Does Segregation Nurture the Sweden Democrats?

*The Political Economy of Segregation*

Andreas Jangmo*

**Abstract**

This thesis studies the relationship of cultural and economic segregation with politics. Based on a theoretical model where the provision of a public good depends on how far apart voters are in terms of preferences, it is suggested that the support for the Sweden Democrats (Sverigedemokraterna) to some extent is driven by segregation. Using municipal level data on incomes and namesday names in zip code areas an index of segregation is created in order to test this hypothesis. The results are inconclusive but there is an indication of a negative association between multidimensional segregation and the election results for the Sweden Democrats.

**Keywords:** Segregation, political economy, the Sweden Democrats (Sverigedemokraterna).

---

*E-mail: andreas@systerdyster.se*
1 Introduction

Segregation is currently a hot topic in Sweden following the 2013 June riots in the suburbs of Stockholm; during a couple of days groups of young (mostly) men burned hundreds of cars and subsequently attacked police and firemen. In essence, these events were not a new phenomenon, but have occurred several times in different Swedish cities in recent years. The public debate that followed focused a lot on the socioeconomic situation of youth in the particular suburbs, mainly housing areas built between 1965-1975 called 'million programme areas' (miljonprogramområden) that to a large extent came to be inhabited by individuals with some form of immigrant background.¹

It is suggested that these individuals are separated as a group both in terms of origin and wealth. 'Segregation' is then referring to a situation where a group of individuals with a common attribute are separated from another group of individuals with a different common attribute. The most straightforward way to visualize the separation is to think of it in terms of geography and that these groups live in different areas. Segregation can be thought of in many dimensions such as culture, wealth, ethnicity etcetera where groups of individuals can be segregated in terms of many dimensions simultaneously. Culture here does not refer to the concept in the popular sense such as ”music, film and arts”, but to the shared traits in a group of people such as language, shared ideas and customs. An economic dimension can contains variables such as personal income or wealth that in themselves can be functions of labour market conditions and public goods such as education, infrastructure and the environment. This is not a claim that none of these dimensions are entirely separate and independent of each other, only that it is useful to have such a distinction in mind since they might be separate to some extent. For instance, immigrants might be relatively poor, but it should be rare that all immigrants are poor.

How does segregation arise? If individuals prefer to live with other individuals with similar attributes as themselves they to some extent prefer a segregated society. This need not imply that this is how the individual understands it since such preferences equates with a situation in which the individual wish to live in a neighbourhood with certain attributes. Naturally then a neighbourhood will attract individuals with similar preferences regardless of whether these have preferences over their fellow neighbours. Segregation in the economic dimension is at least to some extent a consequence of how the economy works; when constructing housing for instance, it is more cost efficient of constructing to keep the variation in production low and to scale the quantity up. In the cultural dimension immigration can be considered an important factor. It is important to emphasize that immigrants do not have a common attribute except having immigrated, such as skin color, the word immigrant only refers to an individual that moves from one area to another.² Immigration as a share of total population has in Sweden risen from around 0.5 percent in 1946 to about 1 percent in 2012. Since this also includes Swedish citizens that have been living abroad for some time,

²Evidently the popular use of the term often is used to denote certain immigrants, but here the word is used in the minimalist definition.
in terms of immigration as an influence of a foreign culture the number of issued permanent 
residence permit (PRP) provides a better estimate. From 1980 (when data is available) the 
number of PRPs have increased from 0.15 percent to about 1 percent in 2012.³ Over time 
the result is that 15.4 percent of the Swedish population were born abroad (4 percent in 
1960) and 20.1 percent of the population have parents that both are born abroad.⁴

What are the political consequences of segregation? It should not be a too far-fetched 
idea that the immigration to Sweden has generated some political resentment in the pop-
ulation. The Sweden Democrats (Sverigedemokraterna) was formed in 1988 (some sources 
say 1988) out of the remnants of an earlier party, the Sweden Party (Sverigepartiet). It 
gained its first mandate in a municipal parliament in 1991 and gained representation in the 
national parliament in 2010. What differentiates this party from others in the assembly is its 
restrictive views on immigration, especially immigration from outside Europe where these 
immigrants are viewed as too culturally distant to be well integrated in the Swedish society. 
The party also explicitly oppose the idea of a multicultural society and wishes the state to 
take an active stance in the assimilation of immigrants, essentially demanding that immi-
grants adapt to the Swedish culture.⁵ Cultural segregation is less frequently pinpointed as 
a problem by the other main Sweden parties where segregation is mainly seen as a problem 
in the economic dimensions; that the relatively poor and rich end up living separated from 
each other and under very different conditions. From the perspective of the voter it is highly 
likely that the voters’ preferences regarding immigration depend on the origin of the im-
igrants and how far apart they identify themselves to be in any dimension. If immigrants 
are also relatively poor, voters might also associate immigrants status with behaviour that 
is associated with poverty. If immigrants then end up separated from natives this tendency 
of identification might be strengthened even more. The response from the voters can then 
be in a social form, “segregation has negative socioeconomic consequences”, or in terms of a 
wish to live close to people of similar attributes. The political implementation of both these 
ideas can lead to support for restrictive immigration policies, but also policies that demand 
immigrants to adapt (assimilation). Secondly, though not in itself related to segregation, 
if immigrants are relatively poor and/or less educated, voters might have, depending on 
the conditions in the labour market, reasons to prefer restrictive immigration policies for 
economical reasons. But preferences are as well highly likely to be formed as a response to perceived consequences of immigration and the challenge at hand when it comes to studying 
the relationship with segregation is to identify these factors when isolating the relationship 
of segregation with politics.

In the subsequent theoretical chapter presents a model of public goods in an economy 
with heterogenic preferences, what previous research has found and how it is possible to mea-
sure segregation. The empirical section presents the strategy for identifying the relationship 
and tests the implications given from the theoretical model.

⁴ Statistics Sweden, "Befolkningsstatistik": http://www.scb.se/Pages/TableAndChart___26040.aspx 
⁵ "Sverigedemokraternas principprogram 2011", p. 11ff: http://sverigedemokraterna.se/wp-content/ 
uploads/2013/08/principprogram_A5_web.pdf
2 Theory

2.1 Public Goods in an Economy With Heterogenic Preferences

Alesina, Baqir and Easterly (1999) develop a model in which the provision of a public good will depend on the preference of the median voter for the type and quantity of the good. If society is polarized into two homogeneous groups, the provision of the public good will be much smaller since each group will be far away from the median voter.

The population in the model is normalized to 1 where individuals are immobile between constituencies. The size and type of the public good is decided in that consecutive order by majority rule. The generic individual has the utility function:

\[ u_i = g^\alpha (1 - l_i) + c \]  

...where the parameter \( \alpha = (0, 1) \), \( g \) the public good, \( l_i \) the preference distance between individual \( i \) and the actual public good and \( c \) private consumption. Income, \( y \), is assumed exogenous and equal for everybody, consumption is equal to disposable income, which is income net taxes \( c = y - t \), where taxes cannot be a function of preferences. Since population size is normalised to 1 the public budget constraint implies that \( g = t \). Individual \( i \) then has the following maximization problem:

\[ \max U_i = g^\alpha (1 - \hat{l}_i) + y - g \]  

...where \( \hat{l}_i \) is the distance from the ideal type of median voter.\(^6\) The solution for this maximization problem, \( g^* \) is given by:\(^7\)

\[ g_i^* = \left( \alpha(1 - \hat{l}_i) \right)^\frac{1}{1 - \alpha} \]  

Aggregating the individuals’ optimal amount of the public good yields:

\[ g^* = \left( \alpha(1 - \hat{l}_m) \right)^\frac{1}{1 - \alpha} \]  

...where \( \hat{l}_m \) is the median distance from the most preferred public good by the median voter.

Even though not explicitly stated in the paper, \( l \) should most likely be a number between or equal to 0 and 1, \( 0 \leq l \leq 1 \), since if \( l > 1 \) there’s a possibility that utility will be a complex number in the plane. The authors also talk about \( l \) as a distance so \( l < 0 \) would neither make any sense.

The question is how this model can be related to segregation. When testing this model empirically the authors emphasize 3 implications of the model for more ethnically fragmented localities.

\(^6\)This equation (5) in Alesina, Baqir and Easterly, p. 1249, is written \( U_i = g^\alpha (l - \hat{l}_i) + y - g \). There is no motivation for why 1 is replaced with \( l \) when going from \( u_i \) to \( U_i \). Since the solution then seem not to follow from the maximization problem it could be a typo. This print seem to support that: http://www.nber.org/papers/w6009.pdf

\(^7\)The derivation of this result can be found in the appendix.
1. The share of public spending is lower.

2. The direction of the correlation need not have any special direction due to the fact that patronage can be related to ethnic fragmentation.

3. The fiscal discipline is lower.

While the first implication follows directly from the mechanics of the model, the second and third comes from the authors’ discussion of the model. They note that parts of the public goods can be directed towards, for instance, ethnic minorities in the form of direct transfers or public employment (patronage). Hence, \( g \) can be divided into a patronage, \( g_1 \), and a public, nonexcludable good, \( g_2 \), that satisfy \( g = g_1 + g_2 \). The argument is that with increased group polarization increases in \( g_1 \) will be larger relative to \( g_2 \), that could decrease. Thus there is stronger implications for the spending on pure public goods as a share, \( g_2/(g_1 + g_2) \) than on their level.\(^8\)

How does the above model translate into political support? A vote maximizing party will have an incentive to model its budget taking into consideration to the heterogeneity of preferences and direct a larger share of the public budget to specific groups (\( g_1 \)) and a smaller share of nonexcludables (\( g_2 \)). This result is derived from a political setting in which decisions are taken by majority rule which may not be a decision supported by the majority of the voters. In many democracies with proportional representation parties have to negotiate with other parties in order to reach a majority in the parliament to take decisions. One implication is that in a proportional system, if a party does not receive a majority of its own, there will be a tendency to take decisions closer to the preferences of the median voter as in the model above though to a lesser degree.

The question then is how voters react to changes in the provision of nonexcludable public goods and patronage? If the heterogeneity of preferences is a function of segregation, given that there are a group of voters that are mainly affected by provisions in the nonexcludable public good their utility will decrease with reductions in that good and will as utility maximizing agents react by voting for parties opposing such reductions or parties that wish to increase patronage.

But segregation also has the potential of driving voter behaviour through other channels. If groups are segregated from each other both in terms of culture, which implies some difference in terms of preferences, and wealth. Problems associated with relative poverty might be perceived by voters as a difference in terms of culture. This issue will be addressed in the empirical section.

### 2.2 Previous research

Alesina, Baqir and Easterly (1999) test their model described in the section above using cross sectional data on public spending in US cities, metropolitan areas and counties. Using an ethnic fractionalization index, that shows the probability that two randomly drawn citizens

\(^8\)Alesina, Baqir, and Easterly, 1999, pp. 1247-1251.
belong to two different groups. The authors especially emphasize the result that ethnic fragmentation is negatively related to public spending on education, roads, sewerage and thrash pickup.9

Åslund & Nordström (2010) show that immigrants in Sweden are both more likely to work in firms where people with similar origins and other immigrants are more frequent comparing with a random distribution.10

In a study by Rydgren & Ruth (2011) the authors test whether the electoral support for the Sweden Democrats can be explained by the theories of social marginalisation or ethnic competition. Their model consist of the percentage votes for the Sweden Democrats as the independent variable and gross regional product (GRP), unemployment, education (post-secondary education of at least 3 years), total, Nordic, EU/EFTA and non-European immigrants, crime (offences per 1000 inhabitants), the proportion of votes for social democrats and the left-wing party, population size and a set of interaction variables between immigrant type, crime and unemployment as independent variables. The most robust, significant variables are GRP and unemployment where the former has a negative coefficient and the latter positive. Education is also significant, but the direction of the effect depends on the econometric model. The authors thereby claim that their results are in line with the theory of social marginalisation.11

Dahlberg, Edmark & Lundqvist (2012) estimates the impact of immigration on preferences for redistribution instrumenting the share of immigrants with the exogenously determined refugee inflow that at the time was determined outside the control of the municipality. Additional variables in this model are unemployment, vacant housing, dummies for municipality size, dummies for socialist majority, green and New Democrats12 representation in the council. They find that immigration has a negative effect on preferences for redistribution.13

In the study by Bertrand & Mullainathan (2004) the authors send out fake job applications signed with black and white sounding names to estimate the discrimination of blacks in the US labour market. The explanation the authors give for their finding that applications with a black sounding name receive a significantly lower callback rate is that since there are often a lot of applications per job, a higher rate of these applications get sorted out due to subconscious associations with black names. This explanation is formed from the psychological theory of implicit discrimination where individuals act in a discriminatory way despite a conscious will to do so. It takes place at a subconscious level where certain visual attributes, such as a name or appearance, are associated with qualitative differences that give rise to negative or positive emotions.14 This might have some bearing on segregation as well. If individuals are separated in terms of culture in a society, the individuals can have a preference to live in a society without segregation but still chose a way of live where the consequence is the same as if they did act in a discriminatory way.

---

10Åslund and Skans, 2010, p. 481.
12This party was represented in the parliament during 1991-93, but can nowadays be considered extinct.
14Bertrand, Chugh, and Mullainathan, 2005.
2.3 Measuring segregation

It is clear that the measurement of segregation is dependant on how the areas are defined, such as that when the earth is divided to a southern and northern hemisphere the measured differences in several dimensions, such as weather, will change if the border moves. Since the definition of such areas always will depend on either societal institutions or some rule set up by an individual, it is more appropriate to think of segregation as clustering, meaning objects of similar type that are grouped together, where there is no need to compare areas. However, performing such an analysis is too time consuming for this thesis.

There are a number of properties a measure of segregation should satisfy. It is appropriate that its range is between zero and one, \([0, 1]\), since it makes it easy to think of the segregation as going from none to complete segregation. The measure should be invariant to population size and symmetrical in the sense that using X or Y as the properties in the population does not render two versions of the index. Furthermore, the ratio of X to Y should not affect the index. Two areas with identical properties could be treated as one area without affecting the index as well as the total size of the populations.\(^{15}\) Another important property to satisfy is the principle of transfers that states that any move of individuals from one area to another should result in a decrease in segregation if the differences between these areas become smaller after the transfer.\(^{16}\)

Say then that we have a set of two areas, \(A_i, i = \{1, 2\}\), each with a population of \(N_i\) individuals. These are of two types, \(\{X, Y\}\), and this dimension can be called \(\alpha\). It is assumed that individuals are identically and randomly distributed in both areas. Pick one individual at random in \(A_1\) or \(A_2\) and the probability of getting a type \(X\) member is \(P(X|A_i) = \frac{X_i}{N_i}\). If there is no segregation, there should be an equal probability of picking an individual of a certain type for both areas, \(P(X|A_1) = P(X|A_2)\). Thus, segregation can be defined as the absolute value of the difference in probabilities: \(|P(X|A_1) - P(X|A_2)|\) gives us a measure of segregation in terms of \(X\)-type individuals. But it is desirable to have a measure that works for any number of areas. If we instead have \(i = 1, 2, ..., I\) areas, the segregation can be defined as:

\[
\Phi_{\alpha}(X) = \frac{1}{I-1} \sum_{i} |P(X_i) - E[P(X)]| \tag{5}
\]

That is simply the degrees-of-freedom adjusted absolute deviation of \(P(X)\). If \(\Phi = 0\) there is no segregation, while if \(\Phi = 1\) there is complete segregation; the probability of picking an individual of a certain type from an area depends only on which area it is. Thus \(\Phi = [0, 1]\). Since the measure basically answers how much the proportion of \(X\)'s deviate on average from the mean proportion, the number can be interpreted as the average percentage deviation. If \(\Phi(X) = 0.3\) the chance of picking \(X\) at random on average differs with 30 percent. \(\Phi\) does satisfy all desirable properties for a segregation index if \(X\) and \(Y\) are exhaustive; \(P(Y) = 1 - P(X)\), thus it is invariant as to whether \(X\) or \(Y\) is used to calculate \(\Phi_{\alpha}\).

\(^{15}\) Blackburn, 2012, p. 187f.

\(^{16}\) Dawkins, 2004.
The index of dissimilarity is quite common in the literature and provides a more straightforward interpretation, that is the proportion of a group that has to move to make the areas equal.\(^{17}\)

\[
D = \frac{1}{2} \sum_i \left| \frac{X_i}{X} - \frac{Y_i}{Y} \right|
\]  

(6)

The problem with this index is that sensitive to the size of the population in an area; even though the proportions in area 1 and 2, \(X_1, Y_1 = X_2, Y_2\), are the same, if \(N_1 > N_2\), segregation will appear worse in area 1. The consequence is that \(D\) puts restraints on the units that are compared since they have to be equal in size to be comparable. \(D\) also violates the principle of transfers if individuals move to an area where the difference in proportions is greater or less than the mean difference; segregation remains constant for all moves of one type of individual above the mean and vice versa. It should be fairly obvious since if we think of the move of one individual from one area to another, decreasing the difference with 1 in the first area and increasing the difference with 1 in the second area.

Since segregation can be thought of in several dimensions, the interaction of two variables might be of interest, say \(\Phi_\alpha\) and \(\Phi_\beta\). Let \(\Phi_\alpha\) be as above with respect to the types \(\{X, Y\}\), but add that they as well can have a high (H) or low (L) income, called dimension \(\beta\). Just multiplying \(\Phi_\alpha\) with \(\Phi_\beta\) has the drawback of not providing any information to answer the relevant question of to what extent the areas share both the property \(X\) and \(H\) or \(L\). There are four possibilities in this two dimensional space: \(X \cap H\), \(X \cap L\), \(Y \cap H\) and \(Y \cap L\). This assumes either that there is information for both dimensions for each individual in all areas or that one of the dimensions are randomly distributed on top of the other, in other words; that any significant correlation between \(\alpha\) and \(\beta\) is spurious or haphazard. Since it is desirable to measure the multidimensional segregation that gives an answer that captures all dimensions simultaneously one can think of it geometrically. Since we are dealing with an exhaustive set of probabilities the following hold true: \(P(X)P(L) + P(X)P(H) + P(Y)P(L) + P(Y)P(H) = 1\) and thus one can think of multidimensional segregation as the area created between two coordinates.

Define the area of the quadrant that can be created from the two coordinates \((X, L)\) in the areas \(i\) and \(j\) as \(S_{i,j} = |P(X_i) - P(X_j)| \times |P(L_i) - P(L_j)|\). This is the purple quadrant in Figure 1 below. Since the comparison now happens between two areas, each area has to be compared with all other to find the average deviation. That is the answer to how many paired combinations one can form from \(I\) objects: \(2C_I = \frac{n!}{2!(n-2)!}\). Calculating \(I!\) quickly becomes difficult, already when \(I = 10\) \(I!\) equals 3’628’800. But it is possible to make a simplification.

\[
2C_I = \frac{I!}{2!(I-2)!} = \frac{I \times (I-1) \times (I-2) \times ... \times 1}{2 \times (I-2) \times (I-3) \times ... \times 1} = \frac{I(I-1)}{2} = \frac{I^2 - I}{2}
\]  

(7)

\(^{17}\text{Duncan and Duncan, 1955, p. 211.}\)
The expected value of $S_{i,j}$ is given by $E[S_{i,j}] = \frac{2}{I^2-I} \sum_{i=1}^{I-1} \sum_{j=i+1}^{I} S_{i,j}$

$$\Phi_{\alpha,\beta}^2 = \frac{2}{I^2-I-2} \sum_{i=1}^{I-1} \sum_{j=i+1}^{I} |S_{i,j} - E[S_{i,j}]|$$ (8)

The interpretation of this area is less straightforward; in Figure 1 is it easy to see that the $S$-part of $Q_1$ is a part of $P_1(Y \cap H)$ and $Q_2$ is correspondingly a part of $P_2(X \cap L)$. So the area is in a sense an estimate of how distant the areas are from each other in terms of probabilities. It is easier to consider the extreme cases; if it on average is 0, for instance, we know that in there is no segregation in at least one dimension (just consider that when $P_1(X) = P_2(X)$ there is still segregation in terms of $\{H,L\}$). It is also easy to see that one can "desegregate" by moving any $X \cap L$ individuals from one area to another given that this share is lower in the area to which it is moved. As can be seen in the figure if such an individual would move from area 1 to area 2, the coordinates $P_1$ and $P_2$ necessarily has to come closer to each other and thus reduce the area of $S$.

From this is it straightforward to define an $n$-dimensional measure since it might be appropriate to use this measure at the one dimensional level as well. First we define $S^n_{i,j}$ as the size given by product of the absolute difference of the probabilities in dimension $k$ to $n$ between area $i$ and $j$.

$$S^n_{i,j} = \prod_{k=1}^{n} |p^k_i - p^k_j|$$ (9)

If one thinks about it geometrically, $S$ goes from the length of a line ($n=1$) then the area of a quadrant to the area of a cube and so forth. The $n$-dimensional measure of segregation is then given by the degrees of freedom adjusted, average deviation from the mean size.

$$\Phi^n = \frac{n!}{\prod_{i=I-n+1} n!} \frac{1}{I-I-n+1} \sum_{i=1}^{I-1} \sum_{j=i+1}^{I} |S_{i,j} - E[S_{i,j}]|$$ (10)

Another attractive feature comes from thinking about the meaning of these indices in a more ordinary sense. For where a measure of the variance of the probability over areas is in essence a comparison against the mean probability it implies that an individual would have to know the mean in order to know the extent of segregation. It seems more likely that any individual has an idea of the average difference between areas since it is itself an inhabitant of one.

### 3 Empirical Setup

What variables are available for testing the model laid out above? What is needed is data on several political units, each possible to divide up in smaller geographical areas. The political unit could be the municipality and the area based on the zip codes. Since it can be suspected that immigration plays some role in the forming heterogeneous preferences the electoral support for the Sweden Democrats can be used as a dependent variable since their political...
By measuring the variation in the intersection, $S$ (purple), between the quadrant $Q_1 \subseteq \text{Area}_1$ (blue) and $Q_2 \subseteq \text{Area}_2$ (red) you get an estimate of the variation in both dimensions. $S$ is the only quadrant that will change size for any movement in $P_1$ or $P_2$, meaning that the variation in $S$ can be used to estimate multidimensional segregation.

An important question is whether voters identify their preferences regarding immigration and segregation with the appropriate political unit. In Sweden the responsibility for the extent of immigration is decided at the national level while the responsibility for receiving refugees is taken in negotiations between the counties, municipalities and the Swedish Migration Board (Migrationsverket).\textsuperscript{18} Voters might cast different votes in the municipal

and national election depending on local circumstances which, but as a large share of the Swedish population lives in big city regions, such as Stockholm, where it can be expected that policies in one municipality also has effects outside the municipal borders.

One way of constructing a measure of segregating is taking a similar approach as Bertrand & Mullainathan (2004) did in a field experiment studying labour market discrimination where they use “black and white sounding names”. One can then use namesday names as a proxy for culture. Statistics Sweden (SCB) explains that approximately 75 percent of 30 000 children born in 2010 got their names from a grandparent within three generations. Of course there is no essential connection between culture and the name of an individual, nor between an immigrant and culture, but it is more likely that the culture of an immigrant or an individual with a non-traditional name diverges from the native culture. It is important to point out that culture is not a very strict concept with clearly defined categories; it is more of a range of values that sometimes overlap.

Regarding the choice of covariates a couple of points need to be addressed. As it has been suggest, immigration plays an important role in cultural segregation and thus the question of how immigration affects the economy is relevant. First, the inflow of immigrants affects the factors of production through an increase in the supply of labour. The standard economic models in trade theory gives different answers depending on whether you assume that the factors of production are mobile between sectors. That labour can flow from one industry to a completely different is a reasonable assumption in the long run, but not in the short run. Thus the answer given from the models is that in wages will be reduced in the short run, but will rise as production adjusts itself. Obviously this ignores the institutional context and the fact that in many countries it is very rare that wages fall. That real wages fall is probably quite common due to inflation, but in a highly unionised country such as Sweden it is questionable how useful these models are as an analytical tool. The consequences for the native working population will depend on the interaction of the labour market’s demand, how well immigrants are able to match this demand and how flexible the production is to adjust to changes in the labour force. Secondly, immigrants increase consumption, which happens instantly, especially for products such as housing where the production process relatively long and complicated comparing to products such as food that rather easily can adjust itself through imports and is probably not something voters can observe even if it had a measurable effect. From previous studies controls for these effects have come in the form of gross regional product, vacant housing, unemployment and education (which can be motivated from other perspectives as well).

In the sociological literature there are a couple of theories that try to explain electoral support for parties such as the Sweden Democrats, either by focusing on ethnic competition or social marginalisation. Some groups that are relatively poor in comparison with a reference group or history, might identify immigrants as the cause of their poverty. The argument is similar when it comes to other economic variables, such as welfare, or crimi-

nality though the groups that feel threatened might be different. Other theories point more to certain values cherished by some groups, such as national identity, culture or religion.21 This implies that variables such as criminality, and education might be relevant as controls which are also used in the study by Rydgren & Ruth (2011) mentioned above.

3.1 Data

Data on incomes for zip-code areas are available for all 290 Swedish municipalities through Swedish Radio (Sveriges radio) for 2009 and are reported as means and medians along with the population in each quartile (which appear to be defined nationally).22 The Swedish zip codes in the sample range from 10005 (Stockholm) to 98499 (Pajala) where there are some numbers not associated with any area (such as numbers starting with 32, 48, 49 and 99) or they are not associated with households, but used by companies, lotteries and post office boxes. As evident from the numbering starting with Stockholm they are not numbered entirely so that with increasing number the region is always further north, but starts with Stockholm (numbers starting with 10-19) and then goes from south to north. A zip code may only be associated with one locality, but can span more than one municipality and county.23 Since the creation of this dataset is not publicly documented it must be pointed out that it’s not entirely clear of how complete it is in all regards.

In total there is data on incomes in 9460 zip code areas where the number of zip code areas in each municipality varies between 1 (Åsele) and 787 (Stockholm) with a median of 22 and a mean of 36.5. The population consist of 5'446'889 individuals in the age span 20-64 years. Since it is not possible to construct a spatial segregation measure using only one region, Åsele has to be dropped. Since the quartiles are constructed in the sample as a whole and they do not take into account that the cost of living can differ considerably depending on the economic conditions in the municipality. This need not be a severe problem since it affects all regions in each municipality equal and Φ measures the variation between regions. But it does pose a problem if the mean proportion of individuals with lower incomes are of interest since that share will be higher in smaller and/or less populous municipalities. This is confirmed by regressing the zip codes on the share of the population with lower incomes; as one moves 10000 step upwards in zip codes the lower income shares increase significantly ($p < 0.001$) with 0.0072 giving approximately a 6 percent difference between localities in the Stockholm area and the ones in the farthest north. However this relationship seems mainly to be driven by the Stockholm region (zip codes starting with 10-19) since excluding these the relationship becomes insignificant ($p \approx 0.9$). There is though a lot of unexplained variation as expected and there might be other patterns of relevance.

Election data for municipalities are from the most recent election in 2010 and consist of the percentage of votes for the Sweden Democrats to the municipal and national parlia-
ment.\textsuperscript{24} Data on education is the share of population between 25 to 64 years of age in a municipality with a postsecondary education of at least 3 years (2011). Crime comes as the number of reported offences on violence and theft per 100’000 inhabitants (2010).\textsuperscript{25} Density the number of inhabitants per square kilometer. Data on the number of refugees accepted by the municipalities is given by the Migration Board which is weighted by the population density of the municipality. Unemployment is the share of population between 20 and 64 years of age that some time during the year were registered as unemployed.\textsuperscript{26}

The method of collecting data on individuals has been through searching 118100.se using zip codes.\textsuperscript{27} The search engine does not allow you to make more sophisticated searches, such as only using zip codes, but returns any match in personal and company details corresponding to the sequence of numbers entered. But it does provide filters enabling you to restrict your hits to individuals, county and municipality. The data must still be filtered more to make it useful since any person that is associated with several phone numbers returns a separate hit for each phone number with some exceptions. Three criteria have been implemented in this filter in order to deal with this problem:

1. To count as an individual it must have a name that consist of at least 2 words, in total 3 alphabetical characters where 1 is a space.
2. Any individual may only occur once associated with one address.
3. An individual’s zip code and city of residence must correspond to the official registry.

Criteria 1 is not a very strict criteria as individuals with such short names ought to be very rare, if existing at all, but it is easy to implement and understand. Criteria 2 ensures that no double counts occur and it economizes on data size. Criteria 3 just checks the validity of the data, that it is correct; if you live in Gothenburg you cannot have a zip code that belongs to Malmö. Applying this filter results in three categories of data; doubles, errors and valids. Doubles are rather numerous making up 17.2 percent of all hits while errors account for about 0.6 percent.

Another program has then been constructed to provide additional filtering capabilities, such as searching for double entries, encoding individuals as having namesday names or not and calculating their shares of the population in each zip code. A list of namesday names, including names that have disappeared in revisions, have been collected from Wikipedia. The classification of what name is to be counted as a namesday name has not been set strict as to that the spelling should be exactly the same, but instead the Loewenshtein algorithm has been used to allow for deviations of up to 1 character in length or content, such as that Karl is sometimes spelled Carl or Mohammed and Muhammed. In total this dataset consist of 6'949'267 observations that are coded as 1 if the individual has a namesday name and 0 otherwise.

\textsuperscript{24}Statistics Sweden: http://www.scb.se/Statistik/ME/ME0104/2010A01v/RI_tabell%20206.xls
\textsuperscript{25}Swedish National Council for Crime Preventions (BRÅ): http://bra.se/bra/brott--statistik/statistik/annalda-brott.html
\textsuperscript{26}Statistics Sweden: http://www.scb.se/Pages/SSD/SSD_TablePresentation____340486.aspx?layout=tableViewLayout1&rrid=0162c172-f381-4db8-a133-27a8c1649e0
\textsuperscript{27}A detailed account of this process can be found in the appendix.
A potentially important source of error is the fact that it is not clear how this database gathers information on individuals. It appears as though the record mainly consist of individuals that have a personal subscription to a phone company with a stationary and/or mobile phone. Persons only using a mobile phone card that is bought anonymously can obviously not be registered. Then, if there is correlation between whether the individual has a namesday name or not and existing in the database, which could run through a correlation with wealth where subscriptions could be less frequent the less wealthy you are, the estimate of segregation will be biased downwards. At last it is important to remember that there still are invalid records present which cannot easily be filtered out. A problem is that they are almost certain to be classified as non-namesday names, but it is questionable whether they are numerous enough to have a significant impact and their distribution across zip code areas are unknown.

It is important to emphasize that the data is dispersed in time with incomes in 2009, votes in 2010 and namesday name proportions from 2013. Needless to say things have changed since the votes were laid and income data were collected and it is impossible know for certain if any changes have occurred that could influence the estimation, but since our interest lies in the aggregate it is seems fairly reasonable to assume that this data gives an approximate reflection of the situation today. A more important source of error could be that the calculation of segregation using namesday names is made with a sample that is almost 30 percent larger which leaves some concern over how the sample is constituted regarding age for example.

### 3.2 Econometric specification

In constructing the segregation measure we take the share of the population in the two lower income quartiles. The covariates consist of crime per capita, population density, the number of refugees weighted by density, the maximum difference between two areas (the maximum of $S_{i,j}$), the mean share of individuals with namesday names and lower incomes in zip code areas.

$$VOTES_m = a + b_1 \Phi_{NN,m} + b_2 \Phi_{Y,m} + b_3 \Phi_{NN,Y,m} + C_m + e_m$$

$VOTES_m$ is the percentage votes in the national election for the Sweden Democrats in municipality $m$. $\Phi_{NN}$ is our namesday names proxy for cultural segregation, $\Phi_Y$ is the economic segregation based on the proportion of individuals in the two lower income quartiles, $\Phi_{NN,Y}$ the two dimensional segregation measure and $C$ is a matrix of covariates. Two approaches will be used in trying to identify the support for the Sweden Democrats that is specific to a certain municipality. First, it is known that the Sweden Democrats has for a long time been established in the southern parts of Sweden, such as the municipality of Höör where the party has had seats since 1991. Using counties or electoral constituencies is thus a viable approach. Since Sweden is divided up into 29 constituencies and 21 counties the former has

28The author himself does not exist in the records, most likely due to that I do not have a phone at home and that I do not have a subscription to any phone company since I use a cash card bought anonymously at a shop.
the benefit of providing a slightly lower level of aggregation of the municipalities. Secondly, it could be possibly to identify the support the party has gained in the municipalities in past elections. One way could then be to take the cumulative percentage votes or seats in municipal parliaments in the elections 1991, 94, 98, 2002, and 2006. The problem is though that this might explain too much of the variation if the party is increasingly successful in those municipalities that are more segregated relative to others. Instead it could be more better to take the number of years that the party has gained seats in the municipal parliament. Since more than 20 years has passed since the first election in which the party participated this figure would give take less consideration of how segregation might have affected the support for the party over time.

A worry regarding this model is that all estimators of $\Phi$ are highly correlated due to the fact that both $\Phi_{NN}$ and $\Phi_Y$ are related to $\Phi_{NN,Y}^2$ since they are essentially measuring the variation in the sides of the quadrant. The consequence, imperfect multicollinearity, for the results is that potentially, at least one of the coefficients $b_1$, $b_2$ and $b_3$ will be imprecisely estimated, meaning that it might render the coefficient insignificant when it actually is.

Figure 2: Segregation in Swedish municipalities

This figure shows the score for each municipality in Sweden (with the exception of Åsele) in the namesday name $\Phi^{1}_{NN}$, income $\Phi^{1}_{Y}$ and the two dimensional measure, $\Phi^{2}_{NN,Y}$, in the colored z-dimension. All axes can be interpreted in terms of percentages. The one dimensional measures are given in the horizontal plane, $x$ and $y$ axes, are the absolute deviations from the mean difference considering all regions.

Figure 2 shows the score of the Swedish municipalities in the one dimensional segregation measures in the $x$, $y$ plane and the multidimensional in the $z$ dimension which is also colored going from low levels in black to high levels in yellow.

4 Results

It is important to emphasize here that none of these specifications are able to disentangle any causal relationship, but it is neither suggested by theory that segregation in itself causes individuals to vote in a certain way. As discussed in the theoretical sections it is a kind of proxy for the heterogeneity of preferences.
Table 1: Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD votes</td>
<td>6.31</td>
<td>5.80</td>
<td>1.70</td>
<td>15.80</td>
<td>2.53</td>
</tr>
<tr>
<td>$\Phi_{1,N}$</td>
<td>2.62</td>
<td>2.12</td>
<td>0.72</td>
<td>11.27</td>
<td>2.49</td>
</tr>
<tr>
<td>$\Phi_{1,Y}$</td>
<td>6.12</td>
<td>5.68</td>
<td>0.46</td>
<td>13.68</td>
<td>1.72</td>
</tr>
<tr>
<td>$\Phi_{2,Y,NN}$</td>
<td>0.58</td>
<td>0.31</td>
<td>0.01</td>
<td>4.58</td>
<td>0.74</td>
</tr>
<tr>
<td>$E[N</td>
<td>N]$</td>
<td>82.39</td>
<td>82.81</td>
<td>67.47</td>
<td>89.64</td>
</tr>
<tr>
<td>$E[L]$</td>
<td>51.40</td>
<td>51.29</td>
<td>34.91</td>
<td>63.51</td>
<td>0.06</td>
</tr>
<tr>
<td>$\max S_{Y,NN}$</td>
<td>5.09</td>
<td>2.25</td>
<td>0.02</td>
<td>54.21</td>
<td>7.34</td>
</tr>
<tr>
<td>education</td>
<td>17.14</td>
<td>15.22</td>
<td>8.27</td>
<td>54.66</td>
<td>7.21</td>
</tr>
<tr>
<td>refugees</td>
<td>4.86</td>
<td>1.51</td>
<td>0.00</td>
<td>102.50</td>
<td>12.22</td>
</tr>
<tr>
<td>theft</td>
<td>42.09</td>
<td>41.22</td>
<td>14.30</td>
<td>92.56</td>
<td>13.34</td>
</tr>
<tr>
<td>unemployment</td>
<td>7.69</td>
<td>7.60</td>
<td>3.60</td>
<td>12.00</td>
<td>1.72</td>
</tr>
<tr>
<td>density</td>
<td>137.3</td>
<td>26.5</td>
<td>0.2</td>
<td>4504.3</td>
<td>474.30</td>
</tr>
<tr>
<td>SD years</td>
<td>0.63</td>
<td>1.00</td>
<td>0.00</td>
<td>5.00</td>
<td>0.74</td>
</tr>
</tbody>
</table>

For models 1 to 6 where there are several measures of segregation an F-test with the hypothesis that all coefficients of segregation are 0 is performed. In models 1 to 3 the focus is laid on one dimensional segregation and the interaction between these. The interaction term gives a score of how segregated a municipality is in terms of both dimensions though without containing any information of whether it is the same areas that are segregated in one dimension are also segregated in the other. The interaction term has a negative coefficient and is weakly significant ($p < 0.1$) in model 1 without any fixed effects, but becomes insignificant when fixed effects are added. Models 4 to 6 have the same approach, but the interaction term is now replaced with the multidimensional measure of segregation. This term is negative in all specifications and weakly significant in models 5 and 6 with fixed effects. In models 7 and 8 the one dimensional segregation measures are dropped since they might be affecting the variance of the estimate of multidimensional segregation. Both estimates are still negative and weakly significant where the the coefficient in model 8 is slightly more precise. As such they imply that a one percent increase in the score on multidimensional segregation is associated with about 0.6 to 0.7 percent reduction in the votes laid for the Sweden Democrats. This implies that as areas become more different in terms of the share of namesday names and income, less individuals in the municipality lay their vote on the Sweden Democrats given the covariates. The problem is that this segregation can also be influenced by covariates such as unemployment and refugee inflow so the translation to an observation that multidimensional segregation has increased cannot be translated directly to the expectation that the Sweden Democrats will receive less votes in the coming election.

The only hypothesis that all coefficients of segregation is 0 that can be rejected is the one in model 1. Since this model is set without fixed effects it is questionable whether this rejection has any relevance.

Of the covariates the most robust estimate is that of education, which has a highly significant, negative coefficient in all specifications. That might be slightly unexpected given that even the direction of the coefficient differed in the estimations by Rydgren &
Table 2: Estimation results

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
<th>Model 7</th>
<th>Model 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>22.16***</td>
<td>10.10*</td>
<td>11.45*</td>
<td>22.25***</td>
<td>10.00*</td>
<td>11.34*</td>
<td>11.56**</td>
<td>12.42**</td>
</tr>
<tr>
<td></td>
<td>(6.06)</td>
<td>(4.71)</td>
<td>(4.68)</td>
<td>(6.10)</td>
<td>(4.69)</td>
<td>(4.66)</td>
<td>(4.32)</td>
<td>(4.27)</td>
</tr>
<tr>
<td>$\Phi_{NN}$</td>
<td>0.14</td>
<td>0.04</td>
<td>-0.01</td>
<td>-0.26</td>
<td>0.15</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.32)</td>
<td>(0.21)</td>
<td>(0.20)</td>
<td>(0.30)</td>
<td>(0.19)</td>
<td>(0.19)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Phi_{Y}$</td>
<td>-0.05</td>
<td>0.04</td>
<td>0.05</td>
<td>-0.15</td>
<td>0.05</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.07)</td>
<td>(0.07)</td>
<td>(0.10)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Phi_{NN} \times \Phi_{Y}$</td>
<td>-0.05°</td>
<td>-0.02</td>
<td>-0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Phi_{Y,NN}$</td>
<td></td>
<td></td>
<td></td>
<td>-0.19</td>
<td>-0.97°</td>
<td>-0.95°</td>
<td>-0.60</td>
<td>-0.68°</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.84)</td>
<td>(0.56)</td>
<td>(0.56)</td>
<td>(0.38)</td>
<td>(0.38)</td>
</tr>
<tr>
<td>$E[NN]$</td>
<td>-0.21**</td>
<td>0.02</td>
<td>0.00</td>
<td>-0.20**</td>
<td>0.02</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.07)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>$E[L]$</td>
<td>0.03</td>
<td>-0.02</td>
<td>-0.01</td>
<td>0.02</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.04)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>$\max S_{Y,NN}$</td>
<td>0.06</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.07°</td>
<td>0.07°</td>
<td>0.08°</td>
<td>0.08°</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.05)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>education</td>
<td>-0.13***</td>
<td>-0.14***</td>
<td>-0.14***</td>
<td>-0.13***</td>
<td>-0.15***</td>
<td>-0.15***</td>
<td>-0.15***</td>
<td>-0.15***</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>refugees</td>
<td>-0.17**</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.17**</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.06)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>theft</td>
<td>0.07***</td>
<td>0.01</td>
<td>0.01</td>
<td>0.08***</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>unemployment</td>
<td>0.06</td>
<td>0.16*</td>
<td>0.12</td>
<td>0.06</td>
<td>0.16*</td>
<td>0.12</td>
<td>0.16*</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.10)</td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>density</td>
<td>0.00°</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00°</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>SD years</td>
<td>0.34**</td>
<td>0.34**</td>
<td>0.34**</td>
<td>0.34**</td>
<td>0.34**</td>
<td>0.34**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
<td>(0.13)</td>
<td>(0.13)</td>
<td>(0.13)</td>
<td>(0.13)</td>
<td>(0.13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rej. hyp. (F)</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>Yes</td>
<td>Yes</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>all $b_0 \Phi = 0$</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.30</td>
<td>0.79</td>
<td>0.80</td>
<td>0.29</td>
<td>0.80</td>
<td>0.80</td>
<td>0.79</td>
<td>0.80</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.27</td>
<td>0.76</td>
<td>0.77</td>
<td>0.26</td>
<td>0.76</td>
<td>0.77</td>
<td>0.76</td>
<td>0.77</td>
</tr>
</tbody>
</table>

The dependent variable is the percentage votes in each municipality for the Sweden Democrats to the national parliament. $\Phi_{NN}$, $\Phi_{Y}$ - Segregation in terms of NN and Y separately. $\Phi_{NN} \times \Phi_{Y}$ - Interaction term. $\Phi_{NN,Y}^2$ - Segregation in terms of dimensions namesday names and income. $E[P(\cdot NN)]$, $E[P(\cdot L)]$ - The mean share of the population with namesday names and lower income in the areas. $\max S_{NN,Y}$ - The maximum two dimensional difference. education - The share of the population with at least 3 years of university education. refugees - The number of refugees per 1000 inhabitants in the municipality. theft - number of offences per 100 000 inhabitants. unemployment - The percentage of the population registered as unemployed during 2010. density - population per km². SD years - The number of elections the Sweden Democrats have gained seats in the local parliament 1991-2006. Rej. hyp. (F) all $b_0 \Phi = 0$ is the hypothesis that all coefficients of the segregation measures is 0; "YES" if this hypothesis is rejected, "NO" otherwise.
Ruth (2011). One could expect that as the share of the population that has a higher education increases, the Sweden Democrats receive less votes. Since we do not know if it is a causal relationship the question is if it is the case that people with certain values proceed to get a higher education or is these values something they attain during their education? Of course, it is a possibility for the humanities, but the latter might be more generally true in the sense that education as an institution promotes a certain set of values.

5 Concluding Discussion

This thesis has in essence tested an implication from a theory of the provision of public goods in an economy with heterogenic preferences developed by Alesina, Baqir & Easterly (1999). There is an indication that the percentage votes the Sweden Democrats receives in the national elections is negatively correlated with multidimensional segregation, which is of the opposite direction that one would expect from theory given that it is correct to interpret the support for this party as a vote for a form of cultural patronage. Leaving that issue aside, one possibility is that since segregation implies that the groups are separate it might be more difficult for such preferences to form. In theory it seems more likely that multidimensional segregation affects preferences through criminality and social unrest since it in essence is measuring the association of culture with relative poverty. Another point that seem relevant is that in recent years a number of policies have been implemented that directly benefits the relatively rich in the form of different tax deductions and public choice implementations in welfare institutions such as education and health care. Tax deductions are in essence not public goods but they are definitely benefiting a group that is quite likely to appear in the namesday name calendar. It is evident from the model and equation 1, \( u_i = g^\alpha (1 - l_i) + c \) where \( c = y - t = y - g \), that lower taxes increases the individual’s utility more than public goods. Public choice implementations allow for greater flexibility when it comes to the type of good which individuals consume. Since each individual will be able to pick a good that is closer to the ideal this could dampen the call for group specific policies. This is also a group that is in general better equipped to benefit from being able to chose school and healthcare. This tendency might have counteracted an increase in the support for the Sweden Democrats. Testing the model in these areas is most likely more appropriate.

This thesis has also left the time dimension of segregation aside completely. Despite the fact that the data is dispersed in time the empirical analysis has aimed to pinpoint a relationship in a window in time. It is highly unlikely that the time dimension of segregation is irrelevant. Voters can, for instance, be more responsive to changes in segregation than levels. The Sweden Democrats is also a relatively new party in the national arena and has increased their share of the votes substantially in each election. The correlation might then be "hidden" and it might then be more appropriate to work with changes in segregation. With the data at hand it has not been possible perform such an analysis, but the historic success of the Sweden Democrats in some municipalities could potentially be explained by
such changes.

There are few reasons to not think that the results presented here are final any way. The measurement of segregation can be redefined with more fine grained data. The effects of adjusting the set of names to proxy culture has been left completely aside. There are other topics related to segregation and maybe this thesis has at least opened a door to other possibilities.

Acknowledgements

I am grateful for valuable advice and comments from my supervisor Peter Fredriksson.
Bibliography


6 Appendix

6.1 Data Collection

Income Data

Data on yearly incomes divided on 20-29 and 30-64 year olds. The mean value of these medians have been used. In the end you get this result of the absolute deviation of incomes. 

\( m_{\text{zips}} \) is the number of zip code areas in the construction of the measure for cultural segregation, \( y_{\text{zips}} \) gives the same number for income. \( y_s \) is the score on economic segregation, \( n_n \) is the score on the segregation measure and \( s_d \) is the percentage of votes laid for the Sweden Democrats in the national election for each municipality.

<table>
<thead>
<tr>
<th>municipality</th>
<th>population</th>
<th>m_zip</th>
<th>y_zip</th>
<th>y_s</th>
<th>n_n</th>
<th>s_d</th>
</tr>
</thead>
<tbody>
<tr>
<td>ale</td>
<td>18721</td>
<td>40</td>
<td>41</td>
<td>0.05</td>
<td>8.1</td>
<td></td>
</tr>
<tr>
<td>alings˚as</td>
<td>34979</td>
<td>43</td>
<td>43</td>
<td>0.03</td>
<td>4.7</td>
<td></td>
</tr>
<tr>
<td>”almhult</td>
<td>15408</td>
<td>21</td>
<td>22</td>
<td>0.03</td>
<td>7.3</td>
<td></td>
</tr>
<tr>
<td>”alvdalen</td>
<td>6356</td>
<td>7</td>
<td>8</td>
<td>0.03</td>
<td>7.9</td>
<td></td>
</tr>
</tbody>
</table>

The Collection of Data on Individuals

The program for collecting data on individuals have been written in PHP\(^{29}\), which is an open-source scripting language with a syntax similar to languages such as C/C++. The program consists of two parts. A class parsePage that provides the methods for splitting up HTML pages based on tags such as \(<\text{div}>\) and a set of functions that identify the desired data. Secondly a procedure for collecting the data based on county and municipality filters.

To be able to run this code you have to install PHP. Normally PHP is executed by a webserver, but it is possible to use it completely standalone, usually by executing \texttt{php yoursript.php}. The three scripts of the program need to be placed in the same folder and the process is started by executing master.php. The data collecting takes several days and is highly likely to get interrupted. But the program knows which zip codes that have been processed and thus if the program gets stuck, it is just to restart it.

Filtering the data

During the gathering of the data the PHP program does some filtering by separating identity doubles, errors and valids. Errors in this case are records that have a zip code that is not associated with the correct city for that zip code. However, since it is impossible to know at start what kind of errors are present in the data and given the size, the data needs to be filter after collecting it aswell and also encoded. Due to the size of the data this process have been implemented in a C++ program. To run this requires to install a couple of libraries, namely the boost libraries for multithreading and multi array.\(^{30}\) The program also needs to be compiled on a platform similar to the one it is used which can be done with the GNU C++ compiler \texttt{g++}\(^{31}\). Compile and build it using the following commands (it cannot be guaranteed that it works regardless of OS/platform):

\(^{29}\)http://www.php.net
\(^{30}\)http://www.boost.org
\(^{31}\)http://gcc.gnu.org
g++ -O3 -Wall -c clean.c

In the second command (the build) clean might have to be changed to clean.exe if you are on a Windows system.

For the most computationally intensive action this program utilises multithreading, which means that if you have a processor with N cores, the program will work with N records simultaneously when filtering doubles. It is rather difficult to construct such programs and make them behave properly regarding the display of messages, thus the output can look a bit scattered at times. Still, on a system with 4 processor cores operating at 2.6 GHz this process takes about 5 minutes in computation time.

clean -(filter/compare/unique) [input file] ...
... [output file proper records] [output file improper records]
clean -match [input file] [match file] [results file]

For instance:
clean -filter mydata.csv filtered_data.csv bad_data.csv

File names may not contain spaces. The input file must obey the following (CSV) structure.

text,#digits,"text containing spaces, or commas",...
...

Encoding Individuals

Running clean with the option -match creates a csv file containing zip codes, municipality and 1 if the name in a record is matched with the contents in the match file, 0 otherwise. The match file contains lines with the words that are supposed to qualify as a match.

6.2 Data

Namesday Names

One way to see whether the data collected is a representative sample of the population as a whole is to examine if the frequency of the most popular names in Sweden are similar to those in the sample. A straightforward way of testing this is to run the following in a BASH-terminal:

cat file.csv | grep '^"Name ' | wc

This basically counts the number of lines in our file that starts with "Name '. Note that this includes a citation and a space. The argument for grep is contained within hard quotes since we want grep to parse this as a regular expression where " means that lines that start with Name are to be returned. The csv file returned from the program above encodes text containing spaces within quotes. Since many names, such as Lars, are also part of common family names, Larsson. As described above some entries in the database
are returned as *Family name Name* and by including a space it is ensured that we do not count both individuals with the name *Lars* and households with the family name *Larsson* as an example. Still this will only be an estimation since if *Lars* for any reason does not appear at the start of the line he will not count. There are ways to be more accurate, but this is sufficient given the purpose.

The result of this procedure is presented below with the top 5 female and male names given from Statistics Sweden (SCB, 2012, accessed 2013-09-15).\(^{32}\) From the sample *Mikael* also includes *Michael, Karl Carl, Karin Carin* and *Kristina Christina*.

![Table of female and male names](http://www.scb.se/Pages/TableAndChart____286717.aspx)

![Table of female and male names](http://www.scb.se/Pages/TableAndChart____31042.aspx)

<table>
<thead>
<tr>
<th>Name</th>
<th>Sample</th>
<th>SCB</th>
<th>(\frac{Sample}{SCB})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Anna</td>
<td>88 672</td>
<td>109 995</td>
<td>0.806</td>
</tr>
<tr>
<td>(1) Lars</td>
<td>100 366</td>
<td>97 063</td>
<td>1.034</td>
</tr>
<tr>
<td>(2) Eva</td>
<td>76 256</td>
<td>92 582</td>
<td>0.824</td>
</tr>
<tr>
<td>(2) Anders</td>
<td>86 771</td>
<td>81 066</td>
<td>1.070</td>
</tr>
<tr>
<td>(3) Maria</td>
<td>71 788</td>
<td>83 553</td>
<td>0.859</td>
</tr>
<tr>
<td>(3) Mikael</td>
<td>72 178</td>
<td>79 272</td>
<td>0.911</td>
</tr>
<tr>
<td>(4) Karin</td>
<td>62 683</td>
<td>77 247</td>
<td>0.811</td>
</tr>
<tr>
<td>(4) Johan</td>
<td>66 003</td>
<td>76 295</td>
<td>0.865</td>
</tr>
<tr>
<td>(5) Kristina</td>
<td>54 496</td>
<td>59 159</td>
<td>0.921</td>
</tr>
<tr>
<td>(5) Karl</td>
<td>47 870</td>
<td>76 176</td>
<td>0.628</td>
</tr>
</tbody>
</table>

As can be seen there is with the exception of *Anders* and *Lars* a lower representation of all names. Since the sample is not covering all individuals in Sweden that is rather natural. That some names are more frequent in the sample might be explained by the fact that the data from SCB only contains first names while some individuals in the sample are listed with all their names.

### 6.3 Derivation of \(g^*\)

\[
\max U_i = g^\alpha (1 - \hat{l}_i) + y - g \\
\frac{\partial U_i}{\partial g} = \alpha g^{\alpha - 1} (1 - \hat{l}_i) - 1 = 0 \\
\alpha g^{\alpha - 1} (1 - \hat{l}_i) = 1 \\
g^{\alpha - 1} = \left(\alpha (1 - \hat{l}_i)\right)^{-1} \\
g = \left(\alpha (1 - \hat{l}_i)\right)^{- \frac{1}{\alpha - 1}} \\
g = \left(\alpha (1 - \hat{l}_i)\right)^{\frac{1}{\alpha - 1}}
\]

\(^{32}\)Male names: [http://www.scb.se/Pages/TableAndChart____286717.aspx](http://www.scb.se/Pages/TableAndChart____286717.aspx)

Female names: [http://www.scb.se/Pages/TableAndChart____31042.aspx](http://www.scb.se/Pages/TableAndChart____31042.aspx)