Stockholm School of Economics
Department of Economics
659 Degree Project in Economics
Spring 2017

Segregation in Stockholm: Birds of a Feather?

Economic Segregation in Stockholm County 2000–2014

Carl-Johan Backman (23076)

Abstract

Rising levels of income inequality in conjunction with increased global flows of migrants have created a need for extending the notion of segregation to include economic measures as well, in addition to ethnicity. In this thesis we explore the economic segregation in the municipalities of Stockholm County, during the years 2000 to 2014. We analyze the variance of income within and between neighborhoods. Moreover, we calculate the dissimilarity index for the highest and lowest income groups, and run a regression on a model with the index as the dependent variable.

We show that the economic segregation has increased for most municipalities during the time period studied. Also, we find indications that migrations patterns, people moving in and out of neighborhoods, tend to reinforce existing economic segregation. People with higher socioeconomic status move from neighborhoods with low socioeconomic status and into areas of higher status. Similarly, inhabitants of lower socioeconomic status tend to move out of areas with higher status and into areas with lower status.

Keywords: Spatial segregation, economic segregation, sorting, income inequality, gentrification, residualisation
JEL: R23, R29

Supervisor:	Anders Olofsgård
Date submitted:	May 15, 2017
Date examined:	June 12, 2017
Discussant:	Erik Stannow Lind
Examiner:	Örjan Sjöberg

Contents

1 Introduction

Our world is getting smaller. Globalization is shortening distances, widening markets and opening up borders; a development which has lead to a greater than ever wealth and prosperity. However, globalization has not come without its challenges. The backlash of globalization has become obvious during recent years, especially in the Western world where political movements against globalization have surged. This seemingly contradictory reaction to a wealthier and healthier world can in part be explained by the gap between the richest and the poorest. People are on average richer, however, the difference between the richest and the poorest is increasing, or in other words, we are experiencing a rise in income inequality (Piketty and Saez, 2003; Piketty, 2014).

In this thesis we investigate if the gap between rich and poor is manifesting itself into a widening spatial separation of people as well, i.e. economic segregation, a phenomenon of great interest for both the academic society and policy makers, as it is believed to be a hotbed for social unrest (Musterd, 2005). In particular, we want to investigate the economic segregation in Stockholm, Sweden.

Sweden is one the most equal countries in the OECD, however, during the last decades income inequality has risen (OECD, 2015). In conjunction with fast urban growth and the relatively large number of migrants entering the country applying for asylum, the topic of segregation in general, including economic segregation, will most likely remain a significant one for both Swedish politicians and the Swedish people, for a long time.

To gain a greater understanding of the dynamics of economic segregation in Sweden, we will in this paper study how economic segregation has developed in the municipalities of Stockholm County, during the years of 2000 to 2014. We approach this with two different methods. First, we compare the variance of income *within* and *between* neighborhoods, in order to see if the gap is widening across the neighborhoods. Second, we calculate a dissimilarity index (DI) for economic segregation, for all municipalities in Stockholm county. Using this index as the dependent variable, we then run a regression analysis in order to find out how certain factors affect the economic segregation. We are chiefly interested in the effect migration flows (people moving in and out of the neighborhoods) have on the economic segregation.

1.1 Background

Segregation is a well-known phenomenon and it has been studied since the beginning of the last century. Historically, segregation has usually meant ethnic segregation, i.e. the physical separation of people from different ethnic groups in a city, commonly measured by calculating DI for two ethnic groups (Taeuber and Taeuber, 1969). However, we would argue that by looking at ethnicity only we will miss important aspects of segregation. The composition of ethnic groups in a person's nearby area will only tell one part of the story of his or hers living situation.

As a response to this, several studies have been published in recent years which broaden the definition of segregation to incorporate socioeconomic segregation as well (Musterd, 2005; Tammaru et al., 2016). Combining ethnic segregation with economic segregation becomes especially interesting since some of the current research indicates that ethnic segregation has stagnated or even decreased in the Western world, while income inequality is increasing at the same time (Musterd, 2005; Malmberg et al., 2016; Andersson and Kährik, 2016).

Segregation is important to assess as it complicates processes of integration and, in addition, is usually associated with less participation in the society in general, be that in the labor market, politically or in other areas. This holds true also in Stockholm, and the topics of segregation and integration are widely debated subjects. Nevertheless, segregation, both ethnic and economic, has proven difficult to evaluate and current research does not give a clear answer on how the segregation has changed in Sweden during the last decade, e.g. Malmberg et al. (2016); Bolling (February 8 2016) claim there are indications of decreasing segregation, while Andersson and Magnusson Turner (2014); Tottmar (November 22 2014,S); Andersson and Kährik (2016) argue the opposite, i.e. segregation is increasing.

By extending the notion of segregation to include an economic aspect as well, we can provide a more comprehensive and accurate view of people's living situation in different neighborhoods. We can add to previous research by shedding more light on a topic which is highly relevant today and will likely be for some time to come.

We have chosen to limit the scope to include Stockholm County only for several reasons. First, segregation is typically accentuated in cities. They are economic engines, while at the same time being home to the greatest diversity and inequalities. In addition, Stockholm is the most densely populated city in Sweden, thus being a good basis for researching the dynamics of segregation. Moreover, Stockholm is divided into a set of geographical areas,

base areas, which are suitable for studying segregation due to their small size. At the end of the day, understanding segregation is about understanding the everyday life of people. Thus, fine granularity of the data and good proxies for a person's living situation are key in gaining that understanding.

1.2 Research Questions

Ahead of stating the research questions we try to answer in this thesis, we will mention some of the definitions we use. First, we define economic segregation as the *spatial separation of people from different income groups*. How these groups are defined is described in Section 3. Next, neighborhoods are the same thing as *base areas* in this thesis. Base area is a small geographic region, typically corresponding to a few blocks, which all municipalities in Stockholm can be divided into. More on base areas in Section 3. Finally, we have chosen to measure economic segregation by looking at the distribution of people in the lowest and highest income group (see Section 3) across different base areas.

With these definitions in mind we are now ready to present the two research questions we try to answer in this thesis.

RQ1 How has the economic segregation in Stockholm changed during the years 2000 and 2014?

RQ2 How do migration flows, in and out of the base areas, affect the economic segregation in Stockholm?

1.3 Outline

In Section 2, we give an overview of the research area of segregation and present the most relevant work conducted within the area. Then, we give a description of the data used in this thesis in Section 3 and the method we use to evaluate our research questions above, in Section 4. Subsequently, in Section 5, we present and comment on our findings and results. The validity of our methodology and results are discussed in Section 6. Lastly, we conclude the thesis with the most important findings and possible future work, in Section 7.

2 Related Work

In this section we present relevant research about segregation and inequality. This section is intended to work as both a background to the subject of segregation and provide an overview of what has been done in the past as well as the current status of the research.

The section is divided into three subsections. The first section deals with the issue of measuring segregation, possible methods as well as the benefits and drawbacks that come with each method. Second, we give a brief history of segregation in the Western world and we conclude with a more focused view on segregation in Sweden.

2.1 Measuring Segregation

Segregation has been studied for quite some time. A starting point for segregation studies could be said to be the research performed by Booth (1903) in the 19^{th} century London. The author carefully accounts for the life and labor of the people of London, showing the residential separation of people in different income classes. Since then, the world has changed significantly and the way segregation studies are performed with it.

How to measure segregation has been debated throughout the entire last century and papers are still published which propose new ways of measuring it. These continuous attempts to develop and improve the methodologies are understandable and are in essence a good practice. Challenging and improving methodologies lies at the core of all research. In addition, segregation is a very complex issue. The underlying observations, people and their living situation, are changing quickly in this highly dynamic and globalized world. This, in turn, translates into changing conditions for what indicators are important to look at. The downside of this great spread in how to measure segregation, however, is a research area which is difficult harmonize and make comparable across cities and countries.

2.1.1 The Chicago School

As mentioned before, historically, the focus in segregation research has been on studying proportions and spatial separation of ethnic groups. The so called *Chicago School* played a major role in this work, where the research focus was the situation between black and whites in urban US. The school was the collective work of researchers from several different university in

the US and became one of the biggest research bodies which conducted work in sociology, especially urban sociology, in the beginning of the last century.

During the first half of the last century the segregation research experienced an inflation of different indices (although using indices is still very common in segregation studies). Duncan and Duncan (1955) contributed by showing, in a rigorous, mathematical manner, that all the information contained in the indices at the time basically measured the same thing and were strongly correlated. In fact, they show that all indices can be described as functions of the what they call the *segregation curve*. Moreover, the authors managed to highlight how this confusion on measurement created difficulties in validating and making the research comparable. The paper concluded that, among the indices they studied, it is sufficient to use the dissimilarity index.

DI measures how evenly distributed two groups are across a set of geographical areas. Meaning that if both groups are somewhat evenly distributed, the index scores low, with 0 being the minimum. Contrary, when the groups are unevenly distributed across the areas the index scores high, with maximum being 1. E.g. if there are 10% of group A in a city and 90% of group B, then the same distribution should hold for all sub-geographical areas within the city as well, in order for the index to score low. The actual value of the index can be interpreted as *the share of the minority group which would have to reallocate in order to achieve an even distribution between the groups*. See Section 4.2.1 for a more detailed description of the index.

The Critique The findings of Duncan and Duncan (1955) had a major impact on the research society. For many years DI was the *de facto* standard way of measuring segregation and it is still frequently used (Duncan and Duncan, 1957; Taeuber and Taeuber, 1969; Malmberg et al., 2016). Nevertheless, the index is not without its drawbacks, which has been pointed out in different studies. The first blow dealt to the DI came some 20 years after Duncan and Duncan (1955)'s paper, with the publication of Cortese et al. (1976). Duncan and Duncan (1955) do mention some difficulties with using DI to measure segregation in their paper, however, these are further elaborated on in Cortese et al. (1976). In addition, the latter paper gives advice on how to deal with the drawbacks.

In short, the critique can be summarized into four parts: i) the inherent assumption of the dissimilarity index that a completely even distribution between two groups represents the least possible segregation is not correct, but instead a random distribution is more accurate; ii) it is difficult to

compare the index between geographical areas, since it is affected by the proportions of the groups in the respective area (i.e. it is not compositionally invariant); iii) the index is affected by the number of observations in the area unit of interest and finally; iv) the index interpretation that the value represents the share of people in the minority group that would have to reallocate is cumbersome, since it does not account for those who will move into the vacant homes left behind the hypothetically reallocated minority group members. The authors argue that it would be more reasonable to look at how the people between the different group should be exchanged between the geographical units of interest instead.

To remedy this, Cortese et al. (1976) present a new index—*the standard score index*—which accounts for the above mentioned drawbacks. However, their methodology and index was never really put into practice and there were also other papers published which replied to the critique and defended DI (Taeuber and Taeuber, 1976; Massey, 1978). Additional critique to index has been that segregation is a too complex phenomenon, thus condensing and describing it by a single number is not sufficient. Also, it is limited to studying the relationship between two groups (Quinn and Pawasarat, 2003).

After Duncan and Duncan (1955), the next pivotal paper from the Chicago School came through Massey and Denton (1988) and their work on the dimensions of segregation. The authors perform a thorough literature review and a cluster analysis of 20 different indices, used for measuring segregation. From this analysis they are able to define five dimensions of segregation. The authors argue that using only one index implies looking at only one of these dimensions, which is not sufficient. The dimensions defined in Massey and Denton (1988) are *evenness*, *exposure*, *concentration*, *centralization* and *clustering*. They are all briefly explained below.

Evenness The most classical of the dimensions. It relates to how minority groups are distributed across different geographical units of interest. Recognized from the discussed above, this is what DI captures. Others indices include the *Gini* coefficient, *entropy* and the *Atkinson index* (Theil and Finizza, 1971; Theil, 1972; Atkinson, 1970; Dixon et al., 1987).

Exposure This dimension deals with interaction between members of different groups. To be more specific, it deals with the possibility or probability of members from the minority group to interact with members from the majority group. Exposure is correlated with evenness, although not dependent on the relative sizes of the groups. Massey and Denton (1988) make

the argument that while a minority group can be distributed evenly across areal units, the inhabitants can still experience segregation if there is no interaction between the groups anyway. Exposure can be measured by the *isolation index* and the *interaction index* (Bell, 1954).

Concentration Concentration looks at the relative size of the area occupied by the minority group in a geographical unit of interest. E.g. the smaller share of the area occupied by the minority group, the more segregated the geographical unit is. Concentration is commonly described by the *delta index* (Hoover, 1941).

Centralization Centralization is similar to concentration, but deals with how close to the urban center members of the minority group reside. To be interpreted as the further away in the periphery the minority group lives, the higher the segregation. Centralization can be measured with the *relative centralization index* as proposed by Duncan and Duncan (1955).

Clustering The fifth dimension, clustering, is different from the four previous dimensions, in the sense that it does not measure how minority and majority groups are distributed across geographical areas. Instead, it measures whether or not areas with minority groups are clustered together. Examples of indices measuring clustering are the *index of absolute clustering* and *index of spatial proximity* (Geary, 1954; Dacey, 1965).

2.1.2 Post-Chicago-School Era

The use of defining segregation along the above five dimensions persists to this day, as does the use of DI. However, like so many before them, Massey and Denton (1988) failed to create a consensus on which index to use for capturing the segregation. That quest is still ongoing (Grannis, 2002; Hutchens, 2004; Echenique and Fryer, 2007; Farber et al., 2015).

One interesting approach of tackling the problem of measurements, which has gained more attention in recent years, is the one used in Malmberg et al. (2016). Instead of creating a new index they look at a different way of defining the neighborhoods used when studying the segregation. Typically, the neighborhoods used in segregation research are geographical areas defined by some statistical authority in the country. What Malmberg et al. (2016) do instead is to define neighborhoods from individuals. More precisely, they define an individual's neighborhood as the k geographically

closest persons. The authors argues that the ordinary way of analyzing segregation, by studying neighborhoods or geographical areas defined by some statistical agency, does not necessarily depict the everyday life of people. Using individualized neighborhood is their solution to bring the research closer to the individual and thus, providing a more accurate understanding.

2.2 Segregation in the West

North America Aside from the pioneering work of Booth (1903), most of the early segregations studies originates from North America and, as mentioned above, the Chicago School. The studies are usually concerned with how the black population was distributed throughout Chicago (which held the second largest population of blacks in the US, after New York), in relation to the white population. Duncan and Duncan (1957) made a detailed study of Chicago and concluded that segregation was mainly driven by a large influx of non-whites to the same areas. However, the study also shows that the influx of people with low socioeconomic standard did not lower the socioeconomic measures of the areas of interest, as expected. Instead other factors managed to cancel out that effect. A similar study was performed in Taeuber and Taeuber (1969). The authors do a vast analysis of segregation across 10 cities in the US. By using DI they are able to show that high segregation between black and whites is not a phenomenon limited to only a few cities. Higher levels of segregation could be found in all the cities studied and was at the time increasing.

However, more recent studies have also been conducted where DI is criticized for not giving a correct picture of segregation and other methods are applied instead. A good example of this is the work of Quinn and Pawasarat (2003). The authors give an extensive overview of the segregation in the US by analyzing block-level data for the 50 largest US cities. They find large differences between the cities. In general, they find that cities in the Midwestern and the Southeastern states experience relatively low levels of segregation, while cities known for being dynamic and diverse such as San Francisco, Chicago and New York, score high on segregation.

Europe Compared to the US, European cities have in general lower segregation, both in terms of ethnic as well as socioeconomic segregation (Musterd, 2005). Europe differs from US since mono-ethnic areas are very unusual in European cities. E.g. there are spatial clusters with large concentrations of non-Europeans, but these clusters are in general comprised of people with very different ethnicity. In his paper, Musterd (2005) also argues that within

Europe, differences in ethnic and socioeconomic are large. In addition, it is not a trivial task to compare the countries due to very different welfare systems, different historical and cultural legacies and so on. However, cities in the UK and Italy are on top of the list of the most segregated, while the Nordic countries and Germany show lower figures.

Musterd (2005) explores correlations between inequality and different types of segregation. In particular, the author finds indications that the correlation between socioeconomic segregation and ethnic segregation is not as strong as it has previously been believed. The correlation between socioeconomic segregation and socioeconomic inequality is higher, but not perfect. Moreover, socioeconomic segregation tends to be lower than ethnic segregation. Musterd (2005) concludes the paper with four important drivers of socioeconomic segregation: i) the welfare system; ii) the labor market and economy; iii) network and stigmatization processes and; iv) the personal characteristics of the individual. These findings are supported by Marcińczak et al. (2015), who argue that the legacy of a strong state and welfare system is the main explanation of the discrepancy in segregation level between Europe and the US.

Recent research which contradicts that of the above, at least in parts, can be found in Tammaru et al. (2016). The book compiles the results of 13 different studies performed in 13 different European cities (Amsterdam, Budapest, Vienna, Stockholm, Oslo, London, Vilnius, Tallinn, Prague, Madrid, Milan, Athens, and Riga). They find evidence of increasing socioeconomic segregation in almost all cities, with primarily two factors driving the development - globalization and growing inequalities. This goes against the previous consensus surrounding the drivers, e.g. that of the size of the welfare system, degree of state intervention in the housing market.

Furthermore, a recent and relevant study has been published by Bailey et al. (2017). They investigate and compare segregation in the Dutch cities of Amsterdam and The Hague, using the *delta index*, which measures segregation along the concentration dimension (Massey and Denton, 1988). In particular, they study the impact that *in situ* social mobility has on segregation as well as migration to and from the city. They find evidence of increasing segregation in both cities but with different drivers. The segregation in Amsterdam was driven by a combination of in-migration and *in situ* social mobility, while The Hague's segregation rose mainly because of in-migration.

2.3 Segregation in Sweden

There have been several studies of segregation in Sweden and especially Stockholm, during the last decades and most of them show evidence of an increasing segregation, both ethnic and socioeconomic, i.e. similar to the rest of Europe (Musterd, 2005; Andersson and Magnusson Turner, 2014; Magnusson Turner and Hedman, 2014; Andersson and Kährik, 2016; Marcińczak et al., 2016). In fact, Tammaru et al. (2016) find Stockholm to be one of the most segregated cities in today's Europe, despite being an egalitarian society with a strong state. Sweden differs from many countries in Europe with its diverse foreign population, as it is more common for the European cities to have a few nationalities which dominates the population of foreign born (Musterd, 2005). However, also when comparing Stockholm and Oslo, which are two rather similar cities in terms of the state, welfare system and so on, there are significant differences to be found (Wessel, 2016).

Murdie and Borgegård (1998) present a model for explaining segregation in Sweden which consists of both global and local factors. Globally, the authors point to increased and greater diversity of migration flows. Depending on these flows the context of Sweden will change, and policies will change with it. Locally, characteristics of the immigrants, attitudes towards immigrants, the labor market and the housing market will determine the spatial segregation. Using their model they show that the segregation in Stockholm increased during the years 1969-95.

Moreover, when trying to understand the housing situation in Sweden it is important to understand a housing policy which was initiated during the 1960's, the so called *Million Program*, as it is a reoccurring explanation for the segregation in Swedish cities. To tackle a severe crisis with lack of housing, the Swedish government made large investments in mass producing new housing. Through this, the city centers were renewed which pushed up prices and thus, made accommodation near the city center difficult for lower socioeconomic classes. At the same time, cheaper accommodation was normally built in the suburbs (Murdie and Borgegård, 1998).

The housing market was historically controlled by the state in Sweden but since 1990's there has been a liberalization of the market, although compared to Europe, a very large share is still public housing (Andersson and Kährik, 2016). In their paper, Andersson and Magnusson Turner (2014) argue that this liberalization is one of the reasons Sweden see such high segregation today. I.e. segregation rose when the public sector changed its policies for supporting socially mixed neighborhoods, while at the same time reducing subsidies for housing.

There have also been recent studies exploring the consequences of segregation in Sweden. An example is the work of Andersson and Malmberg (forthcoming), where the authors show that the neighborhood context into which you are born clearly affect your future life. In particular, they find that the risk of a person to end up in poverty, increases with the share of social allowance recipients present in the neighborhood where one grows up.

In a study of ethnic segregation in Stockholm, Malmberg et al. (2016) present results which partly show that segregation might be decreasing. They study the ethnic segregation both by looking at the variation of neighborhood composition and by calculating DI for individualized neighborhoods. The paper shows somewhat mixed results because when looking at segregation along the dimension of unevenness, using DI, segregation seems to have decreased in Stockholm. At the same time, variation in population composition has increased between neighborhoods, which would be an indication of increased segregation. This again highlights the difficulties of measuring segregation accurately.

Next, we continue with describing the data we have used in this thesis to analyze the segregation in Stockholm County.

3 Data

With this work, we aim at analyzing the economic segregation for the municipalities in Stockholm County. The general methodology applied here for doing that is to aggregate data on neighborhood level, in order to get municipality level results. The importance of drilling down to a micro perspective is grave when studying segregation, since outcomes on micro and macro levels can vary significantly (Schelling, 1978). Moreover, as has been mentioned already, in this thesis we use *base areas* as our definition of neighborhoods. Base areas are small, geographical units, defined by the local statistics department in Stockholm County.

However, important to note when studying different neighborhood effects in some geographical areas, as we do here with base areas, is the possibility of introducing bias through the *modifiable areal unit problem (MAUP)*. Previous studies have shown that neighborhood effects can operate on different geographical levels, yielding different results depending on how neighborhoods are defined and to what level data is aggregated (Openshaw, 1984; Andersson and Musterd, 2010). This limitation is addressed in Section 6.

Stockholm County consists of 26 municipalities and 1164 base areas. Some descriptive data on the size of the geographical areas can be seen in Table 1. From the table we see that the base areas vary in size of population, with the smallest having 20 inhabitants and the largest having around 9500, something which needs to be accounted for when using an index such as DI. More on how we handle caveats in the data is presented in Section 4.

Table 1: *Descriptive data for Stockholm municipalities and base areas, during the years 2000–2014.*

	Municipalities	**Base Areas**
n	26	1164
Mean size	75996	1783
Max size	911989	9568
Min size	8052	20

Furthermore, all data is aggregated from individuals and not from households. The data is panel data, grouped by geographical unit (municipality or base area) and year, in which the time series spans 15 years, from 2000 to 2014. The specified time period has not been chosen in order to study a particular event but is a consequence of limited accessibility to data. Hence, there is a degree of arbitrariness in the analysis, which will be assessed in Section 6. All income data is gross income from work, which includes all

income from employer, self-employment, pension and taxable benefits. Age wise, the subset of the population studied here is people who are 20 years or older.

All of the above have implications for interpreting the results when analyzing the data. These implications and how to handle them will be discussed further in Section 6 and in the next section, where we explain our methodology.

4 Method

Here we will describe how we go about to answer the questions posed in Section 1.2. In short, we investigate how economic segregation has changed during the years 2000–2014 by analyzing the variance of income, within and between base areas. We also calculate the dissimilarity index, based on the distribution of people in the lowest versus the highest income group, for all municipalities and study how it has changed over the same time period. Moreover, in an effort to understand what effect migration flows, in and out of the base areas, have on the economic segregation, we run a regression analysis with the index as the dependent variable.

These approaches, along with possible caveats, limitations and considerations, are described below.

4.1 Analysis of Variance

Our goal with studying the variance of income, within and between the base areas, is to understand how the economic segregation has changed over time. E.g. say that the variance in income on municipality level has increased over time. Using base areas as our fundamental geographical unit, that increase can stem from different sources. It might come from the fact that the variance of income *between* the base areas has increased over time, that the variance of income *within* the base areas has increased or a combination of both. In case of increasing segregation we would expect an increase on municipality level to be driven mainly by an increase between the base areas and not from within. I.e. base areas become more homogeneous in terms of income classes while the gap between the rich and the poor base areas increases.

Important to note, however, is that all values we have for the base areas are aggregated, i.e. we do not have individual data for the inhabitants. This means we cannot calculate the complete variance within the base areas. We are instead forced to use a proxy. For doing this we utilize the distribution of inhabitants across four income groups: low, middle-low, middle-high and high (see definition in Table 2). First, we calculate the mean income for every income group by assuming that the income distribution in every base area is the same as that of the municipality to which it belongs.

With this information the variance within base areas is calculated by

$$\sigma^2_{within} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{4} s_i (x_i - \bar{x}_j)^2}{k}, \tag{1}$$

where k is the number of base areas in the municipality of interest, s_i is the proportion of base area j's population in income group i, x_i is the mean income of income group i and $\bar{x}_j$ is the mean income of base area j.

Using this proxy comes with a price. The variance calculated with Eq. (1) will only be driven by the variation in the income group sizes. This means that if the gap between the mean income for the lowest group versus the highest group changes, that change will not be captured. Adding this effect as well would likely increase the variance, since income inequalities are known to be rising in Stockholm (Andersson and Kährik, 2016). The variance in Eq. (1) is thus a conservative measure of the variance.

The variance between base areas is calculated as

$$\sigma^2_{between} = \frac{\sum_{j=1}^{k} s_j(\bar{x}_j - \bar{\bar{x}})^2}{k-1}, \tag{2}$$

where k is the number of base areas in the municipality of interest, s_j is base area j's share of the municipality's population, $\bar{x}_j$ is the mean income of base area j and $\bar{\bar{x}}$ is the mean income of the municipality.

Table 2: *Definitions of the income groups.*

Income Group	Interval (SEK)
Low	0 - 105 031
Middle low	105 032 - 171 251
Middle high	171 252 - 235 757
High	> 235 757

4.2 Regression Analysis

Here we present everything regarding our regression analysis. Our choice of econometric model, the reasoning behind it and possible limitations. First, however, we start by explaining our dependent variable, the dissimilarity index.

4.2.1 Dissimilarity Index

As mentioned in Section 2, DI measures evenness, i.e. how two groups are distributed across some geographical units. Say we want to see how segregated two groups, H and L, are in an area A, which can be divided into

n neighborhoods. Then, we can calculate the dissimilarity index for A using

$$DI_A = \frac{1}{2} \sum_{i=1}^{n} \left| \frac{h_i}{H_T} - \frac{l_i}{L_T} \right|, \tag{3}$$

where h_i is the number of people in group H in neighborhood i and H_T is the total number of people in group H in the area of interest. In complete analogy, l_i is the number of people in group L in neighborhood i and L_T is the total number of people in group L. The index lies in the interval $[0, 1]$, where 0 represents no segregation (i.e. a perfectly even distribution of the groups across the neighborhoods) and 1 represent maximum segregation.

DI can be calculated for any two groups, commonly for two ethnic groups. However, in this thesis, DI is calculated based on the proportions of inhabitants in the lowest income group versus the proportion of people in the highest income group, defined according to Table 2, within each base area. By aggregating the values for the base areas we get a value of the dissimilarity index for every municipality, which we then use as the dependent variable in our regression analysis. The independent variables are explained in the next section.

There is a myriad of indices for measuring segregation and all of them have limitations. DI remains the *de facto* standard and thus, has the benefit of being easier to compare across studies. Also, as it has been used to such a great extent all drawbacks are well known, which enables us to reason about them. In Section 2 we summarized the critique Cortese et al. (1976) raised against DI and we would argue that we are able handle these caveats to a sufficient degree in this thesis.

The objection of having the zero point of the index as a perfectly even distribution, instead of a distribution which accounts for randomness, is not a major obstacle in this thesis as we are concerned with variations within municipalities over time. Taking randomness into account might make the values more comparable across municipalities, but not necessarily. A better approach in that case would be to make further analysis into any potential randomness prevalent in each municipality, but that is outside the scope of this thesis. Moreover, the dissimilarity index is known to be sensitive to the sizes of the groups considered. However, we mitigate this effect by weighting the values for each base area with its proportion of the municipality's population.

In conclusion, the benefits of keeping DI, i.e. easy to communicate and understand, can be related to other studies and with well-known drawbacks, are higher than using some other, less-known index, with other limitations.

4.2.2 Influx and Outflux Variables

In the previous section we described the dependent variable, DI, in our regression analysis. The key independent variables, however, in our regression analysis are *influx* and *outflux*. With them we try to capture the effect on economic segregation from the flow of people moving in and out of the base areas. The basic idea is to compare the socioeconomic status of those moving in and out versus those who are already living in the base area. We have two dimensions for measuring the socioeconomic status—employment rate and educational attainment (share of people with minimum 3 years of higher education).

We calculate the influx and outflux variables as below.

$$influx_{b,y} = \begin{cases} s_{i,y} - s_{b,y-1}, & s_{b,y-1} > s_{m,y-1} \\ s_{b,y-1} - s_{i,y}, & s_{b,y-1} < s_{m,y-1} \\ |s_{i,y} - s_{b,y-1}|, & s_{b,y-1} = s_{m,y-1} \end{cases}, \tag{4}$$

where $s_{i,y}$ is employment rate (educational attainment) of those who moved in to base area b in year y, $s_{b,y-1}$ is employment rate (educational attainment) of those who lived in base area b by the end of year $y-1$ and $s_{m,y-1}$ is employment rate (educational attainment) of those who lived in municipality m (which base area b belongs to) by the end of year $y-1$.

This means that the $influx_{b,y}$ is positive when: i) people of low socioeconomic status move into a base area which is below the socioeconomic mean of the municipality or; ii) people of high socioeconomic status move into a base area which is above the socioeconomic mean of the municipality. In the same way it is negative when: i) people of low socioeconomic status move into a base area which is above the socioeconomic mean of the municipality or; ii) people of high socioeconomic status move into a base area which is below the socioeconomic mean of the municipality. Hence, we would expect an increasing value $influx_{b,y}$ to correspond to increasing economic segregation.

We calculate the variable for the outflux similarly

$$outflux_{b,y} = \begin{cases} s_{b,y} - s_{o,y}, & s_{b,y} > s_{m,y} \\ s_{o,y} - s_{b,y}, & s_{b,y} < s_{m,y} \\ |s_{b,y} - s_{o,y}|, & s_{b,y} = s_{m,y} \end{cases}, \tag{5}$$

where $s_{o,y}$ is the employment rate (educational attainment) of those who moved out of base area b in year y, $s_{b,y}$ is employment rate (educational

attainment) of those who lived in base area b during year y and $s_{m,y}$ is employment rate (educational attainment) of those who lived in municipality m during year y.

In analogy with the interpretation of Eq. (4) $outflux_{b,y}$ is positive when: i) people of low socioeconomic status move out of a base area which is above the socioeconomic mean of the municipality or; ii) people of high socioeconomic status move out a base area which is below the socioeconomic mean of the municipality. $outflux_{b,y}$ is negative when: i) people of low socioeconomic status move out of a base area which is below the socioeconomic mean of the municipality; ii) people of high socioeconomic status move out of a base area which is above the socioeconomic mean of the municipality. Given this definition, we would expect an increasing value $outflux_{b,y}$ to cause increased segregation as well.

With the influx and outflux variables defined as in Eq. (4) and Eq. (5), we obtain the values for every municipality and year by aggregating the base area values and weighting them with the share of the total population, i.e.

$$influx_{m,y} = \sum_{j=1}^{n_m} w_j \, influx_{j,y}, \tag{6}$$

where n_m is the number of base areas in municipality m, w_j is base area j's the share of the total population in municipality m and $influx_{j,y}$ is the influx for base area j in year y.

Finally, for outflux we have

$$outflux_{m,y} = \sum_{j=1}^{n_m} w_j \, outflux_{j,y}. \tag{7}$$

With these variables in mind, we can now define our regression model.

4.2.3 The Model

The unobserved model for segregation, measured by DI, we try to estimate in this thesis is shown in Eq. (8). In addition to the influx and outflux variables defined above, we also include the share of foreign born inhabitants in the municipality. The value of DI in municipality m in time t is given by

$$\begin{aligned} DI_{m,t} = \beta_0 &+ \beta_1 influx^{edu}_{m,t} + \beta_2 influx^{emp}_{m,t} + \beta_3 outflux^{edu}_{m,t} \\ &+ \beta_4 outflux^{emp}_{m,t} + \beta_5 foreign\_born_{m,t} + \alpha_m + u_{m,t}, \end{aligned} \tag{8}$$

where the $influx$ and $outflux$ variables are those described in Section 4.2.2, $foreign\_born$ is the share of the population born outside Sweden, α is any time-invariant variable and u the residual. t ranges from 2000 to 2014.

We do not claim this model to be a complete model of segregation, but rather an attempt to investigate the relevance of some processes that could be driving segregation. We have included the influx and outflux variables to capture the potential impact the migration flows in and out of the base areas in Stockholm have. Moreover, by including the share of foreign born inhabitants we hope to gain insights in the relationship between economic segregation and ethnic segregation.

Initially, the plan was to include the Gini index as well in order to explore the relationship between economic segregation and economic inequality, however, due to limitations in the data[1] that relationship will be tested separately.

As mentioned, the model is not complete and the risk of omitted variable bias is high. One important example of this is the housing market, which in general is believed to be one of the major structural drivers of segregation (Musterd, 2005; Marcińczak et al., 2015; Andersson and Kährik, 2016). However, as much as we would like to control for that it is difficult to find the right data and the right measure to capture its impact and hence, we have been forced to omit the housing market in our model. The effects of its absence will be discussed further in Section 6.

We estimate Eq. (8) by a fixed effects regression model. Given the fact that the dependent variable is a proportion one might argue that a beta regression, as described by Ferrari and Cribari-Neto (2004), is more suitable. However, even though using a linear regression might suffer compared to a beta regression in terms of a better fit of the data, it simplifies the interpretation of the results and can be easily communicated, as the method is well known. Estimating with fixed effects will yield more conservative results compared to e.g. random effects, although the assumptions allowing for random effects are unlikely to hold. In addition, fixed effects can help account for the limitations of DI not being perfectly comparable across groups and any other bias which might stem from municipality-specific and time-invariant differences.

[1] The Gini index per municipality is only available for 9 of the 15 years studied.

5 Results

Here we comment and depict the results from our analysis of the economic segregation in Stockholm during the years 2000–2014. A discussion of the validity of the results follows in Section 6.

5.1 RQ1: Economic Segregation 2000–2014

Looking at Figures 1 and 2, we see that during 2000–2014 the variance of income has increased both within and between base areas. In Section 4 we described that in the case of increasing segregation we would expect the variance between base areas to increase, while the variance within the base areas would decrease. Therefore, when both are increasing this could naively be interpreted as segregation decreasing.

However, when comparing Figures 1 and 2, we see that the variance between the base areas is larger and increasing faster, than its counterpart between the base areas.

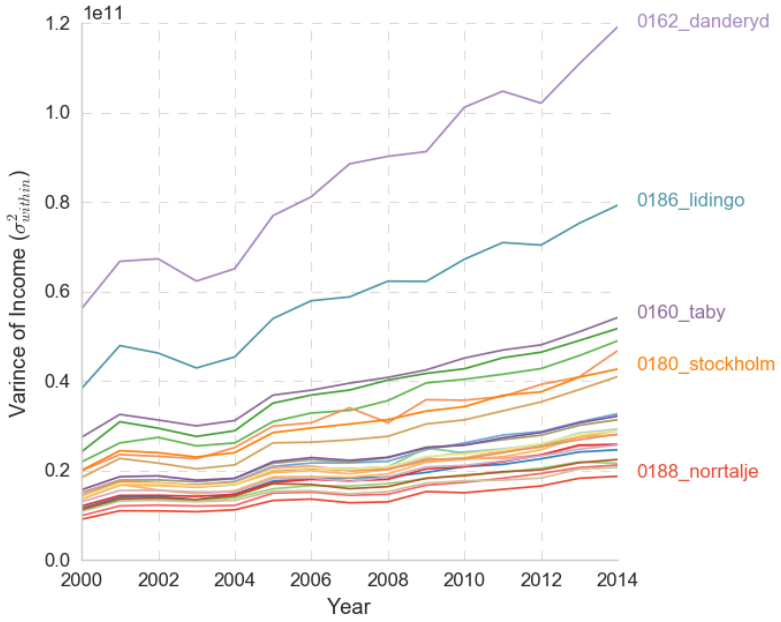

Figure 1: *Variance of income* within *base areas during the years 2000–2014, grouped by municipality. Source: author's calculations.*

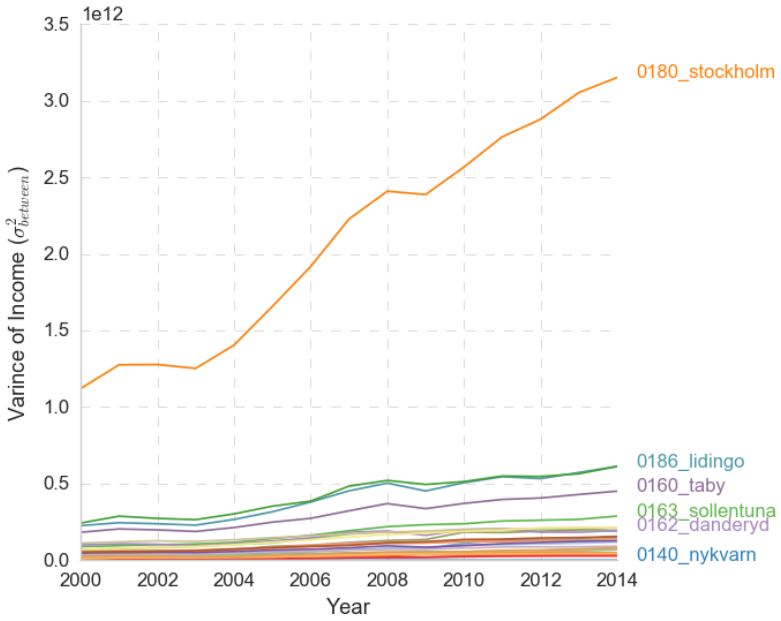

Figure 2: *Variance of income* between *base areas during the years 2000–2014, grouped by municipality. Source: author's calculations.*

Comparing Figures 1 and 2, is somewhat difficult as the variance between the base areas of Stockholm Municipality is increasing several times faster than the other municipalities. Therefore, to simplify the interpretation, we plot the relative difference in Figure 3. The figure depicts the relative difference $\sigma^2_{between}/\sigma^2_{within}$, normalized such that year 2000 is 100. From the figure we can deduce that most municipalities have experienced a 25% to 50% faster growth between the base areas. This could be an indication of increasing economic segregation.

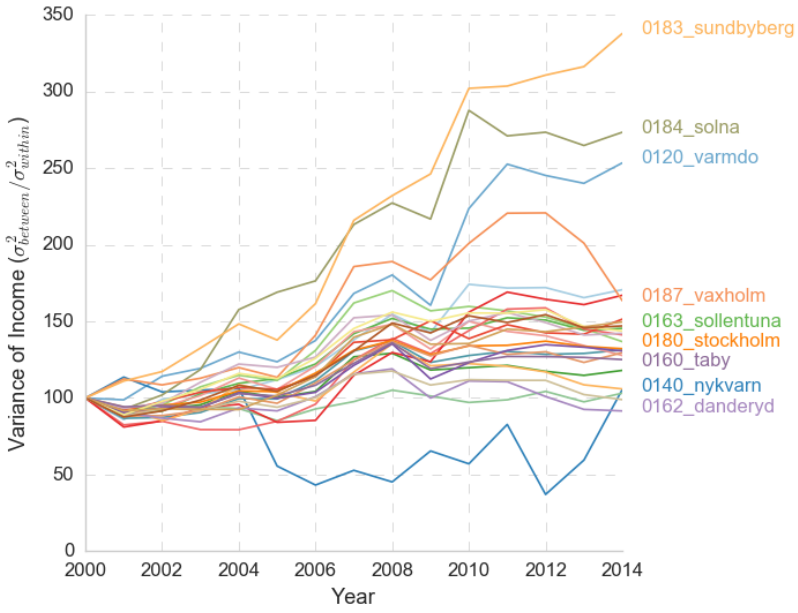

Figure 3: *Normalized (year 2000=100), relative difference ($\sigma^2_{between}/\sigma^2_{within}$) of income variance between and within the base areas, during the years 2000–2014. Source: author's calculations.*

We find more evidence of increasing segregation if we look at Figure 4, which shows the change in DI between the years 2000 and 2014, versus the initial value of the same index in the year 2000. Several conclusions can be drawn from that figure. First, most municipalities are located in the top left corner of the graph, which corresponds to municipalities with a value of DI lower than average in year 2000 but with a negative trend during 2000 and 2014, i.e. a trend of increasing segregation. Second, from the regression line drawn we see that the correlation between change and the initial value is rather weak (-0.296). However, the trend is such that the initially less segregated municipalities are getting more segregated, while the relatively highly segregated areas are getting less segregated. Therefore, would we aggregate to a county level, we might see evidence of decreasing economic segregation, which if it holds shows the importance of looking at both a micro and macro level in segregation studies (Schelling, 1978). Third, two outliers can be observed, Sundbyberg and Nykvarn, which will be analyzed further in the next section.

We make no attempt of comparing the absolute values of DI calculated in this thesis and other studies, as the definitions varies (e.g. different income groups, granularity of data). However, the trend is comparable and our result that the economic segregation in Stockholm County has increased

during 2000 and 2014, is aligned with previously conducted research. Both in Sweden and in Europe (Musterd, 2005; Marcińczak et al., 2015; Tammaru et al., 2016; Malmberg et al., 2016; Bailey et al., 2017; Andersson and Kährik, 2016).

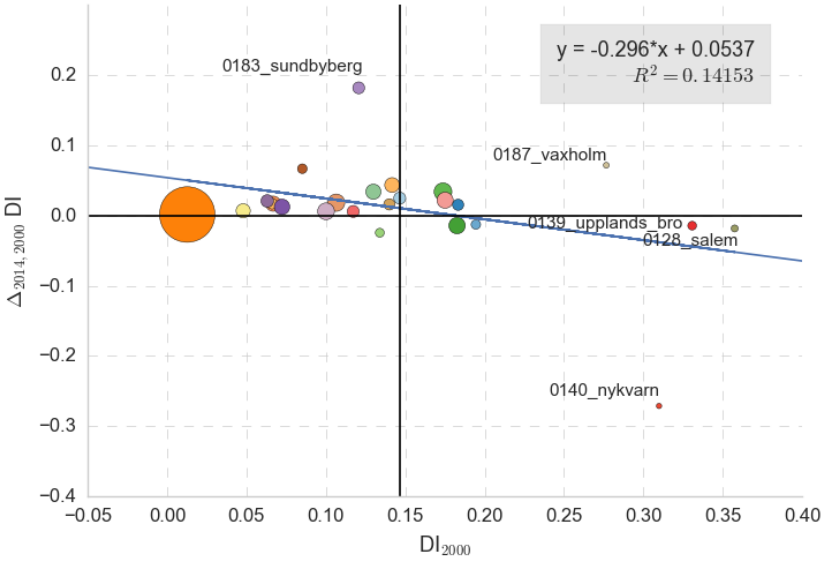

Figure 4: *The change of DI between 2000 and 2014 versus DI year 2000, for the municipalities of Stockholm County. The size of the bubbles are set according to population. Source: author's calculations.*

5.1.1 The Outliers: Sundbyberg and Nykvarn

As we observed in Figure 4, Sundbyberg and Nykvarn were two distinct outliers. During the years 2000 and 2014, Sundbyberg saw an increase of DI with 0.181513, while Nykvarn experienced a decrease of 0.271122. In Figure 5 these trends are seen more clearly.

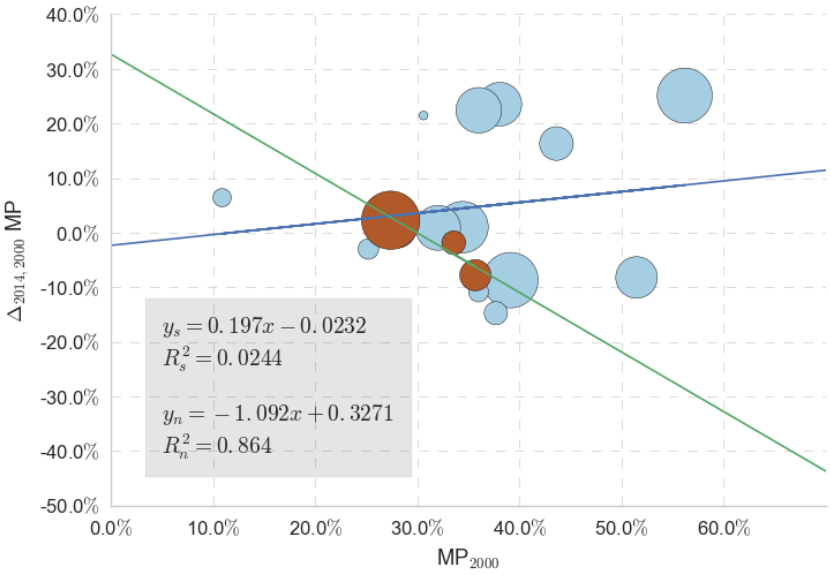

Figure 5: *Change of minority proportion (low income earners) during 2000–2014 versus initial minority proportion at year 2000, for all base areas in Sundbyberg Municipality and Nykvarn Municipality. The three darker bubbles correspond to Nykvarn's base areas and the lighter bubbles are the base areas of Sundbyberg Municipality. Source: author's calculations.*

In Sundbyberg, a few base areas with an initially quite large share of the municipality's low income earners have increased their share significantly over time. As a matter of fact, basically two areas—Hallonbergen and Rissne[2]—accounts for almost all the accumulation of low income earners (compare with Figure 6). Hallonbergen is a legacy of the Million Program in the 1960's (see Section 2) and has not seen much restoration since then. Today it hosts a large share of the municipality's foreign born. The neighboring area of Rissne has a similar history. The rest of Sundbyberg Municipality was either quite well-off from the beginning or has seen restorations after the year of 2000, e.g. Sundbyberg Centrum. In the case of Sundbyberg it would have been interesting to correlate the results with price levels as well. We would assume that house prices have increased significantly in most base areas, excluding Hallonbergen and Rissne.

[2]Rissne is split up into three base areas—Norra Rissne, Södra Rissne and Rissne Norra Arbetsområde.

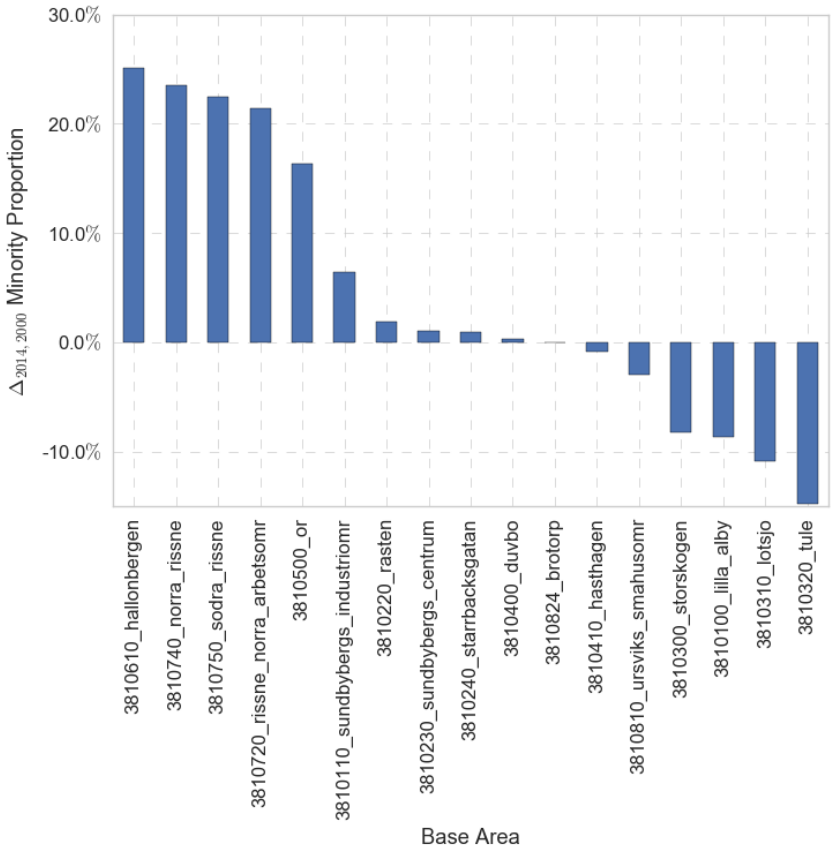

Figure 6: *The absolute change of minority proportion (low income earners) between years 2000 and 2014, in all base areas of Sundbyberg Municipality.*

In Nykvarn, which is a rural municipality south of Stockholm, consisting of only three base areas, the two base areas with the highest share of low income earners have decreased their share over time. In addition, the base area with the lowest share has increased its share slightly (see Figure 5). This has resulted in an almost perfect distribution of low and high income earners and therefore, a very low dissimilarity index.

5.2 RQ2: Migration Flows and Economic Segregation

The results of the regression we ran in order to estimate Eq. (8) are shown in Table 3. The importance of migration flows in explaining segregation has been suggested in both Bailey et al. (2017) and Andersson and Magnusson Turner (2014). As can be seen in Table 3, all variables except $outflux_{emp}$ are significant at a 5% level. $influx_{emp}$ and $outflux_{emp}$ are both positive,

which is the expected sign (see the discussion in Section 4).

If we start with $influx_{emp}$, which is positive, we interpret its value as economic segregation is increasing in the municipalities of Stockholm County partly because people with a higher socioeconomic status move into already well off base areas or people with lower socioeconomic status tend to move into relatively poor base areas. In analogy, when interpreting $outflux_{emp}$, this indicates that those of higher socioeconomic status are moving out of the poorer areas (a phenomenon sometimes called filtering or residualisation) or those of lower socioeconomic status are leaving the richer base areas (sometimes as a consequence of gentrification).

All in all, the migration flows and sorting of people seem to be consolidating already skewed conditions between the base areas and thus, increases the economic segregation. A word of caution, however, there are reasons to be careful with claiming causality, as will be discussed further in Section 6. We also see that for all coefficients, except share of foreign born (which instead has a rather large confidence interval), the effects on DI are small. For all significant variables the effect of increasing them one hundred percentage units will cause the dissimilarity index to change with only a few percent and less than 10% in all cases. Hence, the economic significance can be debated.

Table 3: *Fixed effects regression of Eq. (8).*

VARIABLES	DI
$influx_{edu}$	-0.00685
	(0.0106)
$influx_{emp}$	0.0233***
	(0.00807)
$outflux_{edu}$	-0.0620***
	(0.00821)
$outflux_{emp}$	0.0175***
	(0.00275)
foreign_born	0.535***
	(0.0692)
Constant	0.0701***
	(0.0119)
Observations	364
Number of municipalities	26
R-squared	0.351

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The results discussed above holds for when socioeconomic status is measured as a person having a job or not. If we instead measure socioeconomic status as educational attainment,[3] the sign is negative, which is opposite of the expected sign. In the case of $outflux_{edu}$, which is the only significant coefficient at a 5% level, this can be interpreted as economic segregation is decreasing because people of high socioeconomic status are leaving base areas with a low status ($s_b < s_o$ and $s_b < s_m$) or people of low socioeconomic status moving out of base areas being more well off ($s_o < s_b$ and $s_m < s_b$).

Initially, this result seems very counterintuitive. However, it would be expected in those cases where the normally positive correlation between educational attainment and economic status does not hold. To be more precise, this result is expected if we have groups with: i) high education but low economic status and vice versa, groups with: ii) low educational attainment and high income. If we reason about this we realize these groups probably do exist, and it is likely to be age related. The first group (i) can be found among the younger population, i.e. students and newly graduates. In addition, another group of people which might fit into category (i) is

[3]Minimum of three years of university studies.

migrants, which tend to be overqualified for their jobs to a larger extent than non-migrants. Similar arguments can be made for the other group (ii). People with a low educational attainment but who are still in the higher income group (as defined by Table 2), are likely to be older on average.

Note that these are merely hypotheses which we have not been able to verify due to limitations in data (citizenship and age for the migrant flows between the base areas would be needed). However, an indication is given if we compare the two outliers discussed above—Sundbyberg and Nykvarn. They both have relatively strong, negative correlation between $outflux_{edu}$ and DI (Sundbyberg -0.642858 and Nykvarn -0.35918). Moreover, during 2000 to 2014 the mean age in Sundbyberg has decreased with 1.5 years (39.4 to 37.9), while Nykvarn has increased its with 3 years (36 to 39). Also, Sundbyberg's foreign born population was 28% (increase of 9 percentage points since year 2000) in year 2014 and the same number for Nykvarn was 11% (increase of 0.3 percentage points since year 2000).

Moving on to the coefficient for the share of foreign born in the regression analysis above (Table 3), we see that it is able to explain a relatively large share of the changes in the DI. This result, i.e. areas with a higher degree of foreign born inhabitants also tend to be more segregated, is in line with earlier studies (Malmberg et al., 2016; Tammaru et al., 2016). It also indicates a rather strong correlation between economic segregation and ethnic segregation.

Table 4: *Fixed effects regression with the Gini coefficient as the sole independent variable.*

VARIABLES	DI
gini	0.660***
	(0.184)
Constant	-0.0858
	(0.0681)
Observations	234
Number of municipalities	26
R-squared	0.059

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In Table 4 we see the results of running a simple linear regression, with fixed-effects, where the Gini index is the only independent variable. We ran this regression analysis separately to mitigate the risk of any bias that might

come from the fact that we are missing values for the Gini coefficient during the years 2000-2004 and 2006. Gini has strong, positive correlation with the economic segregation, which is expected as complete equality would imply no dissimilarity. In addition, in the regression above, the covariance of Gini and error term, $cov(gini_{m,y}, u_{m,y})$, is -0.5413; hence, there exists a negative bias and the coefficient is likely underestimated. Nevertheless, in accordance with previous research, the relationship between economic segregation and economic inequality does not appear to be one-to-one (Tammaru et al., 2016; Andersson and Kährik, 2016; Marcińczak et al., 2015).

6 Discussion

6.1 Data

Except for the limitations in the data which have been mentioned throughout the report there are two fundamental properties of the data that deserves special attention and consideration. First, all of the data is individual, i.e. not per household. Using household data would have been desirable as it gives a more accurate view of how people's living situation. There is a risk of exaggerating the segregation when the calculations are based on individual data instead of household data, as a top earner and a bottom earner might share the same household.

Second, in this thesis we have been studying economic segregation, which implies that what type of income we use will affect our results and interpretations. Consistently, throughout the thesis, we have used gross income, including pension and taxable benefits. This was the only data we had at hand. Using income after tax instead would probably lead to lower segregation, as pointed out by Bailey et al. (2017), since Sweden has a progressive tax system and the effect of the tax ought to mitigate inequalities. We argue that this is not a major problem here however, as we are generally interested in change over the last 15 years in this thesis, and not as much the absolute values of DI.

The effect of not including income from capital will mainly have an effect in regards to top earners, as income from capital is usually a more important source of income for people higher up the income distribution. However, as the lower limit for our highest income group (see Table 2) is still relatively low, the effect of not including income of capital ought not to affect the distributions in the income groups to any noticeable degree.

Third, in Section 3 we introduced MAUP (Openshaw, 1984). The issues with MAUP apply also to the work done in this thesis. Essentially, to mitigate or rule out the bias stemming from MAUP the geographical areas chosen for the analysis ought to be homogeneous. It is outside the scope of this thesis to assess the homogeneity of the base areas, however, Amcoff (2012) provides some insights on the subject. The author investigates the homogeneity of the so called *Small Areas for Market Statistics (SAMS)*, which constitute a nationwide grid of small geographical regions, defined by Statistics Sweden in collaboration with Swedish municipalities in 1994. SAMS have been used extensively for studying neighborhood effects and the base areas were part of SAMS division during its creation. Amcoff (2012) argues that the SAMS system does not have an optimal division in terms of homogeneity, however,

this mainly concerns rural areas and not urban. This is something that speaks in favor of using base areas when studying Stockholm. Moreover, the base areas in Stockholm have been revised a few times since their inception, with the latest revision in 2010, and while it does not rule out the presence of bias due to MAUP, it does support the use of base areas over the more commonly used areas in the SAMS division.

Fourth, as was mentioned in Section 3, we look at the years 2000–2014 mainly due to a limited access of data. A longer timespan is preferable when studying residential segregation since processes and drivers behind it often operate slowly. Therefore, we need to be aware that studying the years 2000 to 2014 is in a sense arbitrary. We can assume that the development we have described in this thesis regarding Sundbyberg, where some areas are rebuilt or restored while others are not, we could have seen in other municipalities if the time period was different. E.g. Solna and Stockholm have already gone through the same type of changes as Sundbyberg is currently, a few decades earlier. At the same time, while Sundbyberg is an outlier in negative terms in this thesis, it might not be if the same analysis is performed in a few years. Restoring and rebuilding entire municipalities are slow and costly processes and although the gap between Hallonbergen and Rissne versus the rest of Sundbyberg is currently increasing, tension from this increasing gap might eventually result in a restoration of Hallonbergen and Rissne as well.

In addition, in the midst of the time period studied in this thesis the landscape for housing prices in Stockholm changed by the introduction of presumption rents in 2006. This change of policy allows the rents on new houses to be set according to the market instead of being regulated as the rest of the housing stock. As housing prices were left out of the scope of the analysis in this thesis the effect of this market liberalization has not been assessed, but it is reasonable to assume that this has further increased the gap between the old and the new housing stock.

6.2 Methodology

In this thesis we have chosen to measure segregation using the the dissimilarity index. It has several weaknesses, which have been described and discussed in Sections 2 and 4. DI only captures one dimension of segregation, as described in Massey and Denton (1988). Therefore, the results presented in Section 5 do not give a complete view of the economic segregation in Stockholm, but rather one piece of the puzzle. Another picture might emerge when studying the same phenomenon along other dimensions.

We chose to estimate our model (Eq. (8)), using a fixed-effects regression.

This means we lose the variation across municipalities which could have been explained by our model and made our estimates more precise. In fact, $\rho = 0.94120111$ in the regression shown in Table 3 can be interpreted as 94% of the variation in DI comes from variation across municipalities. However, we believe it is of greater importance for our purpose to control for municipality specific and time-invariant effects. In addition, to account for the limitation of DI not being fully compositionally invariant, it makes sense to investigate only the variation within the municipalities. This is aligned with the focus of this thesis, i.e. to see how the migration flows have affected the situation in the municipalities over time and to compare the development of municipalities to each other, not as much the absolute values of DI.

6.3 Results

Interpreting the results from the analysis of variance (Figures 1–3), it is important to remember that we used a proxy to calculate the variance of income within the base areas (see Section 4), a proxy which estimates the variance conservatively. Thus, the variance within base areas is underestimated and the change in the relative differences overestimated. The results are nevertheless interesting, and especially when we compare them to Figure 4 and we see there is a positive correlation between rising relative difference in income variance and economic segregation. As have been stated earlier, the scope here is to study the change and we would argue the proxy yields results on a sufficient level for that.

When analyzing the results from the regression analysis, we need to be careful before we can claim causality. We have identified three sources which can cause biased estimates. First, there is likely a loop of causality between the independent and dependent variables. In particular, we can assume that areas with already high dissimilarity index might attract a larger share of people with lower socioeconomic status, due to housing prices, network effects among migrants and so on.

Second, the risk of omitted variable bias in the model in Eq. (8) is considerable. Using a fixed effects model mitigates this risk to some degree but not completely. One structural driver of segregation which is omitted is the housing market, a factor which most likely is of significance to understand segregation (Tammaru et al., 2016). However, reasoning about how our results would change, were we to include the housing market, is difficult since it is hard to find a good measure for it and there is not a lot of relevant previous research to compare with.

Third, we chose to estimate our model with a fixed effect linear regression. Would this functional form not be a correct representation of the unobserved model, there is a risk of biased estimates. As was discussed in Section 4, given the fact that the independent variable is a proportion, a beta regression will probably yield a better fit of the data (Ferrari and Cribari-Neto, 2004). Albeit, we would argue that estimating the model with a fixed effects linear regression suffices as a first step into investigating the effect migration flows have had on the economic segregation.

As heteroscedasticity is a common phenomena in most regressions, we repeated the regression in Table 3 with robust standard errors, to control for heteroscedasticity. This reduced the number of significant coefficients at a 5% level to two, namely $foreign\_born$ and $outflux_{emp}$. $outflux_{edu}$ was significant at an 8% level, but neither of the $influx$ variables were significant on a 10% level. This could be an indication that the outflux of people from neighborhoods has a greater effect on the economic segregation than influx.

Moreover, in the regression analysis presented in Table 3 there is a mixture of flow (influx and outflux) and stock (DI and share of foreign born) variables. In order to analyze the sensitivity of this we ran another regression with the change of the share foreign born and the change in DI instead. This resulted in only one significant variable, $outflux_{emp}$, with a slightly lower value (0.011). This might once again be an indication that $outflux$ is of greater significance to explain economic segregation than $influx$ is.

7 Conclusions & Future Work

In this thesis we have explored the economic segregation in the municipalities of Stockholm County, during the years 2000 and 2014. In addition, we have analyzed the impact of migrations flows, people moving in and out of base areas, on economic segregation. By comparing the variance of income within and between the base areas we have found indications of increasing economic segregation during the first 15 years of the 21st century. Moreover, by calculating the dissimilarity index for the lowest income group and the highest, we were able to show further signs of a growing economic segregation in most of the municipalities in Stockholm County.

Furthermore, we have defined variables which take into account the differences in socioeconomic level of the people moving in and out of a base area compared to those who are living there already. These variables, along with the share of the population born in a foreign country, were used as independent variables in a model with the dissimilarity index as the dependent variable. By running a regression on this model we find indications that migration patterns tend to reinforce the economic segregation in Stockholm County. E.g. people of higher socioeconomic status move out of base areas with low status and into those with higher status and people of lower socioeconomic status move out of high status base areas and into those of lower status. However, we are careful claiming causality and the economic significance is debatable, nevertheless, the correlation itself is interesting and gives indications for further analysis. In addition, we show a rather strong correlation between ethnic segregation, income inequality and economic segregation.

This thesis explores the tip of an iceberg. Explaining why people move as they do is not answered here. Moreover, adding variables such as age and citizenship would help gain more understanding into migration patterns and their relationship to economic segregation. Also, this thesis explores segregation along the dimension of evenness. Extending the research by looking at more dimensions would give a more complete picture of the development.

Safe to say, however, is that segregation remains a complex matter and there are no straightforward methods or simple answers. All that is certain is that the topic will remain a relevant one, for the foreseeable future.

Acknowledgments

We are grateful to the following people for resources, discussions and suggestions. Associate Professor Anders Olofsgård: for support and fruitful discussions, which have been essential for the completion of this thesis. Erik Stannow Lind: for comments and remarks imperative for the quality of this thesis. Dr Mattias Lindgren: for initial discussions and guidance. Eric Gratorp, Erik Larsson, Fernanda Drumond and Viktor Eriksson: for investing your precious time into helping me complete this work. Max Berggren: for making stuff with data and sharing it, thus, prettifying the results presented here.

References

Amcoff, J. (2012). Hur bra fungerar samsområdena i studier av grannskapseffekter? en studie av samsområdenas homogenitet. *Socialvetenskaplig tidskrift*, 19(2):93–115.

Andersson, E. K. and Malmberg, B. (forthcoming). Segregation and the effects of adolescent residential context on poverty risks and early income career: A study of the Swedish 1980 cohort. *Urban Studies*, DOI 0042098016643915.

Andersson, R. and Kährik, A. (2016). Widening gaps: Segregation dynamics during two decades of economic and institutional change in Stockholm. In Tiit Tammaru, Szymon Marcińczak, Maarten van Ham, & Sako Musterd (eds.). *Socio-Economic Segregation in European Capital Cities: East meets West*, pages 110–131.

Andersson, R. and Magnusson Turner, L. (2014). Segregation, gentrification, and residualisation: from public housing to market-driven housing allocation in inner city Stockholm. *International Journal of Housing Policy*, 14(1):3–29.

Andersson, R. and Musterd, S. (2010). What scale matters? Exploring the relationships between individuals' social position, neighbourhood context and the scale of neighbourhood. *Geografiska Annaler: Series B, Human Geography*, 92(1):23–43.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory*, 2(3):244–263.

Bailey, N., Gent, W. P., and Musterd, S. (2017). Remaking urban segregation: Processes of income sorting and neighbourhood change. *Population, Space and Place*, 23(3):e2013.

Bell, W. (1954). A probability model for the measurement of ecological segregation. *Social Forces*, 32(4):357–364.

Bolling, A. (February 8, 2016). Ny studie: Segregationen har minskat. *Dagens Nyheter*.

Booth, C. (1903). *Life and Labour of the People in London*, volume 8. Macmillan.

Cortese, C. F., Falk, R. F., and Cohen, J. K. (1976). Further considerations on the methodological analysis of segregation indices. *American Sociological Review*, pages 630–637.

Dacey, M. F. (1965). A review on measures of contiguity for two and k-color maps. Technical report, Department of Geography, Northwestern University, Evanston, IL.

Dixon, P. M., Weiner, J., Mitchell-Olds, T., and Woodley, R. (1987). Bootstrapping the Gini coefficient of inequality. *Ecology*, 68(5):1548–1551.

Duncan, O. D. and Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review*, 20(2):210–217.

Duncan, O. D. and Duncan, B. (1957). *The Negro Population of Chicago: A Study of Residential Succession*. University of Chicago Press.

Echenique, F. and Fryer, R. G. (2007). A measure of segregation based on social interactions. *Quarterly Journal of Economics*, 122(2):441–485.

Farber, S., O'Kelly, M., Miller, H. J., and Neutens, T. (2015). Measuring segregation using patterns of daily travel behavior: A social interaction based model of exposure. *Journal of Transport Geography*, 49:26–38.

Ferrari, S. and Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7):799–815.

Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician*, 5(3):115–146.

Grannis, R. (2002). Discussion: segregation indices and their functional inputs. *Sociological Methodology*, 32(1):69–84.

Hoover, E. M. (1941). Interstate redistribution of population, 1850–1940. *Journal of Economic History*, 1(02):199–205.

Hutchens, R. (2004). One measure of segregation. *International Economic Review*, 45(2):555–578.

Magnusson Turner, L. and Hedman, L. (2014). Linking integration and housing career: A longitudinal analysis of immigrant groups in Sweden. *Housing Studies*, 29(2):270–290.

Malmberg, B., Nielsen, M. M., Andersson, E., and Haandrikman, K. (2016). *Residential segregation of European and non-European migrants in Sweden: 1990–2012*.

Marcińczak, S., Musterd, S., Van Ham, M., and Tammaru, T. (2016). Inequality and rising levels of socio-economic segregation: Lessons from a pan-european comparative study. In Tiit Tammaru, Szymon Marcińczak, Maarten van Ham, & Sako Musterd (eds.). *Socio-Economic Segregation in European Capital Cities: East meets West*, pages 358–382.

Marcińczak, S., Tammaru, T., Novák, J., Gentile, M., Kovács, Z., Temelová, J., Valatka, V., Kährik, A., and Szabó, B. (2015). Patterns of socioeconomic segregation in the capital cities of fast-track reforming postsocialist countries. *Annals of the Association of American Geographers*, 105(1):183–202.

Massey, D. S. (1978). On the measurement of segregation as a random variable. *American Sociological Review*, 43(4):587–590.

Massey, D. S. and Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces*, 67(2):281–315.

Murdie, R. A. and Borgegård, L.-E. (1998). Immigration, spatial segregation and housing segmentation of immigrants in metropolitan Stockholm, 1960 - 95. *Urban Studies*, 35(10):1869–1888.

Musterd, S. (2005). Social and ethnic segregation in europe: Levels, causes, and effects. *Journal of Urban Affairs*, 27(3):331–348.

OECD (2015). Oecd income inequality data update: Sweden. https://www.oecd.org/els/soc/OECD-Income-Inequality-Sweden.pdf. Accessed: March 14, 2017.

Openshaw, S. (1984). The modifiable areal unit problem. CATMOG 38. GeoBooks, Norwich.

Piketty, T. (2014). *Capital in the Twenty-First Century*. Belknap Press of Harvard University Press Cambridge, MA.

Piketty, T. and Saez, E. (2003). Income inequality in the united states, 1913–1998. *Quarterly Journal of Economics*, 118(1):1–41.

Quinn, L. M. and Pawasarat, J. (2003). *Racial Integration in Urban America: A Block Analysis of African American and White Housing Patterns*. Employment

and Training Institute, School of Continuing Education, University of Wisconsin-Milwaukee.

Schelling, T. (1978). *Micromotives and Macrobehavior*. Norton.

Taeuber, K. E. and Taeuber, A. F. (1969). *Negroes in Cities: Residential Segregation and Neighborhood Change*. Atheneum.

Taeuber, K. E. and Taeuber, A. F. (1976). A practitioner's perspective on the index of dissimilarity. *American Sociological Review*, 41(5):884–889.

Tammaru, T., Marcińczak, S., van Ham, M., and Musterd, S. (2016). *Socio-Economic Segregation in European Capital Cities: East meets West*. Routledge.

Theil, H. (1972). *Statistical Decomposition Analysis: With Applications in the Social and Administrative Sciences*. North-Holland Amsterdam.

Theil, H. and Finizza, A. J. (1971). A note on the measurement of racial integration of schools by means of informational concepts. *The Journal of Mathematical Sociology*, 1(2):187–193.

Tottmar, M. (November 22, 2014). Segregationen ökar i stockholm. *Dagens Nyheter*.

Tottmar, M. (September 21, 2015). Stockholm blir en alltmer delad stad. *Dagens Nyheter*.

Wessel, T. (2016). Economic segregation in Oslo: Polarisation as a contingent outcome. In Tiit Tammaru, Szymon Marcińczak, Maarten van Ham, & Sako Musterd (eds.). *Socio-Economic Segregation in European Capital Cities: East meets West*, pages 132–155.