UiT The Arctic University of Norway

School of Business and Economics

Residential Segregation, inequality, and mobility.

A study case from Tromsø

Andreas Knutsen

Master’s thesis in Economics, SOK-3901, June 2023

Preface

This master thesis in Economics was made possible with the help of Rune Ytreberg from the newspaper Itromsø who providing the dataset used in my analysis and would thank him for all the good conversation and exchanging of idea prior to the start of this master thesis. I also want to thank my supervisors Stein Østbye and Mikko Antti Molianen for all the help and guidance during my writing of this paper.

Abstract

Income inequality and residential segregation are growing concerns in European capitals, including Norway, with potential negative societal implications. Income inequalities are increasing in Tromsø, while residential segregation regarding income groups seems to decrease. This study aims to investigate the likelihood of individuals relocating to neighborhoods with similar income levels, drawing upon the framework of the agent-based Schelling Model. The research employs a probit model and an outcome model based on Heckman's Two-step model. This study contributes to the understanding of the factors shaping residential choices and their impact on income-based segregation. By incorporating agent-based modeling and econometric techniques, it offers insights into the decision-making processes that influence neighborhood selection. Several models were tested. The model which fits the data the most was model with a similar income of ±30% range, around the same as the Schelling model found to be a critical threshold. The findings indicate a positive relationship between increased income and the likelihood of relocation, as observed in both the probit equation, which examines individuals' decision to move, and the outcome equation, which identifies the variables influencing the choice of a neighborhood with similar income. If income increase significantly enough, will reduce the probability to move to a neighborhood with similar income. Change in income inequality and house prices have a negative relationship with the dependent variables.

Table of content

1 Introduction

Income inequality in the European capitals have increased, where there is segregation of the rich and poor within these cities. While segregation itself may not be problematic, concerns arise when the lower-income segregates of the population become concentrated, as this can have detrimental effects on their future opportunities and access to services. The limited social networks and social capital resulting from such concentration exacerbate the challenged faces by these individuals. (van Ham, et al. 2016) Correlation between income inequality and health and social problems. Studies shows that individuals experience improved health outcome as they climb the social economic ladder. (Rowlingson, 2011)

Income inequality has been associated with detrimental health outcomes. In Norway, higher levels of income inequality have been found to correlate with heightened rates of mental health problems, chronic illnesses, and decreased life expectancy (Dahl, et al. 2006). These health disparities can be attributed to unequal access to healthcare resources and the influence of lifestyle factors that are associated with lower income. Berg and Ostry (2011) argue that high income inequality could lead to social instability, hinder social mobility, undermine long-term development and impact the distribution of human capital. High inequality can lead to unequal access to education and healthcare, limiting opportunities for individuals at the lower end of the income distribution and hinder their potential economic advancement. This would impact overall economic growth of a country.

In Oslo, economic segregation among the poor has increased up to 11 percantage points between 1993 and 2011. (Wessel, 2015) Income inequality in Norway has witnessed an increase over the period spanning from 2011 to 2018, both before and after tax, as measured by Gini index. The income inequality figures indicate an 18 percent rise before tax and 29 percent increase after tax, with some fluctuation and a decline from its peak in 2016. Similarly, Tromsø have also observed an 8 percent rise in the Gini coefficient after tax between 2004 and 2021. Furthermore, the top 10 percent, 1 percent and 0.1 percent of the wealthiest individuals have experienced respective increase of 5 percent, 5 percent and 4 percent in their share of the total gross income in Norway. (SSB,2021) A study by Wessel and Nordvik(2018) conducted in the Oslo region to examine the relationship between mixed neighborhoods, native out-mobility, and the role of parenthood. The findings reveal a higher likelihood of native Norwegians moving out of mixed neighborhoods. This indicates a concern among native Norwegian parents regarding the potential impact of residing in such neighborhoods on their children's educational opportunities and social networks. Additionally, factors such as housing quality and safety perceptions further contribute to the out-mobility of parents. These findings highlight the significant role of parenthood in shaping residential mobility patterns, with parents prioritizing neighborhood characteristics that they believe will positively influence their children's well-being and social integration.

Residential segregation is a widely study topic, with its roots in segregation between the white and Afro-American populations of the US. The composition of neighborhoods based on different criteria can significantly influence individual’s daily lives. In essence, the geographical location where one lives can have substantial implication in their quality of life and prospect, particularly concerning employment opportunities and social networks. (Wilson, 1996)

The neighborhoods individual inhabit can also serve as a potential indicator of their future socioeconomic attainment, as individual residing in the same area often interacts and mutually influence each other, which can shape the values, beliefs, and other factors that contribute to societal impressions of individuals based on where they live. (Logan & Molotch, 1987) The structure of neighborhoods is constructed through the categorization of population based in characteristics such as race or income groups. (Massey, et al. 1991) These categories can give advantages and opportunities upon individuals. Those from affluent backgrounds encounter dissimilar opportunities pertaining to education, future income prospect, occupational choses and numerous other variables compared to individuals from low-income families.

Residential segregation has been extensively studied in the literature, particularly with regard to its association with race and its implications on future opportunities. The theoretical foundations of residential segregation can be traced back to Thomas Schelling and his agent-based model. (Schelling, 1971) Schelling introduced a checkerboard-like setting populated by two distinct types of agents, or inhabitants. These agents represent any binary social division that may influence the distribution of individuals. Examples of such divisions could include racial categories like black and white, as well as other social divisions such as religious affiliations or recreational preferences (e.g., Christians and Muslims, footballers and skiers).

In Schelling's model, each agent possesses a threshold, or a satisfaction parameter that determines their decision to stay or move. Agents will remain in their current location if they are satisfied with the mix of neighbors. However, if one of their neighbors decides to move and the new neighbor that take its place is dissimilar to them, the agent could also choose to move. The Schelling model predicts that when the proportion of agents of a particular type exceeds a critical threshold, the neighborhood will tend to become segregated.

Residential segregation with the distinctive characteristics and statuses of neighborhoods have for a long time been central to the study of residential segregation. Galester and Killen (1995) argued that neighborhoods influence the opportunities of their residents through different mechanisms, including service provision, the influence of residents' attitudes, and endogenous processes that shape individual decision-making, preferences, and perceived opportunities. Residential segregation creates an opportunity structure in which neighborhoods play a role in shaping the lives of their residents. The dynamics of neighborhoods includes their relative status and characteristics compared to others, as well as the dynamic processes occurring within the neighborhoods that are subject to study.

Understanding the maintenance or change of residential segregation and neighborhood inequality over time requires considering selective patterns of movement at both macro and micro levels. The macro level involves examining selective migration patterns, such as the theory of "white flight" in the United States, which explains changes in ethnic composition and residential segregation when white residents move away once the number of other minority groups reaches a critical tipping point (Clark, 1992). This theory not only highlights strong preferences between whites and minorities but also among different minority groups. Other studies shows how poor areas either emerge or further decline due to selective migration patterns, where more affluent residents leave and are replaced by households with similar socioeconomic status to non-movers (Andersson & Bråmå, 2004).

Low attractiveness of neighborhoods can be inferred from selective mobility patterns, where residents move out more frequently compared to other neighborhoods. The literature also suggests that the younger segment of the population tends to be more mobile than the older population, leading to demographic changes in neighborhood composition rather than changes based on socioeconomic characteristics alone (Bailey & Livingstone, 2007). Gentrification can also contribute to selective mobility patterns, as new investments or construction projects can drive up housing prices, displacing low-income groups and attracting more affluent residents.

To understand the preferences and constraints underlying selective migration patterns and residential segregation, it is important to examine the micro-level decision-making processes involved in individual and household mobility. Brown and Moore (1970) argued that the moving process are two distinct decisions: the decision to move and the subsequent choice of destination. At the micro level, the decision to move often stems from dissatisfaction with the current residence, influenced by household preferences and needs. Recent literature and studies have placed greater emphasis on the role of neighborhoods in shaping individuals' moving decisions. For instance, Lu (1998) demonstrates that individuals who are satisfied with their neighborhood environment are less inclined to initiate a relocation. Empirical findings also indicate that the likelihood of moving increases as the socioeconomic status of one's neighborhood decreases, while the likelihood of moving decreases as neighborhood socioeconomic status improves (Feijten & van Ham, 2009).

Drawing upon Schelling's model, which indicate that individuals tend to move to neighborhoods where their neighbors are similar to them, I will make a model that can examine if empirical evidence supports this. Employing Heckman's Two-Step model (Heckman, 1976), I aim to analyses and identifies the significant variables influencing the decision to move and how these variables contribute to the choice of relocating to a neighborhood with similar characteristics. By using this model, I will try to answer this research question: If an individual's current neighborhood is different from their neighbors regarding their income level, would they be more likely to relocate to an area that is more economically similar? As Schelling's model predicts that segregation will occur once the critical threshold, typically around 0.33, is surpassed. Exploring various variables that capture the range of income similarity to the neighborhood mean can provide valuable insights into individuals' preferences and their inclination to differentiate themselves from others. By examining these variables, the model can gain a deeper understanding of the factors that drive residential segregation and the mechanisms behind it.

By using income and wealth before tax and house transaction data, can give an insight and examine the dynamics of income and wealth inequality in Tromsø. The analysis will use well-established measures such as the Gini coefficient and the P90/10 ratio. Initially, the data will be utilized to generate descriptive statistics, elucidating the variations in income across different areas of Tromsø. In this context, postcodes will serve as proxies for distinct geographical regions within the city, which I will referee as neighborhoods. The descriptive analysis will shed light on the evolving income distribution and provide insights into the spatial concentration of relative affluence and relative deprivation.

Additionally, the research aims to assess residential segregation by categorizing the population into different income groups. The Index of Dissimilarity (Duncan & Duncan ,1955) will be employed to measure the extent of spatial segregation and identify any changes that have occurred between the years 2011 and 2021. This analysis will provide valuable insights into the patterns of socio-economic segregation and the potential shifts in residential composition over time.

2 Inequalities and segregation in Tromsø

The introduction revealed an upward trend in inequality and segregation in Norway over the past few decades. However, in order to gain a more comprehensive understanding of this development, it is important to examine regional variations in greater detail. By focusing on a specific case study such as Tromsø, the largest city in the region Troms and Finnmark, we can uncover the divergences that exist across different regions of Norway, moving beyond the aggregated values that encompass the entire country. This localized approach enables us to discern the nuances and specific dynamics that shape inequality and segregation patterns, shedding light on the heterogeneity that exists within Norway's regional landscape.

2.1 Inequality in Tromsø

To address the level of inequality in Tromsø, this study will use the Gini coefficient and the P90/P10 ratio as measurement methods. The Gini coefficient is a widely used statistical tool for income or wealth inequality, both within a specific and across different countries. It provides insights into how income and wealth are distributed among individuals within a given population. The coefficient ranges between 0 and 1, where a value of 0 indicate perfect income equality, where all individuals possess an equal share of the total income. On the other hand, a Gini coefficient of 1 indicates extreme income inequality, where a single individual holds the entirety of the wealth within the population.

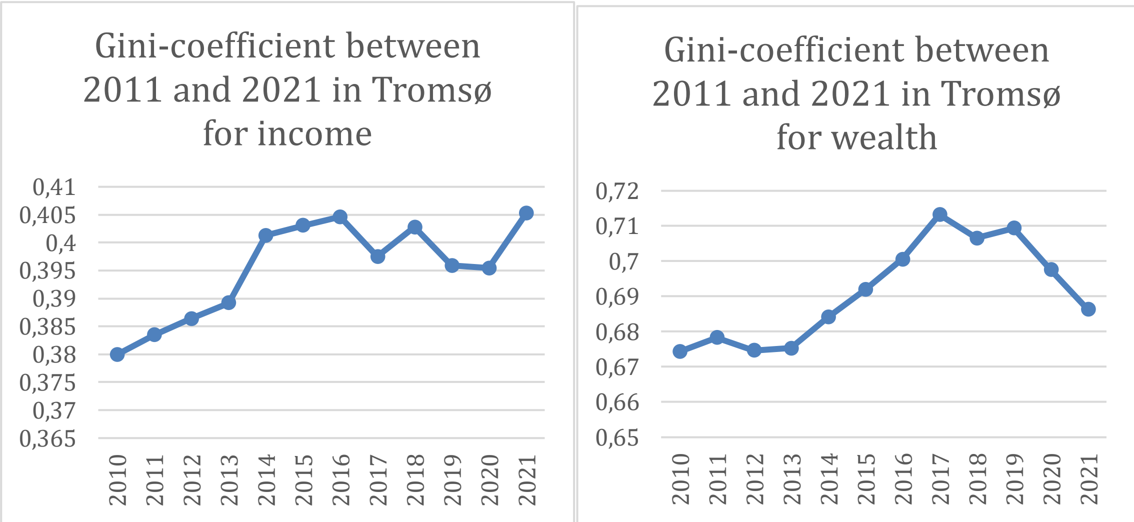

Diagram 1: Gini-coefficients calculated on income and wealth, 2010 to 2021.

Diagram 1 presents the Gini coefficients calculated for wealth and income in Tromsø. The Gini coefficient for income increased from 0.38 in 2010 to 0.405 in 2021, while the Gini coefficient for wealth rose from approximately 0.675 to 0.697. Although both coefficients displayed an upward trend, wealth inequality have declined since 2019. With the exception of a few neighborhoods that exhibited a decrease in the coefficient, the majority experienced an increase in both income and wealth inequality as indicated by the Gini coefficient. Notably, there is no clear association between higher average income or wealth in a neighborhood and a higher Gini coefficient within that neighborhood. Some wealthier neighborhoods displayed lower Gini coefficients than relatively poorer neighborhoods, regardless of whether wealth or income was considered.

Over the period from 2011 to 2021, both income and wealth showed significant growth. Persistent differences were observed between neighborhoods, with some neighborhoods consistently had lower incomes compared to others, a pattern that remained unchanged throughout the study period. Moreover, neighborhoods with lower initial income levels tended to have lower average annual income increases compared to neighborhoods with higher income. The full overview for all neighborhoods regarding income and wealth over time, please look at tables 4 to 11 in section Tables and Diagram. The neighborhoods with the highest income and wealth levels largely remained the same over time, suggesting a certain level of attractiveness or desirability associated with these areas for individuals seeking higher socioeconomic status.

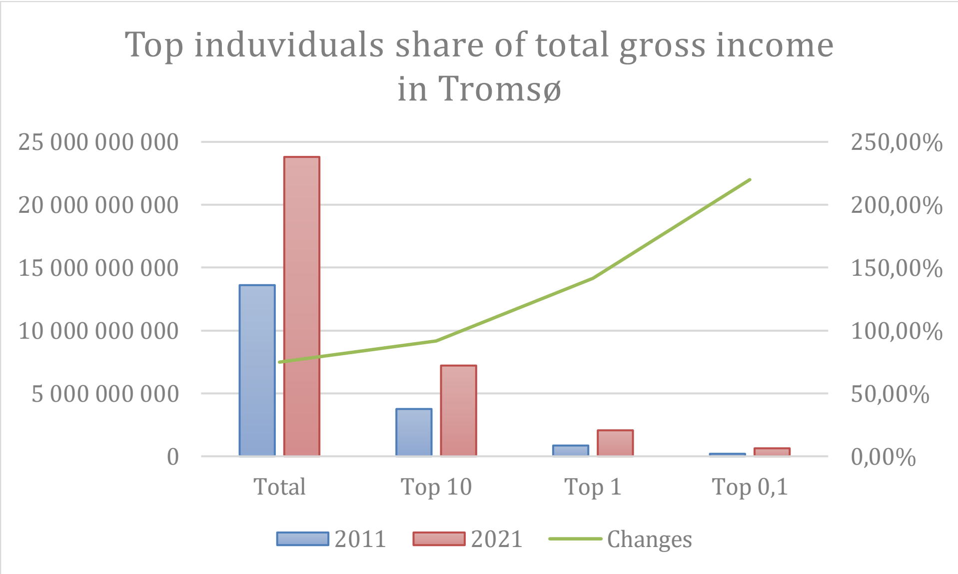

Diagram 2: The share that top 10, 1 and 0,1 percent of the population have in gross income in Tromsø. Changes shows how much they have increased their own gross income in the groups. 2011 shows the total gross income in Tromsø and the share of total gross income for the different groups. 2021 shows the total gross income in Tromsø and the share of the total gross income for the different groups.

Diagram 2 shows the evolution of income in Tromsø from 2011 to 2021 across various income groups. The figure provides insights into the growth of gross income in the region during this period. It highlights the distribution of income among different upper-income brackets. The top 10%, 1%, and 0.1% of individuals with the highest income experienced an increase in their proportionate share of the total income between 2011 and 2021.

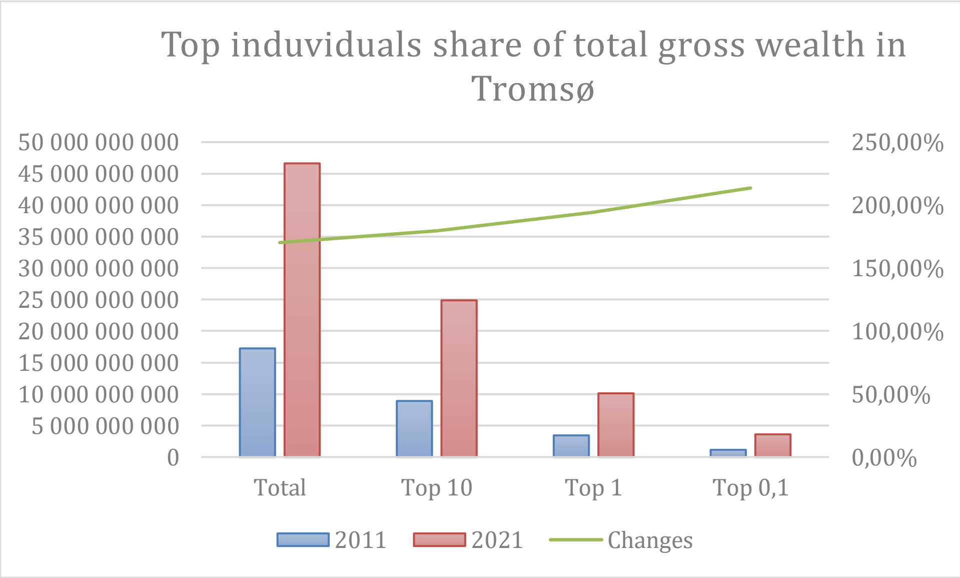

Diagram 3: The share that top 10, 1 and 0,1 percent of the population have in gross wealth in Tromsø. Changes shows how much they have increased their own gross wealth in the groups. 2011 shows the total gross wealth in Tromsø and the share of total gross wealth for the different groups. 2021 shows the total gross wealth in Tromsø and the share of the total gross wealth for the different groups.

Diagram 3 presents the changes in wealth in Tromsø between 2011 and 2021, examining the proportion of wealth held by the top 10%, 1%, and 0.1% of individuals with the highest wealth. The data illustrates a substantial increase in the overall gross wealth in Tromsø during this period. All upper-wealth groups experienced growth in both their total gross wealth and their share of the total wealth in Tromsø.

Given the sensitivity of the Gini coefficient to extreme values, particularly when the sample size is small (Tuv, 2019), an additional measure, P90/P10, will be included as a supplementary analysis. This method measure income inequality by examining the income differential between individuals within the 90th percentile and those within the 10th percentile. The population is divided into various percentiles, and the P90/P10 ratio calculates how much more the top 10% earns compared to the bottom 10%. For instance, if the P90/P10 value is 4, it indicates that individuals in the top 10% earn four times more than individuals in the bottom 10%.

$$2011 = 6{,}59$$

$$2021 = 7{,}33$$

This shows that the top 10% of the population regarding income earns more than the bottom 10% from 2011 to 2021; where in 2011 earned 6,59 times as much, while it increased to 7,33 in 2021. As with the Gini coefficient, income inequality has increased in Tromsø.

According to the report by SSB, both gross income and wealth in Norway, as well as in Tromsø, have an upward trend. This is also reflected in the share of income and wealth held by the top 10%, 1%, and 0.1%, as shown in Diagrams 2 and 3. The data suggest that as gross income and wealth increase, the wealthiest individuals are increasing their share of the total. It is important to note that these findings are based on figures before tax, making it challenging to draw definitive conclusions regarding their actual wealth accumulation. Nonetheless, there are indications that the wealthiest individuals are indeed becoming richer, as supported by the growth in the Gini coefficient reported by both SSB and this master thesis. The increased share of income going to the highest earners and wealthiest individuals contributes to an overall increase in income inequality, particularly when considering percentage-based income growth. If the income of a high-income individual increases by 1%, the same percentage increase for a low-income individual would be relatively smaller, thus magnifying the perception of income inequality even if their average income growth remains the same.

The findings regarding the Gini coefficients for income and wealth inequality in Tromsø raise important discussions about the level of inequality and its dynamics within the city. The increasing observed both coefficients suggests a growing disparity in the distribution of income and wealth among individuals in the population. The increased Gini coefficient for income from 2010 to 2021, and the Gini coefficient for wealth in Tromsø experienced an upward trend during the study period, even with a decline after 2019. This suggests that while income and wealth inequality has been a prevailing issue.

It is noteworthy that the relationship between neighborhood income or wealth and Gini coefficients is not straightforward. The absence of a clear connection between higher average income or wealth in a neighborhood and a higher Gini coefficient within that neighborhood challenges common assumptions regarding income or wealth concentration. This emphasizes the importance of exploring other factors, such as access to resources, educational opportunities, or social networks, that may influence the distribution of income and wealth within neighborhoods.

The persistence of income disparities between neighborhoods throughout the study period suggests the presence of underlying structural factors that contribute to unequal income growth. Neighborhoods with lower initial income levels consistently experienced slower income increases compared to their higher-income counterparts. This indicates the existence of socio-economic dynamics that perpetuate income inequalities and hinder upward mobility for individuals residing in lower-income neighborhoods. The stability of neighborhoods with the highest income and wealth levels over time suggests the presence of certain desirable characteristics or amenities that attract individuals seeking higher socioeconomic status.

The average Gini-coefficient for Tromsø was decreasing, but after the Covid pandemic, it increased to its highest in this period. This can be correlated. (Adarov, 2022) On a global scale, between 2008 and 2010, the income inequality between countries was decreasing. This changed between 2019 and 2021, when it increased. It shows that income inequality increases within-country because of the loss of jobs and income for low-skilled workers, low-income households, and informal workers. Oslo Met(2021) found that through the COVID-19 pandemic, lower-income groups who already were in a challenging position were now in a more challenging position than before, where 40 percent had a lower income than before the pandemic, compared to 13 percent in the part of the population which was not in the lower-income group. This can be a part of the explanation why income inequality seems to have increased on an individual level. However, it is important to note that assessing whether this indicates a increase in income inequality is complex. The data collected and analyzed in this master's thesis primarily focus on individual income before tax. In Norway, a progressive tax system is in place, which means that individuals with higher incomes contribute a greater share of their income in taxes compared to those with lower incomes. As a result, the actual level of income inequality is lower than what is calculated in this particular master's thesis.

The Gini coefficient for wealth has shown an overall increase between 2011 and 2021, although it has decreased from its peak levels prior to the Covid-19 pandemic. One potential factor that can account for this trend is the impact of saving behavior during the pandemic. Research conducted by Oslo Met revealed that 39 percent of lower-income households experienced a decline in their ability to save for the future, and 22 percent depleted all their savings. These findings suggest that the economic challenges brought about by the pandemic, combined with increased costs of fuel, loan rents, and electricity, have affected both individuals and businesses, including their owners. The rise in prices of production factors, goods and services has elevated costs for many businesses, potentially leading to closures. Therefore, business owners who had invested in their stores or factories may have experienced a decrease in their wealth. However, it is worth noting that half of the businesses that received pandemic aid actually achieved better financial results during the pandemic compared to pre-pandemic levels. Some were able to repay the aid without incurring losses in 2020 (Fraser, et al. E24, 2021). This suggests that certain business owners may have seen an increase in their wealth during this period. Nevertheless, it is challenging to obtain a comprehensive assessment of wealth inequality based on this analysis, as the data primarily focus on wealth before tax. In Norway, the existence of a progressive tax system implies that wealth inequality is likely to be lower than what is indicated in this master's thesis.

It is important to note that Norway has relatively low income inequality compared to many other countries. The Norwegian welfare state, progressive tax system, and social policies aim to mitigate the negative effects of income inequality. Nonetheless, income disparities still exist, and addressing income inequality remains an ongoing challenge for policymakers in Norway.

2.2 Residential Segregation

The Index of Dissimilarity (ID), originally proposed by Duncan and Duncan (1955), is a widely used method for quantifying residential segregation between distinct groups. In the context of this study, the ID captures the relative distribution of these two groups across various neighborhoods in Tromsø. Ranging from 0 to 100, the index represents the extent to which members of one group would need to relocate in order to achieve a perfectly even distribution across all neighborhoods. A value of 0 indicates complete integration, with both groups evenly distributed throughout the city. A value of 100 signifies complete segregation, where the two groups are exclusively concentrated in separate areas. The ID is calculated using the following formula:

$$ID = \frac{1}{2}\sum_{i=1}^{N}\left|\frac{a_i}{A} - \frac{b_i}{B}\right|$$

where a_i is the population of group A in an area, and A is the total population in an area where the index is being calculated. b_i is the population of group B in an area, B is the total population in an area where the index is being calculated. In this paper, I will divide the population into three different groups, low-, medium- and high-income groups, which means it will be calculated as such:

1) Low-Medium
2) Low-High
3) Medium-High

Individuals classified as having low income are those whose earnings amount to 60% or less of the median income of the overall population, as defined by the statistical agency (Hattrem, 2022). When categorizing the population into the three income groups, these criteria will be utilized. On the other hand, individuals considered to have high income are those earning 60% or more of the median income of the population.

$$Low = Median\ income * 0{,}60$$

$$High = Median\ income * 1{,}60$$

$$Medium = Low < Median < High$$

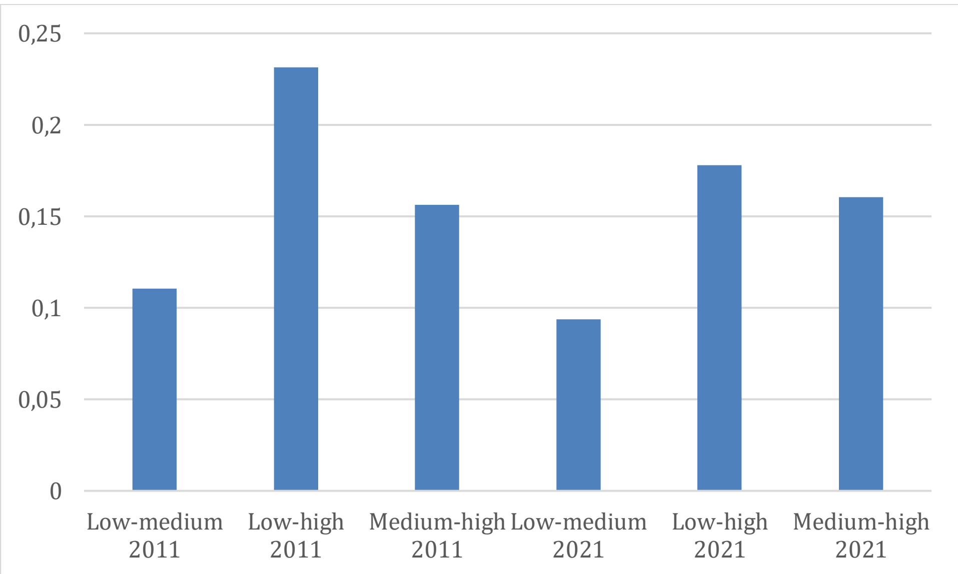

Diagram 4: Index of Dissimilarity

Diagram 4 presents the computed ID values for two time periods, 2011 and 2021, across Tromsø. In 2011, the ID between the low-medium-income groups was 0.1105, between the low-high-income groups was 0.2315, and between the medium-high-income groups was 0.1562. Among these three ID measurements, the low-high-income groups has higher levels of segregation towards each other compared to the medium-high-income and low-medium-income groups.

In 2021, the ID values between the low-medium-income groups decreased to 0.0939, while the ID between the low-high-income groups was 0.1781, and between the medium-high-income groups was 0.1604. Similar to the findings in 2011, the low-high-income groups display a higher degree of segregation towards each other compared to the medium-high-income and low-medium-income groups. The results from the ID analysis reveal changes in income segregation trends between 2011 and 2021. The low-high income group experienced a reduction of 29,9%, indicating a decrease in segregation between them. Similarly, the low-medium income group decreased by 17,67% in segregation. On the other hand, the medium-high income group increased by 2% in segregation during this period. These findings suggest a diminishing income segregation between low-high and low-medium groups, while an increasing segregation between the medium-high income group.

A study conducted by Reardon and Bischoff (2011) provides robust evidence supporting a strong positive association between income inequality and income segregation. The findings reveal that over a forty-year period, average family income segregation increased by 29 percent. However, the situation in Tromsø over the past decade shows a different trend, with a reduction in segregation observed. It is important to acknowledge that the findings from America and Norway can vary significantly due to various factors. One such factor is socioeconomic inequality, which is more pronounced in the United States compared to Norway. Additionally, the divergent welfare systems play a role, with Norway's welfare system providing social support, healthcare, education, and income redistribution to mitigate socioeconomic disparities, while the United States has a more limited welfare system. Despite these differences, studying the specific case of Tromsø in relation to broader research findings is intriguing and can contribute to a deeper understanding of income segregation dynamics.

It is important to note that the results obtained from the ID analysis could differ if the income groups were defined differently from the approach utilized in this study. The definition of the low-income group was established based on individuals earning 60% or less than the median income of Tromsø, which is a standard criterion employed by Statistics Norway (Hattrem, 2022). However, defining the high-income group did not adhere to a standardized framework. In this study, individuals earning 60% or more than the median income of Tromsø were considered part of the high-income group, with the medium-income group encompassing individuals falling between these thresholds. It is worth mentioning that employing different percentages to define the high-income group would yield varying results, as some percentages may align more closely with the concept of high-income. For instance, utilizing a threshold of 80% or higher for the high-income group reveals a different outcome, where only the segregation between the low-medium income group is reduced, while the segregation between the low-high and medium-high income groups increases. Consequently, selecting alternative percentages that better capture the high-income category would impact the results presented in this study, which are based on the utilization of 60% or more as the threshold. It is important to exercise caution when interpreting these findings, as different income group definitions may lead to divergent conclusions regarding income segregation dynamics. Additionally, it is crucial to consider Norway's progressive tax system, which poses a challenge in assessing overall changes in income segregation. The data utilized in this study represent income before tax, making it difficult to draw definitive conclusions regarding the overall change in segregation patterns. However, analyzing individual-level data prior to taxation provides valuable insights into the changes in income segregation over time. To gain insight of income-based segregation in Tromsø, this study employs Schelling's model of segregation as a theoretical framework.

3 Theory: Schelling’s model

The model of segregation introduced by Thomas C. Schelling in 1971 has been widely used to examine the macro-level implications of individual behaviors, incentives, and preferences, particularly in relation to residential segregation. (Schelling, 1971) Originally developed to study ethnic segregation in the United States, Schelling's agent-based model focuses on households as agents and their propensity to reside in neignborhoods to others of the same ethnic group.

The model operates on a grid with dimensions *NxN*, where each cell represents an individual belonging to one of two distinct groups or denotes a vacant space. Each agent possesses a preferred tolerance threshold (t) that signifies the minimum proportion of neighbors belonging to the same group necessary to satisfy the agent's preferences. The model progresses through several stages, during which agents decide whether to relocate based on the current composition of their neighbors. If an agent's threshold (*t*) is not met (i.e., the proportion of same-group neighbors is below *t*), the agent will seek out a vacant cell where this requirement is satisfied ($T \geq t$). This relocation process continues until all agents are content with their respective situations and no further movements occur. However, it is important to note that not all agents may achieve satisfaction, as some may encounter situations where relocation to an improved cell is not feasible.

Schelling's study revealed a critical threshold value, t_{crit}, which corresponds to situations where the two groups are of equal size. Specifically, when $t < t_{crit}$, the model produces a random composition of agents across the grid, while $t \geq t_{crit}$ leads to a segregated composition. The approximate value of t_{crit} is around 1/3. To verify these findings, a simulation was conducted to observe the spatial distribution of agents from two groups under various threshold values.

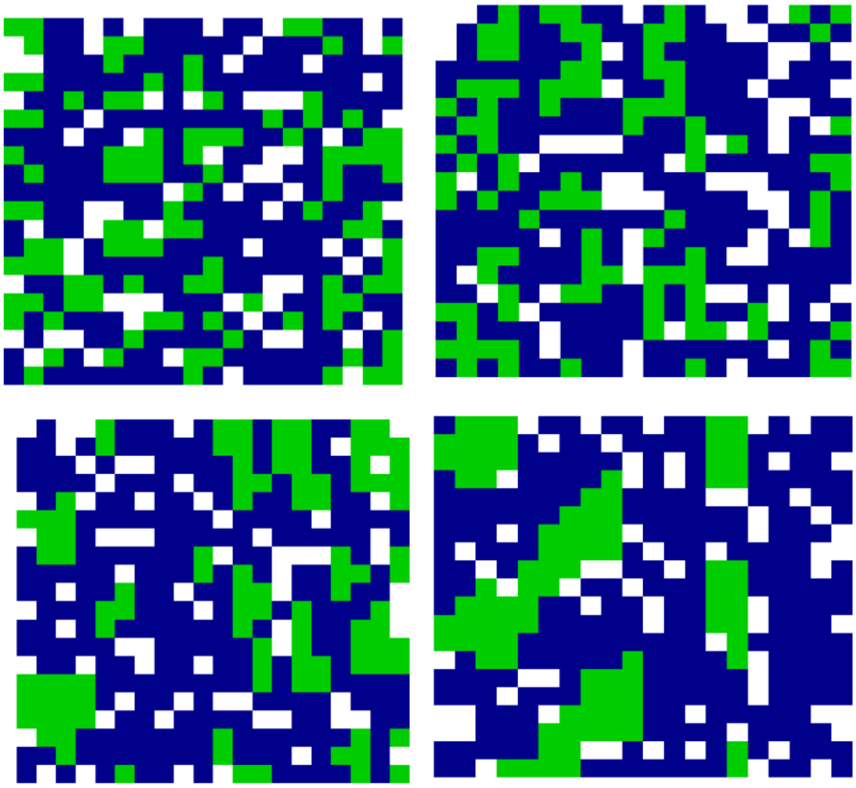

Figure 1: Threshold values for each figure: 10%, 25%, 35%, 50%

The presented figures depict the outcomes of different simulations conducted using Schelling's model. It should be noted that these simulations are based on random configurations, meaning that reproducing the same threshold value will yield different compositions, and the number of stages required to reach the final result may vary. The grid used for the simulations has dimensions of 20x20, with 17% of the cells being vacant (represented by white cells). The agents are divided into two groups: blue and green. Four distinct threshold values were employed in the simulations: 10%, 25%, 35%, and 50%.

In the first simulation, with a threshold of 10%, the composition of located agents was achieved after four stages. The second simulation, with a threshold of 25%, required fifteen stages to complete. The third simulation, with a threshold of 35%, reached its final composition after thirty-six stages. Finally, the fourth simulation, utilizing a threshold of 50%, concluded after twenty-seven stages. The results indicate that higher threshold values, particularly from 35% and above, tend to result in more segregated compositions.

Schelling's model has been extensively employed to investigate various forms of segregation, including race-based segregation and economic disparities among agents. For instance, Benard and Willer (2007) developed a model of residential segregation based on wealth and status, where status represents an agent's desirability as a neighbor according to others' perceptions. The model assumes that agents prefer to reside in neighborhoods with high-status individuals rather than those with low status. High status is attributed to individuals who are popular, famous, or respected in their community. Agents in this model aspire to reside in more desirable neighborhoods, facilitated by their wealth. The study examines the correlation between status and wealth, and how residents' wealth impacts house prices, status, and segregation. The findings suggest that a stronger correlation between status and wealth leads to greater levels of segregation, indicating that residents with higher wealth and status have more opportunities for relocation. On the other hand, some residents face exclusion from relocation due to poverty and low status.

The micro-level behaviors exhibited by individuals can have macro-level implications, particularly in terms of segregation and social dynamics. When agents in the model possess a slight preference for residing in neighborhoods populated by individuals similar to themselves, the resulting aggregation of such preferences can give rise to significant levels of macro-scale segregation. This segregation, in turn, engenders several macro problems. One such problem is social fragmentation, wherein reduced interaction and limited understanding between disparate groups ensues. This phenomenon fosters social divisions and impedes the establishment of cohesive societal frameworks. (Logan & Schneider 2010). Additionally, the perpetuation of stereotypes and prejudice can be observed, as segregated neighborhoods reinforce pre-existing biases, hinder social integration, and generate negative perceptions of other racial or ethnic groups. Segregation acts as a catalyst for the perpetuation of existing inequalities and the manifestation of exclusionary dynamics. These dynamics result in inequal distribution of resources, opportunities, and access to public services across neighborhoods. (Page & Shepherd, 2008) Disadvantaged groups face systemic obstacles and barriers that hinder their social and economic advancement. The consequences of segregation are not solely confined to the social but also extend into the economic. Segregated neighborhoods may experience disinvestment, limited economic prospects, and diminished social mobility, exacerbating pre-existing inequalities and impeding overall economic development. (Ellen & Turner,1997).

It is important to note that this description encapsulates the fundamental dynamics of segregation as elucidated by the model and does not account for the myriad complexities and contextual factors that exist in real-world scenarios. Nevertheless, this model provides valuable insights into the potential ramifications of individual micro behaviors on broader social structures and underscores the importance of addressing segregation as a critical societal concern. (Massey & Denton, 1993)(Reardon & Bischoff, 2011)

To comprehend the decision-making process behind individuals' choices to move or stay, as well as their preference for similar neighborhoods, models are employed to identify and analyze the variables influencing their decisions. One such model is the Heckman two-step model.

4 Method: Heckman two-step

Empirical investigations aiming to establish a causal relationship face several conditions that must be satisfied in order to conclude that a particular variable, *x*, causes another variable, *y*, to occur. These conditions include: the occurrence of *y* following the presence of *x*, the changes in *y* corresponding to changes in *x*, and the absence of any alternative causes that could undermine the relationship between *x* and *y*. In order to meet these conditions, the expectation of the error term, given the independent variables, should equal zero. However, when this expectation is not fulfilled, an endogeneity problem arises. Endogeneity can occur due to various factors, such as the presence of omitted variables, measurement errors in the independent variables, or in the case of this thesis, selection bias(Antonakis et al., 2014).

To address this concern of selection bias, Heckman introduced a two-equation model known as the Heckman Two-stage model (Heckman, 1976), which is also referred to as the Tobit-2 model (Toomet & Henningsen, 2008). This model offers a means to mitigate selection bias and enhance the accuracy of estimations. For a random sample of observations, *I*, an individual, *i*, would have the equations:

$$Y_{1i}^{*} = X_{1i}\beta_1 + U_{1i} \qquad (1a)$$

$$Y_{2i}^{*} = X_{2i}\beta_2 + U_{2i} \quad (1b)$$

$$i = 1, \dots, I)$$

where Y_{1i}^{*} is the latent value of the selection for the given individual i, and $Y_{2i}^{*}$is the latent outcome. X_{1i} and $X_{2i}$are the explanatory variables for the selection outcome equations and can be equal or not equal. ε_{1i} and ε_{2i} are the error terms for the equations. To observe the equations, then:

$$Y_{1i} = \begin{cases} 0 \; if \; Y_{1i}^{*} < 0 \\ 1 \; otherwise \end{cases} \quad (3)$$

$$Y_{2i} = \begin{cases} 0 \; if \; Y_{1i} = 0 \\ 1 \; otherwise \end{cases} \quad (4)$$

Which would mean that outcome can only be observed if the latent selection equations is positive.

The observed dependence between Y_2 and X_2 is written like:

$$\mathrm{E}[Y_2 | X_2 = X_{2i}, X_i = X_{1i}, Y_{1i} = 1] = X_{2i}\beta_2 + \mathrm{E}[\varepsilon_2 | \varepsilon_1 \geq -X_{1i}\beta_1] \quad (5)$$

The error terms in the model are assumed to follow a bivariate normal distribution:

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} =\sim \mathrm{N}\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \vartheta \\ \vartheta & \sigma^2 \end{pmatrix} \quad (6)$$

Since $\mathrm{E}[\varepsilon_2|\varepsilon_1 \geq -X_{1i}\beta_1] \neq 0$ if we calculate equation (5) by using OLS, would give a biased result in general. The only exception is when ε_{1i} and $\varepsilon_{2i}$are mean independent, which means that the ϑ in equation (6) is zero. To prevent an biased result, estimate equations (1) and (3) as a probit model, then insert into (2) which will give us $\mathrm{E}[\varepsilon_2|\varepsilon_1 \geq -X_{1i}\beta_1]$, the control function, giving the following equation:

$$Y_{2i} = X_{2i}\beta_2 + \mathrm{E}[\varepsilon_2|\varepsilon_1 \geq -X_{1i}\beta_1] + \eta_i \equiv X_{2i}\beta_2 + \vartheta\sigma\lambda(X_{1i}\beta_1) + \eta_i \qquad (7)$$

where $\lambda = \varphi/\Phi$ is refers as the inverse Mils ratio, which is defined as the ratio of the standard normal density, φ, devided by the standard normal cumulative distribution, Φ, which can be written as$\frac{\varphi(x)}{\Phi(x)}$. η is the disturbance term, or error term, which indicate the difference between the population mean and the observed value and are independent of $X_1$and X_2. ϑσ is the unknown multiplicator that can be calculated with OLS. This means that λ, or the Mile ratio, is an omitted variable. And because this is unknown, it can be replaced by estimated values based on the estimation of the probit model.

When constructing a model to simulate the decision-making process of individuals regarding relocation, it is important to consider the potential presence of correlated variables that may influence this decision. When relying solely on ordinary OLS regression, there is a risk of obtaining biased and inconsistent estimates that deviate from the true values of the variables. This bias tends to be upward, leading to overestimation of the effects of certain variables. Trying to identify if selection bias is present, one can show this by incorporating inverse Mills` Ratio(IMR) into a probit model. This can be done in two steps. First, by estimating the selection equation, which is a Probit model. Second, estimate the outcome equation. It is the second step one would need to incorporate the IMR. The IMR is calculated as the predicted value from the selection equation divided by its corresponding standard deviation. By comparing the coefficient estimates in the outcome equation with and without the inclusion of the IMR, and if the coefficient estimates change substantially when the IMR is included, it indicates the presence of selection bias and the need to correct for it.

From table 1 in section Tables and Diagrams, when including IMR1 and IMR0 into the equation, the estimated coefficient changes drastically, which indicate that selection bias is present. It can also be confirmed by looking at the significant of the IMR coefficient, which is significant. IMR1 and IMR0 was calculated by using sampleSelection package in R.

To mitigate the issues associated with biased estimation in the model, instrumental variables (IV) are employed. By introducing an additional variable, denoted as *Z,* into the model that is correlated with *X* but uncorrelated with the error term ε, the changes in *Y* can be attributed to variations in *X*. However, the instrumental variable *Z* must satisfy two essential conditions:

$$Cov(Z, X) \neq 0 \qquad 1)$$

The inclusion restriction implies that the instrumental variable Z should exhibit a non-zero correlation with the independent variable X.

$$\mathrm{Cov}(\mathrm{Z}, \varepsilon) = 0 \qquad 2)$$

The exclusion restriction stipulates that there should be no correlation between the instrumental variable and the error term of the model.

To fulfil the second condition, the instrumental variable used should be observed and determined in the selection equation but not in the outcome equation. However, selecting a suitable instrumental variable is challenging, as the two requirements often contradict each other and are contingent upon the quality of the available data. For this study, the chosen instrumental variable will be *Distance*. The description of this variable will be presented in section 6.

5 Data

The data was provided by Rune Ytterberg. The primary dataset utilized in this study comprises tax data from Norway spanning the period between 2011 and 2021. These data were collected and compiled by Skattetaten, the Norwegian tax administration. The dataset have information on individuals' income, taxes, wealth, country of origin, and municipality. The individuals' locations and tax assessments are determined based on their respective postcodes. It is important to note that personal identifiers in the dataset have been encrypted, ensuring the anonymity of individuals. The encryption process involves replacing personal details with a combination of the individual's birth year and a unique person identifier. It is worth mentioning that the 2021 dataset does not include these encrypted identifiers.

The initial edition of the dataset covers the years 2011 to 2016 and represents the first publication by Skattetaten. It is acknowledged that errors may be present within these time series, which are subsequently rectified in later editions. The more recent editions, spanning the years 2017 to 2021, are the most up-to-date publications from Skattetaten and are presumed to have fewer data discrepancies.

A significant portion of the dataset consists of individuals with zero income. The specific reasons for their zero income are not explicitly provided in the dataset. This zero-income category could encompass individuals who were unemployed during the respective year, elderly individuals receiving pensions, or affluent individuals who did not receive regular wages but rather incurred expenses using their accumulated wealth. For calculations pertaining solely to income, only individuals with positive income will be included in the analysis.

The second dataset employed in this study encompasses all property transactions that have occurred in Tromsø over the past three decades. Specifically, the dataset comprises transactions that have undergone official registration processes overseen by the court. Property transactions that have not undergone this formal registration process are excluded from the dataset. The data within this dataset includes various information pertaining to each transaction. This information comprises the transaction value, the usable area of the property, the postcode of the property, and the postcode of the buyer. The dataset also includes information regarding the number of buyers involved in each transaction. However, it does not provide the specific distribution of the payment amounts made by individual buyers, as this data is not available.

It is important to note that transactions involving housing associations are not included in the dataset. This exclusion arises because the value paid by a buyer in such transactions represents a share of the overall value of the property rather than the total value. Furthermore, there is no information available regarding the specific amount paid for each property within housing associations, as the payment structure varies among different associations. Certain variables will be considered based on the house transaction data. Only open sales will be included, which refers to transactions where the sale price corresponds to the market value of the property. Additionally, the analysis will encompass various types of houses, namely semi-detached houses, detached houses, terraced houses, and blocks.

6 Descriptive statistics

This thesis aims to estimate the effects of individual mobility and the selection of similar income neighborhoods using two equations. The first equation examines whether an individual moved between 2011 and 2020. This will be identified by observing a change in postcode within the available data. However, the exact timing and frequency of an individual's moves cannot be determined, and instances where individuals move within the same neighborhoodlike will not be included. The variable *moved* will be a dummy variable with a value of 1 indicating that the individual moved and 0 indicating not observed move from their original neighborhood.

The instrumental variable that will be used in the first equation will be *Distance.* (Carneiro et al. 2003) This variable is a dummy variable that represents the distance of individuals' residences from the city center of Tromsø. The Townhall serves as the reference point, and a series of concentric rings with a radius of 1 kilometer from the Townhall have been created. Each neighborhood located within these rings is assigned a dummy value of 1. Subsequently, for every 2 kilometers, a new ring is formed, and the dummy values range from 1 to 6, with 1 denoting the closest to the Townhall and 6 representing the farthest distance. The inclusion of the distance from the city center as an instrumental variable allows for the examination of its impact on the model's outcomes. By incorporating distance as an instrumental variable, I aim to address potential endogeneity concerns that arise when the relationship between X and Y is confounded by unobserved variables or reverse causality. The use of distance as a proxy variable enables us to capture the variation in X that is independent of unobservable factors, thereby providing more reliable and unbiased estimates of the causal relationship between X and Y.

In the second equation, I will explore whether an individual moved to a postcode with a similar mean income. The independent variable *similar* will also be a dummy variable with the values 0 and 1. The value of 1 for "similar" indicates that the individual moved from their original postcode to a new one with a similar income level, as determined by comparing the individual's income variance with the mean income of the new neighborhood. To capture different degrees of income similarity, I will employ various similar dependent variables, such as *similar10*, *similar20*, *similar30*, and *similar40*. The numerical values represent the percentage range mean income falls relative to the individual's original neighborhood. For example, *similar10* and *similar20* indicate whether the individual moved to a neighborhood with income similarity within ±10% and ±20% respectively. If *similar is* 0 while *moved* is 1, it suggests that the individual moved to a neighborhood outside the defined range of income similarity. The degree of segregation is influenced by the extent to which incomes are similar. Increasing the range of income similarity enhances the sensitivity of segregation. This is related to Schelling's model, where raising the threshold for similarity among an agent's neighbors results in a greater likelihood of segregation.

$$Moved = \beta_0 + \beta_1 Distance + Similar + \beta_2 \log(income2011) + \beta_3 Variation + \beta_4 Pricechange + \beta_5 Gini + U_{1i} \quad (1a)$$

$$Similar = \beta_0 + \beta_1 \log(income2011) + \beta_2 Variation + \beta_3 pricechange + \beta_4 gini + U_{1i} \quad (1b)$$

I will construct four distinct models incorporating different similar variables. These models will follow the same structure as the one just presented, with the only difference being the utilization of alternative similar dependent variables. Instead of *similar1011* the models will use *similar2011* and so forth. For a full description of all variables that will be used in this thesis, please look at table 2 in section Tables and Diagrams.

7 Result from Heckman's two-step model

The Heckman two-step model was calculated in R with package sampleSelection. (Toomet, O., & Henningsen, A. ,2008) This gave the following regression result:

Table 3: Result of Heckman two-step models

Variables	**Probit selection model**	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Probit selection model:					
Moved	1,5834				
	[<2e-16]***				
Distance	-0,0421				
	[7,50e-11]***				
Log(income2011)	0,0890				
	[<2,46e-1e]***				
Variation	-0,0860				
	[5,93e-08]***				
Pricechange	-2,2614				
	[1,37e-06]***				
Gini	-2,1785				
	[<2e-16]***				
Outcome equation:					
Similar		-0,1144	-0,1504	0,2651	-0,3663
		[0,0374]*	[0,0430]*	[0,0001]**	[1,37e-05]***
Log(income2011)		0,0196	0,0250	0,0483	0,0726
		[0,0017]**	[0,0024]**	[2,71e-07]***	[3,65e-14]***
Variation		0,0180	0,0054	-0,0311	-0,0667
		[0,0482]*	[0,6526]	[0,0214]*	[1,48e-06]***
Pricechange		-0,0840	-0,9528	-0,6822	-0,4989
		[0,7584]	[0,0099]**	[0,0918].	[0,2291]
Gini		0,0839	-0,0293	-0,5288	0,6043
		[0,4982]	[0,0787].	[0,0042]**	[0,0014]**
InvMillsRatio		-0,0332	0,1940	0,2288	0,1921
		[0,67]	[0,0644).	[0,0495]*	[0,107]
Number of observation	3181	1691	772	1107	1440
Significant codes:	"***"= 0,001. "**"=0,01. "*"= 0,05. ". "=0,1. " "=1				

Table 3: Result of Heckman two-step regression

The table presented above provides the results obtained from the analysis of four distinct models. These models differ primarily in terms of the chosen similarity variables used. Model 1 employs *similar10*, while Model 2 it uses *similar20*, and so on. The first equation of each model represents the probit selection model, which pertains to the decision-making process of whether to relocate or not. All four models share identical structures in this aspect, as the differences of interest lie in the outcome equation. The dependent variable in this equation is *moved*, and across all models, it exhibits a positive and statistically significant estimate.

Regarding the independent variables, the IV is *Distance*. The coefficient associated with distance is negative and statistically significant, indicating that greater distances from Tromsø's townhall reduce the likelihood of choosing to relocate. *Log(income2011)* is positive and are significant, suggesting that as income increases, the probability of moving increases. The variable *variation* has an negative effect and is statistically significant. This is a dummy variable with values 2,1 and 0 If the value is 2, the individual has higher income than the mean of their neighborhood by 10% and above. Value 1 when the individual has an income between ±10% of the mean of their neighborhood. Value 0 when the individual has income which is from -10% or less than the mean income of their neighborhood. The negative relationship would mean that if the individual increased from 0 to 1, or 1 to 2, would decrease the probability of that individual to move. *Pricechange* are negative and significant, indicating that an increase in average house prices within an individual's neighborhood decreases the probability of moving. *Gini*, representing income inequality in the neighborhood, have a negative and significant relationship, suggesting that an increase in income inequality reduces the probability of moving.

The second equation, which represents the outcome equation, the dependent variable *Similar* is significant across all models, but it is negative in three models, while in model 3 it is positive. *Log(income2011)* is positive and significant association in all models, indicating that an individual's probability of selecting a neighborhood with a similar income increases as their income rises. *Variation* is estimated to have a positive relationship when choosing to move to a neighborhood with similar income in model 1 and 2, while it have a negative relationship in model 3 and 4. It is also significant for three of four models. *Pricechange* is negative in all models, but its statistical significance in model 2 and have a weak significance in model 3. It would mean that when the price of houses on average increases in the neighborhoods, it will reduce the probability to choose a neighborhood with similar income. *Gini* is positive for model 1 and 4, while it is negative for 2 and 3, while it is statistically significant in model 3 and 4.. The error term InvMillsRatio is negative in model 1 and positive in the other three. It is significant in model 3, while it has a weak significance in model 2.

7.1 Discussion

The four different Heckman models employed in this analysis vary in terms of their outcome equations and the estimation of the error term. Notably, Models 2 and 3 has a significant error term, with the error term being statistically significant in Model 3 and weak significance in Model 2. Regarding the remaining estimations, Model 3 have the highest number of significant variables. This suggests that Model 3 provides the best fit to the data when considering all relevant variables and the error term. The model which fits the data was model 3, which had a income similar range of ±30%. It would sems that for the data at hand, they have a threshold which they want the neighborrs to be similar to them around the same range as Shellings model found was the critical threshold for segregation.

In the first equation, the distance from the Townhall of Tromsø have a positive relationship when choosing to move or not. This variable has a negative relationship with the decision to move to another neighborhood. This would mean that the further away one is located from the city center, increases the probability of moving. This would show that the closest areas around the city center of Tromsø could be more accretive areas, and those that live closer to the city center sems to be more satisfied on their location.

Change in income have a positive relationship regarding the decision to move. The higher their income, increases the probability of moving to another neighborhood. According to Ross and Yinger(2002) higher-income induviduals would have better opportunity of moving to a neighborhood with higher quality than lower-income induviduals. Ellen and Turner(1997) found correlation between increased income and increased likelihood of moving to a better neighborhood when regarding crime rates, quality of schools and amenities of the neighborhood. Findings in this equation also shows that an increased income dose increase the probability of moving. Even regarding other findings from studies like Ross and Yinger (2002) and Ellen and Turner (1997), it is important to note that the referenced articles were based on research conducted in the United States, where the dynamics of neighborhood development and societal structures differ from those in Norway. It is possible that in Tromsø, there is no significant variation among neighborhoods in terms of crime rates, school quality, and neighborhood amenities, thus mitigating the impact of income on the decision to move or the reasons they choose to relocate could be because of other factors.

When an individual changes their income from or towards the mean of their neighborhood would reduce the probability of them moving out of that neigborhood. This would indicate when they become more similar regarding income, it would reduce the probability of moving. When the variation variable increases from value 0 to 1 would mean that the individual would have increased their income more than the mean have increased, or the mean of that neighborhood would have been reduced. By looking at the data, there are no such instanced that mean income decreases for a neighborhood. When income increases, the likelihood of moving also tends to increase. However, when the increase in income is substantial enough to change the value of the variation, it actually decreases the probability of moving. This could be attributed to the fact that a significant increase in income enables individuals to better cope with the cost of living in their current area. It alleviates their financial difficulties, removing the need to relocate in order to improve their living situation. With a substantially higher income, individuals are more inclined to invest in enhancing their current living conditions rather than seeking a new neighborhood. This finding bears significant implications within the model and in this master's thesis.

The average change in house prices within an individual's own neighborhood are estimated to have a negative relationship with the decision to move. An increase in the price of one's own residence decreases the likelihood of moving. When property prices rise, homeowners may be motivated to sell their homes due to the potential profit they can realize from their initial purchase. Rising prices can indicate a robust housing market where demand exceeds supply. However, in the case of Tromsø, where the housing market has been strong for the past decade (see tables 12 to 15), the increased price of an individual's own residence may not be a sufficient incentive to relocate, as prices in other neighborhoods have also experienced significant increases. It might be expected that individuals with higher incomes and increased property prices could have a positive inclination to move to a different neighborhood. However, the positive coefficients for income and change in house prices in this equation suggest that, based on the available data, such a relationship does not exist.

When income inequality increases within an individual's neighbourhood, it reduces the likelihood of moving to a different neighborhood. This negative relationship suggests that higher levels of income inequality may discourage individuals from seeking to relocate. It is possible that the presence of income inequality creates socioeconomic disparities and limited access to resources and opportunities, making it less desirable for individuals to move to neighborhoods with higher income disparities. As Gini coefficient have increased for almost every neighborhood in Tromsø(see diagram 3 to 6 in section Diagram and Tables), make it less desirable for individuals to move to these neighborhoods as they may perceive a lack of available resources and opportunities. The fear of facing further socioeconomic disadvantages or challenges in such neighborhoods may discourage individuals from relocating to areas with higher income disparities.

In equation two, which corresponds to the outcome equation in model 3, the dependent variable *Similar* representing individuals whose income is within ±30% of the mean income of their original neighborhood. Increased income has a positive relationship with the outcome of selecting a neighborhood with similar income. As an individual's income increases, they gain more opportunities and resources to choose their desired relocation destination. This finding contrasts with the relationship observed in the probability equation. Once individuals have made the decision to move, higher income positively influences their choice of neighborhoods with similar income levels.

When an individual’s income increases significantly to change the value of variation, it have a negative relationship with the choice to relocate to a neighborhood with similar income. This can be because when they choose to move, they would have other preferences then a neighborhood with similar income would have. When it increases significantly, they are no longer as similar as the neighborhood, making them to choose to relocate another neigborhood instead which may or may not be similar to their new income. This negative relationship can be attributed to the fact that as individuals choose to move, they often have preferences and criteria beyond simply matching the income level of their current neighborhood.

Change in house prices on average continues to have a negative relationship with the dependent variable as observed in the previous equation. This may be attributed to the factors discussed earlier, indicating that an increase in the price of one's own residence may not provide sufficient financial incentive to purchase a new residence in a different neighborhood, especially considering that prices in neighborhoods with similar income levels have also increased. However, it is noteworthy that increased income has a positive relationship with the outcome variable, suggesting that a substantial increase in income, coupled with the rise in house prices, could potentially have a positive overall impact on the choice of neighborhood.

Gini also have a negative relationship in this equation with the dependent variable. When income inequality increases in their original neighborhood will reduce the probability of choosing to relocate to neighborhood with similar income. Again, the same factores that was discussed could have the same impact in this equation, as in general, income inequality in Tromsø have increased in all neighborhoods.

It is worth noting that this model may not capture all factors influencing the decision to move, and there may be other considerations beyond the scope of the analysis. However, the increased income provides individuals with the opportunity and resources to select a neighborhood that aligns with their preferences. Furthermore, a increased income similarity up to a significance degree before they become more non-similary within their original neighborhood decreases the probability of choosing a new neighborhood with a more similar income distribution. It is important to note that the equation and its estimates are based on income before tax. Since individuals do not have access to all the income values present in the datasets, and the reduction in income would be even greater for higher income levels, considering income after tax could yield different results than those presented in this master's thesis.

8 Conclusion

The research question addressed in this master's thesis was: " If an individual's current neighborhood is different from their neighbors regarding their income level, would they be more likely to relocate to an area that is more economically similar?” Based on the estimation results of the Heckman model, it appears that increased income is the only variable exhibiting a positive relationship with the decision to relocate to a neighborhood of similar income. However, when the increase in income reaches a substantial level, it demonstrates a negative effect on this relocation decision, as seen in the *Variation* variable. This variable holds significant importance within the model and demonstrates that as individuals approach a similar income level to the neighborhood mean, and further extend to higher positive income relative to the mean, their likelihood of selecting a neighborhood with similar income diminishes. This would mean that the further away they are from the mean of income on the negative side, increases the probability to choose a neighborhood with a similar income. Other variables like income inequality and average house price change also have a negative effect when choosing to move and where to move. All of this would suggest the possibility of existence of additional factors that significantly influence the choice to move to a neighborhood that aligns with one's income, which the current model fails to capture with the data at hand. These findings imply the presence of complex dynamics and other variables that affect individuals' decisions when selecting a neighborhood of similar income, or that they do not have this information at hand to make such a decision. Further investigation are needed to identify and incorporate these additional factors into the model, allowing for a more comprehensive understanding of the relocation decision-making process.

Tables and diagrams

Table 1: Result of testing for selection bias

Variables	Without IMR	With IMR
Outcome equation:		
Similar	-0,1020	-2,4772
	[0,0081]**	[8,65e-05***
Log(income2011)	0,0199	0,1276
	[2,35e-11]***	[0,0001]***
Variation	-0,0057	0,0261
	[0,1511]	[0,02769]*
Pricechange	0,1026	0,3997
	[0,3736]	[0,0006]***
Gini	-0,0242	-0,8775
	[0,6570]	[0,0001]***
IMR1		-0,8775
		[0,0003]***
IMR0		0.9026
		[1,21e-05]***

Significant codes: "***"= 0,001. "**"=0,01. "*"= 0,05. ". "=0,1. " "=1

Table 1: Result of testing for selection bias using IMR

Table 12 Definitions of variables, means, whole sample and observations

Variables	Definition	Means	Whole sample	Movment observed	Moved to similar observed
SIZE OF DATASET			12340	3181	
DISTANCE	distance of their neighborhood to city center, values 1 to 6	3,248			
INCOME2011	income in 2011 for an induvidual	337410			
MOVED	if induvidual changed neighborhood between 2011 and 2020 then =1, otherwhise=0	0,2577			
VARIATION	induviduals variation from the mean of their neighborgood Var>10%=2 10%>Var>-10%=1 Var<-10%=0				
SIMILAR1011	if the induvidual had a similar income variation -10/+10% to the mean in their neighborhood in 2011, then =1 otherwhise=0				1691
SIMILAR2011	if the induvidual had a similar income variation -20/+20% to the mean in their neighborhood in 2011, then =1 otherwhise=0				772
SIMILAR3011	if the induvidual had a similar income variation -30/+30% to the mean in their neighborhood 2011, then =1 otherwhise=0				1107
SIMILAR4011	if the induvidual had a similar income variation -40/+40% to the mean in their neighborhood 2011, then =1 otherwhise=0				1440
Gini	Gini coefficient in the induviduals origin location in 2011	0,343			
PRICECHANGE	Average yearly change in houseprice for the induviduals origin location in 2011	0,06			

Table 2: Description of variables, its sample size, observations and means.

Postcode	Mean Income 2011	Mean Income 2021	Changes	Percent	Average
9006	290974,851	421804,7126	130829,8615	44,96 %	5,00 %
9007	307665,3391	473369,5776	165704,2385	53,86 %	5,98 %
9008	256102,9956	368413,1882	112310,1926	43,85 %	4,87 %
9009	281733,7795	446013,7084	164279,929	58,31 %	6,48 %
9010	264232,0835	360291,7763	96059,69288	36,35 %	4,04 %
9011	308945,025	465570,754	156625,729	50,70 %	5,63 %
9012	329834,9596	530767,4722	200932,5127	60,92 %	6,77 %
9013	359669,004	520568,0005	160898,9964	44,74 %	4,97 %
9014	318540,8792	479431,0341	160890,1548	50,51 %	5,61 %
9015	274619,8478	393797,8764	119178,0286	43,40 %	4,82 %
9016	376740,1999	543303,8313	166563,6314	44,21 %	4,91 %
9017	303585,619	462265,4804	158679,8614	52,27 %	5,81 %
9018	263492,1467	356736,2625	93244,11575	35,39 %	3,93 %
9019	276793,9692	390926,2879	114132,3187	41,23 %	4,58 %

Table 4: Mean income between 2011 and 2021 on Tromsø island, changes in income, in % and yearly average.

Postcode	Mean Income 2011	Mean Income 2021	Changes	Percent	Average
9020	288071,1777	435143,0885	147071,9107	51,05 %	5,67 %
9022	267491,2032	385794,9357	118303,7324	44,23 %	4,91 %
9024	274967,2342	385367,7091	110400,4749	40,15 %	4,46 %

Table 5: Mean income between 2011 and 2021 on the mainland of Tromsø, changes in income, in % and yearly average

Postcode	Mean Income 2011	Mean Income 2021	Changes	Percent	Average
9100	290380,351	424334,4711	133954,1	46,13 %	5,13 %
9101	298645,6438	404067,5171	105421,9	35,30 %	3,92 %
9102	330032,9498	468541,1173	138508,2	41,97 %	4,66 %
9104	287222,7006	483183,2356	195960,5	68,23 %	7,58 %
9105	287388,9607	439277,992	151889	52,85 %	5,87 %

Table 6: Mean income between 2011 and 2021on Kvaløya. Changes in income, in % and yearly average.

Postcode	Mean Income 2011	Mean Income 2021	Changes	Percent	Average
9027	230042,9366	345441,0399	115398,1033	50,16 %	5,57 %
9030	191688,8308	265550,9444	73862,11368	38,53 %	4,28 %
9040	201670,7253	349832,08	148161,3547	73,47 %	8,16 %
9042	210748,9679	210325,7	-423,2678899	-0,20 %	-0,02 %
9057	216067,9363	325070,5378	109002,6015	50,45 %	5,61 %
9106	222900,5208	344758,0256	121857,5048	54,67 %	6,07 %
9107	267700,25	417329,4211	149629,1711	55,89 %	6,21 %
9108	232543,5915	382744,2525	150200,6609	64,59 %	7,18 %
9110	279894,8405	384390,403	104495,5625	37,33 %	4,15 %
9118	235907,9427	307365,9133	71457,97066	30,29 %	3,37 %
9120	233345,88	304297,2029	70951,3229	30,41 %	3,38 %
9128	180465,5556	138306,4	-42159,15556	-23,36 %	-2,60 %
9130	228970,9459	311726,907	82755,96103	36,14 %	4,02 %
9131	225282,2843	383207,4126	157925,1283	70,10 %	7,79 %
9132	192308,6667	383325,7143	191017,0476	99,33 %	11,04 %
9135	220522,6	357679,8333	137157,2333	62,20 %	6,91 %
9136	158406,5455	253181,75	94775,20455	59,83 %	6,65 %
9140	300360,8621	383875,9231	83515,06101	27,80 %	3,09 %
9141	164201	563334,6667	399133,6667	243,08 %	27,01 %

Table 7: Mean income between 2011 and 2021in the districts of Tromsø, changes in income, in % and yearly average.

Postcode	Mean Wealth2011	Mean Wealth2021	Changes	Percent	Average
9006	956082,1126	1540584,38	584502,268	61,14 %	6,11 %
9007	958341,4792	1858596,841	900255,362	93,94 %	9,39 %
9008	964101,7739	1104248,706	140146,932	14,54 %	1,45 %
9009	920199,1174	2661310,901	1741111,78	189,21 %	18,92 %
9010	602930,0776	1284654,811	681724,734	113,07 %	11,31 %
9011	873067,4706	1690389,275	817321,804	93,61 %	9,36 %
9012	866093,706	2395571,032	1529477,33	176,59 %	17,66 %
9013	1189107,423	2212062,861	1022955,44	86,03 %	8,60 %
9014	836420,7121	1590417,752	753997,04	90,15 %	9,01 %
9015	555284,3609	1185819,713	630535,352	113,55 %	11,36 %
9016	904934,5181	1864789,654	959855,135	106,07 %	10,61 %
9017	559399,0051	1176192,067	616793,062	110,26 %	11,03 %
9018	532979,8351	940107,7732	407127,938	76,39 %	7,64 %
9019	816956,0628	1541086,932	724130,869	88,64 %	8,86 %

Table 8: Mean wealth between 2011 and 2021 on Tromsø island, changes in wealth, in % and yearly average.

Postcode	Mean Wealth2011	Mean Wealth 2021	Changes	Percent	Average
9020	788434,9587	1349954,603	561519,644	71,22 %	7,12 %
9022	582867,9732	1250448,036	667580,062	114,53 %	11,45 %
9024	731145,6342	1315367,668	584222,034	79,91 %	7,99 %

Table 9: Mean wealth between 2011 and 2021 on the mainland of Tromsø, changes in wealth, in % and yearly average

Postcode	Mean Wealth2011	Mean Wealth2020	Changes	Percent	Average
9100	663839,8061	1277098,675	613258,869	92,38 %	9,24 %
9101	801410,8099	1312045,387	510634,577	63,72 %	6,37 %
9102	742149,8103	1345088,238	602938,427	81,24 %	8,12 %
9103	795146,1167	1198007,256	402861,139	50,67 %	5,07 %
9105	723376,0872	1119161,511	395785,424	54,71 %	5,47 %

Table 10: Mean wealth between 2011 and 2021on Kvaløya. Changes in wealth, in % and yearly average.

Postcode	Mean Wealth 2011	Mean Wealth 2021	Changes	Percent	Average
9027	481495,2009	1049670,844	568175,6432	118,00 %	11,80 %
9030	571677,8058	994791,0652	423113,2595	74,01 %	7,40 %
9040	648561,8491	865470,25	216908,4009	33,44 %	3,34 %
9042	511320,3688	1000832,636	489512,2676	95,73 %	9,57 %
9056	927315,1429	753012,8571	-174302,2857	-18,80 %	-1,88 %
9106	628241,5864	1125877,986	497636,3999	79,21 %	7,92 %
9107	773007,932	1452469,378	679461,4459	87,90 %	8,79 %
9108	681132,5089	1264717,685	583585,1763	85,68 %	8,57 %
9110	985581,784	1412968,189	427386,4052	43,36 %	4,34 %
9118	684254,2673	1823444,138	1139189,871	166,49 %	16,65 %
9120	977424,8772	1483054,103	505629,2263	51,73 %	5,17 %
9128	152183,5	301212,3333	149028,8333	97,93 %	9,79 %
9130	690671,15	1262276,421	571605,2711	82,76 %	8,28 %
9131	597900,5882	1225816,371	627915,7826	105,02 %	10,50 %
9132	72008,5	486109	414100,5	575,07 %	57,51 %
9135	450270,6	671109,6	220839	49,05 %	4,90 %
9136	151777,2	679676,6667	527899,4667	347,81 %	34,78 %
9140	2836911,52	2053236,87	-783674,6504	-27,62 %	-2,76 %
9141	456968	696065,5	239097,5	52,32 %	5,23 %

Table 11: Mean wealth between 2011 and 2021in the district, changes in wealth, in % and yearly average.

Postcode	Mean Price 2011	Mean Price 2021	Change	Percent	Average
9006	3 123 568	5 428 124	2 304 556	73,78 %	7,38 %
9007	3 508 750	4 759 136	1 250 386	35,64 %	3,56 %
9008	2 542 688	3 876 319	1 333 631	52,45 %	5,24 %
9009	3 008 133	4 700 419	1 692 286	56,26 %	5,63 %
9010	2 347 888	4 217 193	1 869 305	79,62 %	7,96 %
9011	3 104 537	5 713 367	2 608 830	84,03 %	8,40 %
9012	3 816 000	5 287 723	1 471 723	38,57 %	3,86 %
9013	3 844 848	6 727 408	2 882 560	74,97 %	7,50 %
9014	3 746 047	6 346 867	2 600 821	69,43 %	6,94 %
9015	3 203 125	5 972 791	2 769 666	86,47 %	8,65 %
9016	3 576 077	5 710 373	2 134 296	59,68 %	5,97 %
9017	3 026 763	5 253 272	2 226 509	73,56 %	7,36 %
9019	2 852 206	4 691 941	1 839 735	64,50 %	6,45 %

Table 12: Mean house transaction price 2011 and 2021, Tromsø as a whole and postcodes on Tromsø island, 9006 to 9019

Postcode	Mean Price 2011	Mean Price 2021	Change	Percent	Average
9020	3 126 740	5 241 663	2 114 924	67,64 %	6,76 %
9022	3 628 429	4 985 147	1 356 718	37,39 %	3,74 %
9024	3 213 200	4 562 835	1 349 635	42,00 %	4,20 %

Table 13: Mean house transaction price 2011 and 2021, Tromsø as a whole and postcodes on the mainland of Tromsø, 9020,9022 and 9024.

Postcode	Mean Price 2011	Mean Price 2021	Change	Percent	Average
9100	2 905 556	5 736 887	2 831 331	97,45 %	9,74 %
9101	3 458 095	6 324 821	2 866 726	82,90 %	8,29 %
9102	3 215 909	6 349 216	3 133 307	97,43 %	9,74 %
9103	4 150 000	4 277 083	127 083	3,06 %	0,31 %
9105	3 380 909	5 805 478	2 424 569	71,71 %	7,17 %

Table 14: Mean house transaction price 2011 and 2021, Tromsø as a whole and postcodes on Kvaløya, 9100,9101,9102 and 9105

Postcode	Mean Price 2011	Mean Price 2021	Change	Percent	Average
9027	1 884 318	3 928 571	2 044 253	108,49 %	10,85 %
9030	652 000	1 110 769	458 769	70,36 %	7,04 %
9034	1 850 000	4 600 000	2 750 000	148,65 %	14,86 %
9057	1 700 000	3 231 667	1 531 667	90,10 %	9,01 %
9104	3 159 074	5 510 748	2 351 674	74,44 %	7,44 %
9106	1 762 500	2 806 250	1 043 750	59,22 %	5,92 %
9107	2 428 721	4 752 085	2 323 364	95,66 %	9,57 %
9108	1 598 824	2 578 571	979 748	61,28 %	6,13 %
9109	2 515 938	3 600 694	1 084 757	43,12 %	4,31 %
9110	1 796 429	2 960 647	1 164 218	64,81 %	6,48 %
9118	2 680 000	2 225 000	- 455 000	-16,98 %	-1,70 %
9131	1 485 714	4 300 000	2 814 286	189,42 %	18,94 %

Table 15: Mean house transaction prices 2011 and 2021, Tromsø as a whole and postcodes in the districts of Tromsø.

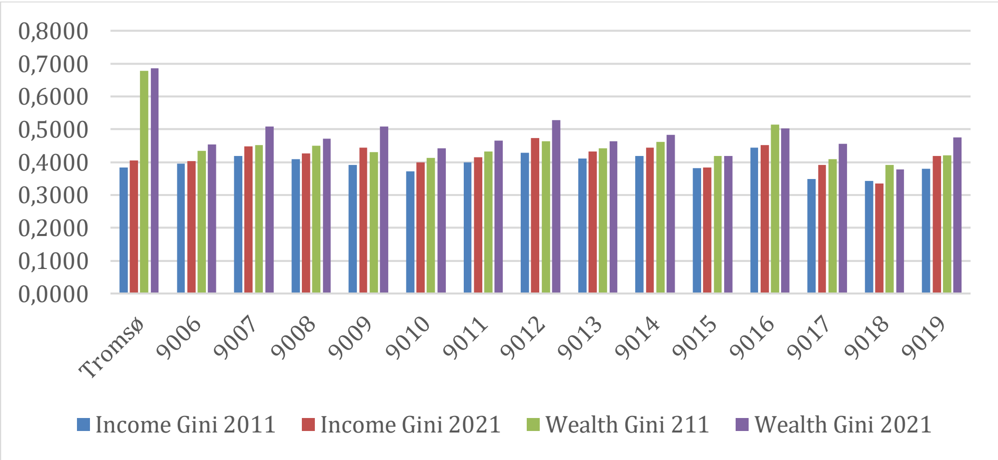

Diagram 3: Gini-coefficient calculated using income before tax and wealth in Tromsø and Tromsø island

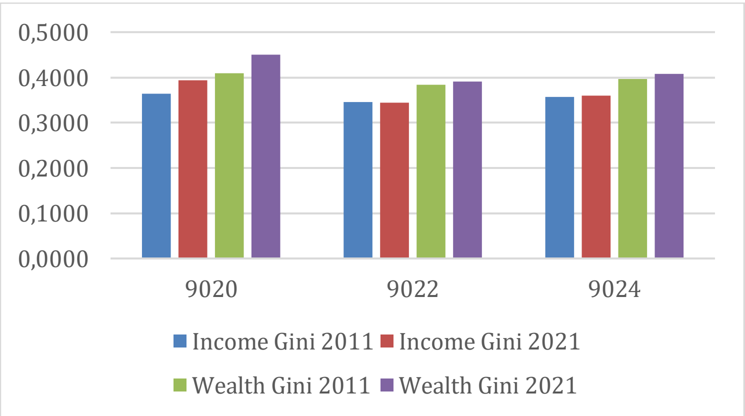

Diagram 4: Gini- the coefficient is calculated with income before tax and wealth mainland of Tromsø

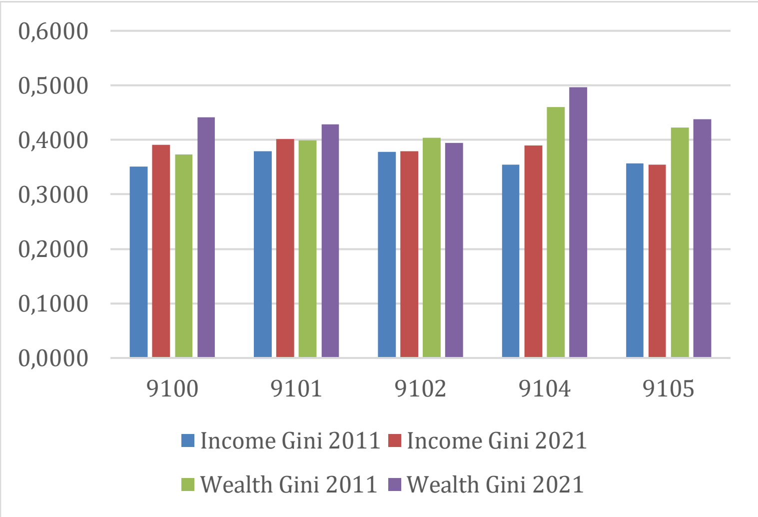

Diagram 5: Gini coefficients calculated by income before tax and wealth on Kvaløya.

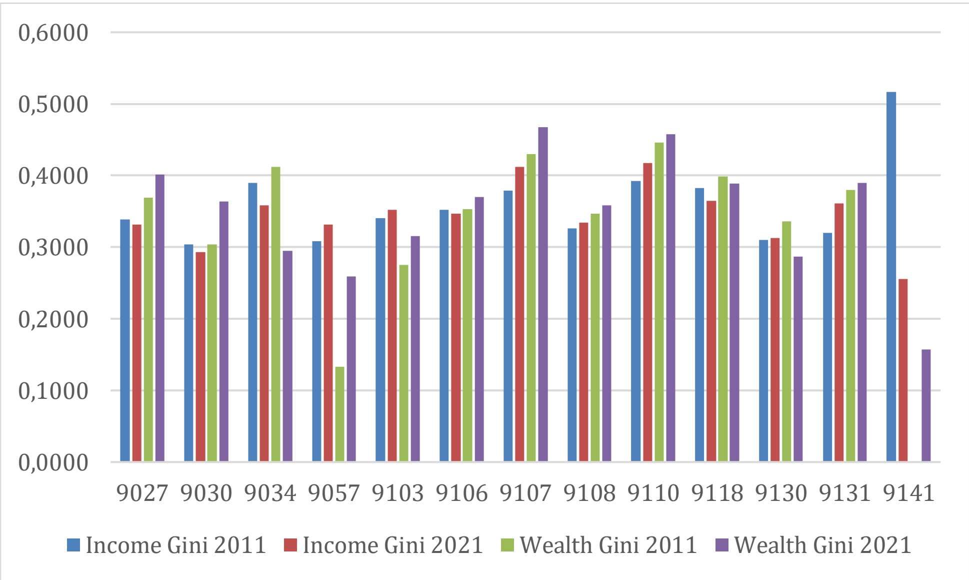

Diagram 6: Gini-coefficient calculated by income before tax and wealth in the districts.

References

Aarberge, R., Mogstad, M., Vedstad, O.L., and Vestre, A.(2021) *Økonomiske ulikheter I Norge i det 21. århundre.* Statistisk sentralbyråd. Retrieved from: https://www.ssb.no/inntekt-og-forbruk/inntekt-og-formue/artikler/okonomisk-ulikhet-i-norge-i-det-21.arhundre

Adarov, A.,(2022). *Global income inequality and the COVID-19 pandemic in three charts.* World Bank Blogs. Retrieved from: https://blogs.worldbank.org/developmenttalk/global-income-inequality-and-covid-19-pandemic-three-charts

Andersson, R. & Bråmå, Å. (2004). *Selective migration in Swedish distressed neighbourhoods: Can area-based urban policies counteract segregation processes?* Housing Studies, 19(4), pp. 517–539.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2014). *Causality and endogeneity: Problems and solutions.* The Oxford Handbook of Leadership and Organizations, pp. 93–117.

Bailey, N. & Livingston, M. (2007*). Population Turnover and Area Deprivation*. Policy Press: Bristol, England. ISBN: 9781861349750

Benard, S and Willer, R (2007). *A Wealth and Status-Based Model of Residental Segregation*. The Journal of Mathematical Sociology, 31(2), pp 149-174.

Berg, A., & Ostry, J. D. (2011). *Inequality and unsustainable growth: Two sides of the same coin?* IMF Staff Discussion Note, No. 11/08.

Carnerio, P. Heckman, J.J. Vytlacil, E. (2003) *Understanding What Instrumental Variables Estimate: Estimating Marginal and Average Returns to Education.* The American Economic Review, 101, pp 2754-2781.

Clark, W. A. V. (1992). *Residential preferences and residential choices in a multiethnic context*, Demography, 29, pp. 451–466.

Dahl, E., Elstad, J. I., Hofoss, D. Mollard, M.M (2006) *For whom is income inequality most harmful? A multi-level analysis of income inequality and mortality in Norway*. Journal of social science & Medicine 63(10), 2562-2574. Retrieved from: https://doi.org/10.1016/j.socscimed.2006.06.002

Duncan, O. & Duncan B. (1955*). A methodological analysis of segregation indexes. The* American Sociological Review, 20(2), 210–217. Retrived from: https://doi.org/10.2307/2088328

Ellen, I. G., & Turner, M. A. (1997). *Does Neighborhood Matter?* Assessing Recent Evidence. Housing Policy Debate, 8(4), 833-866. Retrieved from: https://doi.org/10.1080/10511482.1997.9521280

Fraser, S., Myrvang Ro, H., Solheimsnes, P.A., Sagmoen, I. and Hopland, S.(2021) *Halvparten av bedriftene som fikk pandemi-støtte gjorde det bedre i 2020 enn året før.* E24

Retrived from https://e24.no/naeringsliv/i/5GWdVE/halvparten-av-bedriftene-som-fikk-pandemi-stoette-gjorde-det-bedre-i-2020-enn-aaret-foer

Galster, G. & Killen, S. (1995). *The geography of metropolitan opportunity: a reconnaissance and conceptual framework*, Housing Policy Debate, 6(1), pp. 7–43.

Hatterem, A. (2022) *Færre med lavinntekt*. Statisk Sentralbyrå. Retrived from: https://www.ssb.no/inntekt-og-forbruk/inntekt-og-formue/statistikk/inntekts-og-formuesstatistikk-for-husholdninger/artikler/faerre-med-lavinntekt

Heckman JJ (1976*). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. Annals of Economic and Social Measurement,* 5(4), 475–492.

Logan, J. R. and Molotch, H. L. (1987). *Urban Fortunes: The Political Economy of Place,* University of California Press.

Logan, J. R., & Schneider, M. (2010). *Segregation and the city.* The American Journal of Sociology, 115(2), pp 395-397.

Massey, D. S. & Denton, N. A. (1993). *American Apartheid: Segregation and the Making of the Underclass.* American Journal of Sociology, 92(2), pp 329-357.

Massey, D. S., Gross, A., & Eggers, M. (1991). *Segregation, the concentration of poverty, and the life chances of individuals*. Social Science Research, 20, pp. 397–420.

Oslo Met(2021) *Korona har forsterket den økonomiske ulikheten i Norge.* Retrived from: *https://www.oslomet.no/forskning/forskningsnyheter/korona-har-forsterket-den-okonomiske-ulikheten-i-norge*

Reardon, S. F., & Bischoff, K. (2011). *Income inequality and income segregation.* American Journal of Sociology, 116(4), pp 1092-1153.

Ross, S. L., & Yinger, J. (2002). *The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair-Lending Enforcement.* MIT Press. Retrived from: https://www.jstor.org/stable/41795129

Rowlingson, K. (2011) *Dose income inequaleties causes health and social problem?* Joseph Rowntree Foundation

Statistisk sentralbyråd.(2021) *Inntekts- og formuesstatistikk for husholdninger: 09114: Inntektsfordelingen belyst ved ulikhetsmålene ginikoeffisient og P90/P10, etter statistikkvariabel region og år.* Retrieved from https://www.ssb.no/statbank/table/09114/tableViewLayout1/

Schelling, T. C. (1971*). Dynamic models of segregation*. Journal of Mathematical Sociology, 1 , pp143-186. Retrived from: doi:10.1080/0022250X.1971.9989794]

Toomet, O., & Henningsen, A. (2008). Sample Selection Models in R: Package sampleSelection. *Journal of Statistical Software*, *27*(7), 1–23. Retrived from: https://doi.org/10.18637/jss.v027.i07

Tuv,N(2019) *Disse kommunene har størst og minst inntektsulikhet.* Statisk Sentralbyrå. Retrived from: https://www.ssb.no/inntekt-og-forbruk/artikler-og-publikasjoner/disse-kommunene-har-storst-og-minst-inntektsulikhet

Van Ham, M., Tammaru,T., De Vuist, E. and Ziwers, M. (2016) *Spatial Segregation and Socio-Econimic mobility in European Cities*

Wessel, T. (2015). *Economic segregation in Oslo: polarization as a contingent outcome.* Musterd, M. van Ham, & T. Tammaru (Eds.), Socio-Economic Segregation in European Capital cities: East Meets West (pp. 132-155). Oxford: Routledge.

Wessel, T., & Nordvik, V. (2019). *Mixed neighbourhoods and native out-mobility in the Oslo region: The importance of parenthood.* Urban Studies, 56(5), pp 885–905. Retrieved from https://doi.org/10.1177/0042098018768452

Wilson, W. J. (1996). *When Work Disappears: The World of the New Urban Poor.* New York: Knopf