format. This makes it easier for us to know what datasets to get specific information from when working with leaflet and ggplot2.

For creating maps for the thesis, we used the packages ggplot2, ggmap, leaflet.extras, dplyr, maps, mapproj, RColorBrewer and utf8.