Should I Stay or Should I Go?

A cross-sectional study of refugee’s internal migration pattern in Sweden
Abstract

Background:
Sweden has in the past four years experienced a large influx of refugees. Some of these refugees were arranged housing upon arrival by the Swedish state. Approximately 30 percent of the refugees that were arranged housing have within five years relocated from the initial municipality of residence.

Purpose:
The purpose of this study is to analyze the influence refugee’s networks has on their internal migration in Sweden. Information regarding refugees residential preferences, and if location-specific social capital affect these preferences is important for policymakers to facilitate integration.

Method:
The initial and subsequent residential locations among refugees is studied to analyze how important other immigrants overall, and other immigrants from the individual’s country of birth in a municipality affect if a refugee stays in the initial municipality of residence or not.

Conclusion:
The results suggest that refugees stay in municipalities with high a representation of inhabitants from the individual’s birth country but move away from municipalities where there is a large share of immigrants overall.
Table of contents

1. Introduction.................................................................................................................1

2. Literature Review........................................................................................................4
   2.1 Theoretical Framework ..........................................................................................4
      2.1.1 Micro-Economic Migration Theories ...............................................................4
      2.1.2 Networks and Social Capital ...........................................................................5
      2.1.3 Realistic Group Conflict Theory ......................................................................6
   2.2 Previous Research ..................................................................................................7
      2.2.1 Share of Immigrants .......................................
      2.2.2 Location-Specific Control Variables ...............................................................7
      2.2.3 Individual-Specific Control Variables .............................................................8

3. Methodology ..............................................................................................................10
   3.1 The Empirical Model .............................................................................................10
   3.2 Data .......................................................................................................................10
      3.2.1 Dependent Variable .......................................................................................11
      3.2.2 Variables of Interest ......................................................................................11
      3.2.3 Individual Independent Control Variables ...................................................12
      3.2.4 Regional Independent Control Variables ......................................................12
   3.3 Method ..................................................................................................................13

4. Empirical Analysis ....................................................................................................15
   4.1 Descriptive Statistics ............................................................................................15
   4.2 Correlation Analysis .............................................................................................16
   4.3 Regression Model ................................................................................................17
      4.3.1 Regression Results .........................................................................................18
      4.3.2 Analysis ........................................................................................................19

5. Conclusion ................................................................................................................22

References ......................................................................................................................24

Appendix .......................................................................................................................27
Figures

Figure 1 Number of Asylums Granted ................................................................. 2

Tables

Table 1 Description of Individual and Regional Data ............................................. 15
Table 3 Different Models ...................................................................................... 16
Table 4 Regression Results .................................................................................. 17

Appendix

Figures

Table 2 Correlation Matrix .................................................................................. 27

Tables

Figure 2 Population Size ...................................................................................... 27
Figure 3 Natural Logarithm of Population Size ................................................... 28
Figure 4 Average Income ..................................................................................... 28
Figure 5 Unemployment Rate ............................................................................. 29
Figure 6 Share of Foreign-Born Population ......................................................... 29
1. Introduction

This paper examines how the share of immigrants from the same country of origin, together with the total share of immigrants in a municipality affect the decision of staying in the initial municipality of residence faced by refugees. Refugees that were arranged housing by the Swedish state in 2013 and where they live in 2018 are analyzed to study the effect that networks have on refugees’ internal migration patterns. Refugees from Somalia, Syria, and Afghanistan are examined in this paper. Refugees from these three countries represent approximately two-thirds of all refugees that were arranged housing in 2013.

We will in this paper investigate if migrant’s networks have an effect on the decision of whether to stay or to move faced by a refugee placed in a municipality. This will be done by estimating a logit model where the share of the municipality’s population from the same country of origin is included as an explanatory variable. A variable for the share of immigrants in the municipality together with a set of control variables according to previous research are also included as explanatory variables.

There are three crucial factors that determine people’s motivation and ability to migrate; financial, human and social capital (Massey et al. 1998). Social capital is defined as the aggregate of the actual or potential resources which are linked to the possession of a durable network (Bourdieu, 1985). A migrant network is a location-specific form of social capital (Massey et al. 1998). This paper focuses on the social capital of migrants from a population perspective and therefore include two specific variables. These variables that has been included in previous research to capture the importance of immigrant networks are; the share of the local population that is from the same country of origin and the share of a local population that is foreign-born (Dodson, 2001; Moore & Rosenberg, 1995; Rephann & Vencatasawmy, 1999; Zavodny, 1997; Åslund, 2005). The local population in our case is a municipality’s population.
Looking at the most recent data is relevant as Sweden’s immigration situation has been relatively extreme in the last couple of years. Asylums granted have had a drastic increase. In the first few years after the turn of the century, ten thousand or fewer asylums were granted. In the last four years however, at least thirty-five thousand asylums have been granted annually with 2016 being the peak year with more than seventy thousand asylums granted. Another recent drastic change is Sweden’s asylum laws. In 2016, Sweden went from having EU’s most generous asylum laws to the minimum EU level (Migrationsverket, 2018).

This paper will contribute to policy making by shedding light on how important networks are when predicting moving patterns for refugees. It will also give information on which municipalities that can expect their refugees to stay and which municipalities that can expect their refugees to move. A better understanding of refugee’s internal migration patterns can help to forecast the need for public services at different locations (Åslund, 2005). It can also help policymakers to alter the settlement pattern of immigrants and can potentially facilitate integration, decrease the costs of reception and better utilize available housing and jobs (Rephann & Vencatasawmy, 1999).

The main results from this paper show that refugees tend to stay in municipalities where the share of the population that is from the same country of origin as themselves is larger. It is also shown that the larger the share of immigrants overall in a municipality, the lower is the chance of a refugee staying in that municipality. The share of a municipality’s population that
is from the same country of origin is the statistically strongest variable affecting the decision of staying in the initial municipality of residence.
2. Literature Review

The purpose of this section is to describe the main ideas regarding internal migration and to prepare the reader with both the findings of previous studies on this subject and the relevant theories behind it. The most relevant migration theories are presented with a certain focus on theories associated with determinants of internal migration followed by two sections, one describing the network theory and one on group conflict theory. Finally, a summary of previous research, which will motivate the choice of variables in the model, concludes this literature review.

2.1 Theoretical Framework

Due to increasing migration in general and increasing diversity in types of migration, the number of theories on migration is increasing more than ever. Most theories on migration are examining international migration but they are oftentimes applicable to internal migration as well (King, 2012). This section will begin with an assessment of micro-economic theories connected to internal migration, which are focused on individual migration decisions. This will follow with an assessment of the network theory of migration which has its roots in sociology. Finally, the theory of realistic group conflict is presented.

2.1.1 Micro-Economic Migration Theories

Individual decision-making is the center of micro-economic migration theory, indicating that individual desires and values are more emphasized upon compared to the economic structure of a society (Massey et al, 1993). The main micro-economic theories on migration are the push- and pull-theory (Lee, 1966) and the human capital approach (Sjaastad, 1962).

Lee’s (1966) push- and pull-theory of migration is one of the most commonly referred to models of migration. This theory was the first to suggest that there are other factors than labor market-related reasons to migration. This theory acknowledges both the supply- and demand-side of migration and conceptualizes that there are both negative and positive effects connected to the origin and the destination region of migration. The push-factor serve as the reason for people to leave an area whereas the pull-factor of migration represent the motives of moving to an area (Lee, 1966). The argument behind the theory is that if there are more reasons for leaving an area than there are reasons to stay, that region will experience a net out-migration. Prevailing characteristics that represent push factors are low productivity,
unemployment, and underdevelopment of a region. Typical pull factors are opportunities for better employment, higher income and a greater range of services (Lee, 1966).

Another theory that is associated with internal migration is the human capital approach pioneered by (Sjaastad, 1962). In Sjaastad’s (1962) theory, migration is regarded as an individual investment decision to increase human capital. This theory relies on individuals making rational decisions based on a cost-benefit analysis, migrating only when the expected benefits exceed the expected costs (Hagen-Zanker, 2008). The model further predicts that the age of a migrant is a significant aspect, where young individuals are expected to have a higher probability of migrating. This is because the expected returns to migration are higher for an individual that is young.

Most theories concern the economic motivations behind migration such as employment and wages but internal migration does not only depend on socio-economic factors. Migration decisions also depend on demographic circumstances such as income, unemployment rates, population growth and the demand for higher schooling (Hagen-Zanker, 2008).

2.1.2 Networks and Social Capital
A migrant’s network can be seen as a location-specific form of social capital (cf. Massey et al. 1998). Social capital is defined by Bourdieu (Bourdieu, 1985) as the aggregate of the actual or potential resources which are linked to the possession of a network. Social capital is defined as capital because it is a resource that can be converted into other forms of cultural, human and economic capital (Bourdieu, 1985; Coleman, 1988; Portes, 1998).

The volume of the social capital a person gains from being in a group depends on the size of the network connections and the volume of the economic, cultural and symbolic capital possessed by the other members in that network (Bourdieu 1985). Migrant networks tend to decrease the economic, social and psychological costs of migration (Massey et al. 1998). Böcker (1994) argues that migrants function as "bridgeheads", reducing the risk and cost of other migrants of that group. Hence, an established migrant community at one destination will increase the likelihood of subsequent migration to that place. Therefore, besides financial and human capital, social capital needs to be recognized as a third crucial factor determining people’s motivation and ability to migrate (Massey 1990).
There is no uniform use of the concept of “social networks” in migration studies but a feature of social networks is their homophily (McPherson, Smith-Lovin, & Cook, 2001). Homophily can be explained as the tendency that people have, who share similar personal, social, cultural, political or economic characteristics, to associate with each other (McPherson et al., 2001, p. 415). Shared characteristics and experiences are expected to trigger feelings of comfort, belonging and trust which are prerequisites for social networking and interaction between humans (Avenarius, 2009:23; Ryan et al., 2008).

2.1.3 Realistic Group Conflict Theory

Prejudice and discrimination between groups of a society are oftentimes a product of a conflict of interest according to the realistic group conflict theory (LeVine & Campbell, 1972). Campbell (1965), recognized what seems to be an important factor of inter-group conflicts in determining attitudes and behavior in between them. Attitudes and behavior between groups reflect the interest of the groups and are based on the common goals they share. A positive relationship between groups is expected when goals are compatible, whereas negative intergroup attitudes, often expressed as a conflict, are expected to arise when group goals are incompatible. Campbell’s perspective of group interactions was titled the realistic group conflict to demonstrate that some group conflicts are realistically fought based on competition for scarce resources (Esses, Jackson and Armstrong, 1998).

An important assumption of this theory is that an actual competition over resources not have to exist. The mere perception of an existing rivalry of resources is sufficient for intergroup conflict and hostility. The realistic group conflict theory suggests that as perceived competition over resources increases between groups, so does the conflict between them (Esses, Jackson and Armstrong, 1998). Furthermore, it has been suggested that the more a group can benefit from succeeding in a conflict, the fiercer the conflict becomes and more opposition is expressed toward the conflicting group. The intergroup opposition justifies the conflict and the negative behavior towards members of the conflicting group (LeVine & Campbell, 1972). Finally, LeVine & Campbell (1972) suggest that when there exists competition over resources, geographical closeness increase the aggression between groups rather than reduce it.
2.2 Previous Research

In this section, we will look at previous research and what variables they have found to have an effect on relocation decisions of immigrants. Two papers from the U.S (Dodson, 2001; Zavodny, 1997) one paper from Canada (Moore & Rosenberg, 1995) and two papers from Sweden (Rephann & Vencatasawmy, 1999; Åslund, 2005) are examined. These papers have had different definitions of ethnic groups, have studied different countries and have emphasized on different immigrant groups. What they all have in common is that they previously have been trying to estimate immigrant’s internal moving patterns.

2.2.1 Share of Immigrants

The two variables we want to investigate in this paper is the share of the municipal population from the same country of origin and the share of foreign-born population in a municipality. These variables have been found significant predictors in previous research (Dodson, 2001; Moore & Rosenberg, 1995; Rephann & Vencatasawmy, 1999; Zavodny, 1997; Åslund 2005). There is a higher likelihood of staying in a location or moving to a location where there is a bigger immigrant population overall. Åslund (2005) argues that living close to people of the same ethnic and linguistic background may be valuable as such but can also be a way of establishing oneself in the new country. A high share of immigrants in a region could reflect that the region has experience in dealing with issues related to new immigrants, and/or an acceptance of foreign-born people which may lead to less risk of discrimination and alienation.

2.2.2 Location-Specific Control Variables

Previous research suggests that there are a set of other regional characteristics that affect location preferences among immigrants. These are the unemployment rate, the average income and the population in a region.

According to Rephann & Vencatasawmy (1999), economic circumstances have a strong influence on location decisions among immigrants. Bad labor market conditions have one of the strongest effects on the decision whether to stay or to relocate faced by asylum holders participating in a dispersal program of settlement (Åslund, 2005). Previous research has come up with diverse results regarding the impact of local unemployment. Åslund (2005), finds that it increases the likelihood of immigrants leaving a particular location, other research finds the unemployment rate not to be significant at all (Moore & Rosenberg, 1995; Zavodny, 1997).
High average earnings in general are expected to have a negative effect on the likelihood of moving. Moore & Rosenberg (1995) and Åslund (2005) does, however, find average wages to have a positive impact on the decision of relocating by immigrants. The explanation for this is that people generally prefer living among others with similar economic status. As refugees in general are found in the lower part of the income distribution this may be an explanation for the observed pattern (Åslund, 2005). Zavodny (1997) argues that economic circumstances in general are irrelevant in immigrant’s settlement decisions.

One of the characteristics that most previous researchers find to affect location preferences, both among immigrants and people in general, is the population size. People in general tend to gravitate towards the bigger cities, and immigrants tend to do so even more than native Swedes (Åslund 2005). The population size is a significant component influencing settlement patterns in previous studies (Moore & Rosenberg, 1995; Åslund, 2005). Åslund (2005) also find population size to be one of the strongest factors influencing the decision to stay or move faced by asylum holders. There are several opportunities associated with larger populations that contribute to the urbanization rate. Bigger cities often have more to offer culturally and provide improved consumption opportunities (Clark & Lloyd, 2011). Larger cities generally provide increasing opportunities for employment. A larger number of employers and workers lead to more available jobs, generate opportunities for specialization and are therefore more likely to create effective labor market pools (Glaeser, 1998).

2.2.3 Individual-Specific Control Variables

A number of personal characteristics have been found to influence the decision of resettlement among immigrants. These characteristics are; the age, gender, and country of origin of an individual (Rephann & Vencatasawmy, 1999; Åslund, 2005).

The likelihood to resettle is according to Moore & Rosenberg (1995) and Åslund (2005) quite different between immigrant groups. Moore & Rosenberg (1995) and Åslund (2005) finds that the share of local immigrants from the same ethnic group affects the decision making more for Asians compared to immigrants from other continents. Åslund (2005) also find that people from the Middle East appear to be more responsive to unemployment than other groups. Rephann & Vencatasawmy find that immigrants from Asian and African countries show a greater propensity to resettle compared to immigrants from other continents. Overall most immigrant groups appear to be more mobile than natives (Rephann & Vencatasawmy, 1999).
During young adulthood, the probability to migrate is the greatest. During this period people normally make the decision to enter the labor force or to go to university. Since young people generally have the most to gain from moving, migration is expected to decline with age. (Rephann & Vencatasawmy 1999).

Åslund (2005) finds that relocating from the initial municipality of settlement is significantly less common among women compared to men among program participants, whereas Rephann & Vencatasawmy (1999) in their study of all immigrants in Sweden finds resettlement decisions to be unrelated to gender.
3. Methodology

In this section the empirical model, the econometric method and the data utilized are presented. Furthermore, each variable added to the model is discussed along with their expected effect on the dependent variable. Finally, the method of this study motivating the choice of econometric model is presented.

3.1 The Empirical Model

The decision to stay in or move from the initial municipality of residence for a refugee is affected both by individual and regional attributes according to previous research. This thesis focuses on how the share of immigrants and the share from the same country of origin in a municipality affects residential preferences for the three largest refugee groups that were arranged housing in Sweden in 2013. The regression includes the “share of municipal population from the same country of origin” and the “share of foreign-born population in a municipality” as explanatory variables.

Regression formula:
\[ Y_i = \beta_0 + \beta_1 F_j + \beta_2 N_j + \beta_3 R_j + \beta_4 I_i + \epsilon_{ij} \]

Where:
- \( Y_i \): The outcome of whether an individual (i) has stayed or not.
- \( F_j \): Share of foreign-born population in a municipality (j).
- \( N_j \): Share of the municipal population from the same country of origin (if).
- \( R_j \): Vector of regional variables measured for municipality (j).
- \( I_i \): Vector of individual variables for individual (i).
- \( \epsilon_{ij} \): The error term for individual (i) residing in the municipality (j).

3.2 Data

The empirical analysis in this paper utilizes a dataset from the Swedish Migration Agency (Migrationsverket) identifying all individuals who were arranged housing by the state in 2013, and where they live in 2018. From the dataset, we also have information about the variables sex, age, and country of origin. The original data consisted of 20,332 individuals but only those of working age (18-64) in the year of arrival from the countries Somalia, Syria, and Afghanistan were selected. As a result, the final sample consists of 4401 individuals.
The regional characteristics have been collected from the Swedish Agency for Economic and Regional Growth’s database RAPS and from Statistics Sweden (SCB). The database contains regional statistics for all Swedish municipalities. Data for the year 2013 is used for the regional characteristics; unemployment rate, average earnings, population size and share of foreign-born population. The reason for using data from 2013 is because the immigrants studied in our sample was initially arranged housing in that year.

The individual data is merged with the regional characteristics of the municipalities they were placed in. The regional variables enter the model as continuous variables whereas the individual variables enter as dummy variables.

Since the individual observations within each municipality most likely are correlated with each other, the standard errors in the model are clustered by the initial municipality of residence. The individual observations are merged with regional variables aggregated by geographical location making the assumption of correlation a reasonable assumption. By using the clustering approach, the correlation among same municipality observations is allowed for. This approach is recognized as a necessary measure to avoid incorrect standard errors.

3.2.1 Dependent Variable
A binary variable that takes on the value of 1 if an individual has chosen to stay in its initial municipality of reception after 5 years and 0 if the individual has chosen to move, is used as the dependent variable in this study.

3.2.2 Variables of Interest
Share of foreign-born population
Higher shares of foreign-born in a municipality is expected to be positively related to the dependent variable i.e. increase the likelihood of an immigrant staying at the initial municipality of residence.

Share of population from the same country of origin
The share of the municipal population that is from the same country of origin is expected to be positively related to the decision of staying.
We expect both variables to have a positive relationship with the decision to stay in the municipality. We expect the share of a municipality’s population that share the same country of origin to have a higher impact as the homophily is expected to be stronger.

3.2.3 Individual Independent Control Variables

Country of origin
Dummy variables for two of the nationalities are included in the model indicating everyone’s country of origin. The dummies represent the nationalities of Somalia and Afghanistan where the group Syria act as the group of reference.

Gender
The gender of each individual is included in the model through a dummy variable that takes on the value of 1 if the individual is female and 0 if the individual is male. According to previous research, females are more likely to stay in the initial municipality of residence compared to men. Therefore, this variable is expected to be positively related to the dependent variable.

Age
Two dummy variables indicating age are included in the model in accordance with previous studies claims of decreasing mobility with age. The group of 18-24 act as a reference group for the two other age groups 25-39 and 40-64. The first dummy takes on the value of 1 if an individual is between 25-39 years of age and the second if the individual is 40-64 years old.

As we expect mobility to decrease with age, individuals of age 40-64 are expected to be more positively related to the dependent variable than the other groups. Furthermore, mobility in the age group 25-39 is expected to be lesser than for individuals of age 18-24.

3.2.4 Regional Independent Control Variables

Unemployment rate
The unemployment rate among the working age population (16-64) for each municipality is included in the model. The unemployment rate measures the number of unemployed individuals as a percentage of the labor force and is therefore presented in percentage.
We expect the unemployment rate to be negatively related to the decision of staying in accordance with previous researcher’s findings.

**Average income**
Average income is measured for all employed individuals divided by the number of people of working age. Furthermore, the variable is denoted in thousands of SEK. Previous research finds average earnings both to be a significant and an insignificant predictor. If significant, average income is expected to be negatively related to the dependent variable.

**Population size**
The population size of each municipality is included in the model as an explanatory variable and is expected to be positively related to an individual’s locational preferences. We thus expect the population size to be positively related to the dependent variable.

The distribution of the population size is positively skewed. We will therefore use the natural logarithm of the population size of each municipality.

### 3.3 Method

The goal of an analysis using a logistic regression model is equivalent to that of other model-building techniques used in statistics; to find the most parsimonious one and the best fit. When the dependent variable in an outcome study is dichotomous (i.e. has stayed vs has not stayed), a logistic regression, as opposed to other estimation methods, is particularly appropriate (Hosmer & Lemeshow, 1989). The use of an OLS method would assume the relationship between the dependent and independent variables to be linear, which in the present study does not hold true. The assumptions of normality of errors and homoscedasticity in the OLS method would hence be violated.

In the present study, a logistic regression is used to estimate the probability of a binary response based on independent variables. This method is frequently used in similar studies (Rephann & Vencatasawmy, 1999; Åslund 2005). The predicted probability of an event occurring is presented in the form of logged odds in the logistic model, making the predicted outcomes continuous, linear in its parameters and linear in the independent variables. Furthermore, the logistic regression model guarantees that the predicted probability of an
event occurring lies between zero and one, permitting a non-linear relationship between the independent variables and the predicted probabilities (Gujarati, 2004).
4. Empirical Analysis

4.1 Descriptive Statistics

In Table 1, descriptive statistics for all individual and regional data are presented. There are 4401 observations for the individual variables and 290 observations for the regional variables. Approximately 70 percent of the individuals have stayed in the initial municipality of residence within the studied 5-year interval.

Of the included individuals, approximately 50 percent are between 25-39 years old and 31 percent are in the age of 40-64. Relatively few individuals, 19 percent, are 18-24 years old in the sample. Additionally, over 59 percent of the sample consists of women whereas 41 percent are men. The two largest nationalities in the sample are Somalis and Syrians, representing approximately 43 percent and 36 percent of the sample respectively. 21 percent of the sample individuals are Afghans. The discrepancy in terms of foreign-born population in a municipality is also great. With the highest frequency of 40 percent compared to the low of 5 percent, the mean municipal foreign-born population of 14 percent indicate that a higher number of municipalities in are the lower part of the distribution.

The regional variables Average Income and Unemployment both show a great discrepancy between the lowest and the highest value in the data. The highest municipal unemployment rate of approximately 16 percent is far from the lowest of 3 percent. The mean unemployment rate of 10 percent is approximately right in between the max and minimum indicating that the unemployment rate is normally distributed. The Average income however, ranging between a low of 193,7 and a high of 598,6, indicates a far greater discrepancy between municipalities than the unemployment rate. With an average income of approximately 239,7 thousand SEK annually, it is evident that the average income in most municipalities is lower than the median. The plotted distributions for each regional variable are presented in Appendix 2-6.
Table 1 - Description of Individual and Regional Data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have Stayed</td>
<td>0.70</td>
<td>0.45</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Same Nationality</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td>Share of Immigrants</td>
<td>0.14</td>
<td>0.06</td>
<td>0.05</td>
<td>0.40</td>
</tr>
<tr>
<td>Ln Population</td>
<td>10.62</td>
<td>1.25</td>
<td>7.80</td>
<td>13.71</td>
</tr>
<tr>
<td>Average Income</td>
<td>239.75</td>
<td>27.03</td>
<td>193.70</td>
<td>598.60</td>
</tr>
<tr>
<td>Unemployment</td>
<td>0.10</td>
<td>0.02</td>
<td>0.03</td>
<td>0.15</td>
</tr>
</tbody>
</table>

4.2 Correlation Analysis

The Spearman rank correlation coefficients between the binary dependent variable and the independent variables are shown in Table 2 in the appendix together with the correlations of the independent variables. The regional variables are all significantly correlated to the dependent variable.

The correlation of the Unemployment rate shows a conflicting sign based on expectations from theory. The unemployment rate is negatively related to staying in a municipality, which was not expected. The correlation of 0.188 between Ln Population (Natural logarithm of population size) and the variable Stayed is the highest bivariate correlation of all variables in relation to the dependent variable. The second highest Stayed correlation is the country dummy for Somalis; Somalia with a recorded correlation of 0.1931.

Several of the independent variables are as expected also highly correlated. The high correlation between the population size and the share of immigrants of 0.5118 are according to theory awaited since immigrants to greater extent tend to live in more populated areas. There is also not surprisingly a high correlation between the population size and the average
income. The productivity level of a city depends on the population size and larger cities are generally more productive (OECD, 2015).

### 4.3 Regression Model

The descriptive statistics and the correlation analysis gives an initial hint of the relation between the independent variables and the internal migration patterns among immigrants. To further deepen this study a cross-sectional regression model is estimated to find the magnitudes, the direction of the relationships and the significance levels of the variables. To evaluate the relatively high correlation between the variables and to find out which combination of variables that are better at predicting the dependent variable, the full model (Model 1) is presented along with 5 restricted models (Model 2 - Model 6) in Table 3. One regional variable at the time is removed from the regression creating 5 different combinations. The pseudo R\textsuperscript{2} for each model is reported which is a useful measure when comparing different models from the same dataset. The higher the Pseudo R\textsuperscript{2}, the better the model is in predicting the dependent variable. Furthermore, the individual variables are included in all of the models.

<table>
<thead>
<tr>
<th>Variables</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>
</tr>
</thead>
<tbody>
<tr>
<td>Share same nationality</td>
<td>24.117**</td>
<td>24.535**</td>
<td>27.164**</td>
<td>17.762**</td>
<td>-</td>
<td>16.700**</td>
</tr>
<tr>
<td></td>
<td>(7.045)</td>
<td>(7.128)</td>
<td>(7.470)</td>
<td>(8.935)</td>
<td>-</td>
<td>(7.179)</td>
</tr>
<tr>
<td>Share immigrants</td>
<td>-3.517**</td>
<td>-3.814**</td>
<td>-2.557*</td>
<td>-0.637</td>
<td>-2.493*</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(1.438)</td>
<td>(1.483)</td>
<td>(1.467)</td>
<td>(1.521)</td>
<td>(1.477)</td>
<td>-</td>
</tr>
<tr>
<td>Ln(population)</td>
<td>0.404**</td>
<td>0.346**</td>
<td>0.444**</td>
<td>-</td>
<td>0.3886**</td>
<td>0.341**</td>
</tr>
<tr>
<td></td>
<td>(0.979)</td>
<td>(0.111)</td>
<td>(0.099)</td>
<td>-</td>
<td>(0.994)</td>
<td>(0.998)</td>
</tr>
<tr>
<td>Average income</td>
<td>-0.004</td>
<td>-</td>
<td>-0.089</td>
<td>0.006</td>
<td>-0.04</td>
<td>-0.006**</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>-</td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.006)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>7.589**</td>
<td>10.101**</td>
<td>-</td>
<td>11.767**</td>
<td>8.992**</td>
<td>4.149**</td>
</tr>
<tr>
<td></td>
<td>(3.426)</td>
<td>(3.204)</td>
<td>-</td>
<td>(3.294)</td>
<td>(3.556)</td>
<td>(3.492)</td>
</tr>
<tr>
<td>Numbers of Observations</td>
<td>4401</td>
<td>4401</td>
<td>4401</td>
<td>4401</td>
<td>4401</td>
<td>4401</td>
</tr>
<tr>
<td>Numbers of Clusters</td>
<td>272</td>
<td>272</td>
<td>272</td>
<td>272</td>
<td>272</td>
<td>272</td>
</tr>
<tr>
<td>Pseudo R2</td>
<td>0.0690</td>
<td>0.0682</td>
<td>0.064</td>
<td>0.0485</td>
<td>0.0651</td>
<td>0.0647</td>
</tr>
</tbody>
</table>

* = p < 0.05, ** = p < 0.1
Clustered robust standard errors in parenthesis, clustered by receiving municipalities
Model 1 includes all variables and generates significant results for all variables except Average Income at the 10 percent level. It also provides us with the highest Pseudo $R^2$ of 0.069. In Model 2, where the insignificant variable Average Income is removed, none of the other variables changes substantially in magnitude and all other variables keep the same sign. Model 3 – 6 all provide us with a lower Pseudo $R^2$ compared to Model 1, have several variables that are insignificant at the 5 percent or 10 percent level or does not include our variable of interest. As earlier researchers have found Average Income to be significant and nothing drastic changes when it is removed from the model, we decide to use Model 1 in the proceeding analysis of this thesis.

### 4.3.1 Regression Results

Table 3 – Regression results estimating the propensity to move from the initial municipality of residence

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficients</th>
<th>Clustered Robust S.E.</th>
<th>Marginal Effects</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Share same nationality</td>
<td>24.117</td>
<td>7.046</td>
<td>4.919</td>
<td>0.001</td>
</tr>
<tr>
<td>Share immigrants</td>
<td>-3.517</td>
<td>1.438</td>
<td>-0.717</td>
<td>0.014</td>
</tr>
<tr>
<td>Ln(population)</td>
<td>0.404</td>
<td>0.098</td>
<td>0.082</td>
<td>0.000</td>
</tr>
<tr>
<td>Average income</td>
<td>-0.004</td>
<td>0.006</td>
<td>-0.001</td>
<td>0.496</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>7.589</td>
<td>3.426</td>
<td>1.548</td>
<td>0.027</td>
</tr>
<tr>
<td>Age 25-39</td>
<td>0.346</td>
<td>0.098</td>
<td>0.071</td>
<td>0.000</td>
</tr>
<tr>
<td>Age 40-64</td>
<td>0.657</td>
<td>0.106</td>
<td>0.134</td>
<td>0.000</td>
</tr>
<tr>
<td>Somalia</td>
<td>0.633</td>
<td>0.069</td>
<td>0.129</td>
<td>0.000</td>
</tr>
<tr>
<td>Afghanistan</td>
<td>0.306</td>
<td>0.158</td>
<td>0.062</td>
<td>0.053</td>
</tr>
<tr>
<td>Female</td>
<td>0.284</td>
<td>0.136</td>
<td>0.058</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 4 shows logit estimates for refugees that were arranged housing by the Swedish state in 2013 and their probability of staying in the initial municipality of residence after 5 years. Based on an F-test the overall model is statistically significant at the 1% level. The signs of each coefficient, as argued by Greene (2003), inform us of how the variables influence the direction of the change in the probability, (negatively if the coefficient is negative and positively if the coefficient is positive). To be able to draw conclusions of economic relevance, the marginal effects and the level of significance are also reported.

When we are looking at our variables of interest, the Share same nationality and the Share of immigrants, we find that an increasing share of immigrants in a municipality decreases the
probability of refugees staying in a municipality whereas the share of people of the same nationality increases it.

When it comes to differences between the refugee groups, we find that people from Syria are the most likely to move from the initial municipality, whereas refugees from Afghanistan and Somalia appear to be more likely to stay comparatively. The parameters suggest that Afghans have a 30 percent, and Somalis a 63 percent higher probability of staying in the initial municipality of residence compared to Syrians.

Both Åslund (2005) and our result suggest females to be more likely to stay compared to men. Females in our population are about 28 percent more likely to stay in the initial municipality of residence. We also find that the older a refugee is, the higher is the probability of staying. This was expected and in line with previous research. Since young people generally have the most to gain from moving, migration is expected to decline with age (Rephann & Vencatasawmy 1999).

To control for labor market effects in the decision to relocate or not, each municipality’s Unemployment rate and Average income are included in the model. The results suggest that high unemployment is associated with an increased probability of staying in a municipality whereas average income appears to have an almost non-existing negative effect. For a marginal increase in municipal unemployment, the parameter implies the increase of staying to be 5.48 percent all else equal.

Ultimately, the log of municipal population size is included. This variable indicates a positive relationship to staying in the initial municipality of residence. As depicted by Histogram 1 in the appendix, the population size of Swedish municipalities is non-normally distributed. The natural logarithm of this variable, as portrayed by Histogram 2 in the appendix, is therefore used instead.

4.3.2 Analysis
The share of people of the same nationality represents the fraction of a municipal population made up of people from an individual’s own nationality. As expected when looking at previous research, the larger this fraction is the lower the probability is to relocate (Dodson, 2001, Moore & Rosenberg, 1995, Rephann & Vencatasawmy, 1999; Zavodny, 1997; Åslund 2005). The rationale in this is that a set of common characteristics and a shared background are
assumed to generate feelings of trust and acceptance which are essential in human interaction (Avenarius, 2009, Ryan et al., 2008). According to Massey (1990), social capital is one of three fundamental factors motivating migration. This seems to hold true when looking at our results. The single most important factor affecting the decision of relocating or not is the share of people from a refugee’s own nationality in a municipality.

Our result suggests that the overall share of immigrants in a municipality have a negative effect on whether refugees decide to stay or not. Åslund (2005) suggests that regions which have a higher share of immigrants might be less discriminatory to immigrants. Municipalities with a greater share would therefore be a better place to live for an immigrant initially. We believe that the opposite of what Åslund (2005) believe could also be argued. The counties that have the highest hate crime rate with a racial motive per capita; Stockholm, Södermanland, and Västmanland (Brå, 2010) also have a higher share of immigrants than the average. Also, out of the five counties with the largest share of immigrants, only Stockholm show a lower than average support for Sweden’s anti-immigration party, the Sweden Democrats (Valmyndigheten, 2014).

The negative effect of the share of immigrants in a municipality, but positive effect of people from one’s own nationality might suggest that there is some kind of competition among immigrant groups. They compete for the same jobs since they to a large extent share the same skills and experiences i.e. they belong to the same socio-economic group. As there is competition for the same resources, the recorded pattern of immigrants moving away from municipalities where the foreign-born population is large, is rational. Our notion of existing homophily among immigrants based on certain characteristics defined by McPherson, Smith-Lovin, & Cook (2001) appears to be false, or just to an extent where the negative aspects of competition are bigger than that of the homophily. Instead, the immigrant groups appear to view each other as competitors, or at least have a perception of an existing rivalry. LeVine & Campbell (1972) suggested that geographical closeness would increase competition, and therefore moving from municipalities where there is a large share of immigrants decrease competition on the individual level.

Another reason for the negative relationship between the share of immigrants in a municipality and a refugee’s willingness to stay might have something to do with why they were placed in the municipality in the first place. Because it is easier to find available housing or because
authorities try to battle the net out-migration of municipalities, refugees are often arranged housing in areas already experiencing a declining population. This creates a possible bias in the results as these sparsely populated municipalities experienced a relatively large influx of refugees relative to their population size in 2013. When some of these refugees by 2018 had left these municipalities, the magnitude of this variable might by consequence be exaggerated.

Examples of municipalities with a positive net migration where the share of immigrants is high are Haparanda, Eda, and Övertärne (Statistics Sweden, 2018).
5. Conclusion

The share of people originating from the same country in a municipality appears to play a major role in the decision of whether to stay in the initial municipality of residence faced by refugees from Somalia, Syria and Afghanistan.

This paper examines the effect that social capital generated from an existing network has on refugee’s internal migration patterns in Sweden. A logit regression model is estimated with a binary dependent variable taking on value 1 if an individual has stayed in the initial municipality of residence and 0 if not. This study is based on an individual dataset retrieved from the Swedish Migration Agency with information of where refugees were arranged housing in 2013 and where they reside in 2018. To control for discrepancies between municipalities affecting the decision of staying, a set of regional control variables obtained from Statistics Sweden (SCB) and the Swedish Migration Agency are added to the model. Only the three largest refugee groups in terms of refugees received; Somalis, Syrians and Afghans are used in this study. These three nationalities represent approximately two-thirds of the total amount of refugees that were arranged housing in 2013.

The results generated from this study show that the most important variable affecting the decision to stay in the initial municipality of residence is the share of inhabitants originating from the same country as the refugee in a municipality. The presence of others of the same nationality as oneself in a municipality affect the decision to stay several times as much as the other variables used in this study. The results also show several other variables to be affecting the decision of whether to stay or not. The directional relationships of the independent variables to the dependent variable, except for the overall Share of immigrants in a municipality, are all in line with previous research on this topic. Earlier studies find the Share of immigrants to be a pull-factor for immigrants, while we find the opposite.

We suggest that future researchers examine where the individuals who did not stay decided to move. Our result suggests that refugees tend to stay in municipalities in which the population to a greater extent consist of individuals with the same nationality. Therefore, we suspect there also to be a positive relationship between the municipalities they move to and the share of people from the same country of origin in that municipality. It would also be interesting for future researchers to investigate if the people who arranged housing for themselves also
choose to live in municipalities where the share of people from the same country origin as themselves is large.

Policymakers should take into consideration the role that the share of a local population originating from the same country of origin affect refugee’s internal migration patterns. The results from this paper can be used to predict which municipalities that can expect refugees to stay and where these might come from. We would additionally like to emphasize the importance of not treating refugees as a homogenous group and expect identical migration patterns for each group.
References


Appendix

Table 4 - Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>Stayed</th>
<th>( \ln ) population</th>
<th>Share immigrants</th>
<th>Share same nationality</th>
<th>Average income</th>
<th>Unemployment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stayed</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \ln ) population</td>
<td>0.188</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share immigrants</td>
<td>0.1044</td>
<td>0.5118</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share same nationality</td>
<td>0.1387</td>
<td>0.1229</td>
<td>0.3926</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average income</td>
<td>0.0755</td>
<td>0.6326</td>
<td>0.2182</td>
<td>-0.1717</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Unemployment</td>
<td>0.0921</td>
<td>0.0475</td>
<td>0.2913</td>
<td>0.3782</td>
<td>-0.528</td>
<td>1</td>
</tr>
<tr>
<td>Age 18-24</td>
<td>-0.0871</td>
<td>-0.0397</td>
<td>-0.0494</td>
<td>-0.053</td>
<td>-0.042</td>
<td>0.0069</td>
</tr>
<tr>
<td>Age 25-39</td>
<td>-0.0011</td>
<td>-0.0098</td>
<td>0.0081</td>
<td>0.0593</td>
<td>-0.0141</td>
<td>0.0325</td>
</tr>
<tr>
<td>Age 40-64</td>
<td>0.0753</td>
<td>0.0444</td>
<td>0.0332</td>
<td>-0.0192</td>
<td>0.051</td>
<td>-0.0411</td>
</tr>
<tr>
<td>Afghanistan</td>
<td>0.1318</td>
<td>0.8438</td>
<td>0.4896</td>
<td>0.1465</td>
<td>0.4322</td>
<td>0.157</td>
</tr>
<tr>
<td>Somalia</td>
<td>0.1931</td>
<td>0.7698</td>
<td>0.6113</td>
<td>0.4784</td>
<td>0.3462</td>
<td>0.2421</td>
</tr>
<tr>
<td>Syria</td>
<td>0.1576</td>
<td>0.8175</td>
<td>0.6236</td>
<td>0.3798</td>
<td>0.3557</td>
<td>0.3313</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Age 18-24</th>
<th>Age 25-39</th>
<th>Age 40-64</th>
<th>Afghanistan</th>
<th>Somalia</th>
<th>Syria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 18-24</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 25-39</td>
<td>-0.4871</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 40-64</td>
<td>-0.3221</td>
<td>-0.6699</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Afghanistan</td>
<td>-0.02</td>
<td>-0.0317</td>
<td>0.0514</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Somalia</td>
<td>-0.0487</td>
<td>0.0244</td>
<td>0.0149</td>
<td>0.6623</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Syria</td>
<td>-0.0504</td>
<td>0.0233</td>
<td>0.0175</td>
<td>0.7669</td>
<td>0.7448</td>
<td>1</td>
</tr>
</tbody>
</table>

N = 4401

Figure 2 - Population Size
Figure 3 - Natural Logarithm of Population Size

Figure 4 - Average Income
Figure 5 - Unemployment Rate

Figure 6 - Share of Foreign-Born Population