NORWEGIAN SCHOOL OF ECONOMICS

Bergen, Spring, 2018

Has Norway's Oil Riches Hampered Education?

Oil activity and its relative effects on the educational attainment of affiliated municipalities

Jonathan Curran & Helene Sunde

Supervisor: Associate Professor Torfinn Harding

Master Thesis, Master of Science in Economics and Business Administration, Major in Business and Economics

NORWEGIAN SCHOOL OF ECONOMICS

Abstract

This thesis aims to uncover if communities affected by oil activity have suffered a reduction in their share of higher educational attainment, compared to other municipalities. Using data on oil production and investments for the oil field's main supply bases, together with educational data on a municipal level, we examine this relationship during the period 1980 – 2016. This is estimated through a fixed effects model, where we look at the direct, per-unit effect oil activity has on the share of educational attainment. Secondly, we employ a difference-in-difference (DID) estimation using a roll-out method, where we look at the permanent effects surrounding the first introduction of oil activity. To examine the difference-in-difference assumptions, an event study specification is utilized, at the same time allowing us to observe the yearly effects surrounding the introduction of oil activity.

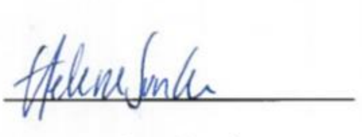
Our findings indicate that the relationship between oil production and the share of higher education for an oil-related municipality is negative between the years of 1998 and 2016 – the most recent half of our inspected time period. Our results suggest that a 10% increase in oil production corresponds to a long-term reduction in higher educational attainment of up to 0.038 percentage points. Most of the power behind these results stem from the effects oil production have on the educational attainment of women. Our estimates show that women experience effects between 159% – 486% larger than their male counterparts. We also find that of the two higher educational levels, undergraduate attainment is the most sensitive to a change in oil production, being subject to reductions twice as large as the attainment of postgraduate degrees. These results have been proven robust to changes in how we define the treatment and control groups.

Acknowledgements

We would like to thank our supervisor, Torfinn Harding, for helpful guidance and encouraging advice during our research and writing. We would also like to thank our family and friends, for thoughtful discussions and support.

Bergen, June 2018

Jonathan N. Curran



Helene Sunde

Table of Contents

LIST OF FIGURES

LIST OF TABLES

1 Introduction

Countries having an abundance of natural resources, typically oil, are often linked to having lower levels of education, in addition to being inversely related to economic development (Gylfason, 2001). Part of this effect stems from these countries often investing less in education. Norway, on the other hand, was in 2009 reported as having the third largest expenditure on tertiary education per student in the world (Global Education Network, p. 28), in addition to being ranked among the top in tertiary educational attainment (OECD, 2011, p.30). We can therefore hardly state that Norway has suffered huge educational losses due to its large petroleum sector.

In spite of this, there might still exist relative educational losses within Norway, due to the effects of the petroleum sector. Theories relating to choices of education usually present them as weighing the marginal benefits of different educations against the alternative cost of not getting a higher education and going straight into work. The alternative cost is then the potential income that could be had in the field not requiring higher education. Norwegian oil workers are consistently rated among the highest paid workers, both compared to oil jobs in other countries (Hays, 2013) and to other sectors within Norway (Statistics Norway, 2018). Although these statistics may also include employees with higher education, there are other statistics portraying similar trends to the operator area on the oil rig (Solberg, 2013). A large part of this remuneration is due to overtime and compensation pay for working offshore.

The high wages for the non-educated jobs offshore could lead to the marginal benefit of education being deemed comparatively lower. This could again lead to a lower educational attainment for the areas especially exposed to work on oil rigs and oil activity in general, compared to areas that are not. Our research question thus becomes: *"Has oil activity negatively affected the educational attainment of related municipalities, compared to municipalities that aren't affected?"*.

A country's level of educational attainment has been shown to have a number of positive effects, among them a positive relationship with GDP-growth and innovation (Olaniyan & Okemakinde, 2008, p.479). Seeing if the Norwegian oil activity has caused a negative change in educational attainment will therefore be of interest, due to the implications it would have on the municipalities closely related to oil activity, in addition to the implications for the country as a whole.

1.1 Outline

This master's thesis consists of eight chapters. Chapter 2 introduces the effects of education, in addition to a brief history of the Norwegian petroleum industry. Next, we present the data used for the empirical analysis in Chapter 3, followed by the empirical strategy in Chapter 4. We then introduce our main findings in Chapter 5, before we examine the robustness of these results in Chapter 6. Our ability to generalize these results and their practical implications are discussed in Chapter 7. Lastly, our concluding remarks are given in Chapter 8.

2 Background

Before we look at the analytical model, it is relevant to briefly give an introduction to the basis of our hypothesis, the importance of education and the Norwegian oil history.

2.1 Education and its Importance

Education is an economic good because it is scarce in relation to its demand, meaning that human effort is required to obtain it. In economics, education is regarded as both a capital and consumer good because it serves as an input into the production of goods and services, and offers utility to the consumer (Olaniyan & Okemakinde, 2008, p.479). As a capital good, education can be used to develop the human resource, which is necessary for economic and social transformation. Having an education of high quality that benefits society is a prerequisite for these benefits. Focusing on education as a capital good relates to the human capital concept which emphasizes the development of skills as one of the main factors in production activities (Becker, 1964).

The importance of education and human capital has been brought out in many studies of economic growth and development. One of them, by Robert J. Barro (1991), have developed a human capital model that shows how education and the creation of human capital was responsible for both the differences in labour productivity and the differences in overall levels of technology that is observed in the world. Similar results were concluded by Benhabib and Spiegel (1994), who find that human capital is positively related to the total factor productivity of a nation.

Today it is widely accepted that education creates knowledgeable citizens and contributes positively to the general standard of living in a society. The increasing belief of education as a factor for positive change in many developing countries has led to a huge educational investment in these countries (Olaniyan & Okemakinde, 2008, p. 479). In addition, the pressure for higher education has been helped by public perceptions of financial reward from pursuing such education. In Norway, students are supported by the Norwegian State Educational Loan Fund, providing them with a stipend for covering living expenses while studying (Ekberg, 2018).

2.1.1 Valuing Education

There are several different explanations in economic theory trying to explain why people choose their chosen education (Raaum, 1999). Card (1999) argues that individuals choose the level of education that maximizes their life’s income. This implies that school attendance ends when the cost of acquiring additional education equals the marginal utility. Since there are variations in the marginal costs and marginal utility between individuals, this can explain why people end up with different levels of education.

Residence and family background are examples of environmental factors that can influence the individual’s marginal costs of education. Place of residence affects the size of travel and moving expenses in order to reach the educational institution. The availability of jobs offered in the area will also affect the cost of either choosing or forgoing education – if there are many jobs requiring higher education, the cost of not getting a diploma will be larger, compared to a place with mostly low-educated jobs. Given that the petroleum sector has a wide variety of jobs not requiring higher education, this could reduce the demand for higher education in areas effected by oil activity. This would make not getting an education less costly and more appealing, using Card’s model.

The opportunity for parents to give economic support to their children and their education vary, depending on the size of the family, the parents level of education and income, to mention a few. Less support from the family leads to an increased cost of education, which can be part of the reason behind the differing marginal costs of education between individuals. Public support for education, on the other hand, can reduce the individual's marginal cost and lead to higher education. Since individuals will have different perceptions of wage levels in a given industry, their expected marginal utility or the return on education will also vary between individuals. Many will claim that higher education provides other benefits of a non-economic nature as well (e.g. knowledge, social status, self-actualization). Appreciation of these non-economic advantages will differ between individuals as well.

2.2 The Norwegian Petroleum Sector

2.2.1 Norwegian Oil History

Few believed that the Norwegian continental shelf (NCS) concealed rich oil and gas deposits in the late 1950s. This changed when a discovery of gas was made at Groningen in the Netherlands in 1959, causing people to revise their opinions on the petroleum potential of the

North Sea. Today, more than 50 years later, the petroleum industry includes workers from more than 98 percent of Norway's municipalities, and is Norway's most important resource, both in terms of revenues to the treasury and investment (Ekeland, 2017).

After Philips Petroleum sent an application to the Norwegian authorities in October 1962, to be allowed exploration in the North Sea, Einar Gerhardsen's government was forced to take a stand regarding the ownership of these areas. This led to the government proclaiming Norwegian sovereignty over the NCS in May 1963. The new regulations determined that the State owns any natural resources on the NCS, and that only the government is authorized to award licenses for exploration and production. The same year, private companies got the opportunity to carry out preparatory exploration. The licenses included rights to perform seismic surveys, but drilling was put on hold until agreements on how to divide the continental shelf between Norway, Denmark and the United Kingdom were reached (Ministry of Petroleum and Energy, 2016).

In March 1965 agreements on dividing the continental shelf in accordance with the median line principle were reached, and the first licensing round was announced. The same year, 22 production licenses for a total of 78 blocks were awarded to different oil companies (Ministry of Petroleum and Energy, 2016). The production licenses gave exclusive rights for exploring, drilling, and production in the license area. In the summer of 1966, the first well was drilled. This well turned out to be dry.

Years without discovery followed, but that was about to change in 1969. December 23rd, 1969, Phillips informed the Norwegian authorities of the discoveries of Ekofisk, which turned out to be one of the largest offshore oil fields ever discovered. Production from the field started on June 15th, 1971. After 1969, a series of major discoveries were made the next few years (Norwegian Petroleum, 2018).

Foreign companies dominated the exploration in the 1970s and were responsible for developing the country's first oil and gas fields (Ministry of Petroleum and Energy, 2016). This caused the government to increase the regulations for oil exploration, in order to develop new knowledge and industries on the basis of petroleum. As a state-owned company, Statoil was funded in 1972, in order to look after the government's commercial interests and pursuing appropriate collaboration with domestic and foreign oil interests.

The continental shelf has been opened gradually, and only a limited number of blocks have been announced in each licensing round. Since the discovery of Ekofisk, production from the Norwegian continental shelf has been dominated by other large fields such as Statfjord, Oseberg, Gullfaks and Troll. These fields are still very important for the development of Norway's petroleum industry (Norwegian Petroleum, 2018).

The oil shock in Norway represents a more permanent income shock due to the steady state of new discoveries, which has lasted for almost 50 years. The initial shocks were still mainly concentrated in the areas surrounding the Ekofisk field and the following early discoveries. This led to the effects of oil activity being more geographically concentrated in the 1970s and 1980s, compared to today.

2.2.2 Validating the Hypothesis

In our initial hypothesis, we propose that the educational attainment might be lower in municipalities affected by oil activity. This is based on the assumption that the educational requirements on average are lower for oil-related jobs. It is also based on the assumption that the uneducated jobs on oil rigs can have a higher net marginal utility, due to high wages and low transportation costs, compared to jobs that require education. It is therefore relevant to examine the validity of these assumptions closer.

2.2.2.1 Education

In Ekeland's (2017) survey of the petroleum industry, he finds a decreasing share of people working in the petroleum sector with primary and secondary school as their highest educational level (see Table 1). This decreasing share of lower education can be explained by an increasing share of the younger generation having higher education, and more jobs that require higher education. Looking at education by gender, women's share of tertiary educational attainment is larger than the men's share, in addition to having experienced a larger increase.

Even though the petroleum sector's level of higher education is rising, it is still considerably lower than the average educational level for Norway's working sectors overall. The average higher educational attainment of the workforce is reported at 40.9%, almost twice that of the petroleum sector (Statistics Norway, 2017a). This helps validate our assumptions regarding the petroleum sector having a relatively low educational level.

Table 1. Educational level for workers in the petroleum sector residing in Norway, by gender

Year	Primary school	Upper secondary school	Tertiary education
		Men	
2003	69.4%	17.6%	13.0%
2015	62.5%	18.4%	19.0%
2016	62.3%	18.2%	19.5%
2016-2013	-7.1 pp	0.6 pp	6.5 pp
		Women	
2003	57.7%	24.6%	17.7%
2015	38.5%	29.5%	32.1%
2016	38.1%	28.6%	33.3%
2016-2003	-19.5 pp	4.0 pp	15.6 pp
		Total (both genders)	
2003	67.4%	18.8%	13.8%
2015	57.5%	20.7%	21.8%
2016	57.3%	20.4%	22.4%
2016-2003	-10.2 pp	1.6 pp	8.6 pp

Note. Men and women in the petroleum sector and their educational level for the time period 2013 to 2016. Numbers in percent and percentage point (pp). Reproduced from Anders Ekeland (2017).

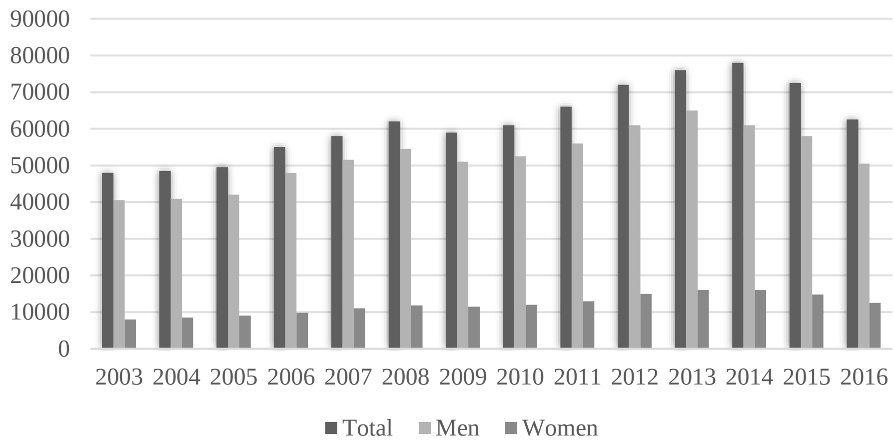
2.2.2.2 Wages

The average annual salary for the oil industry was NOK 765 600 in 2013, according to Statistics Norway (2018), which is the highest average salary in the country. However, this number does not include overtime, non-agreed bonuses and compensation for hazardous working conditions. The global recruitment company Hays Oil & Gas presented an average yearly salary of NOK 903,640 for employees in the Norwegian oil industry in its wage statistics for 2013 (Hays, 2013). Of a total of 54 countries in their survey, Norway was ranked at the top of the payroll for wage earners in the oil and gas industry. This shows how the alternative cost of education can be relatively high if one has the option of earning such considerable amounts in the oil sector nearby. These high wages also help substantiate our initial assumptions of a high marginal utility in the petroleum sector.

2.2.3 Employment in the Petroleum Industry by Gender

Looking at employment data in Figure 1, there is an obvious skew in terms of gender. Men made up almost 80% of the petroleum sector's workforce in 2016. The share of women, however, has increased in recent times, having gone from 16.4% in 2003 to 20.7% in 2016.

Figure 1. Employment in the petroleum sector by gender, for workers living in Norway



Note. Men and women in the petroleum sector and their educational level for the time period 2003 to 2016. Reproduced from Anders Ekeland (2017).

2.3 Prior Research

This thesis is not the first to observe oil activity and educational attainment in unison. Løken (2010) used the early oil booms in Norway in the 1970s as an instrument for an exogenous, isolated income shock, to estimate the causal relationship between educational attainment and parental income. In this natural experiment, Stavanger was used as the treatment, due to its central part in the first oil discoveries, and Sør-Trøndelag was used as a control, due its large distance from Stavanger and similar county traits. She finds that, although there exists a strong correlation between parental income and the children's educational attainment, the relationship is due to selection in to education, and is not causal. As our thesis looks at the effects of oil activity on the educational attainment of the current population, and not the lagged income effects on the next generation, our results will help enrich the research examining the consequences of the Norwegian oil shock.

The US, being a large, longtime producer of petroleum, has been subject to multiple studies examining its effects, among them education. Ratledge and Zachary (2017) look at how primary and secondary education has been affected in the states producing large amounts of oil and natural gas. They find diverging trends between the oil producing states, while their qualitative methods report minimal effects on increased high school dropout rates. The effects of oil on high school educational enrollment has also been estimated by Black, McKinnish and Sanders (2015), who find that a 10% increase in oil production in the 1980s produced negative, long-term reductions in enrollment of up to 5-7%. Looking at test scores and completion rates in

Texas[1], Marchand and Weber (2015) find negative effects on the former, in addition to a decline in the supply and quality of teachers. The effects of shale on higher educational attainment has also been studied, where a negative relationship on both high school and college attainment is established (Rickman, Wang & Winters, 2016). The US has therefore demonstrably been prone to research relating to petroleum and educational attainment. However, to the best of our knowledge, Norway has yet to receive such an inquisition. This thesis will therefore help add to the international literature exploring educational effects of oil production, by looking at a previously unexplored country.

[1] Texas is one of the leading US states in the production of petroleum products.

3 Presentation of the Dataset

In this section, we present the data used in our analysis. Our dataset can be divided into two broad parts: oil rig and oil-related activities, and education. The data on oil is sourced from Norwegian Petroleum, which has been composed and constructed by Lene Bonesmo Solberg, former research assistant at the Norwegian School of Economics. The data on education is retrieved directly from Statistics Norway's public report: "Educational attainment of the population" (Statistics Norway, 2017b).

3.1 Oil

Our dataset on oil activities encompasses nearly all petroleum activity that has transpired under Norwegian jurisdiction. The first observation in our data is therefore also the first discovery made on the Norwegian continental shelf in 1969 by the Ekofisk oil rig.

3.1.1 Fields, Main Supply Bases and Base-Municipalities

Norwegian Petroleum provides us with information on 109 oil fields, whereas 74 are defined as connected to a main supply base located on the Norwegian mainland. In addition, there are 11 fields that aren't connected to a main supply base by Norwegian Petroleum but have been assigned a main supply base by Solberg. She does this by using the supply bases of connected fields as a proxy for the fields missing supply bases. The supply bases and their connected fields are summarized in Table 2 below. The remaining 24 oil fields are either directly or indirectly connected to bases outside of Norway and are not further included in our analysis.

Table 2. The main supply bases for oil fields in Norway

Bases connected to fields	Stated fields	Proxies	Total fields	Municipalities
Dusavik	13	2	15	Stavanger
Florø	8	0	8	Flora
Hammerfest	2	0	2	Hammerfest
Kristiansund	10	2	12	Kristiansund
Mongstad	12	3	15	Austrheim and Lindås
Sandnesjøen	7	0	7	Alstahaug
Sotra	2	0	2	Fjell
Sotra/Florø	5	0	5	Fjell/Flora
Tananger/Dusavik	1	0	1	Sola
Tananger	13	4	17	Stavanger/Sola
Ågotnes/Mongstad	1	0	1	Fjell
	74	11	85	

Note. Table showing how many oil fields the different supply bases are connected to, and what municipality they are located in. The stated oil fields are the fields clearly defined by Norwegian Petroleum belonging to certain bases. The proxy oil fields are defined by Solberg, based on the relations of connected oil fields.

As stated in our hypothesis, we want to examine the effects on municipalities associated with the production of oil. This requires us to assign each unit of oil activity to its rightful municipality, which calls for some modifications. We start by isolating the oil activity of each field to a supply base. In some cases, multiple supply bases are connected to the same oil field. The activity for these oil fields are split between their connected supply bases, so each unit of activity is only assigned to a single base, to avoid overlapping.

Having each field assigned to a supply base, we now connect these to their correct municipalities. The supply base in Mongstad is technically located in two municipalities, Austrheim and Lindås. Because it only has a very small part residing in Austrheim, and also has its postal address in Lindås, we have chosen to classify this as belonging to Lindås. Lastly, we have a case of two bases residing in the same municipality; the Sotra base and Ågotnes base both lie in Fjell. Through a dialog with Statoil, Solberg concludes that Ågotnes and Sotra operate as the same place, and therefore combines the two, classifying them as a unified Sotra.

This leaves us with each base having an assigned oil field, its proportion of oil activity and an assigned municipality, called the *base-municipality*. The base-municipality will serve as the basis for deciding the location of initial impact made by the supply bases. The result of the process described above is summarized in Table 3 below. We have also included the year of the first production and investment for each supply base. It should be mentioned that for most of our analysis we widen the scope of impact beyond looking at single municipalities. This means that exactly which municipality is treated as a base-municipality is somewhat arbitrary and should have little effect on our results.

Table 3. Base municipalities, production and investment start

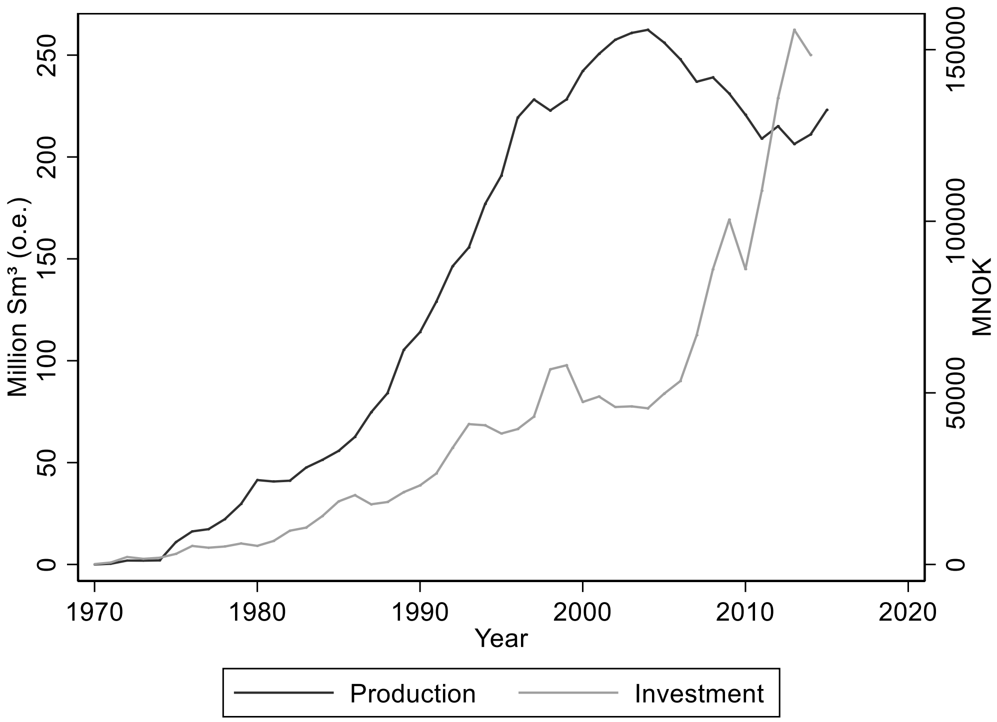
Main supply base	Base-municipality	Production start	Investment start	Years from investment to production
Dusavik	Stavanger	1990	1986	4
Florø	Flora	1971	1970	1
Hammerfest	Hammerfest	1979	1974	5
Kristiansund	Kristiansund	1986	1981	5
Mongstad	Lindås	1979	1974	5
Sandnessjøen	Alstahaug	1993	1987	6
Sotra	Fjell	1997	1994	3
Tananger	Sola	2007	2002	5
Mean				4.25

Note. The main supply bases and their assigned municipality. The year of first production and investment is also included, in addition to the years in between and the total mean time.

3.1.2 Oil Activity

Our oil dataset has three main variables relating to the oil fields: investment, production and reserves. Investments is given in million nominal NOK and includes investments made both before and during the production phase of the oil rig. There are four different types of production: oil, gas, natural gas liquids (NGL) and condensate. In the production variable, all four types have been converted into the most commonly used metric for different petroleum products, standard cubic meters of oil equivalents (abbreviated as Sm^3 o.e.) (Norwegian Petroleum, n.d.). The different amounts of each petroleum product are then merged together, to form a single production variable. We therefore make no distinction between what type of petroleum product a field produces.

Figure 2. Norway's total production of oil and investment



Note. The total amount of oil production and investment in Norway. The left axis shows Sm^3 oil equivalents (o.e.) given in millions. The right axis shows investments in MNOK.

Oil production has been steadily increasing since its initiation in 1969, as can be seen in Figure 2, hitting a peak in 2004. Investments have been expanding, but at a slower pace, before a sharp increase starting in 2005. The amount of registered oil reserves is also rising at a steady pace, as more and more have been discovered (see Figure A1 in the appendix). At its peak, the oil reserves are more than 10 times higher than the total oil production.

The data on oil reserves are not as complete as the data on production and investments, with several years prior to 2000 being missing. The effect of reserves on education is also less intuitive; the fact that large amounts of petroleum may lie in the ground does not necessarily engage a large working crew. For these reasons we have chosen not to focus on reserves as a variable and will for most of the analysis look at production and/or investments.

3.2 Education

Statistics Norway provides us with yearly observations on education for municipalities in Norway between 1980 and 2016, with an additional year of data for 1970. Each observation is divided into a male and female category, in addition to a combined male and female category. The dataset operates with five different levels of education: primary school, upper secondary education, tertiary vocational education, tertiary education short and tertiary education long. There is also a category for unknown or no completed education which makes up a very small part of the observations. This category is removed from the analysis, when calculating the different shares of education.[2]

The dataset does not contain the level of education that a person is currently attending. Instead, it contains the highest level of education that a person has *completed*. The exact mechanics of this will be explained separately for each level of education. The data only involves people over the age of 16, with an exception for 15-year-old students that have completed primary school, who are also a part of the statistic.

3.2.1 Primary School

Primary school is the first instance of official schooling in Norway and traditionally lasts from years 1 through 10. This usually involves an elementary school from years 1 through 7 and an intermediate school the following three years. Primary school is mandatory for children ages 6 to 16. A student enrolled in primary school will not be in our observations, due to this person not having a completed degree. A student that has completed primary school will be registered the same year.

[2] When presenting shares of educational attainment, Statistics Norway uses the educational shares, less the category for unknown registries. We follow their example, for fluency in comparing our data to other sources, as most research uses Statistics Norway's data as a basis for analysis.

3.2.2 Upper Secondary Education

Upper secondary education lasts three years, from years 11 to 13. The chance to complete upper secondary education is not mandatory but is a legal right every Norwegian possesses (Thune, Reisegg & Askheim, 2015). A student enrolled in upper secondary education will be registered under the primary school category until graduation, when they will be registered as having completed upper secondary school.

3.2.3 Tertiary Vocational Education

Tertiary vocational education consists of shorter studies of up to two years. They do not require to have achieved the general university admission certification required to attend the traditional tertiary institutions. These studies are directly related to a profession and are not accredited as higher education. Vocational studies make up a very small part of the educational population, with only 2.84% having this registered as their highest attained education in 2016. This is in fact the only year for which we have data on this field, due to Statistics Norway registering it as upper secondary education up until 2015. For simplicity, we choose to classify vocational studies as belonging to upper secondary education for 2016 as well, removing this level of education from our dataset and further analysis.

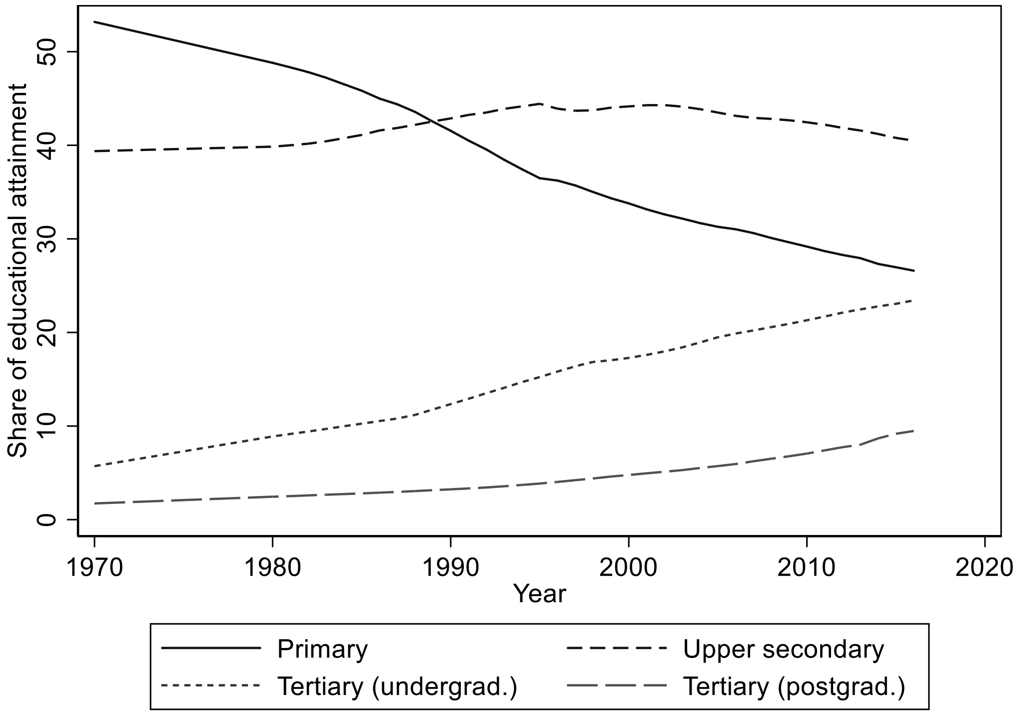
3.2.4 Tertiary Education Short (Undergraduate) and Long (Postgraduate)

Tertiary education involves all forms of higher education. It is divided into two groups categorized as short and long, or undergraduate and postgraduate. Short comprises higher education up to 4 years, while long is anything above 4 years. A tertiary education of short length therefore involves both one-year studies, two-year studies, bachelor's degrees and four-year studies. It will also encompass any student who has completed 120 credit points (usually 2 years) or more, even though their current degree is uncompleted. This means that students will be registered under the undergraduate level after having completed 2 years, regardless of enrollment in an undergraduate or postgraduate degree. Once a student has completed a five-year study, they will appear in the tertiary postgraduate variable. All students undertaking and/or completing doctoral degrees will be registered at the postgraduate level.

3.2.5 Measuring Education

Just looking at the raw values can give us a skewed impression of how the level of education has evolved. If there has been a large increase in the number of educated individuals in a municipality, these numbers will in isolation seem like the population has become more educated. However, if the number of individuals with lower education has increased even more, we would have a relatively less educated population, in terms of percent. The raw values can therefore be misleading, due to increases in the population leading to deceptive interpretations of increased relative educational levels. To account for this, we look at the *proportions* or *shares* of each educational level, having the denominator be the total population of interest. This is also the same way Statistics Norway calculates share of education.[3]

Figure 3. The development of attainment for different levels of education



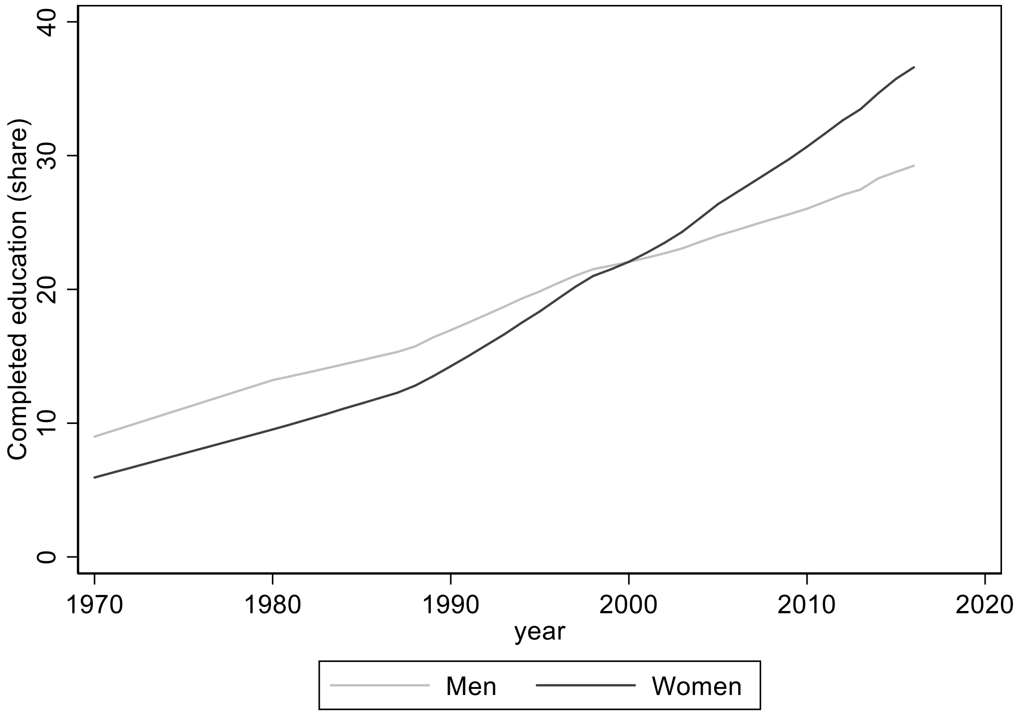
Note. The shares of educational attainment for all levels.

Aggregating our data up to a national level in Figure 3, we can observe a significant change in the educational trend. Primary school being a person's highest educational achievement has gone from covering more than half of the population in 1980 to only being applied to 26% in 2016. This contrasts with the attainment of higher education, which has steadily been increasing over the entire period. In 2016 the total share of higher education, both short and long, has reached a record high 32.6%, encompassing nearly a third of the population. There has also been a change in the gender composition of the population of higher education. As seen in Figure 4, women have gone from being the gender with the smallest share of higher education,

[3] An alternative approach would be to calculate shares of educational attainment with the entire population in the denominator. This would exclusively reduce the shares of educational attainment, as it would add individuals under the age of 16, individuals that aren't registered as having any educational attainment (as per how Statistics Norway collects their data). We find that this adds nothing to the analysis, in addition to serving as a poor comparison to other sources and research, assuming that most research uses Statistics Norway's data as a basis.

to becoming the most educated gender. Since 1990, women have consistently been increasing their proportion at a higher rate than men, until the end of our registered time period.

Figure 4. Attainment of higher education, by gender



Note. The share of men and women attaining higher education, consisting of both under- and postgraduate attainment.

3.3 Data Cleaning

When using municipal data spanning several years, a typical problem is the merging and splitting of municipalities, due to different political initiatives over time. Having over 40 years of data, this was especially true for us. We have used the municipalities that were in effect in 2016, the most recent year of our data, and traced these back so that all observations reflect the 2016-municipality. How this affects the number of municipalities we study and the number of observations in our dataset is summarized in Table 4 below.

Table 4. Cleaning the data

Stages of the data cleaning	Observations	Municipalities	
Original dataset	231 699	507	100.0 %
Splitting and merging municipalities to match 2016-status	204 706	449	88.6 %
Removing undeclared municipalities	195 586	429	84.6 %

Note. Each column shows the remainder of observations after each stage. The last column shows how much of the original dataset remains, given in percent.

For municipalities that were split in 2016, but have previously been merged together, we attempt to estimate the share for each municipality prior to the split, based on their shares the first year after the split. We then remove the merged municipality that no longer exists from our dataset. For municipalities that were merged in 2016, but were previously split apart, we simply add the individual municipalities' observations to the current merged municipality. We then remove the individual municipalities that no longer existed in 2016.

4 Empirical Approach

4.1 Main Model

When estimating a general effect for many different units of observation, a central part in making a causal link is that the observed effects are caused by factors estimated in our model, and not factors hidden in the error term. Due to our panel data structure, we follow the same units of observation over time and therefore can't rely on the advantages a randomized sample provides (Angrist & Pischke, 2008, p. 12). There might be inherent factors or abilities in some municipalities that will get credited to our model unless we are able to account for them. The fixed-effects estimator manages to remove such *time-invarying* factors, by assuming they are time-constant and not part of the error term, through time-demeaned differencing the equation (Woolridge, 2015, p. 467). Even with a fixed effects estimation, there is still a concern that the error terms within our panels are correlated (Cameron & Miller, 2015). To be able to make our standard errors relevant for inference, we cluster them at the municipal level.

There might also be time-specific factors on a country-wide level interfering with our ability to correctly estimate our model. To handle this, we include dummy-variables for each year of data. Due to education having experienced a large, consistent, country-wide upward trend for the last 40 years, the level of educational attainment is highly correlated with time. Adding these time-dummies will therefore lead to very high R-squared estimates, the metric used to gauge how much of the variation in Y is explained by X. It is, however, important that we remove these country-wide effects, to be able to distinguish what part of the variation is due to the effects of oil activity. This leaves us with the following general model:

$$\mathrm{Y}_{m,t} = \alpha + \beta_1 \mathrm{X}_{m,t} + \gamma_t + \lambda_m + \mathrm{u}_{m,t} \qquad (1)$$

The model's intercept is captured in *α,* and our dependent variable *Y* will for most of our analysis be the share of educational attainment. The municipal fixed-effects λ are removed through the within transformation. Our explanatory variables are captured in *X* and will generally take two shapes. First, we want to estimate the direct, per-unit effect oil activity has on education. This assumes that the relationship can be modeled linearly. The intuition being that as oil production increases, the demand for labor on oil rigs increases, also increasing demand for supporting sectors. This will be done with *X* as oil activity, where oil activity will be either production or investment. The control group, by definition not being affected by oil activity, will always have values equal to zero for the explanatory variable.

Secondly, we want to estimate the more general effect the presence of oil activity has on education. We do this by deploying a difference-in-difference (DID) model with a roll-out approach, similar to the one used by Bütikofer, Løken and Salvanes (2016, p. 12).

The roll-out method allows us to account for the introductions of oil activity happening in different years for the different municipalities. We measure the introduction of oil activity in two ways; 1) whenever production is positive and 2) whenever investments are positive. The first year of introduction is therefore also the first year that production or investment starts, where investments normally comes first. We estimate this by letting *D* be an indicator-variable equal to one when oil activity is positive and zero when it is not. The indicator variable will therefore always be equal to zero for the control group. Looking at oil activity aggregated to a supply-base level, activity seldomly reaches zero once it has been initiated. In practice, this means that once the indicator variable has switched to one, it stays this way for the remainder of our time period:

$$\mathrm{Y}_{m,t} = \alpha + \beta_1 \mathrm{D}_{m,t} + \gamma_t + \lambda_m + \mathrm{u}_{m,t} \qquad (2)$$

The identifying assumption for a difference-in-difference model is that pre-treatment trends are parallel (Angrist & Pischke, 2008, p. 163). To test this assumption, we use an event study specification, where we visually inspect the pre- and post-effects of the introduction of oil activity, similar to Jacobsen, LaLonde and Sullivan (1993, p. 693). If pre-treatment differences are close to zero, this implies little to no difference prior to the introduction of oil and supports the identifying assumption. We do this by creating an indicator variable *E* for each of the *q* years of the event, starting at year *T* prior to the introduction of oil activity:

$$\mathrm{Y}_{m,t} = \alpha + \sum_T^q E_{m,t}^T \beta_1 + \gamma_t + \lambda_m + \mathrm{u}_{m,t} \qquad (3)$$

E is equal to one for the treated municipalities during the *q* years of the event. The omitted or normalized year, that serves as the reference for the event-year dummy variables, is *T=-1*, one year prior to the introduction of oil activity. The effect of this year will be equal to zero in the event study. By plotting the regression coefficients of each dummy variable and their respective confidence intervals along *T*, we can observe the differences in the pre-treatment trends and their statistical variance. We can also observe post-treatment patterns, if for example the effects are increasing, decreasing or constant.

In our event study we use both the treatment and control group. The control group, never being treated, always returns event-year dummies equal to zero, and is therefore also placed in the normalized year *T=-1*. This means that the point estimates not only compare within municipalities not yet having experienced oil activity, but also compare to municipalities never experiencing oil activity (i.e. the control group). This gives us a mix of effects from the contrasts between the treated and the never treated, as well as an effect from the timing of the tested activity. An alternative could be to not include the control group and only observe the changes within our treatment, observing the effects of the timing of oil activity. As the focus of this thesis is the relative effects between the treated and control, we choose to include the control group in our event study.

4.2 Treatment and Control

A central part in being able to properly test our hypothesis is defining what municipalities are being affected by oil activity and what municipalities that are not; defining our treatment and control groups.

4.2.1 Treatment

The municipalities we define as being affected by oil activity will make up our treatment group. The petroleum industry is not a small sector, and one could argue that the effect of oil activity has in some way influenced the entire country. Defining the boundaries of such an activity is therefore far from an exact science. If we define the treatment group too broadly, we risk diluting the effects we are interested in studying. If our scope is too narrow, we risk not capturing the extent of the relevant effects, for not to mention statistical issues of having a small sample size.

Defining a group as treated involves assigning the relevant oil activity to the treated municipalities. We have chosen to equally assign oil activity within the groups connected to a supply base. An alternative approach could be to differentiate the assigned amount based on some parameter, for example distance from the supply base.

With the assumption that distance from the supply base influences how much a municipality is affected by its related oil activity, we try to construct a treatment group. The ideal approach would be to create a circle of impact around the epicenter of oil activity, defined as the supply base, and study the individuals within. To be able to do this we would need data on the exact location and educational attainment on an individual level, something we don’t possess. As a simplification, we confine our treatment group to municipalities that border to the base-municipality, including the base-municipality itself.

Table 5. List of municipalities within treatment, supply bases and their affiliated base-municipality

Main supply base	Base-municipality	No. of municipalities
Dusavik	Stavanger	6
Tananger	Sola	3
Sotra	Fjell	7
Mongstad	Lindås	8
Florø	Flora	5
Kristiansund	Kristiansund	7
Sandnessjøen	Alstahaug	7
Hammerfest	Hammerfest	5
		48

Note. Overview of the treatment group, showing supply bases, their related base-municipality and number of municipalities in each group.

We have summarized the number of municipalities included for each supply base in Table 5 (see Table A1 for a complete list of municipalities). Having a relatively concise area of treatment will help isolate the effects of each individual supply base, while we at the same time avoid having to aggregate effects for any overlapping supply bases. Even though none of the oil activity for the supply bases are added together, this treatment group still has a drawback; Sola and Stavanger share a border. Due to them having large gaps between their production start, we decide to have them in separate groups. This could lead to a spillover effect, if the oil activity in each group affects the other.

Comparing our treatment group to the rest of the country's municipalities, we find some similarities and some differences. First and foremost, the educational attainment between them for the earliest years of our data, 1970 and 1980, which is mostly prior to any large oil procurements, is not statistically different from each other (see Figure A2 in the appendix). That our dependent variable is relatively homogenous prior to the introduction of our explanatory variable is a good starting point for the analysis.

Looking at some demographical factors, there are some clear differences that should be addressed. Based on data gathered from Kartverket (2016), we find that the treatment group's municipalities are smaller than the remaining municipalities, having on average roughly half the area of land in square kilometers (see Figure A3 and Table A2). In addition, the treatment group has more than twice the populational density. How important is this to our analysis? Area of land is a time-constant factor, and the effects of this will be differenced out through the fixed effects estimation. Population density will technically differ over time, as long as there is at least one more or less individual registered in the area, something that is highly likely. That being said, as long as the change in population density remains constant across municipalities, this should not affect our estimates, as they will represent the time-constant factor of area of land. Visually inspecting this in Figure A4, we can see that most of the municipalities in both groups share a similar, upward trend, although some of the municipalities in the treated group experience a declining trend. If population density affects the higher educational attainment of municipalities, this could lead to biased estimates in our main model.

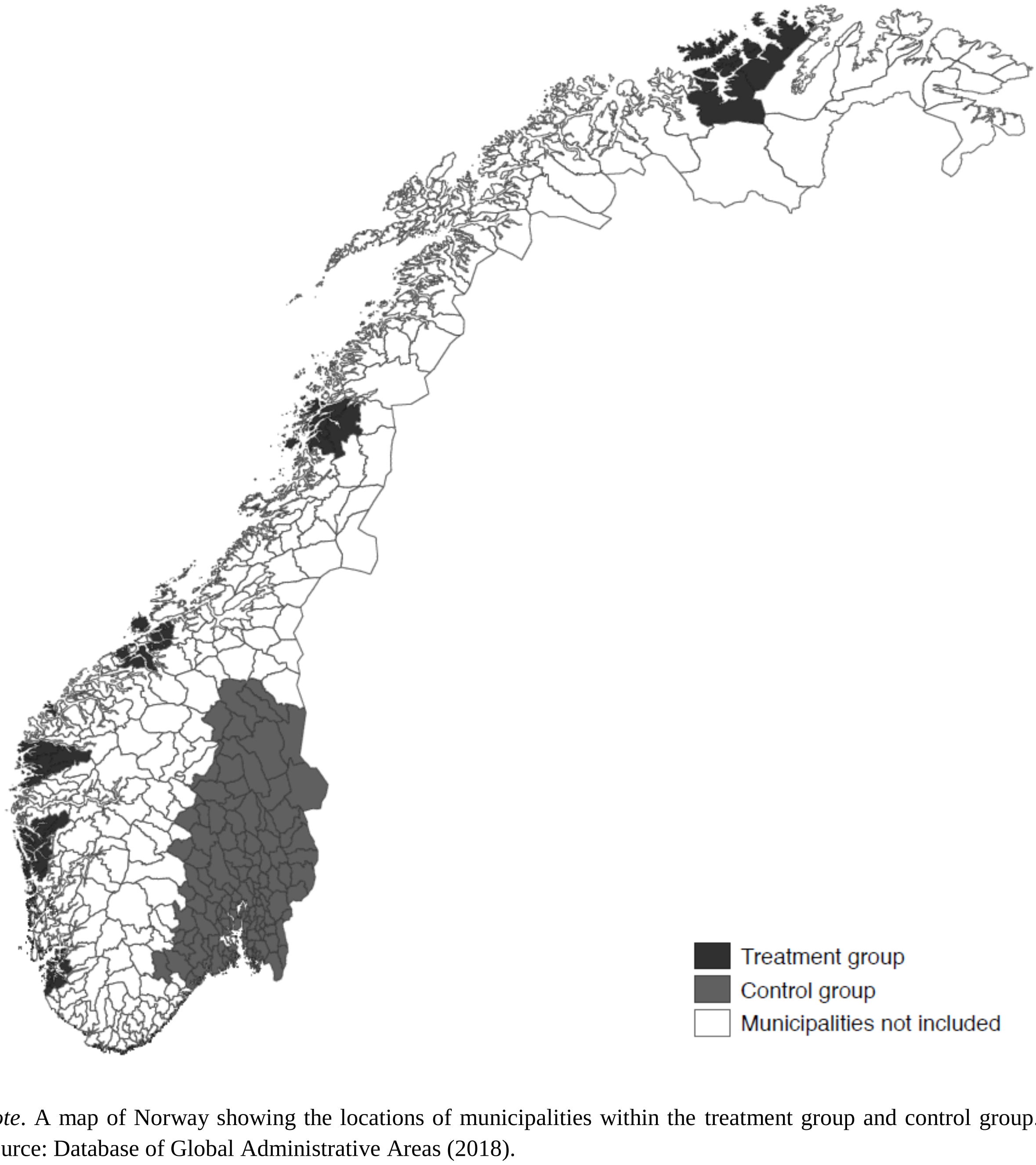
4.2.2 Control

To be able to state that any effects we observe are caused by our explanatory variables, and not a result of other factors, it is essential that we establish a valid comparison group (Angrist, 2003; Rosenbaum, 2005). The ideal control group will be equal to our treatment group in every regard, except the presence of oil activity. As we have stated earlier, most municipalities will likely be affected in some way by oil activity, so finding a group of municipalities with no connection will be difficult.

Following the logic from the treatment group containing bordering municipalities, we choose municipalities that are geographically distant from the supply bases. All the supply bases are located along the Norwegian western coast line, where most are in the southern and western parts of the country, while Hammerfest is far north. We therefore choose municipalities that are located to the east of Norway as a control group (see Table A3 for an exhaustive list of control municipalities). Both the treatment group and control group's geographical locations can be found in Figure 5. Using population density as a means of comparison, we can see in Figure A5 that both treatment and control display fairly similar trends. However, when using a two-sampled t-test on the log of population density for all years of our data, the treatment and control group prove to be statistically different from each other. The average population density is lower in the treatment group than in the control group, which can be visually inspected in Figure A6 in the Appendix.

Using kernel density plots to estimate how balanced the treatment and control group are in Figure A7, we look at some initial factors in 1970, before most of the oil production had begun in Norway. The treatment group has 16.7% lower mean area of land, 29.3% lower average number of inhabitants and slightly lower population density, none of which are statistically different from each other. The educational attainment, however, is a lot lower in the treatment group, being 1.62 pp lower, corresponding to being 28.4% less, compared to the control group's share of educational attainment of 5.7%. This difference in initial educational attainment is statistically different from zero. Having a balanced educational attainment prior to our treatment would be preferable, but the essential factor for inference is that the trends are parallel.

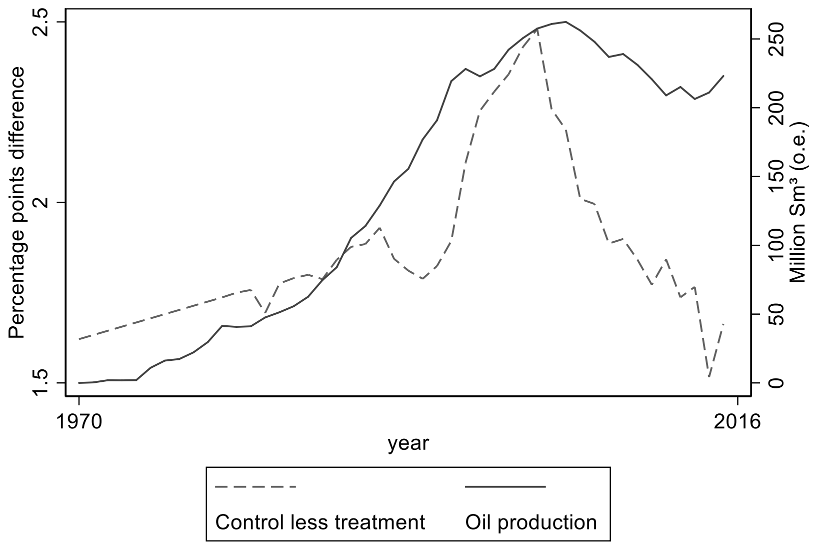
Figure 5. Map of treatment and control municipalities



Note. A map of Norway showing the locations of municipalities within the treatment group and control group. Source: Database of Global Administrative Areas (2018).

In Table A4, we can observe the changes in the average educational attainment for the municipalities within the treatment, in addition to a group mean for the entire treatment and control. The control group always has a higher average educational attainment, with an increased difference from 1.8% in the first half of the dataset, to 2.0% in the second half. Plotting the differences between the two over the entire period we can see in Figure 6 that the differences steadily increase to a height of 2.48 pp in 2002, after which they start decreasing closer to the differences of the 1970s. This rise and fall of the gap in educational attainment between the control and the treatment, closely resembles that of the total oil production's rise and fall, which is also added to the figure. This gap is calculated by taking the difference between the educational attainment of the control less the treatment. A positive number therefore indicates that the control group has a higher educational attainment, compared to the treatment group. When this number falls, it means that the treatment group is increasing its educational attainment, relative to the control group.

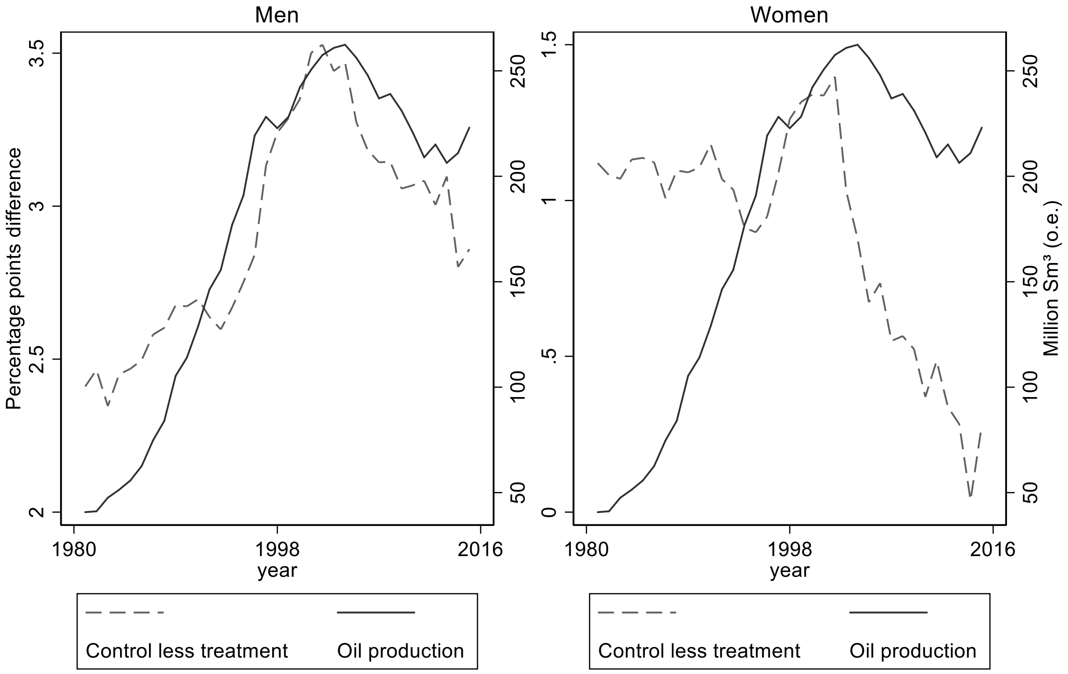
Figure 6. Gap in educational attainment between treatment and control, together with oil production



Note. The dashed line plots the difference in higher educational attainment for the control and treatment group over time. The difference is calculated by taking the control less the treatment, a positive number therefore indicating the control group having a higher educational attainment. The solid line shows the total oil production, which closely resembles that of the educational gap between the control and treatment group.

Looking at the same plot in Figure 7, now divided by gender, we can see that men and women are affected somewhat differently. The gap for men follows the shape of oil production closely, having its peak gap the year before the peak of oil production. The gap between the treatment and control for women is relatively stable the first 10 years of the period, and almost seems unaffected. Then around the mid-90s the gap starts mirroring the oil production's relative rise, followed by a steep fall when the oil production starts to decline.

Figure 7. Gap in educational attainment between treatment and control, together with oil production, by gender



Note. The dashed line plots the difference in higher educational attainment for the control and treatment group over time, separated by gender. The difference is calculated by taking the control less the treatment, a positive number therefore indicating the control group having a higher educational attainment. The solid line shows the total oil production. The left axis shows the difference in the share of educational attainment between control and treatment, the right axis shows the oil production.

Although the relative change in the educational gap between the control and treatment group for women is larger than men, the gap for men is larger in absolute terms (note that the axes for men and women have different ranges in Figure 7). This is made clear in Figure A8, where we can see that men in general have quite a large gap, compared to women. Looking closer at the differences in Table A5, we can see that while men in the treatment experience an increased difference compared to the control between the first and second half of the period, women experience a decrease. The difference between periods is increased for men by 0.57 pp, while the difference is reduced for women with 0.36 pp, having less than one percent difference compared to the control group.

Although these differences can give indications to how oil activity has affected the treatment group, they are not an accurate representation. As the different supply bases within the treatment group experience oil activity at different times, averages spanning the entire group will attenuate any local variances. The educational development of each group is shown in Figure A9.

4.2.3 Variables

4.2.3.1 Dependent Variable

Our dependent variable is higher educational attainment, and as mentioned in chapter 3, we will look at the *share* of education. This means that we combine the data we have on short and long tertiary education, to be able to observe the total change in higher education. Educational attainment is a slow-moving variable, especially at the higher levels. Due to our variable observing attainment and not enrollment, once a higher level is attained, it does not change unless that person achieves an even higher degree. This means that once a person is registered in the postgraduate category, the highest level of attainment in our data, the observation will stay constant for the remainder of the individual's lifetime. On an individual level, this makes tertiary education less inclined to become reduced relative to primary or secondary education, due to reduction only arising from the passing of an individual. On a municipal level, the share of postgraduate attainment, or any other level of attainment, can also be reduced by the individual moving away from the municipality and changing his or hers registered address. In addition, while enrollment in a line of study usually can be done within a year, attaining a degree fundamentally must take longer. If a person has signed up for a postgraduate degree, it can at most take up to four years before they are even registered at the undergraduate level, taking an additional year for them to end up in their enrolled level.

The factors that show this variable is slow-moving might also be indicative of educational attainment depending on the share of the previous year. A model with a self-dependent, dependent variable usually calls for an autoregressive term; a lagged dependent variable. Estimating an autoregressive model, when we are already using fixed effects introduces the possibility of a dynamic panel bias (Nickell, 1981), through correlation between the error term of the lagged dependent variable and the individual fixed effects of our panels. Since we are already including time-dummies to capture the variance related to yearly trends, much of the dependence of the educational attainment variable should be captured within these. We therefore choose to not include an autoregressive term, to reduce the chance of an asymptotic bias.

As Marchand and Weber (2015) find, the allure of the high wages in the petroleum sector is not limited to students, but also affected the teachers. They find that the number of available teachers declined, in addition to the remaining teachers being relatively inexperienced. They attribute the Texas students' decreased degree of achievement to this decline of teachers. As we do not have data on employment, we cannot observe what part of our reported results are due to a change in the supply (i.e. teachers) or the demand (i.e. students)[4]. We therefore make no such distinction, although this would admittedly be an interesting point of analysis.

To keep the education attainment variable's coefficients from becoming too small, we multiply them by 100, so that it theoretically can range from 0 to 100.

4.2.3.1 Explanatory Variables

As we lead with in our main model, oil activity is our explanatory variable, where it can be utilized as either oil production or investment. To better fit the model, we are using the natural logarithms of both activity measures (Wooldridge, 2015, p. 183).

Due to education being a variable that changes slowly, we believe that including lags of our explanatory variable is necessary, to capture the full effects of oil activity. Our preliminary analysis indicates that two lags is preferable. This is also somewhat intuitive, as a person enrolling in a bachelor's or master's degree will be recorded under the higher education variable after two years. This leaves us with the following equation to be estimated for the main analysis:

$$\mathrm{Y}_{m,t} = \alpha + \beta_1 \mathrm{X}_{m,t} + \beta_2 \mathrm{X}_{m,t+1} + \beta_3 \mathrm{X}_{m,t+2} + \gamma_t + \lambda_m + \mathrm{u}_{m,t} \quad (4)$$

For our model to infer causality, we must be able to state that our explanatory variable is exogenous (i.e. uncorrelated with the error term). If the educational attainment of a municipality decided the amount of oil production that would take place, we would have an issue of reverse causality. Although this seems unlikely, we might still be susceptible to other biases through factors included in our error term. If there are other factors influencing where the supply bases are placed, for example cheap labor, which in turn is caused by low educational attainment, we would again be faced with reverse causality. Fortunately for our model, the supply bases need to be placed in close proximity to the oil fields they aim to supply. We would therefore believe that the supply bases are endogenous in relation to oil fields, but exogenously placed in relation to educational attainment, and in turn, our model. The placement might still coincide with certain factors, due to most of the municipalities being placed in similar areas – the coast. If there were certain factors within the coastal municipalities (e.g. climate, topography, culture) that affected educational attainment, this would initially bias our estimator. Through our fixed

[4] High wages could attract both teachers and students. A decline in teachers could reduce the number of available spots for students, which would lead to a decline in enrollment, and subsequently educational attainment. If the number of interested students declined, the amount of educational attainment would directly be reduced.

effects estimation, we are removing time constant factors and thus also removing a lot of the natural variation between municipalities. With an exogenously determined explanatory variable and a fixed effects estimation removing time constant factors, we argue that our model does not require control variables to be unbiased.

4.3 Time Periods

Having a dataset spanning 36 years, there have undoubtedly been a lot of changes during this time. Workforce mobility has steadily increased for the past 40 years, through better roads and infrastructure. The more static nature of the workforce in earlier days could mean that oil activity affected educational attainment more, due to people being less inclined to move away from their hometown.

Another factor that has changed significantly during the last few decades is digitalization and the introduction of the so-called age of information. With home computers becoming a commodity for most households at the start of the 21st century, information that previously had to be physically moved was now readily available. If oil related information was previously less visible, but has become more salient in newer times, this could be a factor leading to educational attainment becoming more affected now. The view of women on oil rigs has also become more positive in modern times. The fact that oil production has affected women differently in different times also seems apparent in Figure 7, presented earlier in this chapter, where we can see a sharp contrast in how the educational attainment changes for women between the first and last period of the plot.

Whatever effect shows to be the most dominating one, we find it relevant to split our dataset into two subset periods, to see if there are any significant differences.

5 Empirical Results

For the main empirical results, we use the distributed lag model presented in Equation (4), including two yearly lags of the explanatory variable, in addition to the current form of the explanatory variable. We have also added dummy-variables for each year and clustered the standard errors at the municipal level. For the main model we have used oil production as the *X*-variable for oil activity, which is estimated in log form. We separately use investment as well, when we examine the introductory effects of oil activity found in the difference-in-difference estimation.

Our explanatory variable is the share of higher educational attainment. We therefore have what is called a level-log model. This means that the intuition for each coefficient (using β_1 as an example coefficient) is the following: if we increase X by 1%, we expect Y to increase with (β_1/100). Since our dependent variable is given in shares or proportions, measured in percentages, the coefficients indicate a *percentage point (pp)* change, and not a proportional percentage change, as one would have in a log-log model.

5.1 Main Model

The results from our main model are presented in Table 6 below. We have looked at the years between 1980 and 2016. These are all the years for which our dataset provides observations, in addition to one year of data, the year 1970. We have chosen to omit this single year, in order to avoid any gaps in our dataset and achieve a strongly balanced panel. We have also split the period from 1980 and 2016 into two subsets. These subsets will be referred to as Period 1 and Period 2, encompassing the years 1980 – 1997 and 1998 – 2016, respectively. For each result, we will look at both the short-run and long-run effects.

5.1.1 Short-Run Effects

Looking at Panel A in Table 6, we can see that almost every coefficient has a negative sign, although the statistical significance varies widely between periods and lagged explanatory coefficients. Comparing Period 1 and Period 2, they portray very different trends. In Period 1, the effect starts off positive and becomes increasingly negative through the first and second lag. In Period 2, the effect starts off with a large negative effect, which decreases over the next two years. For the entire period, all coefficients are small and have large standard errors. Of the single coefficients, only the one-year lagged coefficient for the first period and the current lag for the second period are significant at the 5%-level. A direct interpretation of the current coefficient for Period 2 would state that increasing oil activity by 10% would decrease the share of tertiary educational attainment by 0.025 pp the same year. This reduction is relative to the municipalities in the control group.

Table 6. Short- and long-run coefficients for the main model

	(1) Period 1 (1980-1997)	(2) Period 2 (1998-2016)	(3) Period 1 + 2(1980-2016)
	Panel A: Coefficients for the main model (Fixed Effects)		
β_1 (Current)	0.0253	-0.251**	-0.0761
	(0.640)	(0.012)	(0.298)
β_2 (1-year lag)	-0.0428**	-0.118*	-0.0110
	(0.039)	(0.097)	(0.754)
β_3 (2-year lag)	-0.0637	-0.0084	-0.0099
	(0.329)	(0.942)	(0.910)
R^2	0.902	0.945	0.948
N	2672	2839	5845
	Panel B: Long-run effects for the main model (t-test)		
LRP	-0.0813	-0.3772 **	-.0970
	(0.244)	(0.011)	(0.372)

Note. Coefficients for the short- and long-run effects for the main model, for different time periods. Short-run effects are reported through the coefficients from Equation (4). Long-run effects are presented as the long-run propensity (LRP), attained through Equation (5). P-values, derived from standard errors clustered at the municipal level, are presented in parentheses below each coefficient. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Knowing that our dependent variable takes time to change, this significant current explanatory variable would mean that the individuals anticipated the oil increase and reacted in advance. We can also see that we have very large R-squared values, which is to be expected due to our inclusion of yearly indicator variables.

5.1.2 Long-Run Effects

To more easily extract meaning from these coefficients we estimate the long-run effects through calculating the long-run propensities (LRP) for each period, seen in Panel B of Table 6. The LRP should indicate what happens permanently, after a change in X. It is calculated by taking the sum of the explanatory variable's coefficients with a time specification:

$$LRP = \beta_1 + \beta_2 + \beta_3 \quad (5)$$

First off, the LRPs are negative for all periods. This is in line with our hypothesis, that oil production has a negative effect on the share of higher education. The LRPs, however, vary both in size and in standard errors. The long-run effect for the most recent period is more than four times higher than Period 1, while also being the only significant coefficient at the 5%-

level. A direct interpretation indicates that by increasing the production of oil by 10%, the average affected municipality will experience a permanent reduction of their educational attainment by 0.038 pp, relative to municipalities not affected.

Both the short-run and the long-run effects indicate that oil production has had a significantly larger impact on educational attainment in more recent times, having an LRP more than four times larger in Period 2. One reason for this could be the increased flow of information in newer times, mentioned in Chapter 4, might be the reason for such differing effects. This could also explain why the entire period has such insignificant p-values, due to the combination of two changing trends nulling each other out.

5.2 Heterogeneity by Gender and Educational Length

The results above are combined both for men and women, and short and long higher education. To see if there are any differences between these groups, we isolate and study the effects of each. We do this by establishing dependent variables for both genders, for both educational lengths, and for both genders and their different educational attainments.

For the gender isolated educational attainment variable, the shares are calculated with the gender of the educational attainment in the numerator and the total amount of the given gender in the municipality in the denominator. This means that when comparing the gender isolated long-run effects to the total long-run effects, the total effects are a *weighted average* of the gender isolated ones. This contrasts with the interpretation of the educational length isolated effects. The educational length isolated dependent variables are calculated by having the amount of the relevant educational attainment in the numerator, and the *total* municipal population in the denominator. This means that total long-run effect will be approximately equal to the *sum* of the undergraduate and the postgraduate long-run effects.

5.2.1 Men and Women

Looking at Panel A in Table 7, the short-run effects for men and women are somewhat similar to their aggregated results in Table 6: Period 1 starts out positive and ends up negative, Period 2 starts out with large, negative coefficients and experiences diminishing effects over the next two years. Looking at men and women, the only period with two statistically significant point estimates for both genders is the current lag between 1998 and 2016. This shows that women experience a 40% larger decrease in educational attainment in response to increased oil production the same year, than men. The one-year lag for men in Period 1 and the entire period is also statistically significant at the 1%- and 5%-level, respectively.

In Panel B, we have one long-run effect significant at the 5%-level. This is for women in more recent times and shows a relatively large 0.053 pp negative response to a 10% increase in oil activity, compared to municipalities defined as not being affected. The effect for women is also more than twice that of men and is clearly the dominating force behind the long-run effect observed in Table 6. This is also in line with the aggregated visual differences shown in Figure 7, where women experience a much larger relative difference than men, in the face of oil production.

Table 7. Short- and long-run coefficients for the main model, by gender

	Panel A: Coefficients for men and women (Fixed Effects)					
	Period 1 (1980-1997)		Period 2 (1998-2016)		Period 1+2 (1980-2016)	
	(1) Male	(2) Female	(3) Male	(4) Female	(5) Male	(6) Female
β1	0.0229	0.0287	-0.225**	-0.314**	-0.053	-0.11
	(0.739)	(0.570)	(0.014)	(0.017)	(0.458)	(0.187)
β2	-0.109***	0.0263	-0.0691	-0.128	-0.0764**	0.0706
	(0.000)	(0.433)	(0.359)	(0.283)	(0.021)	(0.124)
β3	-0.0218	-0.106	0.0903	-0.0926	-0.023	0.0158
	(0.743)	(0.158)	(0.420)	(0.580)	(0.777)	(0.878)
R^2	0.852	0.904	0.854	0.958	0.898	0.96
N	2672	2672	2839	2839	5845	5845

	Panel B: Long-run effects for men and women (t-test)					
	Male	Female	Male	Female	Male	Female
LRP	-0.1082	-0.0509	-0.2039	-0.5342**	-0.1524	-0.0235
	(0.139)	(0.513)	(0.177)	(0.012)	(0.131)	(0.847)

Note. Coefficients for the short- and long-run effects for the main model, for different time periods. Short-run effects are reported through the coefficients from Equation (4). Long-run effects are presented as the long-run propensity (LRP), attained through Equation (5). P-values, derived from standard errors clustered at the municipal level, are presented in parentheses below each coefficient. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

These findings somewhat contrast with results found by Cascio and Narayan (2015), who look at the effects of hydraulic fracturing (fracking) on high-school drop-out rates. They find that men are far more affected than women. This is, however, related to secondary educational attainment, which is not necessarily comparable to the effects we have reported on tertiary educational attainment.

The intuition for why women are affected more than men in newer times isn't inherently obvious. As shown in Chapter 2, men far outweigh women on oil rigs. These statistics, however, only show the women working directly on the platform. This could mean that there are jobs not requiring higher education that aren't on the oil rig, but directly or indirectly support or otherwise get affected by oil activity, that attract women.

Another explanation could be related to men in general having more non-educated jobs, than women. This is supported by the fact that women on average have a larger share of higher education than men (see Chapter 3). The large LRP for women could mean that women prone to working in oil-related, non-educated professions would, on average, choose to educate themselves more often than men, if living in a municipality not related to oil fields. Since our coefficients show how much the treatment group is affected relative to the control group, if women in the control group chose to educate themselves more often than men, this would lead to higher relative effects for women in our model. Following the same line of logic, this would cause men to have lower estimates than women, due to more men in the control group choosing jobs not requiring education, thus leading to lower differences. Said more plainly, it could indicate that the men would have chosen another low-educated job, if the oil-jobs weren't available. Women, on the other hand, would have choose education in the absence of oil, which is why the differences are larger.

Another explanation could be related to cultural differences between the mostly western treatment group and eastern control group. If views on marriage are more traditional in the treatment group, it could lead to more women tending to the home, if men are the sole source of family income. As oil production decreases, and the men earn less, this could have led to women fearing for their family's financial stability. This could in turn lead to women acquiring part-time jobs, but could also cause them to enroll in higher education with the hopes of getting a better job in the future. This could help explain why women seem to be relatively unaffected for Period 1, while being widely affected in Period 2. If the traditional grasp has become looser over time, women might be more prone to take action for the family's financial future in more modern times, while in older times initiative would be viewed as less acceptable. This is however mostly speculative and would require additional research to firmly assert.

5.2.2 Undergraduate and Postgraduate

Observing the differences between the two levels of higher educational attainment, there are some clear dissimilarities. Of the short-run effects, we have two effects that are below the 5%-level and one at the 10%-level. This can be seen in Table 8, Panel A below. For period 1, the postgraduate attainment has a highly significant, but very small negative effect. Undergraduate attainment in more recent times experiences negative effects for the current and one-year lagged explanatory variables, at the 5%- and 10%-level, respectively.

Looking at the long-run effects in Panel B, undergraduate attainment in Period 2 is the only statistically significant LRP and has a relatively large effect with a point estimate of 0.265. This is more than twice that of the postgraduate attainment of the same year and means that the negative effects from the undergraduate attainment accounts for more than two thirds of the total long-run effect observed in Table 6. The negative effect on undergraduate attainment in the long-run is larger in all cases, compared to the postgraduate attainment.

Table 8. Short- and long-run coefficients for the main model, by educational level

	Panel A: Coefficients for postgrad. and undergrad. (Fixed Effects)					
	Period 1 (1980-1997)		Period 2 (1998-2016)		Period 1 + 2 (1980-2016)	
	(1) Undergrad.	(2) Postgrad.	(3) Undergrad.	(4) Postgrad.	(5) Undergrad.	(6) Postgrad.
β_1	-0.0007	0.0260	-0.132^{**}	-0.119	-0.0495	-0.0266
	(0.984)	(0.293)	(0.048)	(0.138)	(0.273)	(0.507)
β_2	-0.0203	-0.0225^{***}	-0.136^{*}	0.0184	-0.0082	-0.0029
	(0.249)	(0.003)	(0.056)	(0.701)	(0.708)	(0.904)
β_3	-0.0375	-0.0262	0.0029	-0.0113	0.0057	-0.0156
	(0.474)	(0.189)	(0.973)	(0.883)	(0.925)	(0.713)
R^2	0.917	0.605	0.943	0.763	0.964	0.745
N	2672	2672	2839	2839	5845	5845

	Panel B: Long-run effects for postgrad. and undergrad. (Nonlinear test)					
	Undergrad.	Postgrad.	Undergrad.	Postgrad.	Undergrad.	Postgrad.
LRP	-0.0585	-0.0227	-0.2649**	-0.1123	-0.0520	-0.0450
	(0.303)	(0.256)	(0.012)	(0.408)	(0.348)	(0.496)

Note. Coefficients for the short- and long-run effects for the main model, for different time periods. Short-run effects are reported through the coefficients from Equation (4). Long-run effects are presented as the long-run propensity (LRP), attained through Equation (5). P-values, derived from standard errors clustered at the municipal level, are presented in parentheses below each coefficient. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

One could argue that the amount of years a person is willing to invest in their education is connected to how dedicated they are to their chosen field or profession. Using this logic, it makes sense that students considering studying an undergraduate degree are more prone to changing their career path, than students considering studying for a postgraduate degree. This could be a reason for the high and significant coefficient for the undergraduates long-run effect, compared to the lower coefficients of their postgraduate counterparts; the potential undergraduates are more easily swayed by the allure of the immediate high wages in the oil industry, while the potential postgraduates are more determined on taking their degree. These coefficients are of course the change relative to the municipalities not affected by oil production. We note that once a postgrad. student has completed their fourth year, they go from being registered under the undergrad. variable to being registered under the postgrad. variable. Looked in isolation, this would lead to a reduction of the undergrad. category. This effect, however, will be the case for both the treatment and control group, and shouldn't affect the relative difference between them.

5.2.3 Male and Female, Undergraduate and Postgraduate

To keep the tables from being cluttered with too much information, we here only present the long-run effects. These are shown in Table 9.

Table 9. Long-run coefficients for the main model, by educational level and gender

	Panel A: Long-run effects by tertiary level and gender (Undergraduate)					
	Period 1 (1980-1997)		Period 2 (1998-2016)		Period 1 + 2 (1980-2016)	
	Men	Women	Men	Women	Men	Women
LRP	-0.0945*	-0.0166	-0.1077	-0.4100**	-0.1459***	0.0617
	(0.092)	(0.809)	(0.259)	(0.036)	(0.009)	(0.372)

	Panel B: Long-run effects by tertiary level and gender (Postgraduate)					
	Period 1 (1980-1997)		Period 2 (1998-2016)		Period 1 + 2 (1980-2016)	
	Men	Women	Men	Women	Men	Women
LRP	-0.0137	-0.0343*	-0.0962	-0.1242	-0.0065	-0.0852
	(0.62)	(0.097)	(0.458)	(0.455)	(0.92)	(0.228)

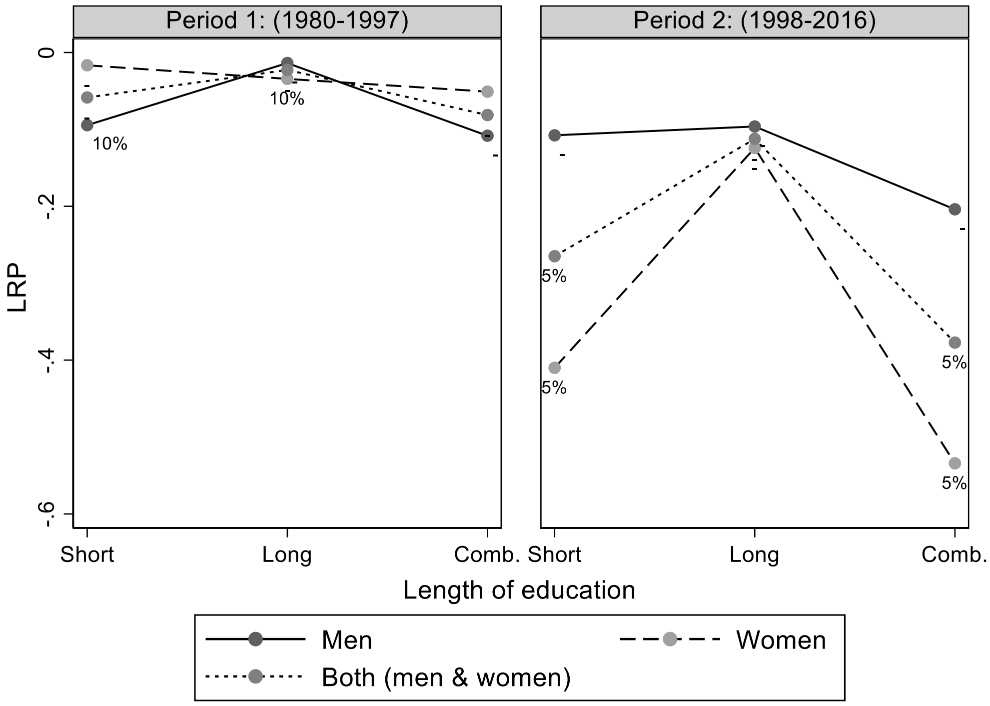
Note. Long-run effects presented as long-run propensities (LRP), attained through Equation (5). They are estimated for postgrad. and undergrad. attainment, for men and women separately. The p-values of the t-tests are shown below each coefficient in parenthesis'. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Comparing to the long-run effect for undergraduate attainment in Period 2, found in Table 8, we can now see that the driving force behind this effect was women, having an LRP roughly four times larger than men. The women's LRP is also significant at the 5%-level. Another significant LRP, this time at the 1%-level, is men's undergraduate attainment in the entire period. Out of all the long-run effects estimated for the entire period, this is the first one that has produced a significant p-value and goes to show how heterogeneity analysis can reveal new aspects to the model. We can also see that the general trend of undergraduate attainment being more affected by oil production than postgrad. attainment still holds for most of the point estimates. In addition, we can see that out of the twelve different estimates, only one has a positive sign. This supports our hypothesis that oil production has a negative effect on the related municipalities, compared to non-oil affected municipalities.

5.2.4 Summary of Heterogeneity

Very few of the long-run effects for the entire period have been proven to be statistically significant. As mentioned earlier, this could be due to the effects of oil activity on education being inherently different from earlier years to the present, making inference difficult over such a large time span. This change of pattern can be seen in Figure 8 below, where we show the long-run effects for Period 1 and Period 2. We have not included the effects for the total period, from 1980 to 2016, due to lack of significant results.

Figure 8. Long-run propensities (LRP)



Note. The graph visualizes the long-run propensities (LRP) presented in Table 6-9, for Period 1 and 2. Each dot represents an LRP-value measured on the y-axis. Along the x-axes are the different lengths of education; short, long and combined (short + long). Each gender, in addition to both genders combined, are represented by lines with different line patterns. The significance level for each t- test affiliated with the LRP is shown next to the dots, where a hyphen is shown for LRPs with a p-value higher than 10%.

The graph shows every long-run effect (LRP) estimated for the two periods individually as dots, concatenating information given in Table 6, 7, 8 and 9. The lines are drawn for men, women and both genders combined. There are three parameters along the x-axis, one for each educational length and one for the lengths combined (total tertiary educational attainment). As we would expect, the line for both genders is always in between the lines for men and women, following the logic initially presented in this chapter; the gender aggregated effect is a weighted average of the gender individual effects. Following logic from the same paragraph, the combined tertiary educational attainment point is always lower than the individual educational lengths, due to it being the sum of the two.[5]

Looking at the graph broadly, we can see that every long-run effect is below zero, which is in line with our initial hypothesis, that oil activity has a negative effect on educational attainment in related municipalities, compared to unrelated ones. Comparing the two periods, we can clearly see a change in how oil production has affected educational attainment differently between Period 1 and Period 2. Between 1980 and 1997, all long-run effects are close to zero, with an average LRP of -0.05, and mostly show high p-values. We can also see fairly homogenous effects between genders, all LRPs being within a few points of one another.

In period 2, the long-run effects are a lot larger, having an average LRP of -0.25, five times higher than the average of Period 1. There is also a lot of heterogeneity between genders. Men have an average LRP of -0.13, three times lower than women's average LRP of -0.36. Men also experience fairly similar effects on both types of higher education, while the share of women's educational attainment of undergrad. degrees are reduced by four times the size of postgrad. degrees. The effects on postgraduate attainment are however very homogenous, both between genders and time periods, also displaying mostly insignificant p-values. This is in line with our earlier reasoning, that individuals inclined towards postgraduate degrees are less swayed by external changes in their environment and thus less affected by the increased labor demand caused by increased oil production.

5.3 Rolling Out Differences-In-Differences

In the above main model, we have assumed that the relationship between oil and education can be explained linearly, where a %-change in X corresponds to a certain amount of change in Y. This allowed us to measure how the *size* of oil activity affected education. It could be that the relationship isn't quite as refined as this, but rather depends more broadly on the *introduction* of oil. To measure these effects, we use the differences-in-differences roll-out method described in Chapter 4, in Equation (2).

To define what time oil activity started to affect the level of education of the connected municipality, we look at two different introduction points: 1) the time of the first produced unit

[5] This would not be the case, if the two educational lengths had coefficients with either positive or different signs; the sum of two positive coefficients would be larger than their individual parts, and the sum of a positive and negative coefficient would be somewhere in between the two.

of petroleum, and 2) the time of the first investment made. Every production start happens in Period 1, except for the production for Hammerfest. All years for both production and investment start can be seen in Table 3, Chapter 3.

5.3.1 Estimates

In table 10, we have summarized the results for the difference-in-difference model, when introduction is related to both production and investment. In column (2) and (5), we have confined the model to include a time span in which all of our treatment groups are active, 9 and 10 years, respectively. In column (3) and (6), we have estimated an unconfined effect, where the number of active years range from 9 to 36 years. For reference, the pre-activity average share of educational attainment is also shown.

Interpreting these results directly, the introduction of oil production has permanently reduced educational attainment in the treatment group by 0.5 pp, while the introduction of oil investment has reduced educational attainment by 0.4 pp. None of these estimates are statistically significant, but they are both close to a significance level of 10%. During the first nine years after production has started, the average permanent reduction in educational attainment is shown to be 0.23 pp, compared to the control group. Having larger effects in the long-run than in the short-run indicates that the impeding effects oil activity have on education increases as time passes. The effects for investments are more than half that of production.

Table 10. Roll-out estimates

	Production start			Investment start		
	(1) Mean pre-production	(2) Confined effect	(3) Total effect	(4) Mean pre-investment	(5) Confined effect	(6) Total effect
Educational attainment	9.282	-0.230	-0.500	8.707	-0.093	-0.404
		(0.283)	(0.111)		(0.697)	(0.120)
Post-activity years		9	26.9		10	29.9
Yearly Dummies		YES	YES		YES	YES
N		6,179	6,179		6,179	6,179

Note. Column (1) and (4) show the average higher educational attainment one year prior to the start of production or investment, respectively. Column (2), (3), (5) and (6) contains the results from the difference-in-difference estimator from Equation (6), using production start as an indicator. The effects are an average change over the time period. Column (2) and (5) consist of a balanced panel, where there are observations for all nine years. Column (3) and (6) have a longer, but unbalanced panel, where every year for which production or investment is positive is included. P-values attained from the robust standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

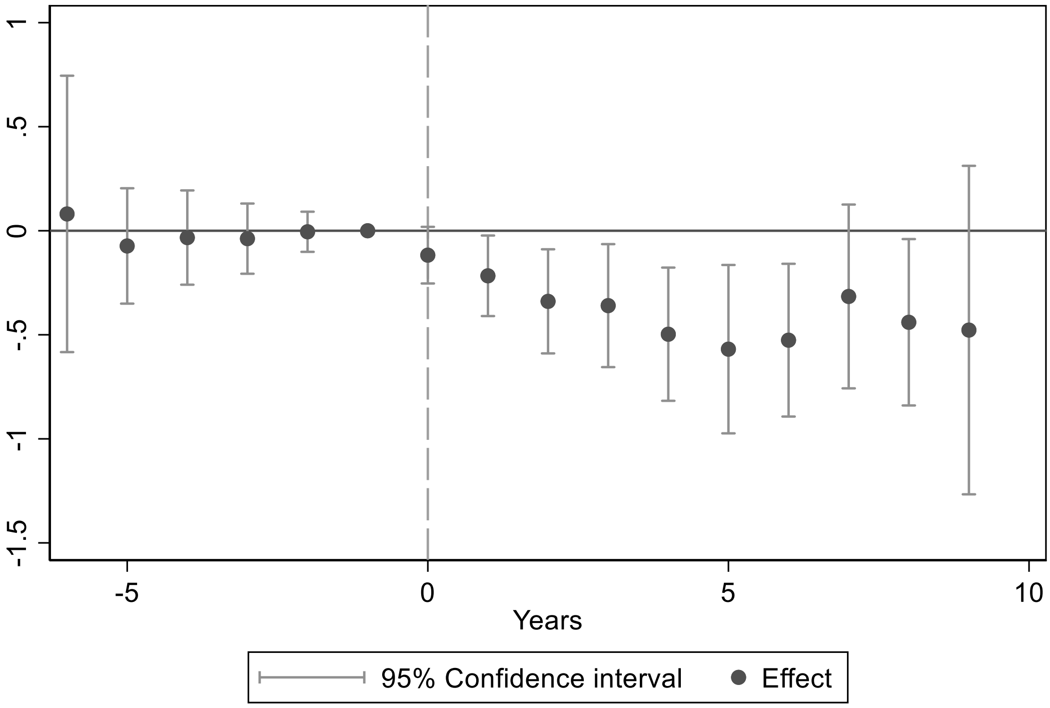
5.3.2 Event Study

We now try to inspect the different yearly effects of the introduction of oil activity, with an event study specification. To achieve a balanced panel along our event study, we only include years for which we have observations for every treatment group. Due to our yearly educational data starting in 1980, we are forced to omit the three treatment groups, specifically those regarding Florø, Sotra and Mongstad, due to production starting before 1980. This leaves us with a pre-treatment horizon of 6 years, due to Mongstad starting production in 1986, and a post-treatment horizon of 9 years, due to Hammerfest starting production in 2007. Translating these parameters to Equation (3) gives us $T=-6$ as the first event-year and $q=16$ as the event lasts for 16 years. The first year before the introduction ($T=-1$) is normalized to zero and serves as the base-year.

5.3.2.1 Production

In Figure 9 below, the pre-treatment effects display little difference from zero, while all of the post-treatment estimates are below zero, with only two years, year 7 and 9, having confidence intervals breaching zero. This indicates that the introduction of oil production to a municipality has a negative effect on tertiary educational attainment, compared to municipalities not affected yet or at all. The point estimates also indicate that the effect is increasing for the first five years, before becoming weaker and less statistically significant.

Figure 9. Event study estimates for the start of oil production's effect on educational attainment

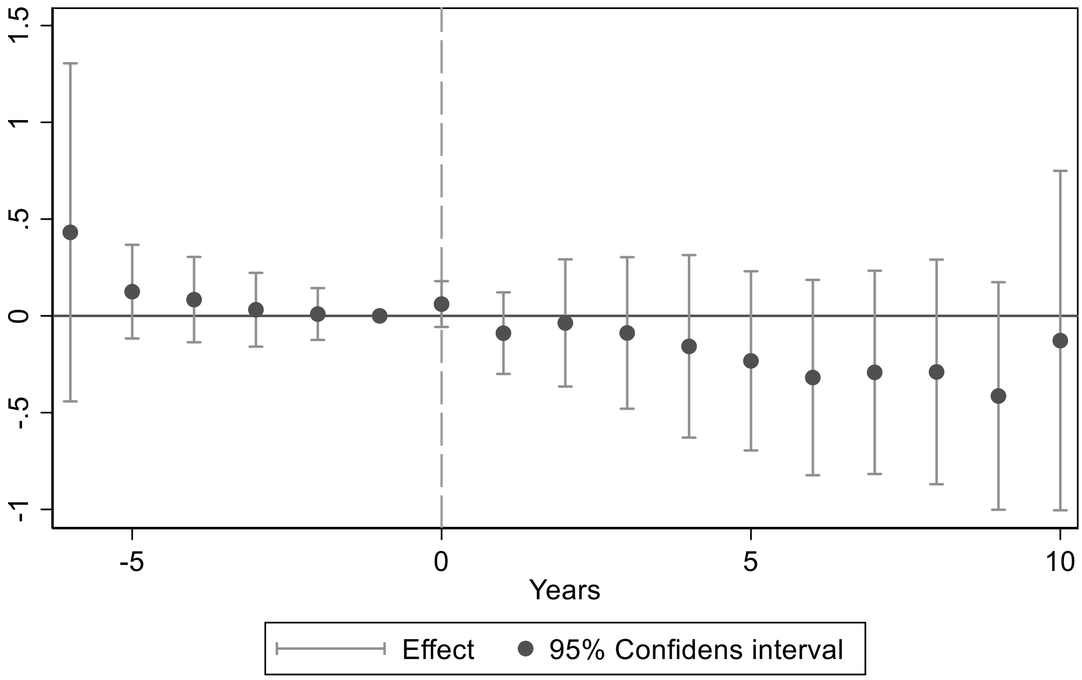


Note. Event study estimates and confidence intervals for each year before and after the production of oil has started, where year 0 indicates production start. The control group is placed in the normalized year prior to production start and set to zero, to serve as the baseline for the changes in education.

5.3.2.2 Investment

The event study estimates for the introduction decided by the first investment is shown in Figure 10. We can see clear similarities between the event study for investment and production. This is to be expected, as the first investment occurs on average 4.25 years before the first instance of production (see Chapter 3). This seems to be replicated in the figures, the largest coefficient for production happening in year 5, while investment has its largest coefficient in year 9, four years later. While the trends and direction of the estimates for investment are similar to production, none of the post-treatment estimates are significant, having every confidence interval above zero.

Figure 10. Event study estimates for the start of oil investment's effect on educational attainment



Note. Event study estimates and confidence intervals for each year before and after the investment of oil has started, where year 0 indicates the year of investment start. The control group is placed in the normalized year prior to investment start and set to zero, to serve as the baseline for the changes in education.

6 Robustness

Since our control group is subject to some asymmetry with our treatment group, in terms of initial educational attainment, size and population, due to our nonrandomized approach, our results could be subject to selection bias (Heckman, 1979). When we have such confounding factors, it makes it challenging to infer that differences reported in our model are due to our explanatory variable alone. To combat these issues, we attempt to create a new control group for which these factors are as close to identical as possible to our treatment group. If our results remain unaffected to the changes in these factors, it increases the robustness of our model (Brewer & Crano, 2000, p. 19) and subsequently improves the validity of our explanatory variable being responsible for the reported effects.

6.1 Balanced Control Group

Using propensity score matching (Rosenbaum & Rubin, 1983), we establish a new control group balanced across population density, area of land and higher educational attainment in 1970 (see Figure A10 for balancing tests). Using area of land together with population density allows us to balance across size of population as well, due to the way these three factors are connected. Additionally, comparing the population density from 1970 to 2016, based on more than 3,400 observations, the two groups still aren't significantly different from zero (see Figure A11). Following the recommendations of Caliendo & Kopeinig (2008, p. 42), we impose a caliper restriction to avoid bad matches. We choose 0.1 as an acceptable range of propensity score, which forces us to omit two treatment municipalities and subsequently two control municipalities.[6] Since we used nearest neighbor matching with no replacements, this leaves us with a treatment and control group consisting of 46 municipalities each (see Table A6 for an exhaustive list of chosen municipalities). The pool for choosing control municipalities consisted of all municipalities less the treatment group, except municipalities along the coast and western part of Norway.[7] The geographical location of both the treatment and balanced control group can be seen in Figure A12.

The reason for omitting coastline and western municipalities follows the logic of the treatment group; proximity to supply bases is the factor determining treatment or not. Although the closest

[6] The omitted treatment municipalities were Kvitsøy and Fedje, which both have unusually small areas of land. Having respectively 6 km^2 and 9 km^2 of land, they are the smallest and third smallest municipalities in the country. This gave rise to bad matches and high propensity scores.
[7] Municipalities numbered 1000-1599 are defined as belonging to the coastline and western part of Norway.

municipalities are already included in our treatment group, the surrounding municipalities may still be subject to some effects of oil activity, though most likely on a smaller scale. Including such municipalities will move our estimates closer to zero, as the differences between the treatment and control are diminished. To mitigate this risk, we therefore aim to omit these semi-affected municipalities, and use the coastline and western municipalities as a proxy for this.

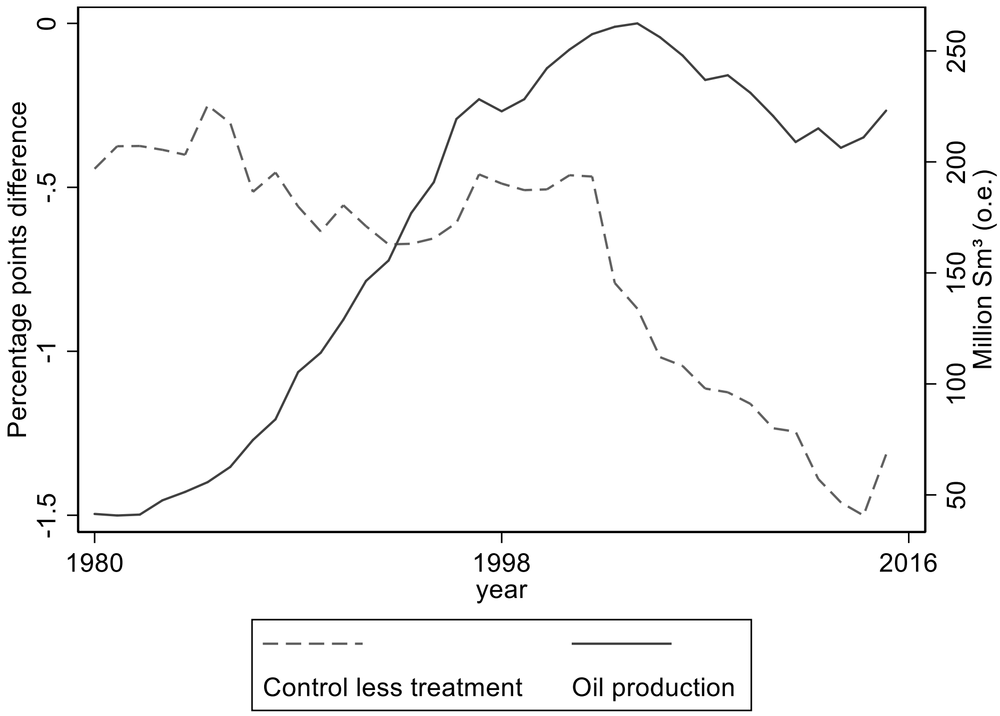
6.1.1 Main Model Retested

Looking at the difference between the average educational gap for the treatment group and the balanced control group in Figure 11, we now find a more ambiguous connection to oil production than previously reported in Figure 6 in Chapter 4. As oil production increases, the educational gap stays relatively constant for the first 15 years. This is followed by a minor increase in the line tracing the gap as production keeps rising. The last 15 years, the educational gap becomes larger, moving the line of the gap downwards. The decline of oil production follows shortly after. While the control group used previously had a higher educational attainment, returning a positive difference in educational attainment, the balanced control group has a lower educational attainment on average, producing negative differences for all years.

Looking at the differences by gender shown in Figure A13, the trend for women is similar to the one reported in Figure 7 (see Chapter 4); as oil production declines, the educational attainment for the women in the treatment group increases more than the control group. For men, although the slope for the educational gap line is similar to the one reported in Chapter 4, the size of the gap is a lot smaller, staying close to zero for most of the time period. The largest gap for men is reported at 0.43 pp, which contrasts with the size of the educational gap between the treatment and control for women. At its peak, their gap is measured at 2.77 pp, more than six times the size of their male counterparts.

This difference is also made clear in Figure A14 in the appendix, plotting the educational attainment for the treatment and control group over time divided by genders. The men in both groups have similar shares of educational attainment, while women in the treatment group always have a higher attainment, experiencing an increased gap the last half of the period. As mentioned in Chapter 4, these graphs aren’t an accurate representation of how education has been affected by oil activity, due to the treatment municipalities being affected differently, both in terms of timing and magnitude of production. That being said, they do make indications toward the connection between oil and education being less clear than previously assumed.

Figure 11. Gap in educational attainment and oil production, using a balanced control group



Note. The dashed line plots the difference in higher educational attainment for the control and treatment group over time. The difference is calculated by taking the higher educational attainment of the control less the treatment, a negative number therefore indicating the treatment group having a higher educational attainment. The solid line shows the total oil production.

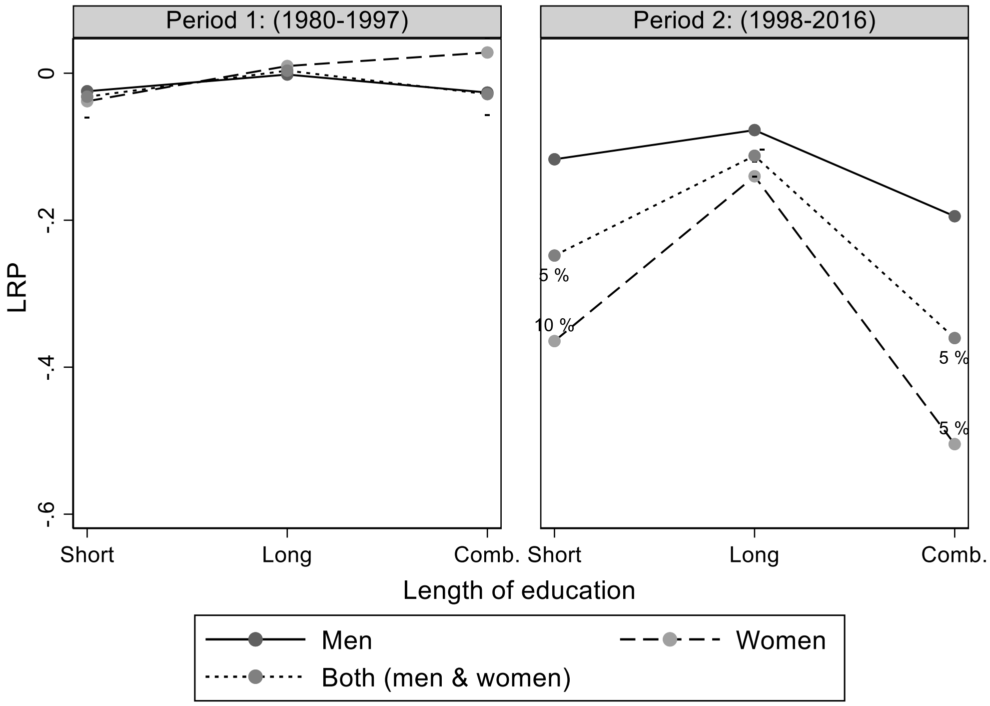
In Table A7, we have repeated the analyses from our main model in section 5.1, this time with our balanced control group. The table is structured so that each panel is a remake of the long-run effects presented in Tables 6, 7, 8 and 9, respectively, where Panel D1 and D2 in Table 9 are the same as Panel A and B in Table A7.

Looking broadly at the table, we can see that using the balanced control group returns less negative long-run effects, having 17 out of 27 estimates below zero, compared to the 26 negative estimates reported previously. The magnitude of the long-run effects are also on average smaller, having gone from eight estimates with a coefficient smaller than 0.05, to now reporting 16 estimates smaller than 0.05. An LRP of 0.05 indicates that a 100% increase in oil production changes the share of educational attainment by 0.05 pp, meaning the effect is relatively minor. Comparing the p-values, 21 of the 27 point estimates are reported with larger p-values than before, leaving us with 6 improved point estimates. A part of this general reduction in significance and increased standard errors stems from the reduction in observations, due to a control group more than the half the size of the original. All things being equal, this reduces our statistical power and thus our significance.

In general, the analysis performed with the balanced control group has returned fewer negative estimates and smaller long-run effects from oil production on educational attainment, in addition to being less statistically significant on average. However, looking at the estimates that are statistically significant, the differences are less severe. Out of the seven statistically significant results from our main analysis, four of the estimates are still significant and close in size to the previous estimates. The three estimates that are no longer statistically significant were pertaining to the educational attainment of undergraduate degrees for men, in Period 1 and the entire period, in addition to women's attainment of postgraduate degrees in Period 1. These estimates have gone from significance levels between 10% and 1% in the main analysis, to being very insignificant and close to zero in the current model. This indicates that these specific results were probably caused by differences between the treatment and previous control group.

Looking at the remaining statistically significant long-run effects, they are all in Period 2 and they all include some form of women and undergraduate attainment. For both genders and educational levels, the estimates in Panel A show that a 100% increase in oil production is associated with a decrease in the share of higher educational attainment by 0.36 pp, compared to municipalities in the control group. The effects for the same period for women in Panel B are even larger, showing a relative decrease of 0.50 pp, while the undergraduate attainment for both genders in Panel C report decreases of 0.25 pp per every 100% increase in oil production. Plotting all the point estimates from Period 1 and 2 over the different educational levels in Figure 12 and separating by gender, as we did in Figure 8, we see that a lot of the results are similar to the results presented in Chapter 5. The long-run effects for Period 1 are still low and insignificant, while the effects are a lot larger in Period 2, women having the most difference in effect between educational levels.

Figure 12. Long-run propensities (LRP) for using the balanced control group



Note. The graph visualizes the long-run propensities (LRP) presented in Table A7, for Period 1 and 2. Each dot represents an LRP-value measured on the y-axis. Along the x-axes are the different lengths of education; short, long and combined (short + long). Each gender, in addition to both genders combined, are represented by lines with different line patterns. The significance level for each t- test affiliated with the LRP is shown next to the dots, where a hyphen is shown for LRPs with a p-value higher than 10%.

6.1.2 Difference-In-Difference and Event Study Retested

Continuing with the robustness check of the main analysis, we look at the broader effect of introducing supply bases, or more specifically starting oil activity. This is done by using Equation (2), where positive production is used as an indicator for oil activity.

We now use the new, balanced control group, to see if the previous results are robust to changes in the demographical factors between the treatment and control. If the results are different from before, it implies that our previous analysis might be caused by the variance in factors between the treatment group and the initial control group.

From Table 11, we can see that the difference-in-difference estimates have changed compared to our main analysis. The total long-run effects of introducing oil production is reduced by more than a third and has gone from being almost significant at the 10%-level to being distinctly insignificant. This contrasts with the confined effects, measured only for the length of which we have a balanced panel. The confined effects have increased, in addition to being accompanied with lower standard errors, making the estimate significant at the 10%-level.

Table 11. Roll-out estimates using the balanced control group

	Production start		
	(1) Mean pre-production	(2) Confined effect	(3) Total effect
Educational attainment	8.944	-0.369*	-0.148
		(0.094)	(0.656)
Post-activity years		9	26.9
Yearly Dummies		YES	YES
N		3,404	3,404

Note. Column (2) and (3) contain the results from the difference-in-difference estimator from Equation (2), using production start as an indicator. The effects are an average change over the time period. Column (1) is the average educational attainment before production started. Column (2) consists of a balanced panel, where there are observations for all nine years. Column (3) has a longer, but unbalanced panel, where every year for which production is positive is included. P-values, derived from standard errors clustered at the municipal level, are presented in parentheses below each coefficient. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

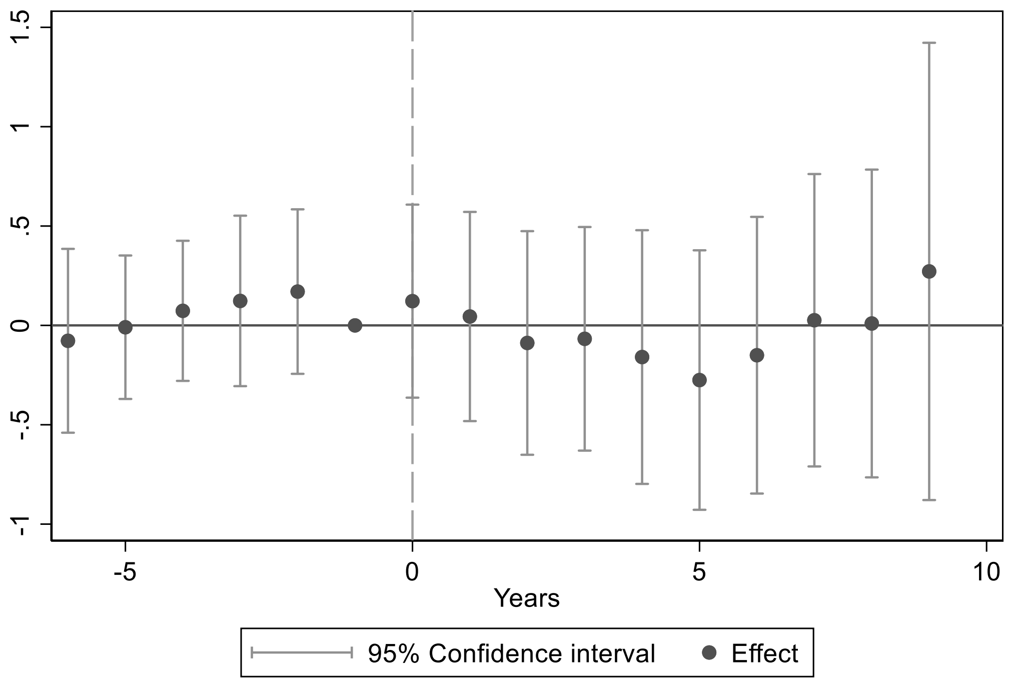
The confined effect being larger than the total effect suggests that the effect oil production has on education in the related municipalities are lessened over time. The fact that the total effect of oil production has lost most of its statistical significance implies that the results presented in our main analysis should be interpreted with caution.

6.1.3 Event Study

To test the identifying assumption of the difference-in-difference estimates, we again use the event study specification for the balanced control group.

The yearly effects shown in Figure 13 are quite different from the ones previously shown in Chapter 5. Although the pre-treatment effects aren’t statistically different from zero, they now seem to form an upward pattern. The post-treatment effects have a slightly similar pattern to the post-treatment effects reported in Figure 9, although it starts with a positive response to the start of production. In addition, all the estimates have very large standard errors, evident from their large confidence intervals, while at the same time never being statistically different from zero. The magnitudes of the yearly changes are also lower than before, which is surprising, considering the confined estimates reported in the previous section were larger than what was initially shown in the main analysis. In total, this deems the difference-in-difference estimates presented in the previous section to be erroneous. It also indicates that the same estimates from the main analysis are an unreliable source for inference, regarding the effects oil production has on education. Due to other factors varying at the same time, to claim that oil production is the sole cause of the reported differences is hard to defend.

Figure 13. Event study estimates for the start of oil production's effect on educational attainment, using a balanced control group



Note. Graph plotting changes in education in relation to the year of production start, in addition to their 95% confidence intervals. The changes in education are the average differences across the treatment municipalities, compared to municipalities not yet experiencing oil activity and municipalities never experiencing oil activity, the control group. The control group is placed in the normalized year prior to production start and set to zero, to serve as the baseline for the changes in education.

We have chosen not to include investment as an explanatory variable, as we previously did in the main analysis. As we explained earlier in Chapter 5, the effects of investments closely resemble that of oil production, with the difference that they lag a few years behind and are less significant. We therefore feel that they add little to this chapter and choose to omit it as a part of the analysis. The event study, using positive investment as an indicator variable, can be seen in Figure A15 in the Appendix.

6.2 Area of Impact

As our treatment group only consist of the base-municipality's neighboring municipalities, our area of impact is quite isolated. To see if our findings are robust to the chosen area of impact, we retest our main analysis with a new assignment of treated municipalities. For this we need to use a treatment division which is larger than before, while at the same time being relevant to our analysis, reflecting areas that would be directly or indirectly related to oil rig activity.

M. S. Bhuller (2009), as part of Statistics Norway, aimed to form an alternative to the geographical division of *economic regions*, which had been previously used as an intermediary grouping of municipalities between municipal and county level (Hustoft, Hartvedt, Nymoen, Stålnacke & Utne, 1999). The economic regions were restricted to only include municipalities within the same county, a restriction that not always reflected the working patterns of the individuals living in the assigned region. Based on commuting data, Bhuller wanted to create a unit of analysis that better reflected regions in which individuals worked. His efforts produced what he called *work regions*. As we aim to study the effects that may arise from oil activity absorbing a larger part of the workforce, regions that capture working patterns are relevant, and we therefore use these work regions as a basis as our second treatment group.

To form our new groups of treatment, we use the base-municipality for each supply base as the foundation for choosing the relevant work regions. We then assign the oil activity for each supply base to the municipalities within the work region. There are two cases of more than one base-municipality residing in the same work region; Fjell and Lindås reside within Bergen work region no. 44, and Stavanger and Sola reside in Stavanger work region no. 41. The perhaps intuitive approach to deal with this would be to aggregate production for the supply bases within each region, which would return the work region level of production. This would, however, give the municipalities within these work regions artificially large production values when both supply bases are active, in turn creating very small relative effects on educational attainment. This would downward bias our estimates and produce larger standard errors. If we use the mean of production, we will keep the correct work region level when only one base is active and will divide the aggregate production in two when two bases are active, keeping it closer to realistic levels. We therefore choose work region mean as the level for our explanatory variable, finding it preferable to using aggregation.

This leaves us with six treatment groups, in total containing 80 municipalities (see Table A8 in the appendix). For all intents and purposes, this new treatment group acts as an enlargement of the original treatment area (see map in Figure A16) and therefore helps us answer the question initially posed in this section. If the estimates are roughly similar to our results in the main analysis, it indicates that the range of which oil production affects educational attainment is larger than what we previously expected. If the effects are reduced, it indicates that the effects are diminishing and might not be applicable to a larger area.

For the counterfactual control group, we use propensity score matching as in the last section. Using the same caliper range, we now only omit one treatment and subsequently one control municipality, leaving us with 79 municipalities in each group[8].

The retested results for the main model with the work region treatment group and its associated balanced control group can be viewed in Table A9. Looking at the coefficients broadly, most of the point estimates are within range of the results presented with the balanced control group earlier in this chapter (see Table A7 in the appendix), both in terms of statistical significance and magnitude. We note that the differences compared to the main analysis are somewhat larger than the previous comparison for the retested balanced control group. Looking at the long-run effects that were deemed statistically significant in Period 2 from Section 6.1.1, they are all still statistically significant, in addition to every estimate experiencing decreased standard errors. They are also all very close in magnitude to their earlier counterparts, except for the long-run effects on women's undergraduate attainment in Period 2. This has increased quite a bit, revealing it to be the clear dominating force of the significant results regarding women in Panel B, now being responsible for 97% of the total long-run effect.[9] In addition, the effects on women's undergraduate attainment are also responsible for essentially the entire effect behind the undergraduate attainment effect in Panel C. For the entire period, the long-run effects for postgraduate attainment for both genders and postgraduate attainment for men are also statistically significant. These specific long-run effects have never been close to statistical significance prior to this analysis.

In conclusion, we can state that the results of any significance are relatively robust to the size of impact and assignment of municipalities defined as treated. The statistically significant estimates just being slightly lower than our previous estimates suggest that the range at which oil production influences educational attainment is larger than what we initially anticipated.

6.3 Summary of Robustness

Judging by these results, the effects of oil production on educational attainment seem more confined than our analysis in Chapter 5 initially suggested. Where we previously reported

[8] Fedje was omitted this time as well, due to its unique characteristics.

[9] The long-run effect for women in Period 2 shown in Panel B (-0.4786), consists of both the undergraduate and postgraduate attainment. This is the sum of these two estimated separately, found in Panel C (-0.4642) and Panel D (-0.0144). This allows us to measure the relative effects for each level of educational attainment. For genders, the operation is slightly different, due to their aggregated effects being a weighted average, and not the sum of their individual effects. The mechanisms behind this are explained more closely in Section 5.2.

significant long-run propensities regarding all periods, genders and educational attainment, using a more balanced control group has left us with results driven exclusively by women and undergraduate attainment, only for the second period. Looking at the descriptive statistics, it seems that women in more recent times experience a slight relative reduction in education when oil production increases. However, the same statistics indicate that the differences become more salient during a *decrease* in oil production, when women's higher educational attainment sharply *increases* compared to the balanced control group.

The same results are shown to be significant using a different assignment of treated municipalities. This treatment group, being both larger in size and in number of municipalities, indicates that the effective range oil production has on educational attainment is even greater than what our main analysis implied.

Our difference-in-difference estimates regarding the introduction of oil production cannot be inferred as having an effect on educational attainment, between the treatment and balanced control group. This is however somewhat in line with the results from our long-run propensities (LRPs) only being significant for the later years of our model. The production start for the supply bases, which is what the difference-in-difference model uses as a point of analysis, all happen before 1998, with one exception (see Table 3, Chapter 3). Although some of the estimates from the event study cross over to the second period, most of the estimates are contained within the first. Therefore, we would argue that even though the roll-out and event study have proven to be statistically insignificant for the balanced control group, it does not disprove the significant estimates of the long-run propensities from the second period, attained through our main model.

7 Implications

Our research question and subsequent models aim to identify if there has been a reduction in the higher educational attainment in municipalities affected by oil activity, compared to municipalities that aren't affected. Due to our models examining higher educational attainment as a *share* of the entire population in the given municipality, our estimates can show how the *composition* of educated individuals has changed. What the model does not show is if this change in educational attainment is due to the current population educating themselves less, or if it is due to a difference in the educational attainment of inhabitants moving to or from a municipality[10]. This affects how broadly we can credit the reductions found in our model. If the inhabitants of the municipalities affected by oil indeed do educated themselves less, we can make the case that the reduction in absolute terms has reduced the country's higher educational attainment. If the results stem from individuals with lower education moving in to the municipalities or individuals with higher education moving out, the location of the educated population is simply being changed around within Norway. In this case we can present how the educational composition within the treatment group has been affected, but the country as whole will still have the same total amount of educated individuals.

Looking at the average populational changes in the treatment and balanced control group, there has been a higher influx of inhabitants to the municipalities in the treatment group over the course of our examined period, both in absolute and relative terms. The supply bases require large amounts of manpower and supporting businesses, so the fact that this has led to an increased populational growth is to be expected. As our hypothesis assumes, that oil rigs and supply bases require a large amount of low educated workers, one would expect a large part of the influx to consist of people with lower educational levels. However, knowing that the effects aren't statistically significant for men, being the largest part of the directly oil related workforce, the increase of oil workers moving in to the municipalities does not seem to be the complete answer. In addition, we know that women experience the largest increase in educational attainment when oil production decreases. Could this mean that women with lower educational levels move out, during such a decline in production? If indeed the municipalities within the treatment group have a more traditional view on marriage than the control group, this could be a part of the explanation. If less men were employed by the oil rigs due to a decrease in production, this could force them to move out of the municipality to find work. If these low educated men are also married to women of the same educational level, something which studies have shown often is the case (Buss, 1985), this would lead to a decrease in lower educational levels for women. A decrease in lower educational levels would conversely increase the share of higher educational levels. However, this implies that the amount of men with lower educational attainment would also be decreased, an effect that has not been shown as statistically significant.

[10] We are assuming that individuals that change their permanent place of residence are also changing their registered address, as per the Norwegian law regarding national registry (Folkeregisterloven, § 6-1, 2016).

Without knowing more about the actual in- and outflow of individuals, looking at the growth of registered inhabitants alone is not sufficient to answer whether oil production has caused more or less educated individuals to move in our out. Using such data would provide more nuance to the results presented in this thesis and would be an interesting basis for future research.

To better understand how oil production affects the educational attainment of a municipality, it would also be interesting to combine our results with data on employment. This would allow us to see how the working sector changes with the changes in oil production and subsequent changes in the educational attainment. If the share of women with higher education decreases in response to an increase in oil production, what channels are affecting this change? Do the women choose jobs in supporting industries of oil rigs, as these experience more demand because of the increased oil activity? Do they choose to not work and stay at home? Do they simply move away from the municipality? These questions could be answered with data on employment and would be an interesting topic for further research.

This master thesis has focused on bases and nearby municipalities overall, at an aggregate level. It is, however, natural that there would exist differences between these bases. There can be differences in climate, demography, economy and culture, to name a few. At an aggregate level these differences will be diluted. Ratledge and Zachary (2017), look at differences between states in the US experiencing booms in the production of petroleum. They find that there is a clear heterogeneity between how these states react to their connected oil booms and warn against generalizing broadly for geographically scattered regions. Although US states and Norwegian municipalities aren’t perfect equivalents, US states on average being many times larger in both size and population, it still suggests that different communities experience oil shocks differently. Looking more closely at the effects each individual supply base has on the local economy could reveal heterogeneity that would be interesting to follow up.

7.1 Impact

In spite of our limited ability to assert educational reductions on a country wide level, we will still attempt to illustrate what type of changes our estimates imply, within our treatment municipalities. We will here focus on the results that have consistently garnered statistical significance, meaning we will use the long-run effects established for Period 2, between 1998 and 2016. We will also only use results aggregated for both educational levels, meaning the long-run effects for genders and their combined estimates.

We now have three different estimates on long-run propensities (LRP): one derived from the main analysis in Chapter 5 and two from different robustness checks in Chapter 6. Since these estimates are relatively close in range, we will use an average of the three for the sake of concision. The previous estimates and their significance levels, their collective mean and standard deviations are shown in Table 12. We can't claim any significance levels for these mean long-run effects. The significance levels of the estimates they are based on, still imply that the mean effects on genders combined and women are statistically significant, while the estimates for men are not.

Table 12. All estimated long-run effects from the main model and their mean

	(1) LRP	(2) LRP	(3) LRP	(4) Mean LRP	(5) Standard deviation
Genders combined	-0.3772**	-0.3604**	-0.3097***	-0.3491	0.0351
Women	-0.5342**	-0.5047**	-0.4786***	-0.5058	0.0278
Men	-0.2039	-0.1946	-0.2479	-0.2155	0.0285

Note. Long-run propensities collected from previous chapters, in addition to their collective mean and standard deviation. Column (1), (2) and (3) are collected from Tables 6 and 7, Table A7 and Table A9, respectively. The significant levels from the previous estimations are also shown by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

From the LRP we get the long-term effect of an increase of 1% of oil production. To interpret the effect in terms of actual reduction in share of educational attainment and individuals affected, we need to know how much the production of oil has increased in percent in the relevant period. We can see how oil production has developed over the years in Figure 2 in Chapter 3. Although there has been a huge increase since the 80s, the changes are not as large in Period 2. The increase in oil production from 1998 to 2016 was 0.137%, making it almost unchanged looking at the two periods in isolation. A change this small will naturally lead to very small changes in the educational attainment implied by our model. There has, however, been both an increase and a decrease in the period, production reaching its all-time peak in 2004. To extract more meaningful and tangible results, we will also look at how the educational attainment was affected before and after the peak. We do this by using the long-run propensities for Period 2, for the relative change from 1998 to 2004, then from 2004 to 2016, and lastly the total change for the entire period, 1998 to 2016. This can be seen in Table 13 below.

Our estimates imply that during the time leading up to the peak of oil production, the municipalities affected reduced their number of individuals with higher education by 334. This is relative to the municipalities in the control group. Almost 75% of these were women.[11]

Table 13. Effects of oil production on share of educational attainment and educated individuals

	Panel A. Estimates and implied effect on share of educational attainment								
	1998-2004			2004-2016			1998-2016		
	(1) Men	(2) Women	(3) Total	(4) Men	(5) Women	(6) Total	(7) Men	(8) Women	(9) Total
LRP	-0.2155	-0.5058	-0.3491	-0.2155	-0.5058	-0.3491	-0.2155	-0.5058	-0.3491
Production change	17.78%	17.78%	17.78%	-14.98%	-14.98%	-14.98%	0.137%	0.137%	0.137%
Attainment change (pp)	-0.038	-0.090	-0.062	0.032	0.076	0.052	0.000	-0.001	0.000
	Panel B. Population in treatment municipalities and the implied effects on educated individuals								
Population	266 057	271 605	537 662	333 213	324 189	657 402	333 213	324 189	657 402
Educated individuals	65 264	71 897	137 161	101 030	121 741	222 771	101 030	121 741	222 771
Change in educated individuals	-102	-244	-334	108	246	344	-1	-2	-3

Note. The table shows the effects of the average long-run-propensities (LRP) on educational attainment. Panel A shows the effects in terms of the share of attainment, given in percentage points (pp). The change in share of attainment is calculated by multiplying the LRP with the oil production's percentage change. Panel B shows how this affects the actual individuals in the treatment municipalities. The columns are divided by periods, Columns (1)-(3) leading up to the peak of production, Columns (4)-(6) succeeding it, and Columns (7)-(9) encompassing the entire period.

The inverse effects can be seen for the period following the peak of oil production, where our estimates imply that the treated municipalities increased their number of educated individuals, in comparison to the control group. We note here that the estimates, being derived from the relative change in oil production, do not show the number of educated individuals gained or lost, compared to a reality without oil production. Rather, it shows how the educational population changes, compared to situation where the level of production was held constant. This concept is somewhat abstract, so to make it more tangible, we have tried to illustrate the effects after the peak of oil production in Table 14.

We can see that the educational attainment in the treatment group was 33.89% in 2016, after oil production experienced a 14.98% decrease. What our estimates imply, is the number of individuals that would have kept their current, lower educational level (i.e. not increased their educational level), if the level of production had not decreased (i.e. stayed constant). We can see from the table that in the absence of a decreased oil production, 344 more individuals would

[11] We note that the sum of the columns for men and women in Table 13, showing individual's educational change, in theory should equal the total. Due to this change being the product of both very small and very large numbers, rounding has produced a slight difference between the total and sum of genders.

have obtained a lower educational level, compared to the case of a decrease. This is also shown in the share of educational attainment reported at 33.83%, 0.05 pp lower than the actual case.

Table 14. Change in educational attainment, with or without a change in oil production, in the period 2004 – 2016

	Actual change in oil prod.		Change in educational		Constant oil prod.	
	Population	Educational attainment	Attainment (pp)	Individuals	Population	Educational attainment
Lower education	434 631				434 975	
Higher education	222 771	33.89 %	0.052	344	222 427	33.83 %
Total	657 402				657 402	

Note. The table shows the effects of the change of oil production and the outcome if the level of production was held constant. The oil percent change is from the period between 2004 and 2016, the period following the peak of oil production. The populational numbers are based on 2016 data.

Oil and oil production is today an important part of the Norwegian economy, but being a finite, non-renewable, natural resource, its production will eventually come to a halt. Our estimates give indications as to how this will affect the municipalities in the areas around the supply bases in the future. If, for instance, oil production experienced a reduction of 90%, this would on average increase the share of educational attainment in the surrounding municipalities by 0.31 pp. Using the populational numbers from 2016, this would imply an increase of over 2000 educated individuals.

8 Conclusion

The aim of this master's thesis has been to uncover if communities affected by oil activity have suffered a reduction in their population's higher educational attainment, compared to other communities. We estimate this through a fixed effects estimation, where we look at the direct, per-unit effect oil activity has on the share of educational attainment. In this model we estimate both the short- and long-run effects. We follow this up with a difference-in-difference roll-out method, where we look at the permanent effects surrounding the first introduction of oil activity, in form of the first produced unit and the first Norwegian krone invested. We also use an event study specification to examine the model's assumptions, at the same time allowing us to observe the yearly effects surrounding the introduction of oil activity.

Our findings indicate oil production has a negative relationship with educational attainment in oil-related municipalities, between 1998 and 2016. For every 10% increase in oil production, the share of educational attainment decreases with up to 0.038 percentage points (pp). From 1998 to 2004, our estimates suggest that 334 less individuals had higher educational attainment in the municipalities effected by oil, compared to municipalities outside of the range of oil activity. Most of the power behind these results stem from the effects oil production has on the educational attainment of women. Our estimates show that women experience effects between 159%-486% larger than their male counterparts. We also find that the effects are heterogenous to how they influence the levels of higher educational attainment, affecting undergraduate attainment more than twice as much as postgraduate attainment on average. These results have proven to be robust, both for multiple treatment groups and control groups. This contrasts with our attempts to estimate how the introduction of oil activity influences educational attainment, through a difference-in-difference estimation, as these results varied widely when exposed to different control groups.

In conclusion, our findings can confirm our initial research question, that oil production has a negative relationship with educational attainment. Although the effects are admittedly not of a huge magnitude, they could still help municipalities under the influence of oil activity to anticipate how their educational population might change, when facing a change in the production of oil. Finally, as our results indicate that oil-related municipalities will increase their share of educated inhabitants as oil production decreases, this thesis can provide a more positive silver lining to the otherwise grim forecasts surrounding the decline of the Norwegian oil riches.

References

Angrist, J. D. (2003). Randomized trials and quasi-experiments in education research. *NBER Reporter Online*, (Summer 2003), 11-14.

Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Barro, R. J. (1991). Economic growth in a cross section of countries. *The quarterly journal of economics*, *106*(2), 407-443.

Becker, G.S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Columbia University Press, New York

Benhabib, J., & Spiegel, M. M. (1994). The role of human capital in economic development evidence from aggregate cross-country data. *Journal of Monetary economics*, *34*(2), 143-173.

Bhuller, M.S. (2009). Inndeling av Norge i arbeidsmarkedsregioner. Statistics Norway. Notes 2009:24.

Black, D. A., McKinnish, T. G., & Sanders, S. G. (2005). Tight labor markets and the demand for education: Evidence from the coal boom and bust. *ILR Review*, *59*(1), 3-16.

Brewer, M. B., & Crano, W. D. (2000). Research design and issues of validity. *Handbook of research methods in social and personality psychology*, 3-16.

Buss, D. M. (1985). Human mate selection: Opposites are sometimes said to attract, but in fact we are likely to marry someone who is similar to us in almost every variable. *American scientist*, *73*(1), 47-51.

Butikofer, A., Løken, K. V., & Salvanes, K. G. (2016). *Infant Health Care and Long-Term Outcomes* (CEPR Discussion Paper No. DP11652). Retrieved from the Centre for Economic Policy Research.

Cascio, Elizabeth U., and Ayushi Narayan. 2015. “Who Needs a Fracking Education? The Educational Response to Low-Skill Biased Technological Change.” NBER Working Paper

No. 21359. Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w21359.

Card, D. (1999). The causal effect of education on earnings. In *Handbook of labor economics*: *Vol. 3*, (pp. 1801-1863). Elsevier.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, *22*(1), 31-72.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, *50*(2), 317-372.

Database of Global Administrative Areas (2018). *Shapefile for Norway*. Retrieved from https://gadm.org/download_country_v3.html

Ekberg, E. (2018). Statens lånekasse for utdanning. In *Store Norske Leksikon*. Retrieved June 11, 2018 from https://snl.no/Statens_l%C3%A5nekasse_for_utdanning

Ekeland, A. (2017). Sysselsatte i petroleumsnæringene og relaterte næringer 2016. *Statistics Norway, Reports*, *2017*(27).

Folkegisterloven. Law December 19, 2016, nr. 88 of national registry. (2016). Retrieved from https://lovdata.no/dokument/NL/lov/2016-12-09-88#KAPITTEL_7

Global Education Network Europe. *Global Education in Norway*. Retrieved from https://gene.eu/wp-content/uploads/Gene_NationalReport-Norway.pdf

Gylfason, T. (2001). Natural resources, education, and economic development. *European economic review*, *45*(4-6), 847-859.

Hays. (2013). Oil & Gas Global Salary Guide 2013. Retrieved from http://www.hays.com/cs/groups/hays_common/@og/@content/documents/promotionalcontent/hays_1269348.pdf

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, *47*(1), 153-161.

Hustoft, A. G., Hartvedt, H., Nymoen, E., Stålnacke, M., & Utne, H. (1999). Standard for økonomiske regioner. *Statistics Norway, Report*, *99*(6).

Jacobson, L. S., LaLonde, R. J., & Sullivan, D. G. (1993). Earnings losses of displaced workers. *The American economic review*, 685-709.

Kartverket (2016). *Arealstatistikk for Norge 2016*. Retrieved from https://kartverket.no/

Ministry of Petroleum and Energy (2016, February 29). Norsk oljehistorie på 5 minutter. Retrieved from https://www.regjeringen.no/no/tema/energi/olje-og-gass/norsk-oljehistorie-pa-5-minutter/id440538/

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, 1417-1426.

Norwegian Petroleum (2018, May 16) Norsk petroleumshistorie. Retrieved from https://www.norskpetroleum.no/rammeverk/rammevilkarpetroleumshistorie/

Norwegian Petroleum. (n.d.) *Conversion.* Retrieved April 26, 2018 from https://www.norskpetroleum.no/en/calculator/about-energy-calculator/

Olaniyan, D. A., & Okemakinde, T. (2008). Human capital theory: Implications for educational development. Pakistan Journal of Social Sciences, *5*(5), 479-483.

Raaum, O. (1999). Inntektseffekter av utdanning i Norge: en litteraturoversikt. *Arbeidsnotat*, *2*, 1999.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55.

Rosenbaum, P. R. (2005). Observational study. *Encyclopedia of Statistics in Behavioral Science*, *3,* 1451-1462.

Løken, K. V. (2010). Family income and children's education: Using the Norwegian oil boom as a natural experiment. *Labour Economics*, *17*(1), 118-129.

Marchand, J., & Weber, J. (2015). The labor market and school finance effects of the Texas shale boom on teacher quality and student achievement. *University of Alberta, Working Paper*, *15*.

Ratledge, N., & Zachary, L. (2017). *Impacts of Unconventional Oil and Gas Booms on Public Education*. Retrieved from http://www.rff.org/files/document/file/RFF-Rpt-ShaleReviews_Education.pdf

Rickman, D. S., Wang, H., & Winters, J. V. (2017). Is shale development drilling holes in the human capital pipeline? *Energy Economics*, *62*, 283-290.

Solberg, S. (2013, February 15). Lønnsandel i Norge – og i verden. *Nettavisen24*. Retrieved from https://www.nettavisen.no/

Statistics Norway. (2017a). *Labour force survey*. Retrieved June 11, 2018 from https://www.ssb.no/statbank/table/08338/?rxid=824e2d8b-f425-4bee-bf8a-d533daf3de4e

Statistics Norway. (2017b). *Educational attainment of the population*. Retrieved April 26, 2018 from https://www.ssb.no/en/statbank/table/09429/?rxid=a56896e5-834c-4cdc-9d1b-ed0115464eaf

Statistics Norway. (2018). *Earnings of all employees*. Retrieved May 5, 2018 from https://www.ssb.no/en/arbeid-og-lonn/statistikker/lonnansatt/aar

Thune, T., Reisegg, Ø. & Askheim, S. (2015). Skole og utdanning i Norge. In *Store Norske Leksikon*. Retrieved April 27, 2018 from https://snl.no/Skole_og_utdanning_i_Norge

Wooldridge, J. M. (2015). *Introductory Econometrics: A Modern Approach*. Canada: Cengage Learning.

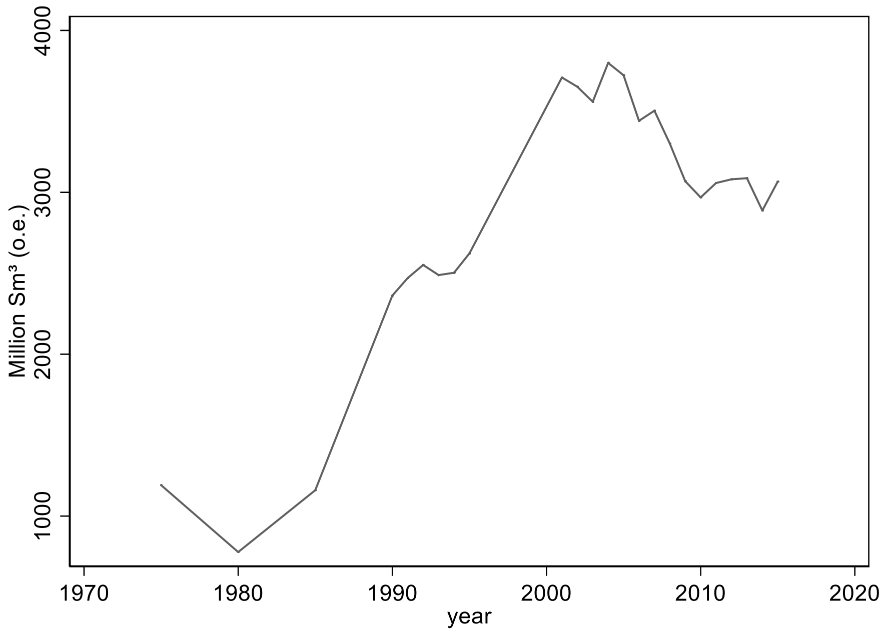
Appendix

List of Supplemental Figures

List of Supplemental tables

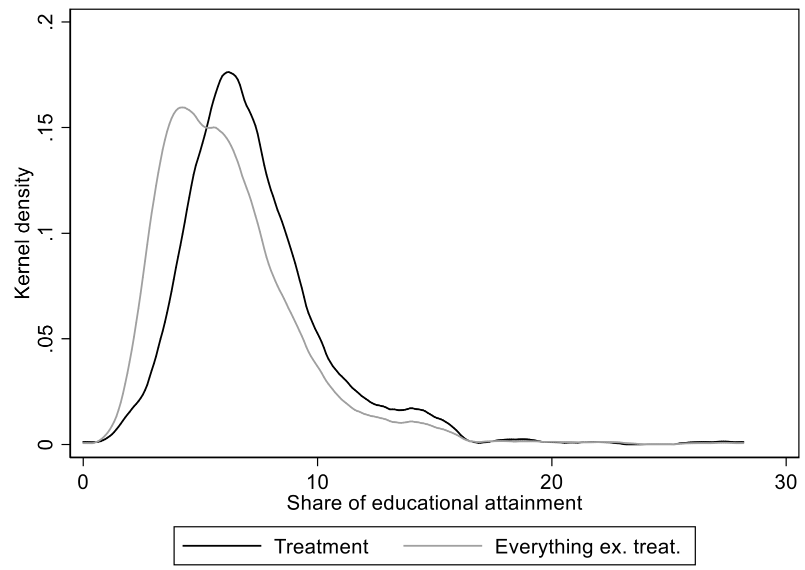
Supplemental figures

Figure A1. Norway's total oil reserves



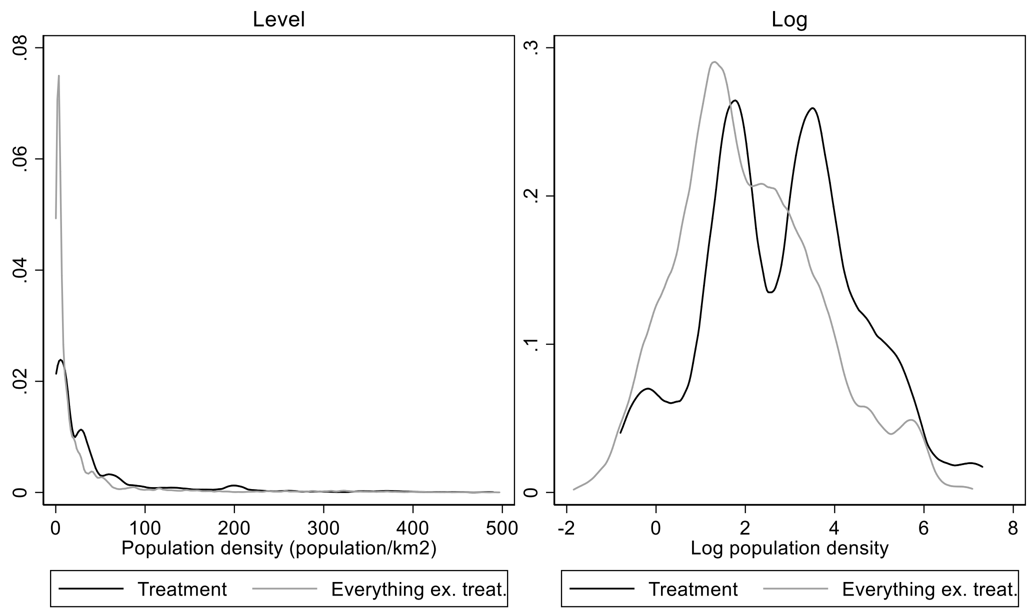
Note. The amount of oil reserves in Norway, given in MNOK. There are several missing years of data. Years for which data are included are: 1975, 80, 85, 90-95 and 00-15.

Figure A2. Educational attainment between treatment municipalities and the remaining municipalities in 1970 and 1980



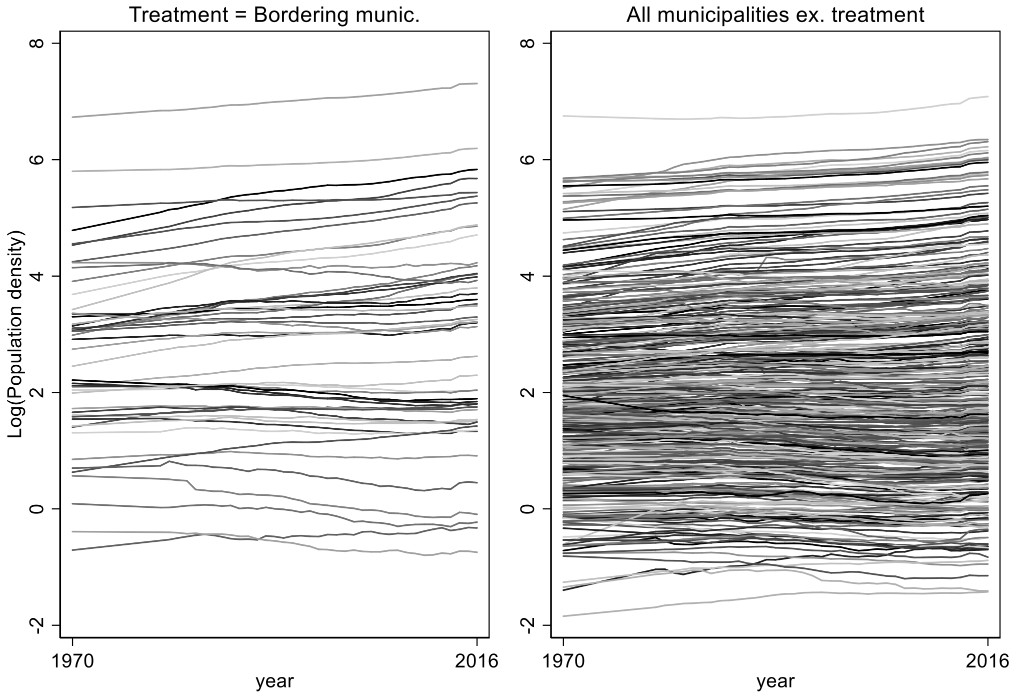
Note. Epanechnikov kernel density plot for the tertiary educational attainment of the treatment group and all other municipalities, in 1970 and 1980. Their mean difference is not considered statistically different from zero, with a p-value of 0.1567.

Figure A3. Population density between 1980 and 2016, comparing treatment and remaining municipalities



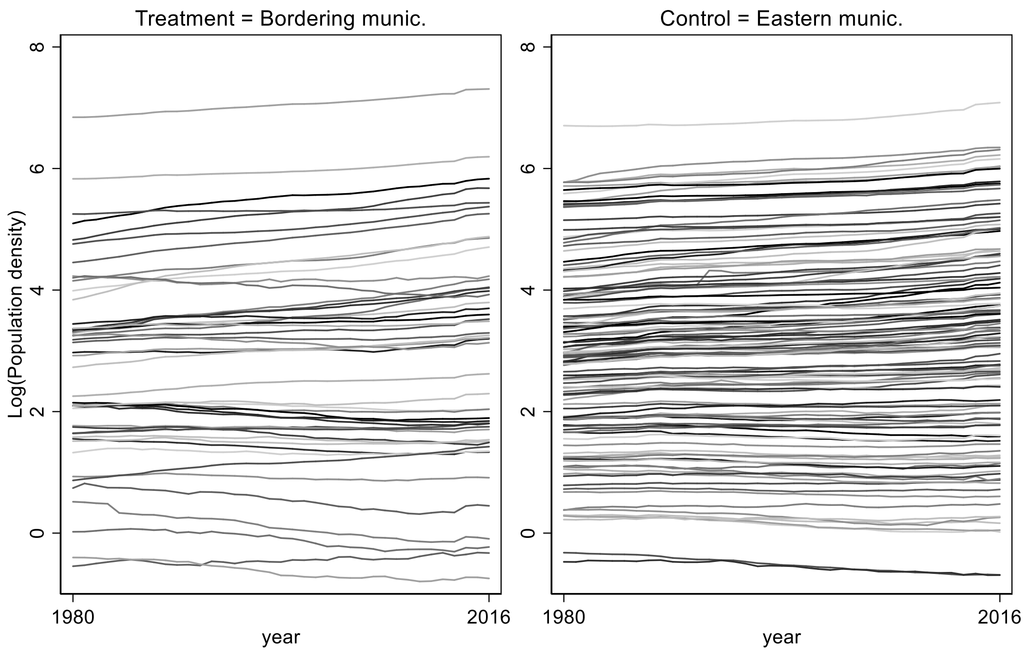
Note. Shows the level and log of the population density for both the treatment group and every other municipality, using an Epanechnikov kernel density plot. All points range from 1980 to 2016. For the level of population density, the top 1% of municipalities were removed, for presentational purposes.

Figure A4. Evolution of population density over time, for treatment and remaining municipalities



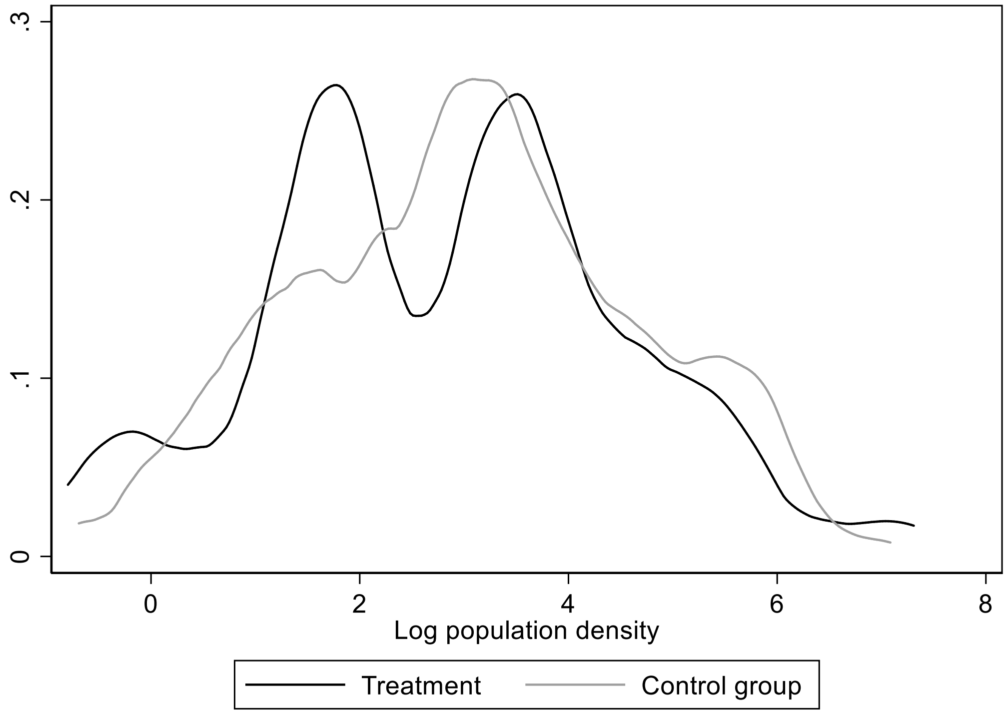
Note. Showing how the log of population density has evolved over time, separated by treatment municipalities and untreated municipalities.

Figure A5. Population density per municipality, for treatment and control



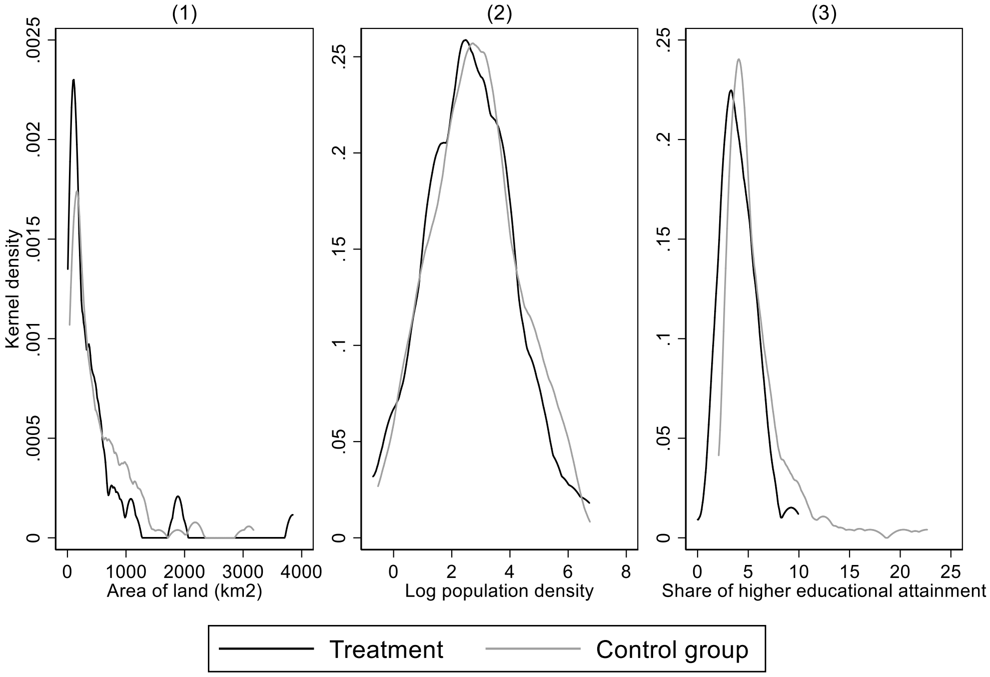
Notes. Figuring showing the development of population densities for the municipalities of the treatment and control group. The population densities are used in log-form. Each line represents a municipality.

Figure A6. Balancing test for population density between 1970 and 2016, comparing between treatment and control group



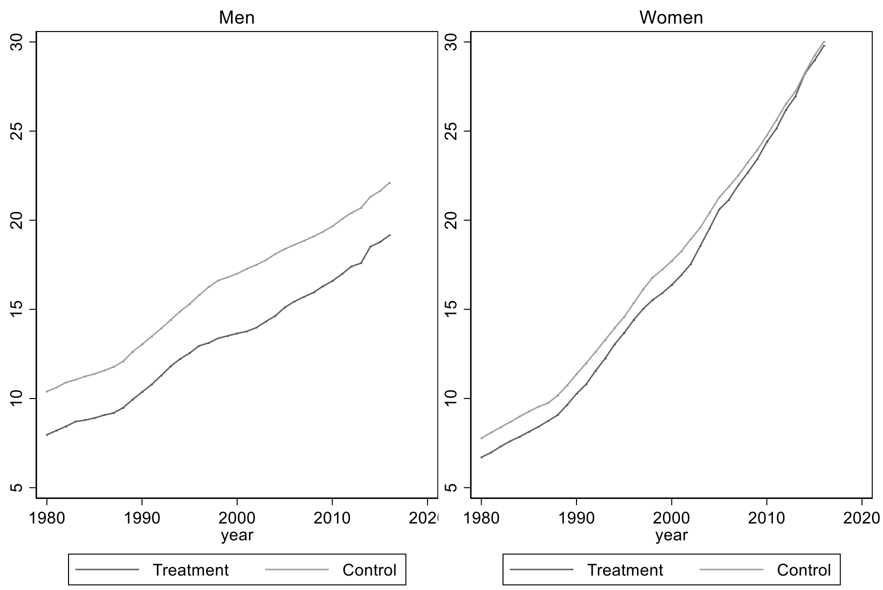
Note. Shows the log of the population density for the treatment group and balanced control group, using an Epanechnikov kernel density plot. All points range from 1980 to 2016, in addition to a year of observations from 1970. Using two-sampled t-tests to check for differences between the groups reports that they are statistically different from each other, with a p-value of 0.0016.

Figure A7. Balancing tests between the treatment and control group in 1970, for area of land, log population density and educational attainment



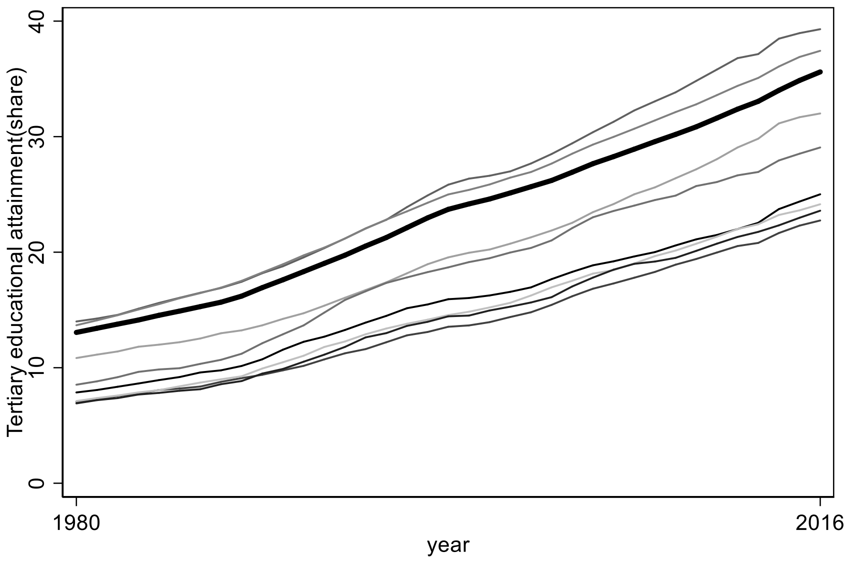
Note. Shows an Epanechnikov kernel density plots for area of land, log population density and share of higher educational attainment in 1970, in graph (1), (2) and (3) respectively. The figure shows how the treatment and control group are balanced along these three factors. Using two-sampled t-tests to check for differences between the groups, (1) and (2) are not statistically different from each other, returning p-values of 0.379 and 0.497. The educational share of attainment for the treatment group is statistically different from the control group, returning p-values of 0.0016.

Figure A8. How the share educational attainment differs between control and treatment group



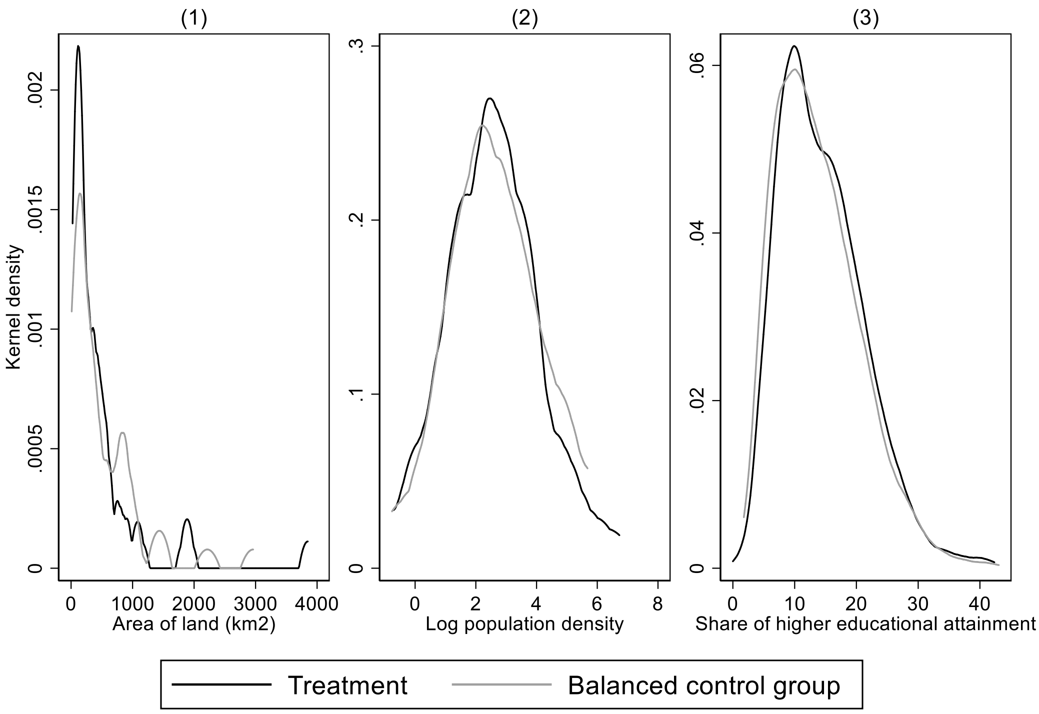
Note. Showing the average tertiary educational attainment over time by gender between the treatment group and control group.

Figure A9. How the share educational attainment differs between control and treatment group



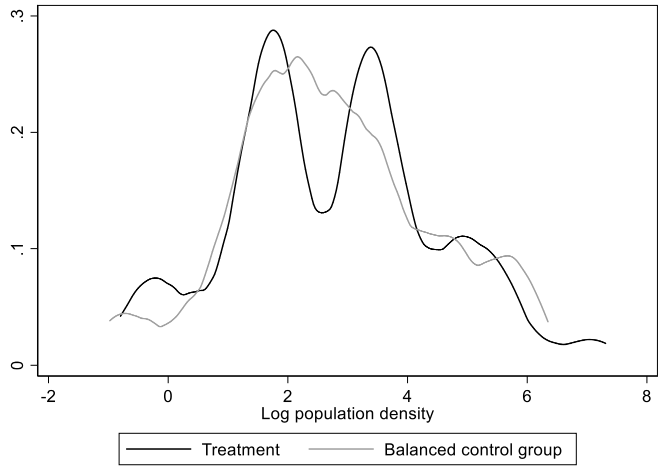
Note. Shows the aggregated means of tertiary educational attainment for the different supply bases within the treatment group and the control group. The control group is shown with a black line.

Figure A10. Repeated balancing tests between the treatment and new balanced control group in 1970, for area of land, log population density and educational attainment



Note. Shows an Epanechnikov kernel density plots for area of land, log population density and share of higher educational attainment in 1970, in graph (1), (2) and (3) respectively. The figure shows how the treatment and control group are balanced along these three factors. Using two-sampled t-tests to check for differences between the groups, we cannot reject the null hypothesis that the average means are different from zero for any of the three factors, the p-values ranging from 0.74 to 0.91.

Figure A11. Repeated balancing test for population density between 1970 and 2016, comparing between treatment and new balanced control group

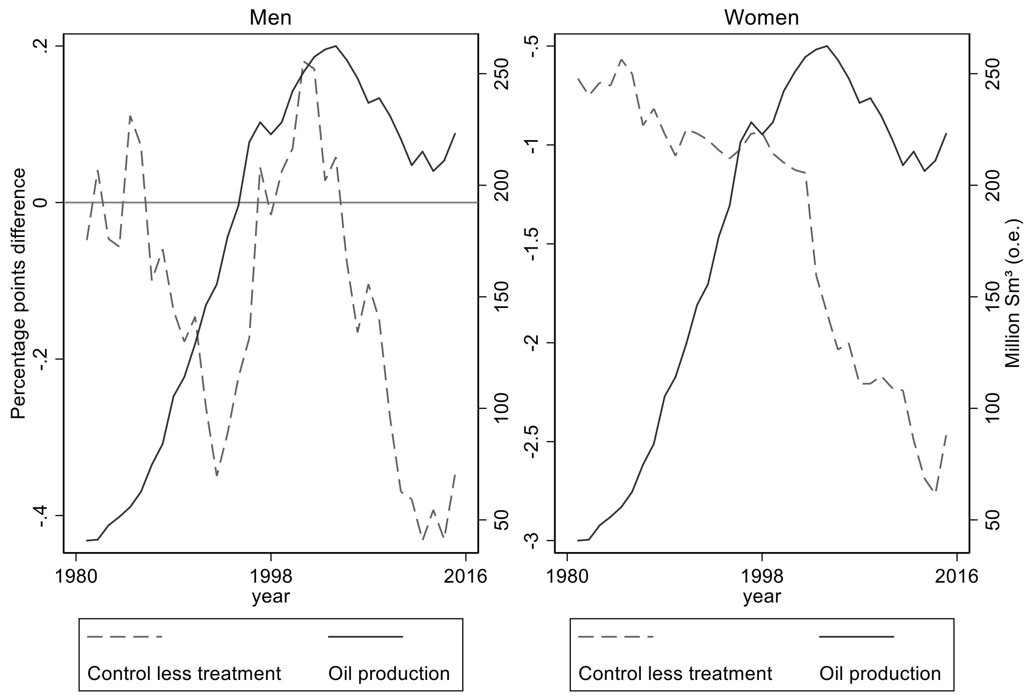


Note. Shows the log of the population density for the treatment group and balanced control group, using an Epanechnikov kernel density plot. All points range from 1980 to 2016, in addition to a year of observations from 1970. Using two-sampled t-tests to check for differences between the groups reports that they are not statistically different from each other, with a p-value of 0.669.

Figure A12. Map of treatment and balanced control municipalities

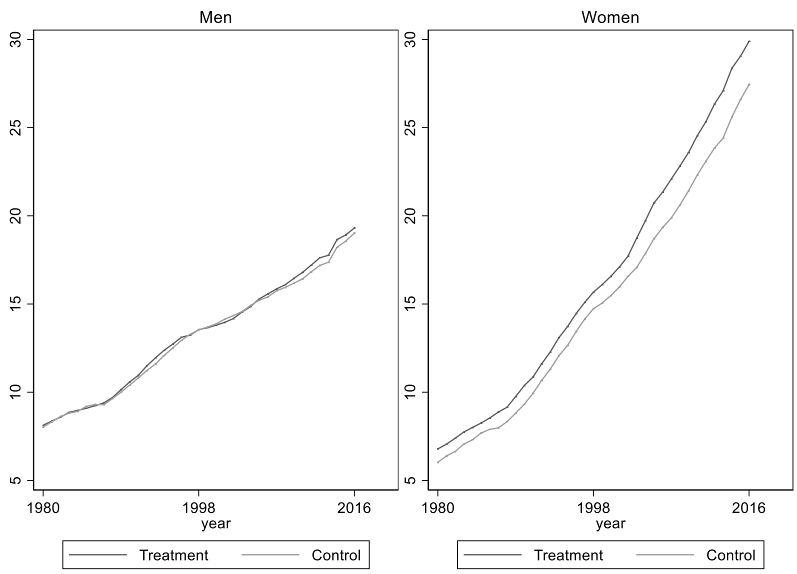
Note. A map of Norway showing the locations of municipalities within the treatment group and the balanced control group. Source: Database of Global Administrative Areas (2018).

Figure A13. Gap in educational attainment and oil production by gender, using a balanced control group



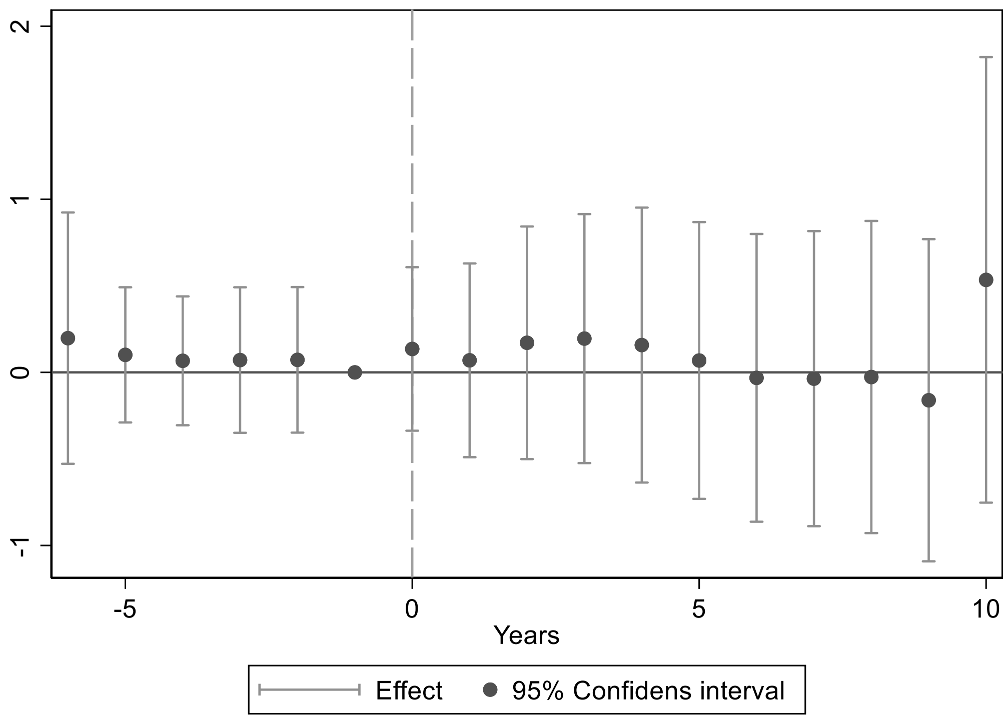
Note. The dashed line plots the difference in higher educational attainment for the control and treatment group over time. The difference is calculated by taking the higher educational attainment of the control less the treatment, a negative number therefore indicating the treatment group having a higher educational attainment. The solid line shows the total oil production.

Figure A14. Higher educational attainment for the treatment group and balanced control group, divided by gender



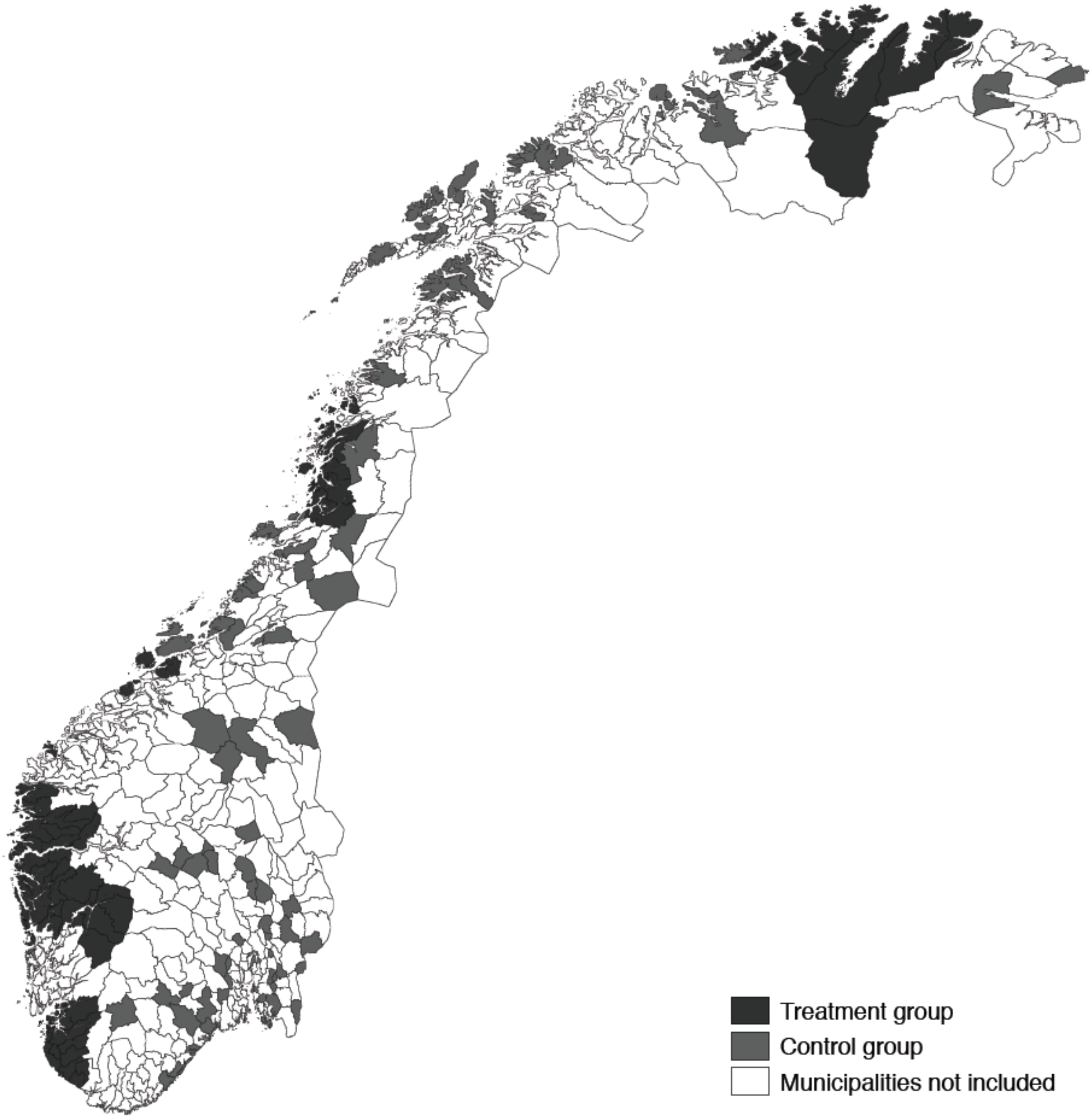
Note. Mapping average aggregate tertiary educational attainment by gender between the treatment group and the balanced control group.

Figure A15. Event study estimates for the start of oil investment's effect on educational attainment, using a balanced control group



Note. Graph plotting changes in education in relation to the year of investment start, in addition to their 95% confidence intervals. The changes in education are the average differences across the treatment municipalities, compared to municipalities not yet experiencing oil activity and municipalities never experiencing oil activity, the control group. The control group is placed in the normalized year prior to investment start and set to zero, to serve as the baseline for the changes in education.

Figure A16. Map of work region treatment and accompanying balanced control



Note. A map of Norway showing the locations of municipalities within the work region treatment group and the accompanying balanced control group. Source: Database of Global Administrative Areas (2018).

Supplemental tables

Table A1. The municipalities included in the treatment group, by base

Dusavik	
1103	*Stavanger*
1127	Randaberg
1130	Strand
1141	Finnøy
1142	Rennesøy
1144	Kvitsøy

Tananger	
1102	Sandnes
1120	Klepp
1124	*Sola*

Sotra	
1201	Bergen
1243	Os
1244	Austevoll
1245	Sund
1246	*Fjell*
1247	Askøy
1259	Øygarden

Mongstad	
1252	Modalen
1253	Osterøy
1256	Meland
1260	Radøy
1263	*Lindås*
1264	Austrheim
1265	Fedje
1266	Masfjorden

Florø	
1401	*Flora*
1428	Askvoll
1433	Naustdal
1438	Bremanger
1445	Gloppen

Kristiansund	
1505	*Kristiansund*
1554	Averøy
1557	Gjemnes
1560	Tingvoll
1571	Halsa

Sandnessjøen	
1815	*Vega*
1816	Vevelstad
1818	Herøy
1820	Alstahaug
1822	Leirfjord
1824	Vefsn
1827	Dønna

Hammerfest	
2004	*Hammerfest*
2012	Alta
2015	Hasvik
2017	Kvalsund
2018	Måsøy

Note. List of municipalities included in the treatment group, divided by supply base. Each supply base is shown in bold, while each base-municipality is shown in italics. Both the municipal numbers and municipal names are included.

Table A2. Comparing treatment group to remaining municipalities

Variable	Treatment group	Everything excluding treatment	St. error	p-value
Mean area of land	443.50 km2	759.99 km2	136.73	(0.010)
Mean population density (1980-2016)	70.47 per km2	32.01 per km2	2.41	(0.000)

Note. Comparing the mean of area of land and population density for the municipalities in the treatment group and the remaining municipalities. The standard errors and p-values of a two-sample t-test for the two means being statistically different from zero are shown on the right-hand side. The p-values indicate that they are statistically different from each other.

Table A3. The municipalities included in the control group, by county

Østfold		**Oslo**		**Buskerud**	
101	Halden	301	Oslo	602	Drammen
104	Moss	**Hedmark**		604	Kongsberg
105	Sarpsborg	402	Kongsvinger	605	Ringerike
106	Fredrikstad	403	Hamar	612	Hole
111	Hvaler	412	Ringsaker	615	Flå
118	Aremark	415	Løten	621	Sigdal
119	Marker	417	Stange	622	Krødsherad
121	Rømskog	418	Nord-Odal	623	Modum
122	Trøgstad	419	Sør-Odal	624	Øvre Eiker
123	Spydeberg	420	Eidskog	625	Nedre Eiker
124	Askim	423	Grue	626	Lier
125	Eidsberg	425	Åsnes	627	Røyken
127	Skiptvet	426	Våler	628	Hurum
128	Rakkestad	427	Elverum	631	Flesberg
135	Råde	428	Trysil	632	Rollag
136	Rygge	429	Åmot	**Vestfold**	
137	Våler	430	Stor-Elvdal	701	Horten
138	Hobøl	432	Rendalen	702	Holmestrand
Akershus		434	Engerdal	704	Tønsberg
211	Vestby	436	Tolga	706	Sandefjord
213	Ski	437	Tynset	709	Larvik
214	Ås	438	Alvdal	711	Svelvik
215	Frogn	439	Folldal	713	Sande
216	Nesodden	441	Os	714	Hof
217	Oppegård	**Oppland**		716	Re
219	Bærum	501	Lillehammer	719	Andebu
220	Asker	502	Gjøvik	720	Stokke
221	Aurskog-Høland	519	Sør-Fron	722	Nøtterøy
226	Sørum	520	Ringebu	723	Tjøme
227	Fet	521	Øyer	728	Lardal
228	Rælingen	522	Gausdal	**Telemark**	
229	Enebakk	528	Østre Toten	805	Porsgrunn
230	Lørenskog	529	Vestre Toten	806	Skien
231	Skedsmo	532	Jevnaker	807	Notodden
233	Nittedal	533	Lunner	811	Siljan
234	Gjerdrum	534	Gran	814	Bamble
235	Ullensaker	536	Søndre Land	815	Kragerø
236	Nes	538	Nordre Land	817	Drangedal
237	Eidsvoll	540	Sør-Aurdal	819	Nome
238	Nannestad	541	Etnedal	821	Bø
239	Hurdal	542	Nord-Aurdal	822	Sauherad
				830	Nissedal

Note. List of municipalities included in the control group, divided by county. Counties are shown in bold.

Table A4. Comparing educational attainment between treatment and control

Group	Main supply base	No. of Munic.	Mean higher educational attainment 1980-1997	1998-2016	1980-2016	2016
1	Dusavik	6	12 %	22 %	17 %	28 %
2	Tananger	3	13 %	24 %	19 %	30 %
3	Sotra	7	11 %	21 %	16 %	27 %
4	Mongstad	8	9 %	17 %	13 %	22 %
5	Florø	5	11 %	19 %	15 %	24 %
6	Kristiansund	7	9 %	17 %	13 %	22 %
7	Sandnessjøen	7	8 %	16 %	12 %	22 %
8	Hammerfest	5	10 %	18 %	14 %	23 %
Treatment group mean		48	10.1 %	18.9 %	14.6 %	24.3 %
Control group mean		119	12.0 %	20.9 %	16.6 %	26.0 %
Difference			-1.8 pp	-2.0 pp	-1.9 pp	-1.7 pp

Note. Tertiary educational attainment given in shares for different periods and years. The averages for each group within the treatment are measured per municipality, and not aggregated. This means that each municipality is weighted equally, in contrast to the aggregate approach, where the size of municipalities will affect the mean. The differences are given in percentage point (pp) change, the difference in absolute terms.

Table A5. Comparing educational attainment between treatment and control, by gender

Group	Main supply base	No. Of munic	Mean higher educational attainment 1980-1997		1998-2016		2016	
			Men	Women	Men	Women	Men	Women
1	Dusavik	6	11.72 %	11.70 %	18.69 %	24.88 %	22.82 %	33.95 %
2	Tananger	3	14.55 %	11.97 %	22.09 %	25.93 %	26.55 %	34.63 %
3	Sotra	7	12.46 %	10.55 %	18.72 %	23.31 %	22.24 %	31.29 %
4	Mongstad	8	9.38 %	9.19 %	14.26 %	20.17 %	17.06 %	27.09 %
5	Florø	5	10.12 %	11.01 %	15.35 %	22.71 %	18.33 %	30.23 %
6	Kristiansund	7	8.71 %	9.26 %	13.97 %	20.51 %	16.87 %	27.26 %
7	Sandnessjøen	7	8.38 %	8.24 %	13.22 %	19.42 %	16.66 %	27.26 %
8	Hammerfest	5	8.71 %	10.54 %	13.84 %	23.19 %	16.89 %	30.73 %
Treatment group mean		48	10.21 %	10.08 %	15.83 %	22.09 %	19.16 %	29.79 %
Control group mean		119	12.82 %	11.14 %	19.01 %	22.81 %	22.10 %	30.01 %
Difference			-2.61 %	-1.06 %	-3.18 %	-0.72 %	-2.94 %	-0.22 %

Note. Tertiary educational attainment for different periods and years, divided by gender. The averages for each group within the treatment are measured per municipality. The differences are given in percentage point (pp) change, the difference in absolute terms.

Table A6. The municipalities included in the balanced control group

No.	Municipality	No.	Municipality	No.	Municipality
104	Halden	702	Holmestrand	1755	Leka
106	Fredrikstad	706	Sandefjord	1813	Brønnøy
111	Hvaler	711	Svelvik	1835	Træna
128	Rakkestad	723	Tjøme	1840	Saltdal
213	Ski	807	Notodden	1856	Røst
217	Oppegård	811	Siljan	1857	Værøy
221	Aurskog-Høland	815	Kragerø	1865	Vågan
519	Sør-Fron	901	Risør	1867	Bø
534	Gran	912	Vegårshei	1870	Sortland - Suortá
536	Søndre Land	926	Lillesand	1928	Torsken
538	Nordre Land	1620	Frøya	1938	Lyngen
602	Drammen	1717	Frosta	1941	Skjervøy
621	Sigdal	1718	Leksvik	2027	Unjárga - Nesseby
623	Modum	1738	Lierne	2028	Båtsfjord
625	Nedre Eiker	1749	Flatanger		
628	Hurum	1750	Vikna		

Note. A list of the included municipalities in the balanced control group created through propensity score matching. Both the municipal number and name of each municipality is listed.

Table A7. Long-run effects (LRP) for different subsets using the propensity matched control group

	Panel A: Long-run effects by period		
	Period 1 (1980-1997)	Period 2 (1998-2016)	Period 1 + 2 (1980-2016)
LRP	-0.0284	-0.3604**	0.0308
	(0.686)	(0.015)	(0.799)

	Panel B: Long-run effects by gender					
	Period 1 (1980-1997)		Period 2 (1998-2016)		Period 1 + 2 (1980-2016)	
	Men	Women	Men	Women	Men	Women
LRP	-0.0263	0.0282	-0.1946	-0.5047**	-0.0127	0.0908
	(0.715)	(0.732)	(0.192)	(0.017)	(0.910)	(0.498)

	Panel C: Long-run effects by tertiary educational level					
	Period 1 (1980-1997)		Period 2 (1998-2016)		Period 1 + 2 (1980-2016)	
	Undergrad.	Postgrad.	Undergrad.	Postgrad.	Undergrad.	Postgrad.
LRP	-0.0319	0.0035	-0.2481**	-0.1123	0.0188	0.0120
	(0.593)	(0.853)	(0.024)	(0.411)	(0.747)	(0.870)

	Panel D1: Long-run effects by tertiary educational level and gender (Undergraduate)					
	Period 1 (1980-1997)		Period 2 (1998-2016)		Period 1 + 2 (1980-2016)	
	Men	Women	Men	Women	Men	Women
LRP	-0.0245	-0.0381	-0.1171	-0.3645*	-0.0297	0.0852
	(0.676)	(0.599)	(0.218)	(0.070)	(0.603)	(0.252)

	Panel D2: Long-run effects by tertiary educational level and gender (Postgraduate)					
	Period 1 (1980-1997)		Period 2 (1998-2016)		Period 1 + 2 (1980-2016)	
	Men	Women	Men	Women	Men	Women
LRP	-0.0018	0.0098	-0.0774	-0.1402	0.0170	0.0056
	(0.948)	(0.610)	(0.526)	(0.423)	(0.805)	(0.943)

Note. Table containing long-run propensities (LRP) for 27 different regressions, using the propensity score matched control group presented in Chapter 6. The exogenous variable used has been the log of oil production. The panels are divided by having different subsets of the dependent variable from Equation (4). Panel A uses higher educational attainment for both genders, for both levels of higher educational attainment. Panel B also shows for the aggregated educational levels but separates between genders. This is the opposite of Panel C, where the educational levels are separated, while the genders are combined. In Panel D1-2, the dependent variable is split for both educational level and gender. Panel D1 shows the undergraduate attainment for both genders, while Panel D2 shows the same for postgraduate attainment. All regressions control for firm fixed effects and time fixed effects. All p-values are derived from robust standard errors clustered at municipal level shown in parentheses below the relevant regression. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A8. Overview of municipalities in the work region treatment group

	Work regions	Main supply base	Base-municipality	No. of municipalities
41	Stavanger	Dusavik; Tananger	Stavanger; Sola	11
44	Bergen	Sotra; Mongstad	Fjell; Lindås	18
51	Sunnfjord	Florø	Flora	12
58	Kristiansund	Kristiansund	Kristiansund	27
64	Ytre Helgeland	Sandnessjøen	Alstahaug	4
82	Hammerfest	Hammerfest	Hammerfest	8
				80

Note. List of the treatment group for work regions, connected supply bases, their base-municipality and number of municipalities within each region. When more than one base and/or base-municipality lie within a region, they are separated by a semicolon.

Table A9. Long-run effects (LRP) for different subsets using the work region treatment group and new balanced control group

Panel A: Long-run effects by period

	Period 1 (1980-1997)	Period 2 (1998-2016)	Period 1 + 2 (1980-2016)
LRP	0.0508	-0.3097***	-0.1013
	(0.625)	(0.008)	(0.253)

Panel B: Long-run effects by gender

	Period 1 (1980-1997)		Period 2 (1998-2016)		Period 1 + 2 (1980-2016)	
	Men	Women	Men	Women	Men	Women
LRP	0.0362	0.0584	-0.0985	-0.4786***	-0.0951	-0.0990
	(0.719)	(0.641)	(0.572)	(0.005)	(0.258)	(0.385)

Panel C: Long-run effects by tertiary educational level

	Period 1 (1980-1997)		Period 2 (1998-2016)		Period 1 + 2 (1980-2016)	
	Undergrad.	Postgrad.	Undergrad.	Postgrad.	Undergrad.	Postgrad.
LRP	0.0451	0.0057	-0.2479**	-0.0618	-0.0276	-0.0737*
	(0.606)	(0.839)	(0.015)	(0.487)	(0.664)	(0.056)

Panel D1: Long-run effects by tertiary level and gender (Undergraduate)

	Period 1 (1980-1997)		Period 2 (1998-2016)		Period 1 + 2 (1980-2016)	
	Men	Women	Men	Women	Men	Women
LRP	0.0430	0.0431	-0.0009	-0.4642**	-0.0209	-0.0273
	(0.606)	(0.701)	(0.994)	(0.014)	(0.750)	(0.766)

Panel D2: Long-run effects by tertiary level and gender (Postgraduate)

	Period 1 (1980-1997)		Period 2 (1998-2016)		Period 1 + 2 (1980-2016)	
	Men	Women	Men	Women	Men	Women
LRP	-0.0068	0.0153	-0.0976	-0.0144	-0.0742**	-0.0717
	(0.848)	(0.599)	(0.223)	(0.897)	(0.050)	(0.116)

Note. Table containing long-run propensities (LRP) for 27 different regressions, using the work region treatment group and its associated propensity score matched control group. The exogenous variable used has been the log of oil production. The panels are divided by having different subsets of the dependent variable from Equation (4). Panel A uses higher educational attainment for both genders, for both levels of higher educational attainment. Panel B also uses the aggregated educational levels but separates between genders. This is the opposite of Panel C, where the educational levels are separated, while the genders are combined. In Panel D1-2, the dependent variable is split for both educational level and gender. Panel D1 shows the undergraduate attainment for both genders, while Panel D2 shows the same for postgraduate attainment. All regressions control for firm fixed effects and time fixed effects. All p-values are derived from robust standard errors clustered at municipal level shown in parentheses below the relevant regression. $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$