Aalto University
School of Business

THE DETERMINANTS OF INTERNAL MIGRATION IN FINLAND

Master's Thesis
Janne Palkama
Aalto University School of Business
Economics
Spring 2018

Author	Janne Palkama
Title of Thesis	The determinants of internal migration in Finland
Degree	Master's Degree in Economics and Business Administration
Degree Programme	Economics
Thesis advisor(s)	Manuel Bagues
Year of approval 2018	**Number of pages** 42 **Language** English

Abstract

Many Finnish regions are slowly emptying, as their young working age populace leave their home municipalities and congregate in large cities to study or in the hopes of better career prospects. It is my purpose in this thesis to find out to what extent it is possible to predict which municipalities will be on the winning and which on the losing side of Finland's internal migration flows in the future by looking at municipal characteristics such as average income, unemployment and population. Additionally I analyse how distance between municipalities is associated with the size of the migration flows between them. I use a logarithmic fixed effects regression model to analyse migration flows from 1996 to 2015, as well as from the periods of 1995-2005 and 2006-2015 separately.

I find that the municipalities that experience a lot of in-migration over the period of 1995-2015 have high average education, low municipal tax rate, low unemployment, and high population, although the significance of education and unemployment has lowered over time. Additionally my results indicate foreigners and working age adults are more numerous in these municipalities, even more so during the later time period.

Municipalities that experience a lot of out-migration have generally higher unemployment from 1995 to 2015, although the significance of this variable has lowered over time. Additionally my results indicate these areas have more foreigners and more working age adults, to a degree that from 1995 to 2015 areas with a high share of working age adults had lower net migration than otherwise similar areas with low share of working age adults. This has changed over time however, and my results predict that in the future municipalities with a high share of working age adults can be expected to have higher net migration.

I find the overall ability of the municipal characteristics used in my model to predict net migration to be quite poor. The ability of these characteristics to predict in- and out-migration is somewhat better, with approximately 6 percent of variation in in-migration and 21 percent of variation in out-migration explained by the model. However, most of the predictive ability of the model on out-migration comes from the inclusion of population as a control variable, and should this variable be removed the predictive ability of the model regarding out-migration drops to a level similar that of in-migration. I find distance alone to be a very strong predictor of migration flows between two municipalities, with approximately 24 percent of variation between municipalities explained by this variable.

Tekijä	Janne Palkama
Työn nimi	Sisäisten muuttoliikkeiden määräävät tekijät Suomessa
Tutkinto	Kauppatieteiden maisteri
Koulutusohjelma	Taloustiede
Työn ohjaaja(t)	Manuel Bagues
Hyväksymisvuosi 2018	**Sivumäärä** 42 **Kieli** Englanti

Tiivistelmä

Monet kunnat Suomessa ovat tyhjenemässä, kun työikäinen väestö muuttaa kotikunnistaan isompiin kaupunkeihin opiskelemaan tai parempien työmahdollisuuksien perässä. Tavoitteeni on tässä pro-gradu työssä tutkia, missä määrin muuttoliikkeiden voittaja- ja häviäjäkuntia on mahdollista ennustaa kuntien ominaisuuksien kuten tulotason, koulutuksen, työttömyyden ym. perusteella. Tämän lisäksi tutkin, millainen yhteys kuntien välisellä etäisyydellä on niiden välisten muuttovirtojen suuruuden kanssa. Analysoin tutkimuksessa logaritmisella kiinteät vaikutukset sisältävällä regressiomallilla Suomen sisäisiä muuttoliikkeitä vuodesta 1995 vuoteen 2015, sekä erikseen vuosina 1995-2005 ja 2006-2015.

Löydösteni mukaan kunnissa, jotka vastaanottavat välillä 1995-2015 paljon muuttajia, on suhteessa korkea koulutustaso, matala kunnallisverotus ja matala työttömyys, vaikkakin koulutustason ja työttömyyden merkitys on vuosien myötä laskenut. Lisäksi näissä kunnissa on suhteessa enemmän ulkomaiden kansalaisia ja työikäistä väestöä.

Kunnissa, joista muutti paljon väestöä pois välillä 1995-2015, on suhteessa korkea työttömyys, vaikkakin tämän merkitys on vuosien myötä laskenut. Lisäksi näillä alueilla on suhteessa enemmän ulkomaan kansalaisia ja työikäistä väestöä jopa siinä määrin, että välillä 1995-2015 nettomuutto oli heikompaa paljon työikäistä väestöä omaavissa kunnissa verrattuna muuten samanlaisiin kuntiin, joissa työikäisen väestön osuus oli pienempi.

Löydösteni mukaan mallissani käytettyjen kuntaominaisuuksien kyky ennustaa kuntien nettomuuttoa on huono. Ominaisuuksien kyky ennustaa sisäänmuuttoa on hieman parempi ja kyky ennustaa ulosmuuttoa paljon parempi. Suuri kyky ennustaa ulosmuuttoa johtuu kuitenkin suurelta osin väestömuuttujan sisällytyksestä malliin. Tämän muuttujan poistaminen laskeekin mallin kyvyn ennustaa ulosmuuttoa samalle tasolle sisäänmuuton kanssa. Kuntien välinen etäisyys pystyy ennustamaan erittäin vahvasti niiden välisten muuttovirtojen suuruutta. Etäisyys yksinään kattaa noin viidenneksen kuntienvälisestä muuttoliikkeiden suuruuden vaihtelusta.

Table of Contents

1. Introduction 1
2. History of Migration Research 3
   2.1 Ravenstein's Laws of Migration 3
   2.2 Spatial Gravity Models of Migration 4
   2.3 Systemic models 6
3. Review of empirical literature 8
   3.1 A Conditional Logit Approach to U.S. State-to-State Migration by Davies et al. 8
   3.2 The Gravity Model of Migration: The Successful Comeback of an Ageing Superstar in Regional Science by Poot et al. 10
   3.3 A long-run analysis of push and pull factors of internal migration in Italy. Estimation of a gravity model with human capital using homogeneous and heterogeneous approaches by Piras 11
4. Data and descriptive statistics 14
   4.1 Gender differences 15
   4.2 Net migration 15
   4.3 Average Household Income per Consumption Unit 16
   4.4 Municipal Income Tax 17
   4.5 Unemployment rate 19
   4.6 Higher education rate 20
   4.7 Share of Working Age Adults 21
   4.8 Share of Foreigners 22
5. Empirical methodology 24
6. Results 27
   6.1 In-to-Out-Migration Ratio 27
   6.2 In-Migration 29
   6.3 Out-Migration 32
   6.4 Municipality-to-Municipality Migration and Distance 36
7. Conclusions 38
8. References 41

1. Introduction

During the last couple of decades urbanization has steadily increased in nearly every country in the developed world. As more and more people congregate to live around large population centres, whether keeping the whole country occupied should be a goal in policy has been a hot topic in Finland for some time now.

While innovation and progress are being fostered in cities, so is the price of living in these areas skyrocketing, leading to a situation of inequality where only the prosperous can afford to live close to the concentration of jobs and services. At the same time the population centres are able to gather a lion's share of all tax revenue and are thus able to offer their citizens the best public services, while many more remote areas are forced to rely on transfer payments to be able to offer even basic healthcare to their constituents. The price of real-estate in these areas has collapsed as well, leaving many citizens trapped with a worthless house, unwilling to sell as the sales price would be close to nothing compared to the price once paid for the accommodation.

The factors governing people's migration choices is a rather widely studied subject. Many empirical studies look into how factors related to economic well-being are associated with migration patterns (Leisure & Lewis, 1977; Liang Z. & White, 1996; Piras, 2015). Some of these focus on personal characteristics of the migrants (Thomas, 1963), while some attempt to still extend their research to account for differences in amenity offerings between regions (Liu & Shen, 2014).

My aim in this paper is to analyse Finnish intermunicipal migration flows from 1995 to 2015 and determine to what extent it is possible to predict which municipalities are on the winning and which on the losing side of these flows based on municipal characteristics such as income, municipal tax rate, population, etc. I compose a logarithmic fixed effects regression model to analyse net migration, as well as in- and out-migration separately. I also look at the periods of 1995-2005 and 2006-2015 separately to determine whether the associations found have significantly changed over time. Additionally, I compose a separate model to analyse the association of distance with municipality-to-municipality migration.

The rest of the paper is constructed as follows: In chapter three, I briefly go over some aspects of migration research history to give the reader a better idea on how the field has moved from making simple observations into using more and more complex empirical

models over the last hundred and forty years. In chapter 4, I look at some examples of empirical migration research to showcase the use of the models mentioned in chapter 3 and to present some results gained in the field. In chapter 5, I discuss the data used in this thesis, as well as illustrate the development of several key variables over the years in Finland. In chapter 6, I discuss in detail the empirical methodology used. In chapter 7, I present the results of the analysis. Chapter 8 concludes by briefly reviewing the results.

2. History of Migration Research

In this part of the thesis, I go over some key aspects of migration research history. Namely, I briefly introduce the research done by E. G. Ravenstein, as his is one of the first documented attempts to analyse internal migration flows. His research lay the groundwork for more formal models historically used to analyse migration, of which I also introduce two in this chapter.

2.1 Ravenstein's Laws of Migration

One of the first systematic studies on migration was done in the late eighteen-hundreds by E. G. Ravenstein. In this study (Ravenstein 1885) Ravenstein examined the British census data on nativity of the population and place of residence. From a compilation of net migration figures, Ravenstein was able to identify places of net in-migration and net out-migration. The sex detail in his data also allowed him to describe the sex composition of these migration patterns. The study was purely observing in nature and void of any more advanced empirical methodology. It also relied on lifetime migration data, the limitations of which Ravenstein never came to grips with during his lifetime. He discussed some data shortcomings in the beginning of his paper, such as different sizes of counties not being accounted for. However, he entirely missed the major criticisms of lifetime migration data, such as their inability to pick up multiple migrations and return migration (Greenwood & Hunt 2003). Despite these short-comings, from his various and detailed analysis Ravenstein came to infer seven "laws of migration", which lay considerable groundwork for future migration studies.

Ravenstein's laws incorporate three separate notions: 1. Ceteris paribus, distance appears to deter migration; 2. Large cities are the direction of much migration; 3. Migration flows between two places depend on the population sizes of both the origin and the destination. A review of Ravenstein's discussions also gives the notion that he viewed differential economic opportunities as important pull and push factors in the context of rural-to-urban migration flows, and considered spatial variation in economic opportunities to be a key determinant of migration in general (Greenwood & Hunt 2003). While rather self-evident, these insights lay the groundwork for gravity models of spatial interaction, discussed in the next section.

2.2 Spatial Gravity Models of Migration

For almost half a century, the gravity model has been the most common method to portray gross migration flows between regions. This model is intuitively consistent with earlier migration research, easy to estimate in its simplest form, and provides a relatively good fit in most applications (Poot et al, 2016).

Although the idea of the gravity model predates Ravenstein's work, its application to migration research occurred after his time, in the 1920s. While studying rural migration in New York, E.C Young (1924) hypothesized that migration flows might be predicted by a model similar to what's used to predict gravitational pull in physics. He ended up with a model that focuses on the structure of the migration response in terms of the relative sizes and locations of regions. As in physics, gravitational pull in the model is directly proportional to mass and inversely proportional to distance. In the case of migration, mass equals the number of people living in a population centre or area. The purpose of the distance parameter is to capture the difficulty or cost of relocation, and as such has taken many forms over the years. Possible proxies are for example straight line distance, driving distance, railway distance or travel time. This model works in effect to quantify some of the associations noticed by Ravenstein 40 years earlier. In its most commonly applied form, the gravity law of migration states that:

$$M_{ij} = G\frac{P_i^{\alpha} P_j^{\beta}}{D_{ij}^{\delta}}$$

where:

M_{ij} = The amount of people to move from region *i* to region *j* inside the time period being scrutinized.

G = A proportionality constant depending on the context where the migration occurs, influenced by variables such as geography or the length of the time period.

P_i = Population of region *i*.

P_j = Population of region *j*.

α, β, δ = Parameters to be estimated.

D_{ij} = The distance between municipalities *i* and *j*.

The model can be estimated by ordinary least squares after its transformation into the following logarithmic form:

$$\ln M_{ij} = \ln G + \alpha ln P_i + \beta ln P_j - \gamma ln D_{ij} + \varepsilon_{ij}$$

While the model in its original form is already rather good at predicting migration, it wasn't until the 1960s that it began to truly gain traction in the form of modified gravity models that sought to improve the model by adding additional variables expected to significantly influence the decision to migrate, such as ratios of income, unemployment rate, degree of urbanization, various climatological variables, measures of public expenditures and/or taxes. Like the unmodified version of the model, modified gravity models are frequently estimated in log-log form, as the coefficients obtained this way can be directly interpreted as elasticities of migration response to changes in the independent variables of the model. An example of a form commonly used is demonstrated below:

$$ln\ M_{ij} = ln\ \beta_0 + \beta_1 lnD_{ij} + \beta_2 lnP_i + \beta_3 P_j + \beta_4 lnY_i + \beta_5 lnY_j + \sum_{n=1}^{m} \alpha_n lnX_{in} + \sum_{n=1}^{m} \varphi_n lnX_{jn} + e_{ij}$$

where:

Y_i = The average income in municipality *i*.

Y_j = The average income in municipality *j*.

X_{in} = A vector of spatial variables such as unemployment, urbanization rate, tax rate, etc. in municipality *i*.

X_{jn} = A vector of spatial variables such as unemployment, urbanization rate, tax rate, etc. in municipality *j*.

This approach has garnered some criticism. For example, Schultz (1982) argues that the approach fails to account for the differences in geographic sizes of the regions for which migration is measured. If all regions had the same land area, migration between areas would be tied to the costs and benefits of the various location choices. However, as regions differ in size, a larger share of all moves tend to occur within the boundaries of the larger regions. As a consequence, more non-migration will appear to exist for such regions. This leads to migration being spuriously correlated with population size. The concerns raised by this criticism can probably be countered by controlling for land area or by using panel data with a fixed effects model to absorb the effects of any non-changing spatial variables (such as land area). However, modified gravity models can still be criticised for only focusing on two municipalities at a time, and thus failing to take into account the possible effects of all other

competing migration destinations in the system. Systemic models are formulated to counter this failing.

2.3 Systemic models

Gravity models treat migration decisions purely in terms of origin and destination. In contrast, systemic models take into account not only the characteristics of origin and destination, but also the characteristics of all other possible and relevant destinations in the geographical region under study. Some prominent examples of systematic migration models are Alonso's (1978, 1986) systemic model and McFadden's (1978, 1981) nested logit model. Alonso's model to incorporate systemwide effects takes the following form:

$$M_{ij} = v_i D_i^{\alpha_i - 1} w_j C_j^{\beta_j - 1} t_{ij}$$

where

M_{ij} = The amount of people to move from municipality *i* to municipality *j*.

v_i = A function of the characteristics of municipality *i* or its population.

D_i = The pull of the system on the population of *i*, a function of the relation of *i* to the rest of the system.

α_i = The elasticity of response of migration from municipality *i* to municipality *j* to the pull (D_i)

w_j = A function of the characteristics of municipality *j* or its population.

C_j = The competitive force exerted by the system on the potential movers from municipality *i* to municipality *j*, a function of the relation of *j* to the rest of the system.

β_j = elasticity of response to C_j.

t_{ij} = ease of transitioning between *i* and *j*.

It's noteworthy that if D_i and C_j are set to equal one and t_{ij} is determined as the inverse function of distance, this equals a modified gravity model. The modified gravity model stated in the previous section of this chapter can be rewritten as a systemic model as follows:

$$ln\, M_{ij} = ln\, \beta_0 + \sum_{j \neq i} \beta_{1j} lnD_{ij} + \beta_2 lnP_i + \sum_{j \neq i} \beta_3 P_j + \beta_4 lnY_i + \sum_{i \neq j} \beta_5 lnY_j + \sum_{n=1}^{m} \alpha_n lnX_{in} + \sum_{i \neq j} \sum_{n=1}^{m} \varphi_{jn} lnX_{jn} + e_{ij}$$

At the time of its creation, Alonso's model was interpreted as a theory of movement. However, Alonso himself later indicated that he preferred to view it not as a theory with specific falsifiable predictions, but as an encompassing structure for a variety of spatial interaction models (Greenwood & Hunt, 2003).

According to McFadden's nested logic model, migration response consists of several layers. In a two-level setup, the top layer models a migrant's unconditional decision to migrate from or stay in the origin. The lower layer models the choice of destination, conditional on a positive migration decision having been reached. The destinations and origins in the model include all the areas in the geographic system under scrutiny. The model does not imply that migrants first decide whether to migrate and then decide where to migrate. At the top layer, the decision to migrate depends on area characteristics of both the origin and all possible destinations. Systemic information about the possible destination areas is passed from the lower layer to the top one, and so it is incorrect to assume that the top layer decision of whether to migrate is dependent only on the characteristics of the origin and not on the characteristics of other areas in the system in question.

3. Review of empirical literature

Several empirical papers have researched internal migration in various countries around the world. Some good examples are Liang and White's Internal migration in China, 1950-1988 (1996), Leasure and Lewis' *Internal migration in the USSR: 1897-1926 (1967)*, Thomas' *Internal migration in Sweden: a recent study (1963)*, Greenwood's *An analysis of the determinants of geographic labor mobility in the United States (1969)*, Davies et al's *A Conditional Logit Approach to U.S. State-to-State Migration (2001)*, Poot et al's *The Gravity Model of Migration: The Successful Comeback of an Ageing Superstar in Regional Science (2016)*, and Piras' *A long-run analysis of push and pull factors of internal migration in Italy. Estimation of a gravity model with human capital using homogeneous and heterogeneous approaches (2016).*

In this section of the paper I briefly go over the studies by Davies et al., Poot et al. and Piras as a few concrete examples on the methods used and the findings made by internal migration research.

3.1 A Conditional Logit Approach to U.S. State-to-State Migration by Davies et al.

In the paper by Davies et al. the researchers use a conditional logit model motivated by a random utility model to analyse factors associated with the probability of a person moving to a particular US state instead of moving to some other state or staying put. Their analysis spans eleven years from 1986 to 1997. Theirs is a good example of a systemic model, taking into account all the possible migratory destinations in the system. An individual at area *i* faces *J* migration choices, including staying at their current location. The utility level for this individual of choosing area *j* is

$$U_{ij} = \beta' X_{ij} + \varepsilon_{ij}$$

where X_{ij} is a vector of location specific attributes and β constant across choices. The individual chooses the destination with the highest utility. Davies et al. model the probability of an individual at area *i* choosing destination *j* as

$$P(y_i = j) = \frac{e^{\beta' x_{ij}}}{\sum_k e^{\beta' x_{ik}}}$$

when there are K destinations and *i* = *j* for non-movers. The log-likelihood for all individuals combined moving from any area *i* to a specific area *j* is

$$\ln L = \sum_i m_{ij} \ln P(m_{ij} = 1)$$

where $m_{ij} = 1$ if an individual at area *i* chooses destination area *j*. In their study Davies et al. have 47 source states and 47 potential destination choices. The corresponding log-likelihood function is

$$\ln L = \sum_{i=1}^{47} \sum_{j=1}^{47} N_{ij} \frac{e^{\beta' x_{ij}}}{\sum_{k=1}^{47} e^{\beta' x_{ik}}}$$

where N_{ij} is the number of people moving from state *i* to state *j*. To capture unobservable economic and non-economic state characteristics that might play a role in the migration decision, they also include a set of dummy variables for potential destinations. A dummy variable is also included to capture the likelihood of simply not migrating.

As prediction variables Davies et al. use the ratios of populations, unemployment rates and per capita incomes between states, distance between states, and distance-squared between states. They find that in 1996, a one-standard deviation change in the destination-to-origin unemployment ratio reduces the probability of migration by between 18 to 20 percent, and a one-standard deviation change in the destination-to-origin per capita income ratio increases the probability of migration by between 16 to 17 percent. The proportional changes in the probability of migration resulting from a hundred-mile increase in distance between origin and destination states vary widely across the pairs of states, but the gist of the findings is that increased distance between states decreases the probability of migration.

Davies et al. also analyse trade-offs between variables in terms of per capita income. They find that to offset an increase of one in the ratio of unemployment between states, the corresponding ratio of per capita income between states should increase by 0,76. For example, if the unemployment rate in both the origin and destination is at first 4 % and the per capita income $10 000, after an increase of a 100 percent in the unemployment rate of the destination, to 8 %, an increase of $7 600 to $17 600 in the per capita income of the destination would be required for the probability of migration to stay the same. The trade-off between distance and per capita income is found to decrease as the distance grows longer. For example, the income compensation for an additional mile from Tennessee falls from $47 at 500 miles from the origin to $28 at 2,000 miles from the origin.

3.2 The Gravity Model of Migration: The Successful Comeback of an Ageing Superstar in Regional Science by Poot et al.

In the paper by Poot et al. the researchers analyze internal migration patterns in New Zealand. They modify a traditional gravity model to include international migration, as well as to be able to analyze migration flows separately for inter-urban migration and urban-rural migration. They use data from the New Zealand Census, in which migration data is recorded by means of a question on 'usual residence five years ago'. The data for the study is assembled from six censuses starting from 1986 and ending in 1997. The population is restricted to individuals aged 25 to 54 in order to focus the analysis on labor migration and to exclude students and retirees.

The model used by Poot et al. takes the following form:

$$\ln M_{ij} = U_{ij}\,(\delta_1 + \alpha_1 \ln P_i + \beta_1 \ln P_j - \gamma \ln D_{ij}) + E_{ij}(\delta_2 + \alpha_2 \ln P_i) + O_{ij}(\delta_3 + \alpha_3 \ln P_i) + I_{ij}(\delta_4 + \beta_2 \ln P_j) + R_{ij}(\delta_5 + \beta_3 \ln P_j) + \varepsilon_{ij}$$

where:

U_{ij} = 1 if both the origin *i* and destination *j* are urban areas, and 0 otherwise.

E_{ij} = 1 if origin *i* is an urban area and the destination *j* is abroad, and 0 otherwise.

I_{ij} = 1 if origin *i* is abroad and the destination *j* is an urban area, and 0 otherwise.

O_{ij} = 1 if origin *i* is an urban area and the destination *j* is rural, and 0 otherwise.

R_{ij} = 1 if origin *i* is a rural area and the destination *j* is urban, and 0 otherwise.

$P_{i(j)}$ = the population in the origin (destination).

D_{ij} = the distance between origin *i* and destination *j*.

'International' and 'rural' areas in the model don't have a specific location, so the distance between these areas is considered undefined. Similarly, the population of these areas is considered undefined as the international migrants come from all over the world and while the total rural population is known, it is spatially dispersed rather than a compact mass. To test the sensitivity of the model to the exact measure of distance used, different specifications are estimated for three different proxies: straight line distance, road travel distance in kilometers and road travel time in minutes.

Poot et al. find that in all periods, the association between distance and migration is the largest when distance is measured as road travel distance in kilometres. Distance in minutes yields almost similar estimates, while straight line distance suggests associations that are about 0.1 lower than the other two proxies. The estimates for distance as road travel

distance in kilometres range from -0.714 to -0.845, meaning a one percent increase in road travel distance is associated with a 0.714-0.845 decrease in migration. The estimates for travel time in minutes range from -0.706 to -0.834 and the estimates for straight line distance from -0.623 to -0.747. The estimates are the lowest for the period of 1981 to 1986. They rise until peaking at 1996 to 2001, after which they start to come down again. The authors note that this goes against the expectation that greater connectivity between urban areas would have reduced the deterrence effect of distance over time.

On the estimates for association between origin population and migration, Poot et al. find coefficients that range from 0.883 to 0.959. This indicates a one percent increase in origin population is associated with a 0.883-0.959 percent increase in migration. As these estimates are below zero, it can be concluded that the total share of outmigration from the total population declines as population increases. An exception to this is emigration. The estimate for the association between origin population and migration in emigration ranges from 1.301 to as high as 2.072. This indicates the relative share of the population to move abroad increases when the population does.

On the estimates for association between destination population and migration, Poot et al. find coefficients that range from 0.831 to 0.954. This indicates a one percent increase in destination population is associated with a 0.832-0.959 percent increase in migration. The estimates are higher for immigration, indicating immigrants are more likely to be drawn into big cities, and lower for rural-urban migration, indicating inter-urban migration is more likely to flow into the direction of the biggest cities than rural-urban migration. The authors also find that large urban areas generate relatively more inter-urban than urban-rural migration compared to smaller urban areas.

3.3 A long-run analysis of push and pull factors of internal migration in Italy. Estimation of a gravity model with human capital using homogeneous and heterogeneous approaches by Piras

The paper by Piras analyses internal migration flows across Italian regions from 1970 to 2005 using a gravity model modified to include systemic effects to determine push and pull factors associated with migration. The model the author uses assumes that both supply (push) and demand (pull) factors together with other restraining and/or aiding factors determine migration flows. The starting point for the model takes the following form:

$$F_{ijt} = A_{ij}^{a^0} \left(\frac{S_{it}^{a_1} D_{jt}^{a_2}}{R_{ijt}^{a_3}}\right)$$

where

F_{ijt} = Migration from origin *i* to destination *j* at time *t*.

S_{it} = Supply (push) factors at origin *i*.

D_{jt} = Demand (pull) factors at destination *j*.

R_{ijt} = Other variables restraining and/or aiding migration associated with the specific origin-destination pair. Typically represents the costs associated with moving from origin *i* to destination *j* at time *t*

A_{ij} = A constant catching all time-invariant region-pair effects linked with migration.

a_k = Structural elasticities.

The author further assumes that at time *t*

$$S_{it} = b_i y_{it}^{b_1} n_{it}^{b_2} h_{it}^{b_3}$$

$$D_{jt} = c_j y_{jt}^{c_1} n_{jt}^{c_2} h_{jt}^{c_3}$$

where

y_{kt} = Expected income at region *k* at time *t*.

n_{kt} = Size of the population at region *k* at time *t*.

h_{kt} = The average level of human capital at region *k* at time *t*.

After substituting the equations for the supply and demand factors into the equation for migration and taking natural logarithms, the authors get the following model:

$$\ln F_{ijt} = a_0 \ln A_{ij} + a_1 \ln b_i + a_2 \ln c_j + a_1 b_1 \ln y_{it} + a_1 b_2 \ln n_{it} + a_1 b_3 \ln h_{it} + a_2 c_1 \ln y_{jt} + a_2 c_2 \ln n_{jt} + a_2 c_3 \ln h_{jt} - a_3 \ln R_{ijt}$$

Expected income is assumed to be a function of per capita income and unemployment. To disentangle the role of these two factors, in the empirical implementation of the model per capita income and unemployment at origin (pcy_{it} and u_{it}) and destination (pcy_{jt} and u_{jt}) will be simultaneously considered. Average years of schooling is used as a proxy for human capital. Thus the empirical specification can be estimated in the following form:

$$\ln F_{ijt} = \gamma_{1ij} \ln pcy_{it} + \gamma_{2ij} \ln pcy_{jt} + \gamma_{3ij} \ln u_{it} + \gamma_{4ij} \ln u_{jt} + \gamma_{5ij} \ln n_{it} + \gamma_{6ij} \ln n_{jt} + \gamma_{7ij} \ln h_{it} + \gamma_8 \ln h_{jt} + \lambda_{ij} + \varepsilon_{ijt}$$

where λ_{ij} is a dummy for the region pair *i-j* intended to capture all region-pair specific fixed factors such as distance or common borders. This makes the model used by the author notably different from those in the other two papers looked at in this section, as the estimation of the association between migration flows and distance is completely left outside the confines of the model.

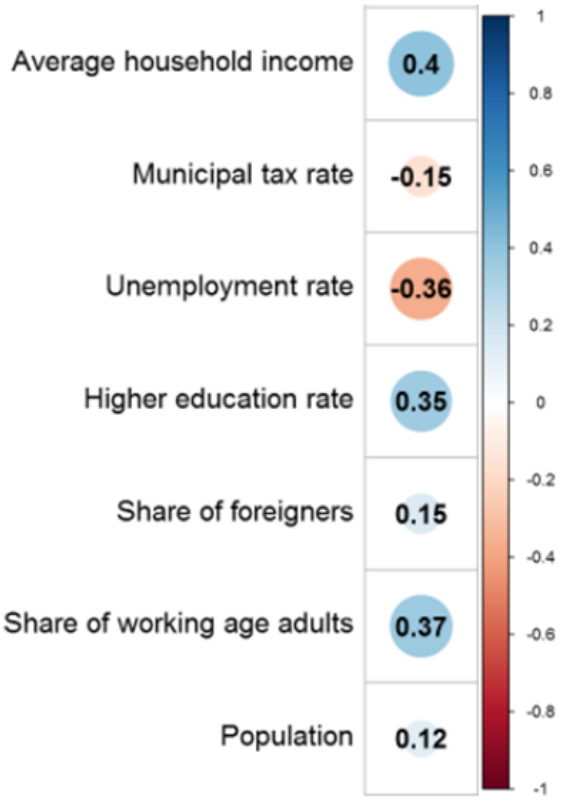
The author runs numerous regressions for multiple origin-destination pairs. As it would be exhaustive for the reader to go through all of them, I present the findings the author made for migration flows from South to Centre-North Italy.

The author finds per capita GDP at origin to have a negative association with migration (from -0.268 to – 0.505) and per capita GDP at destination to have a positive association with migration (from 0.681 to 0.875). He finds unemployment at origin to have a positive association, although this is significant for only some of the origin-destination specifications (0.159). Unemployment at destination has a negative association, although this is also significant only for some origin-destination specifications (from -0.046 to -0.209). Interestingly, the population at origin has a positive association for some origin-destination pairs and a negative for others. These associations also range from very highly positive (4.12) to highly negative (-2.253). Population at destination is found to have a highly positive association with migration (from 1.101 to 3.058). Average years of schooling is found to have no significant association with migration, although it should be noted this variable gains significance for some of the other origin-destination pairs in the study.

4. Data and descriptive statistics

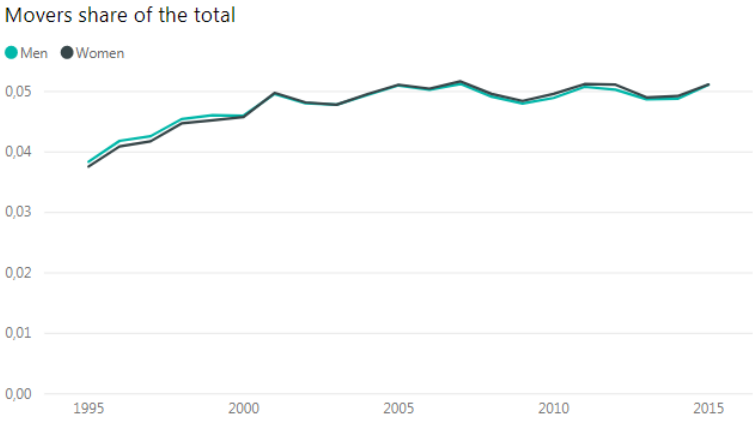
The data used in this thesis is a combination of statistics provided by Statistics Finland and the Association of Finnish Local and Regional Authorities. The data from Statistics Finland include a time series dataset on Finnish internal migration from 1995 to 2015, another on the average usable income of a housing unit per consumption unit from 1995 to 2015, and a third comprising of numerous municipal key metrics for Finnish municipalities during this same time period. The data from the Association of Finnish Local and Regional Authorities comprises of the Finnish municipal income tax rates from 1995 to 2015.

The municipal division used in the data is that of 2017. The data contains the amount of men and women that moved from each municipality of origin to every destination municipality each year. From these figures are calculated in- and out-migration for each municipality for each year, as well as ratios between in- and out-migration. Other metrics used in the study are *Average Household Income per Consumption Unit*, *Municipal Income Tax*, *Unemployment Rate*, *Higher Education Rate*, *Share of Foreigners, Share of Working Age Adults* and *Population*. These metrics are available for each municipality of origin and destination for each year.



The correlation plot on the left illustrates the correlation coefficients between the ratio of in- to out-migration and each metric for the whole time period under scrutiny. The larger and more intensely coloured the circle, the stronger the correlation. It appears that municipalities with high in-to-out-migration ratio also have high average household income, high education rate, high share of working age adults, high share of foreigners, high population, low unemployment and low municipal tax rate. The correlation appears the strongest for average household income, higher education rate, unemployment rate and share of working age adults. This is largely what one would expect. I discuss development of these metrics over the years and their distribution between Finnish municipalities in more detail in the following section.

4.1 Gender differences



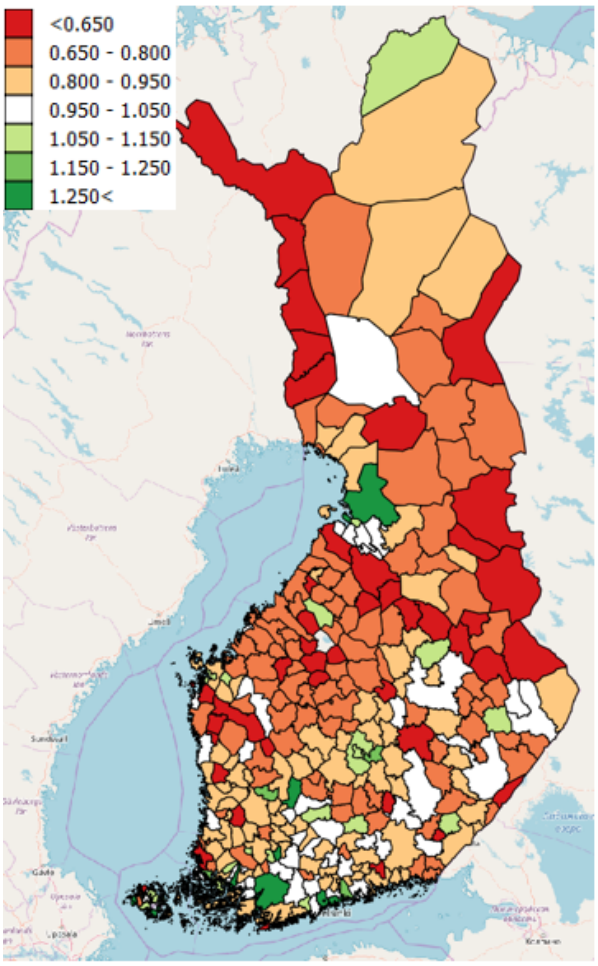
Plot 1

As shown in Plot 1, the share of migrants of the total population increases somewhat from 1995 to 2000, but has remained at roughly 5 percent from that year onwards. There is no real difference between the genders in this respect.

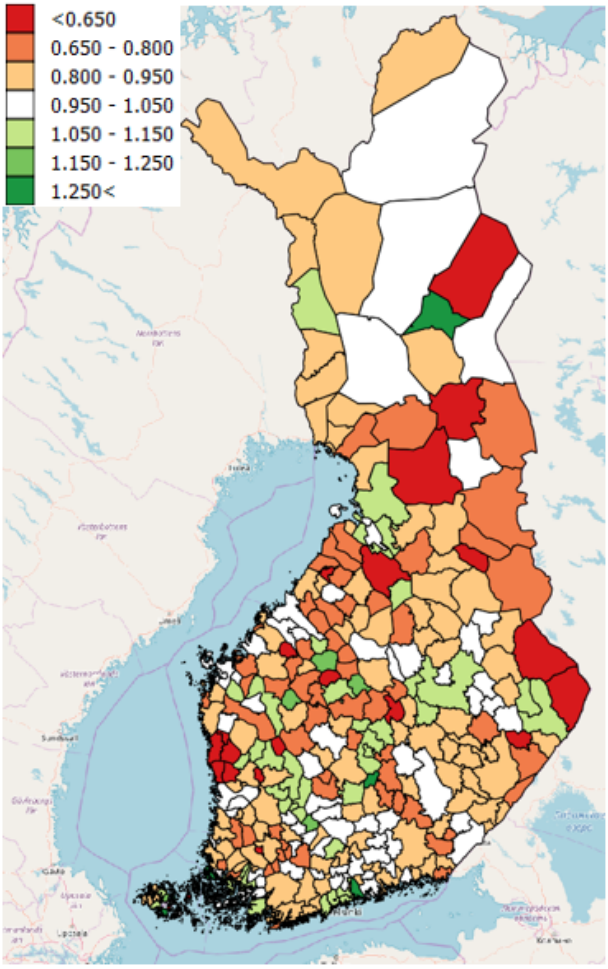
4.2 Net migration

Maps 1 and 2 illustrate the distribution of in-to-out-migration ratio by municipality in 1995 and 2015, respectively. In 1995 most municipalities in the country experienced negative net migration, with migration flows concentrated on few large cities such as Helsinki, Turku and Oulu. Of the 311 municipalities in Finland, 251 experienced negative net migration, 57 experienced positive net migration, and 3 had exactly zero net migration. While migration is still quite concentrated into a few municipalities in 2015, this contrast between them and the other municipalities in Finland is far less stark than two decades earlier. In 2015, 231 municipalities experienced negative net migration, 74 experienced positive net migration, while 6 had exactly zero net migration.

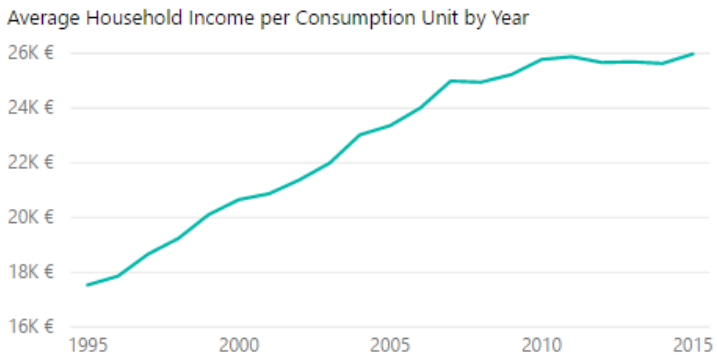
Map 1: In-to-out-migration ratio distribution by municipality in 1995



Map 2: In-to-out-migration ratio distribution by municipality in 1995



4.3 Average Household Income per Consumption Unit

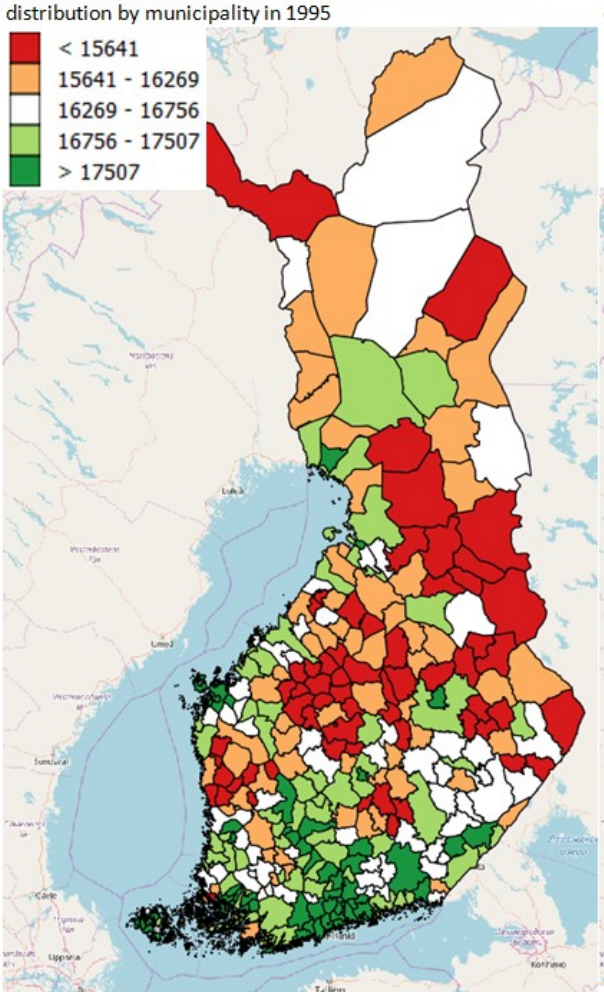


Plot2

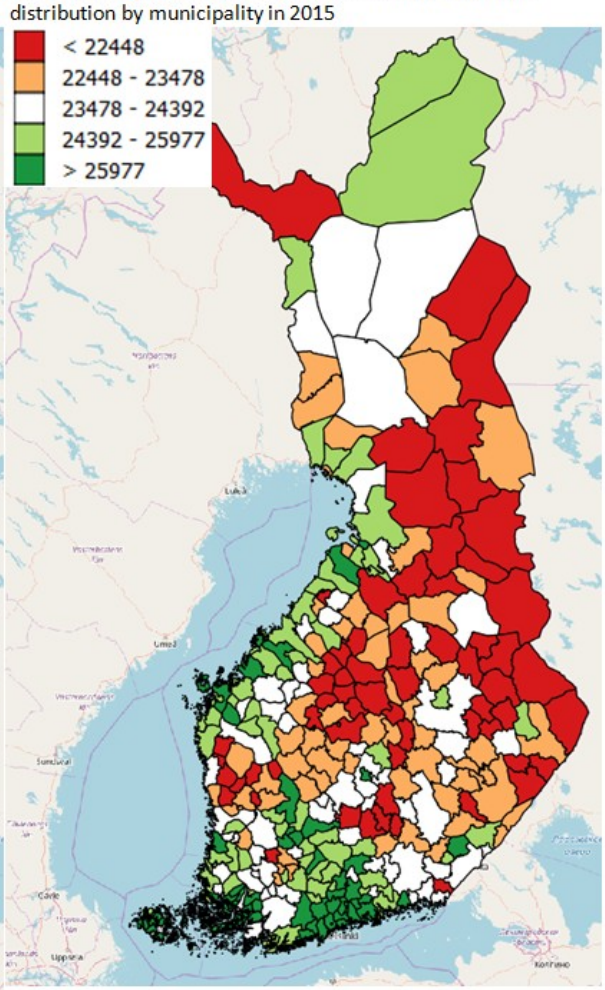
As shown in plot 2, the nominal average household income per consumption unit has increased by approximately 2 % per year from 1995 to 2015. A consumption unit is an attribute used to compare housing units differing in size and structure to each other. The first adult in a household is given the weight of 1 unit, the rest of household personnel over the age of 13 the weight of 0.5, and children the weight of 0.3. The increase was somewhat steady from 1995 to 2008, after the rate of increase lessened likely due to the

depression that began at the time. Maps 3 and 4 show the distribution of average household income per consumption unit by quintile between years 1995 and 2015, respectively.

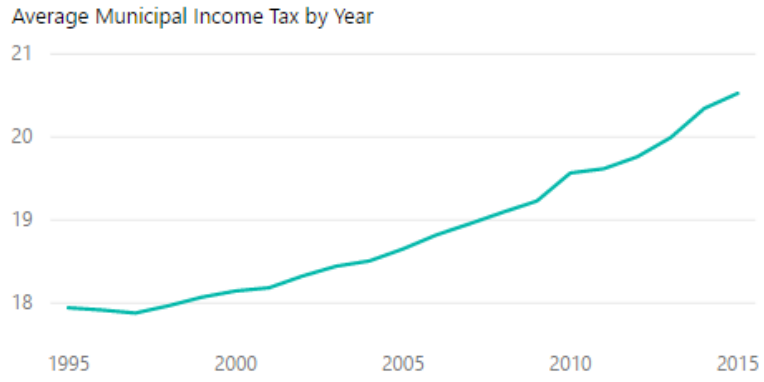
Map 3: Average household income per consumption unit distribution by municipality in 1995



Map 4: Average household income per consumption unit distribution by municipality in 2015



4.4 Municipal Income Tax

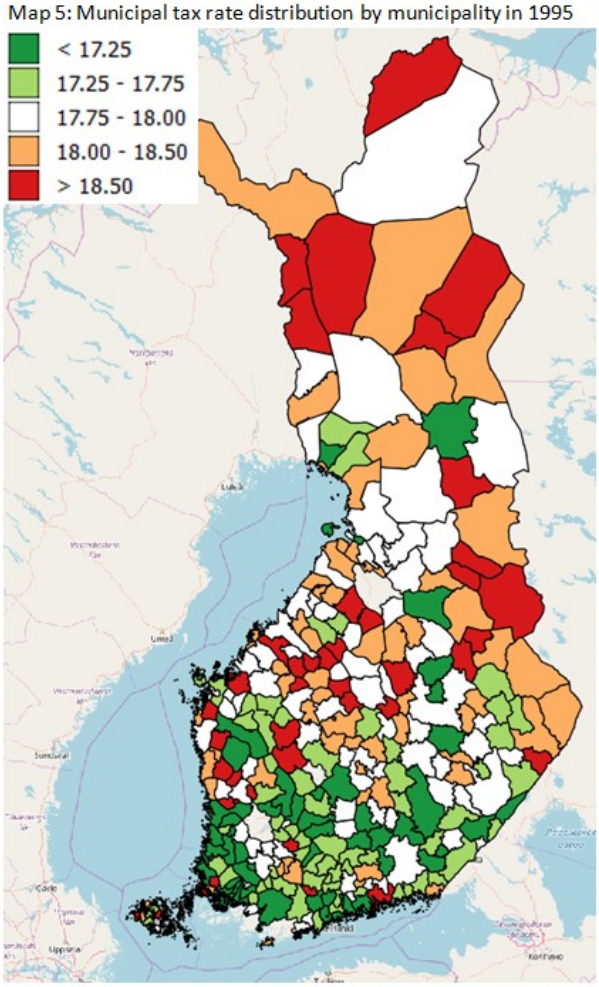


Plot 3

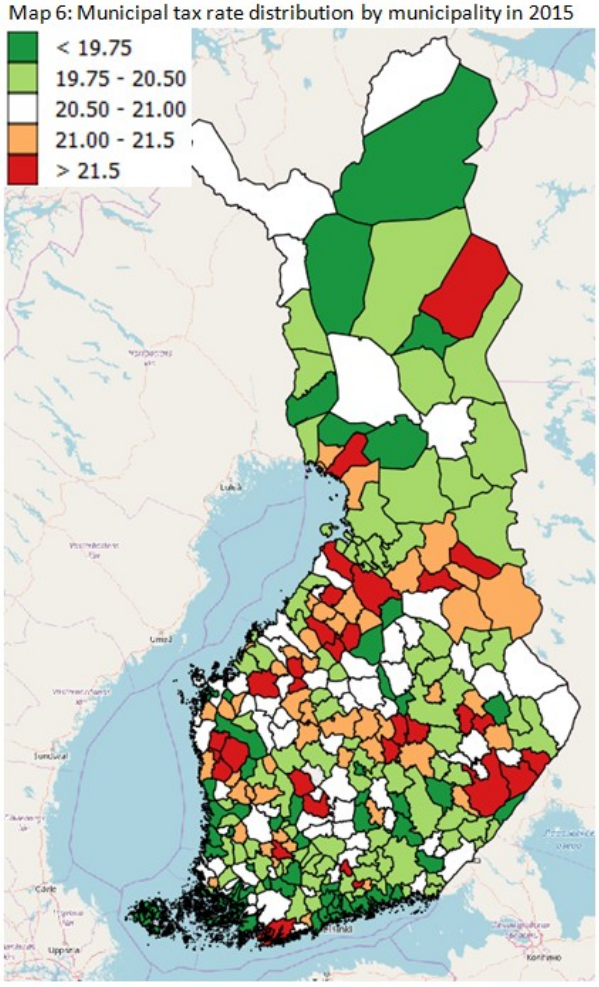
A shown in plot 3, the average municipal income tax has increased steadily from 1995 to 2015. Maps 5 and 6 show the distribution of municipal income tax by quintile in 1995 and 2015, respectively. In 1995

there was a clear divide between southern Finland and the rest when it comes to municipal tax levels. In 20 years, this divide has changed to be more between the central Finland and the rest.

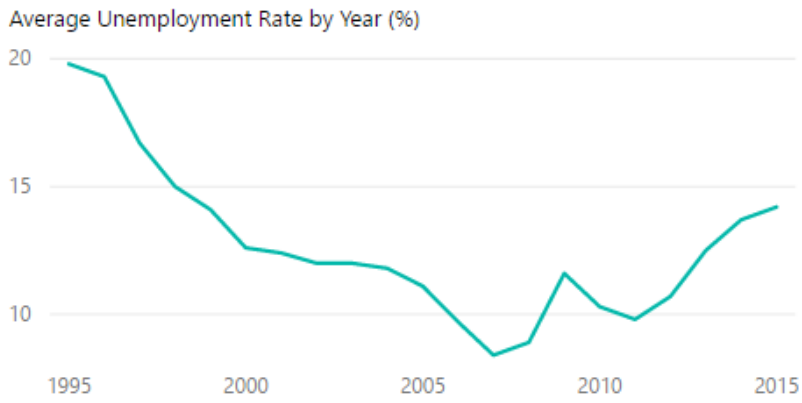
Map 5: Municipal tax rate distribution by municipality in 1995



Map 6: Municipal tax rate distribution by municipality in 2015

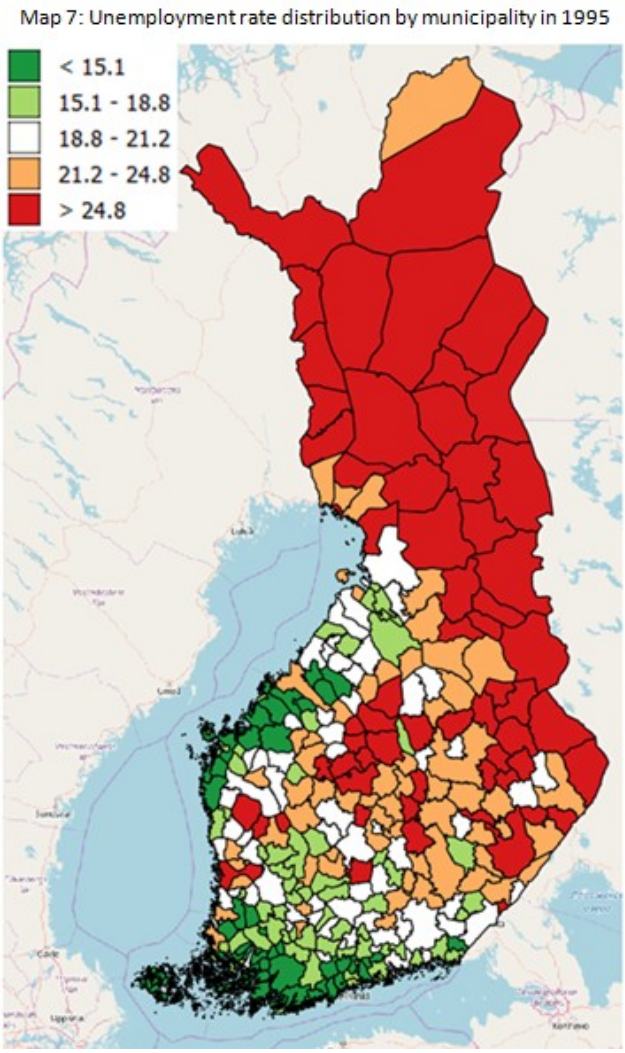


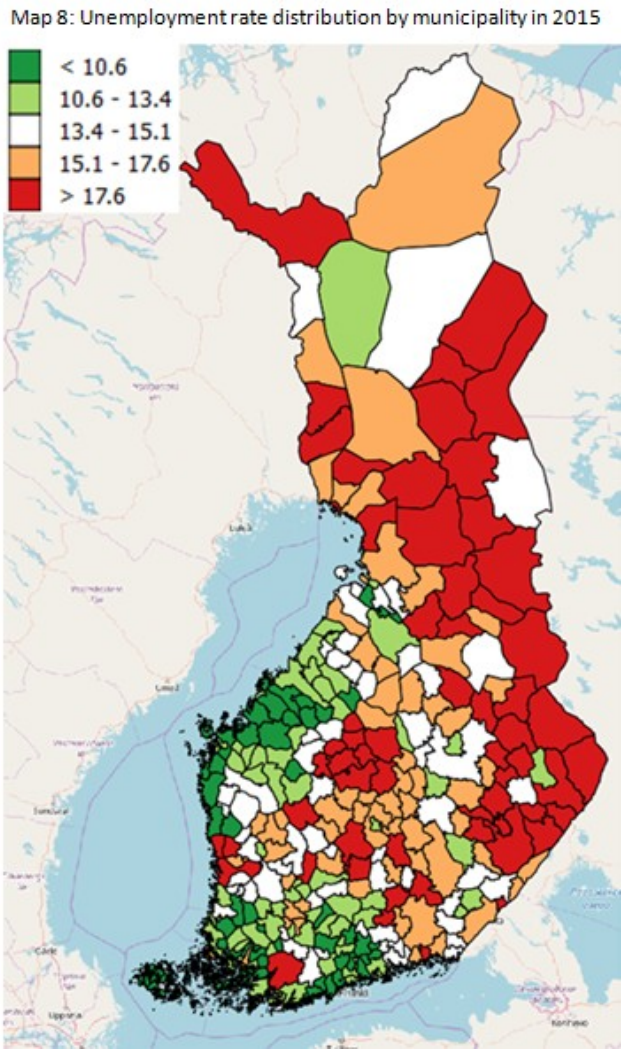
4.5 Unemployment rate



Plot 4

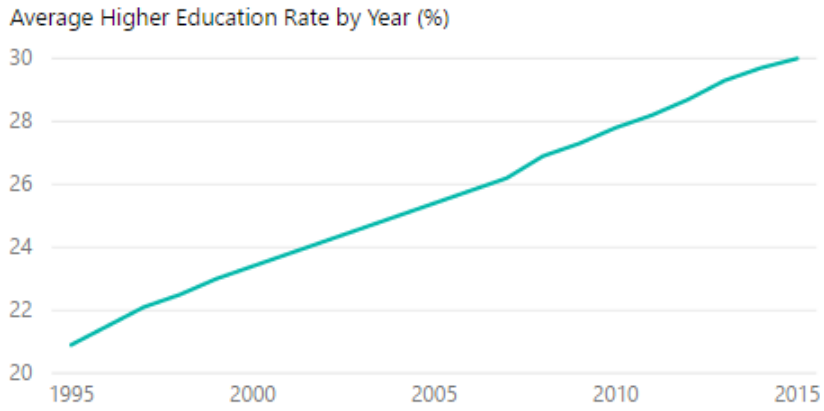
As shown in plot 4, the average unemployment rate in Finland came down from approximately 20 percent to below 10 from 1995 to 2007. The effects of the depression beginning in 2008 caused the rate to start rising again that year. There was brief recovery in 2010 and 2011, after which unemployment continued to rise for the rest of the whole time period under scrutiny. Maps 7 and 8 illustrate the distribution of unemployment between municipalities by quintile in 1995 and 2015,





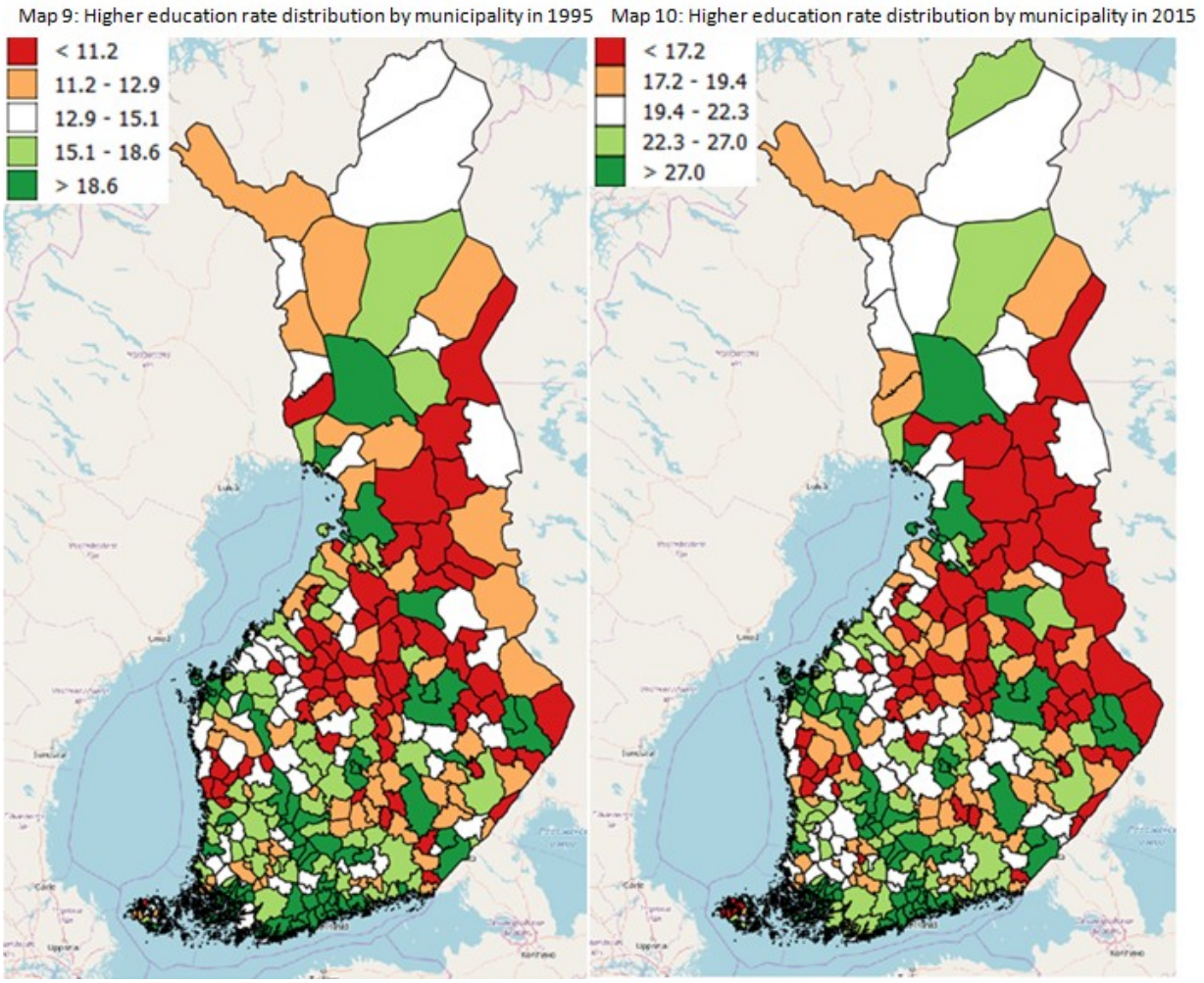
respectively. The distribution has remained largely the same with eastern and northern Finland suffering from unemployment rates that are significantly higher than those of the west and south. Some municipalities have made a relative recovery in Lapland however.

4.6 Higher education rate



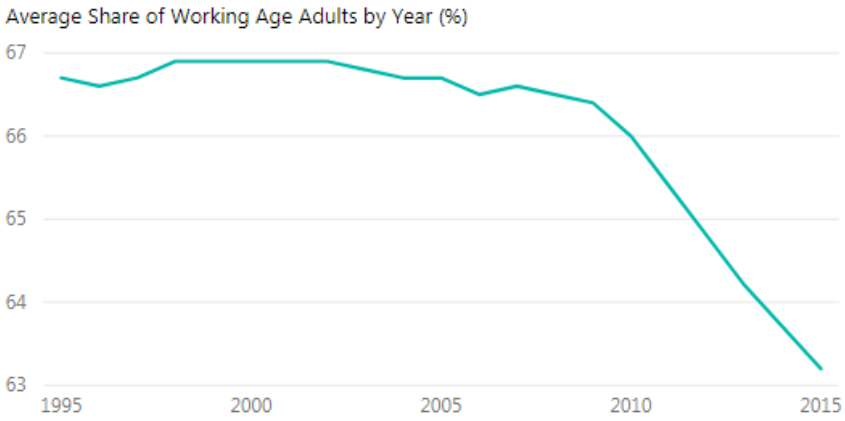
Plot 5

As shown in plot 5, higher education rate has steadily increased in Finland from 1995 to 2015. The higher education rate measures the percentage of the population over 15 years old that have obtained a higher education degree. Maps 9 and 10 illustrate the distribution of



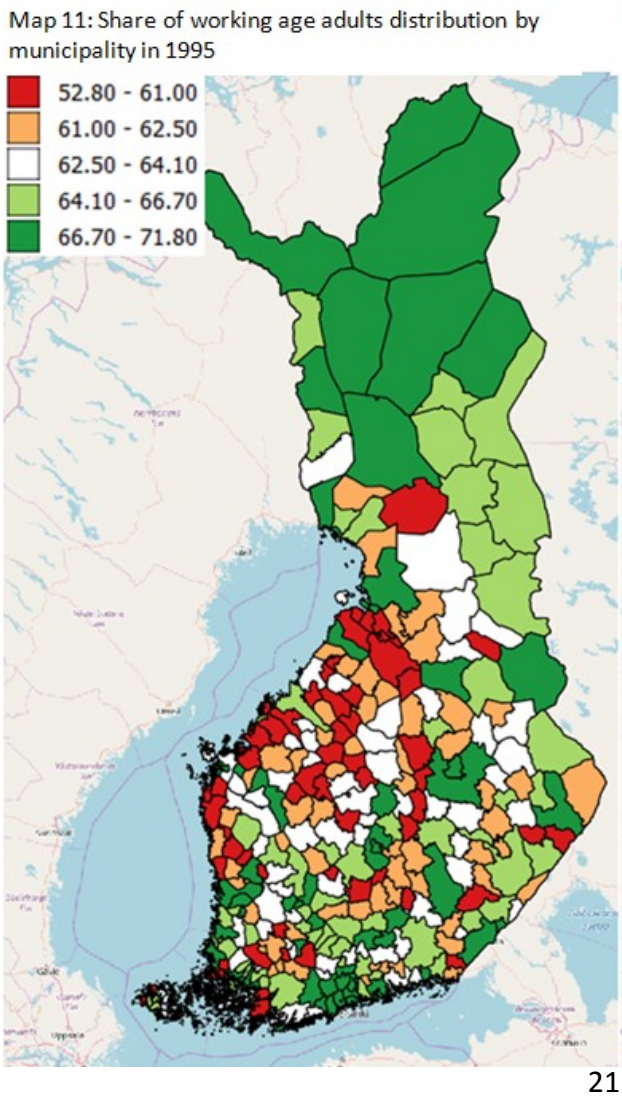
higher education rate by quintile in 1995 and 2015, respectively. This distribution has not dramatically changed in 20 years, with the central and eastern parts of Finland remaining relatively less educated.

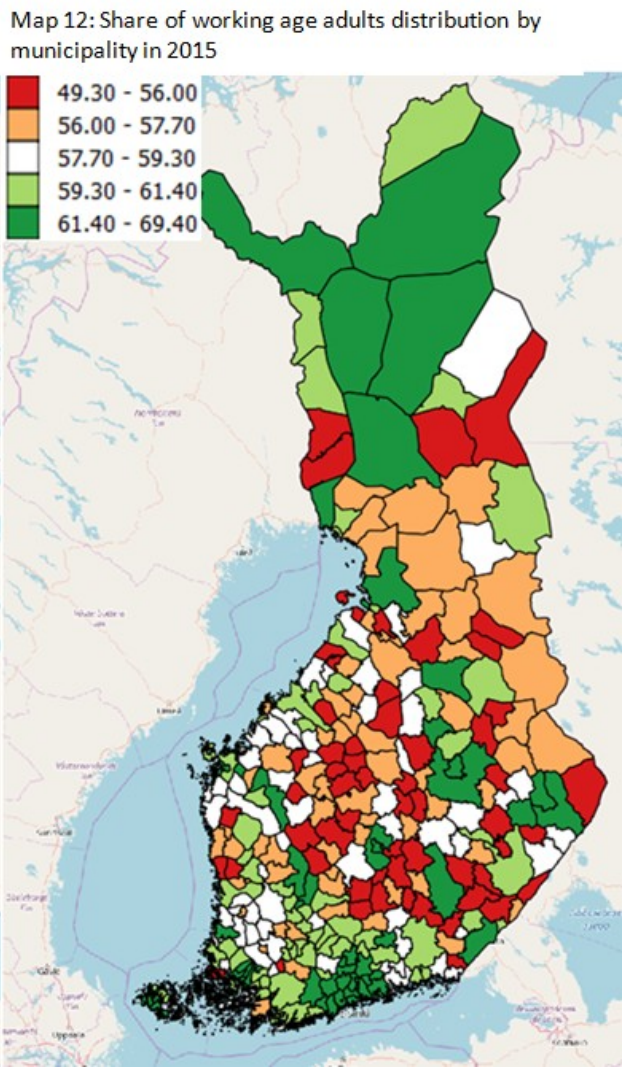
4.7 Share of Working Age Adults



Plot 6

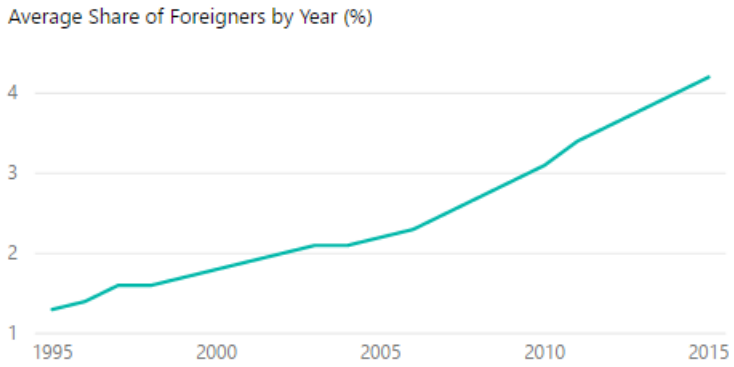
As shown in plot 6, the share of working age adults remained between 66 and 67 percent from 1995 to 2010, after which it began to quickly plummet. Maps 11 and





12 illustrate the share of working age adults distribution by quintile between municipalities in 1995 and 2015, respectively. While Finland as a whole has begun to suffer from an aging of the populace, this aging has hit relatively harder in many parts of Eastern Finland.

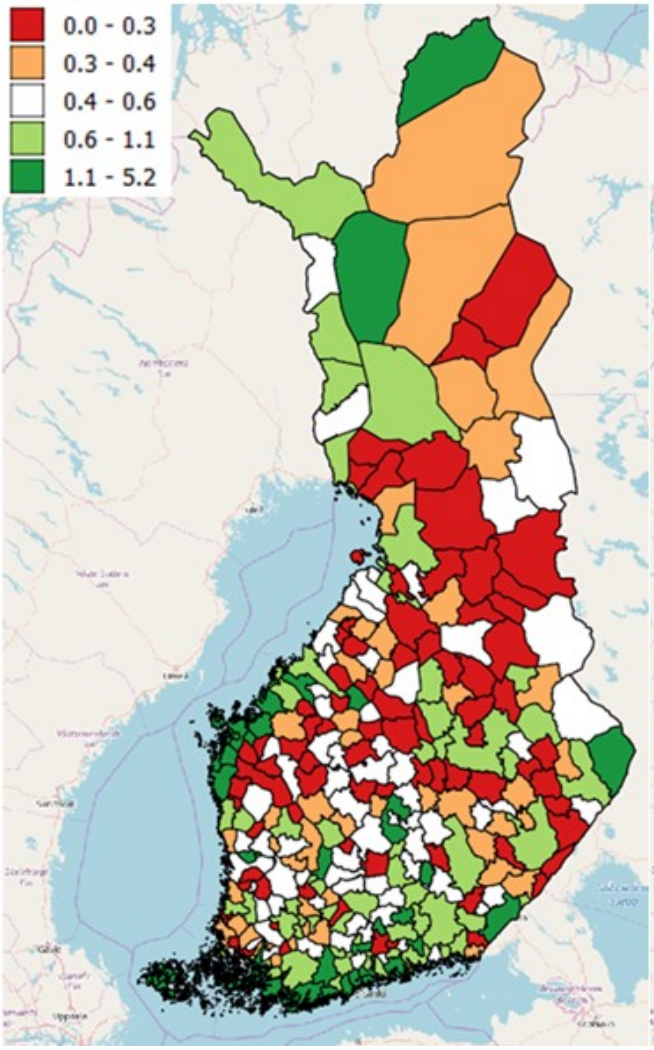
4.8 Share of Foreigners



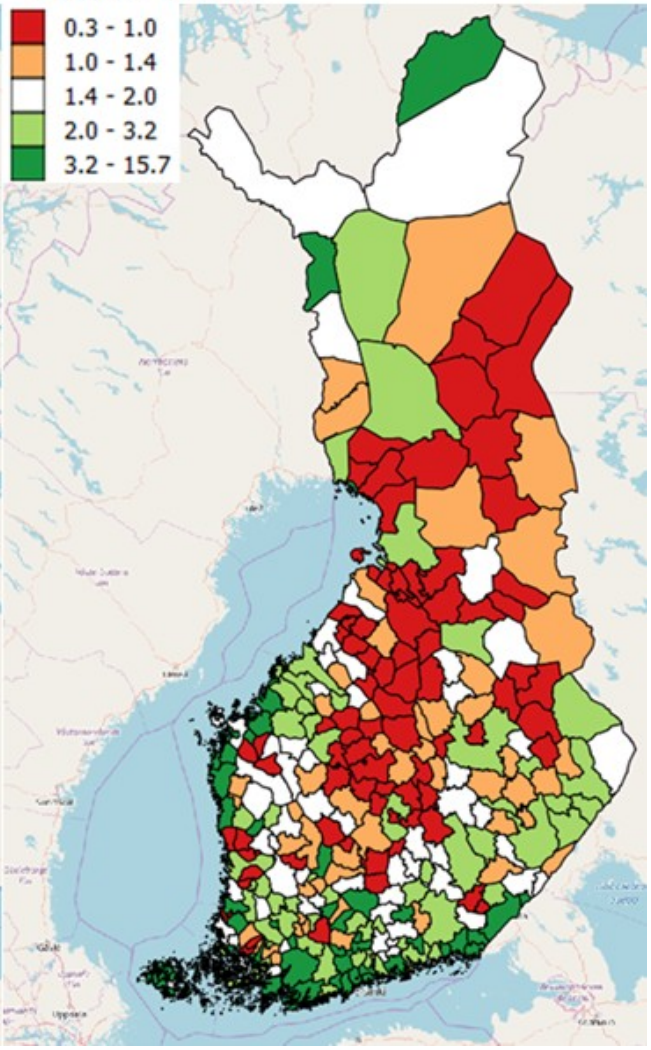
Plot 7

As shown in plot 7, the share of foreigners of the Finnish population has experienced a fourfold increase from close to one percent in 1995 to four percent in 2015. Maps 13 and 14 illustrate the foreigners share distribution by quintile between municipalities in 1995 and 2015, respectively. At both times foreigners have been relatively concentrated in municipalities close to Finnish borders, although this phenomenon appears a bit more pronounced in 2015.

Map 13: Share of foreigners distribution by municipality in 1995



Map 14: Share of foreigners distribution by municipality in 2015



5. Empirical methodology

The empirical strategy I use in this thesis consists of two parts. In the first part, I aim to find municipal characteristics associated with net migration and whether these associations have changed between the time periods of 1995-2005 and 2006-2015. After performing the analysis on net migration, I look at in-migration and out-migration separately, to better understand which characteristics act as push factors and which as pull factors. I also run the analysis separately for both women and men to see whether there are any differences in associations between the sexes.

I use the ratio between in-migration and out-migration as a proxy for net migration. The reasons for this are two-fold. Firstly, in-to-out-migration ratio allows easy conversion with natural logarithm for simple interpretation of the results. Secondly, in-to-out-migration ratio accounts for the population of a municipality without being spuriously correlated with it. If the log-log associations of population between in-migration and out-migration are the same, we can expect the association between population and in-to-out-migration ratio to be zero. This is necessary, as I wish to accurately capture the association between net migration and population.

The model I use in the first part of the analysis is the following fixed effects log-log regression model:

$$lnM_{ij} = ln\beta_0 + \beta_1 lnY_{ij} + \beta_2 lnI_{ij} + \beta_3 lnU_{ij} + \beta_4 lnE_{ij} + \beta_5 lnF_{ij} + \beta_6 lnW_{ij} + \beta_7 lnP_{ij} + \sum_{i=1}^{311} \alpha_i X_i + \sum_{j=1}^{m} \varphi_j Z_j + e_{ij}$$

where:

M_{ij} = In the regression analysing net migration, this is the ratio of in- and out-migration in municipality *i* on year *j*. In the regression analysing inwards migration flows this is the total amount of internal migrants that moved to municipality *i* on year *j*. In the regression analysing outwards migration flows this is the total amount of internal migrants that moved from municipality *i* to some other municipality on year *j*.

Y_{ij} = The average usable income (€) of a housing unit per consumption unit in municipality *i* on year *j*.

I_{ij} = The municipal income tax in municipality *i* on year *j*.

U_{ij} = The unemployment rate in municipality *i* on year *j*.

E_{ij} = The percentage of the population in municipality *i* on year *j* with a higher education degree.

F_{ij} = The ratio of population with a foreign nationality to those with a Finnish one in municipality *i* on year *j*.

W_{ij} = The ratio of working age population to the total population of municipality *i* on year *j*.

P_{ij} = The population of municipality *i* on year *j*.

X_i = A dummy variable for municipality *i*.

Z_j = A dummy variable for year *j*.

I run the model twice, first as presented, and then with a decade interaction variable, which takes the value of 0 for 1995-2005 and 1 for 2005-2015, interacted with all the independent variables. In the first run I find the coefficients for the whole time period. In the second I find the coefficients for the period of 1995-2005, as well as whether these coefficients have significantly changed over time.

In the second part of the empirical strategy, I use a modified gravity model to analyse municipality-to-municipality migration flows between individual municipalities in Finland. It is the point of this part to determine how distance is associated with migration flows, and whether this association has significantly changed over time. The model I use in this part of the analysis also takes the form of a fixed effects log-log regression:

$$lnM_{iyj} = \beta_0 + \beta_1 lnD_{iy} + \sum_{n=1}^{7} \theta_n X_{in} + \sum_{n=1}^{7} \nu_n X_{jn} + \sum_{n=1}^{311} \alpha_n O_n + \sum_{n=1}^{311} \gamma_n N_n + \sum_{n=1995}^{2015} \varphi_n Z_n + e_{iyj}$$

where:

M_{iyj} = The amount of people to move from municipality *i* to municipality *y* on year *j*.

D_{iy} = The distance in kilometres between municipalities *i* and *y*.

X_i = A vector consisting of the municipality characteristics employed in part one of the empirical analysis for municipality *i*.

X_j = A vector consisting of the municipality characteristics employed in part one of the empirical analysis for municipality *j*.

O_n = A dummy variable for the origin municipality *n*

N_n = A dummy variable for the destination municipality n.

Z_n = A dummy variable for the year n.

As in the previous part, I run the model with and without a decade interaction variable. Additionally, I run the model with and without municipality characteristics as control variables to see whether these have any impact on the associations found between migration and distance. Finally, I perform the same analysis with zero-migration municipality pairs removed from the data for the purposes of more relevant comparison with previous migration literature.

6. Results

In this section I present the results of the study. I start by looking at the results on the in-to-out-migration ratio, and then move on to in- and out-migration separately. Finally, I present the results of the second part of the empirical strategy to see how intermunicipal distance is associated with municipality-to-municipality migration.

6.1 In-to-Out-Migration Ratio

The tables presenting the regression results consist of the following parts: The left-most part (1995-2015) shows the results for the whole time period. The parts in the middle and the right (1995-2005 and 1996-2016) show the results for the first time period of 1995-2005, and the change to the coefficients from the first time period to the second. The results on in-to-out-migration ratio are show in table 1.

In-Migration / Out-Migration												
Predictors	1995-2015				1995-2005				1996-2015 (difference to 1995-2005)			
	Estimate	*Std. Error*	*p-value*		*Estimate*	*Std. Error*	*p-value*		*Estimate*	*Std. Error*	*p-value*	
Average household income per consumption unit	-0.16	0.09	0.0700	.	-0.17	0.10	0.0969	.	0.10	0.10	0.3289	
Higher education rate	**0.37**	0.07	<0.001	***	**0.47**	0.10	<0.001	***	**-0.15**	0.04	<0.001	***
Municipal tax rate	**-0.53**	0.11	<0.001	***	**-0.50**	0.14	<0.001	***	-0.22	0.15	0.1469	
Share of foreigners	**-0.06**	0.02	<0.001	***	-0.03	0.02	0.1709		-0.02	0.02	0.1971	
Share of working age adults	**-0.48**	0.14	<0.001	***	**-0.64**	0.17	<0.001	***	**0.82**	0.16	<0.001	***
Unemployment rate	**-0.18**	0.02	<0.001	***	**-0.19**	0.02	<0.001	***	**0.08**	0.02	<0.001	***
Population	**-0.28**	0.04	<0.001	***	**-0.19**	0.05	<0.001	***	0.00	0.01	0.6538	
R-Squared / Adj. R-Squared	0.04957 / -0.00279				0.06487 / 0.01222							

Table 1

Like all the other models used in this paper, the model I use in this part of the study is log-log in nature, and thus the coefficients should be interpreted roughly as percent changes in in-to-out-migration ratio for one percent changes in the predictor variables. The overall ability of the model to predict in-to-out migration ratio seems very poor. The r-squared value of the model without the interaction variable for decade is around five percent, and the adjuster r-squared value is less than zero. The model with the interaction variable fairs only slightly better, with an r-squared value of around 6.5 percent and an adjusted r-squared value of 1.2 percent.

The predictor with the highest association with in-to-out-migration ratio is the municipal tax rate. For the whole time period, a one percent change in the municipal tax rate is associated with a 0.53 percent decrease in the in-to-out-migration ratio. Looking at the two time periods separately shows there is no significant change in this association from 1995-2005 to 2006-2015.

The predictor with the second highest association with in-to-out-migration ratio is the share of working age adults. A one percent increase in this predictor is associated with a 0.48 percent decrease in the dependent variable for the whole time period. Looking at the two time periods separately shows that this association has been even more strongly negative from 1995 to 2005 (-0.64) but has actually reversed and turned slightly positive (0.18) from 2006 to 2015. This indicates the relative strengths of the associations between the share of working age adults and in- and out-migration have changed from the first time period to the next. This is confirmed in the next two sections of this chapter.

According to the results, a one percent increase in the higher education rate is associated with a 0.37 percent increase in in-to-out-migration ratio for the whole time period. When looking at the two time periods separately, we can see that this association was stronger during the first time period (0.47) and grew weaker during the second (0.32). Perhaps the overall rise in higher education rate for the whole country as well as the slightly diminished difference in education between the highest and lowest quintiles of municipalities has over time caused average educational attainment in a region to play less of a role in migration decisions.

According to the results, a one percent increase in population is associated with a 0.28 percent decrease in in-to-out-migration ratio. Looking at the two time periods separately shows that the strength of this association has not significantly changed over time. The fact that this coefficient is negative tells something about the relative strengths of the associations between population and in- and out-migration. It is somewhat of a given that larger population means larger out-migration, as the population acts as a pool of potential migrants. Whether the association between population and in-to-out-migration ratio is positive or negative depends on whether migrants are drawn to population centres to a sufficient degree to counter-act this. The negative coefficient on population here indicates this is not the case, which is also confirmed in the next two sections of this chapter.

A one percent increase in the unemployment rate is associated with a 0.18 percent decrease in in-to-out-migration ratio for the whole time period. Looking at the two periods separately shows the association was stronger during from 1995 to 2005 (-0.19), and grew

weaker from 2006 to 2015 (-0.11). A possible reason for this change could be simply that unemployment was significantly higher in Finland during the first time period than the second. Perhaps as unemployment has come down it has also become less of a factor in migratory decision making.

The final predictor associated with in-to-out-migration share is the share of foreigners. For the whole time period, a one percent increase in the share of foreigners is associated with a 0.06 percent decrease in the dependent variable. Looking at the two time periods separately shows that the association is not significant for the first period. Neither is the change towards the second period significant. This is probably due to the smaller sample size available when calculating the coefficients for each period separately instead of for the whole time period at once. Anyway, the association appears to have grown slightly stronger from 1995-2005 to 2006-2015. This is likely due to the fact that the number of foreign citizens has greatly increased in Finland from the first period to the second.

It should be mentioned that average household income appears to have a very marginally negative association with in-to-out-migration ratio. This could be due to collective bargaining being very prominent in many fields of work in Finland, and I'll expand upon this hypothesis in the concluding section.

6.2 In-Migration

Next I'll look at the results analysing municipal in-migration, to get a better understanding of which municipal characteristics are particularly associated with a municipality's ability to pull migrants in. The results of the regressions for in-migration can be seen in tables 2, 3 and 4. Table 2 shows the results for all in-migrants combined, whereas tables 3 and 4 show the results for men and women, respectively. As with the in-to-out-migration ratio, the model used in this part of the analysis is log-log in nature, and the results should thus be interpreted as percent changes in in-migration per one percent change in the predictors.

In-Migration: Both Sexes

Predictors	1995-2015				1995-2005				1996-2015 (difference to 1995-2005)			
	Estimate	*Std. Error*	*p-value*		*Estimate*	*Std. Error*	*p-value*		*Estimate*	*Std. Error*	*p-value*	
Average household income per consumption unit	-0.08	0.07	0.2555		-0.02	0.08	0.8391		-0.05	0.08	0.5170	
Higher education rate	**0.41**	0.05	<0.001	***	**0.43**	0.06	<0.001	***	**-0.11**	0.03	<0.001	***
Municipal tax rate	**-0.41**	0.09	<0.001	***	**-0.38**	0.12	<0.001	***	0.03	0.12	0.8343	
Share of foreigners	**0.03**	0.01	0.0241	*	0.00	0.02	0.8538		**0.04**	0.01	0.0058	**
Share of working age adults	**1.18**	0.11	<0.001	***	**1.01**	0.13	<0.001	***	**0.28**	0.13	0.0306	*
Unemployment rate	**-0.04**	0.01	0.0135	*	**-0.05**	0.02	0.0049	**	**0.03**	0.02	0.0431	*
Population	**0.53**	0.03	<0.001	***	**0.68**	0.04	<0.001	***	0.00	0.01	0.5995	
R-Squared / Adj. R-Squared	0.10412 / 0.05479				0.11223 / 0.06226							

Table 2

In-Migration: Men

Predictors	1995-2015				1995-2005				1996-2015 (difference to 1995-2005)			
	Estimate	*Std. Error*	*p-value*		*Estimate*	*Std. Error*	*p-value*		*Estimate*	*Std. Error*	*p-value*	
Average household income per consumption unit	-0.10	0.08	0.1967		0.01	0.09	0.9531		-0.14	0.09	0.1194	
Higher education rate	**0.40**	0.06	<0.001	***	**0.39**	0.07	<0.001	***	-0.05	0.04	0.1190	
Municipal tax rate	**-0.47**	0.10	<0.001	***	**-0.46**	0.13	<0.001	***	0.11	0.14	0.4350	
Share of foreigners	**0.05**	0.02	0.0013	**	0.00	0.02	0.9452		**0.06**	0.01	<0.001	***
Share of working age adults	**1.45**	0.13	<0.001	***	**1.36**	0.15	<0.001	***	-0.01	0.14	0.9549	
Unemployment rate	-0.01	0.02	0.5862		-0.02	0.02	0.4025		0.02	0.02	0.3881	
Population	**0.46**	0.03	<0.001	***	**0.59**	0.04	<0.001	***	0.00	0.01	0.8708	
R-Squared / Adj. R-Squared	0.08345 / 0.03298				0.08900 / 0.03772							

Table 3

In-Migration: Women

Predictors	1995-2015				1995-2005				1996-2015 (difference to 1995-2005)			
	Estimate	*Std. Error*	*p-value*		*Estimate*	*Std. Error*	*p-value*		*Estimate*	*Std. Error*	*p-value*	
Average household income per consumption unit	-0.09	0.08	0.2382		-0.05	0.09	0.5740		0.00	0.09	0.9570	
Higher education rate	**0.44**	0.06	<0.001	***	**0.50**	0.06	<0.001	***	**-0.17**	0.03	<0.001	***
Municipal tax rate	**-0.34**	0.09	<0.001	***	**-0.29**	0.13	0.0200	*	-0.05	0.13	0.7296	
Share of foreigners	0.00	0.01	0.8044		-0.01	0.02	0.5981		0.02	0.01	0.1733	
Share of working age adults	**0.87**	0.12	<0.001	***	**0.56**	0.15	<0.001	***	**0.61**	0.14	<0.001	***
Unemployment rate	**-0.06**	0.02	<0.001	***	**-0.07**	0.02	<0.001	***	**0.04**	0.02	0.0368	*
Population	**0.59**	0.03	<0.001	***	**0.75**	0.04	<0.001	***	0.00	0.01	0.7098	
R-Squared / Adj. R-Squared	0.09306 / 0.04311				0.10420 / 0.05378							

Table 4

First off, the ability of the predictors to predict in-migration is a lot better than their ability to predict in-to-out-migration ratio. The model without the interaction variable for decade reaches an r-squared value of 10.4 percent and an adjusted r-squared value of 5.5 percent. The model with the interaction variable once again fares a bit better, with an r-

squared value of 11.2 percent and an adjusted r-squared value of 6.2 percent. The predictive capability of the model with the interaction variable in regard to in-migration is, while not huge, roughly 5 times that of its predictive capability in regard to in-to-out-migration ratio. The predictive ability also appears to be higher for women than men.

The variable with the strongest association with in-migration is the share of working age adults. A one percent increase in the share of working age adults is associated with a 1.18 percent increase in in-migration for the whole time period. This association appears to be particularly strong for men (1.45), as opposed to women (0.87). However, looking at the two time periods separately shows that this is only true from 1995 to 2005. During the second time period, the association between the share of working age adults and in-migration more than doubled for women and rose to a level only slightly below that of men.

As shown in the previous section, the share of working age adults is negatively associated with in-to-out-migration ratio. For these results to be consistent, the share of working age adults must be even more strongly associated with out-migration. Indeed, the association with in-migration appears to have been weaker from 1995 to 2005 (1.01), after which it grew by 0.28 to 1.29. This is consistent with the finding that the negative association between the share of working age adults and in-to-out-migration ratio turned slightly positive from 2006 to 2015.

The predictor with the second highest association with in-migration is population. This comes as no surprise really, since migration research already described people moving to large population centres over a hundred years ago. A one percent increase in population is associated with a 0.53 percent increase in in-migration. The strength of this association hasn't significantly changed from the first time period to the second and is slightly stronger for women than men.

A one percent increase in higher education rate is associated with a 0.41 percent increase in in-migration for the whole time period. Looking at the two time periods separately shows that the association was slightly stronger during the first time period and suffered a drop from 0.43 to 0.32 during the second. Looking at the two genders separately shows this drop was caused by the association between female in-migration and higher education rate growing smaller, whereas the association for men suffered no such change. While the association was larger for women (0.50) than men (0.39) from 1995 to 2005, there was no remarkable difference between the sexes from 2006 to 2015.

A one percent increase in the municipal tax rate is associated with a 0.41 percent decrease in in-migration for the whole time period. This association was slightly larger for

men (-0.46) than women (-0.29) during the first time period, and neither saw any significant change from the first period to the second.

The predictor with the second lowest but significant association with in-migration is the unemployment rate. A one percent increase in the unemployment rate is associated with a 0.04 percent decrease in in-migration for the whole time period. Looking at the two time periods separately shows that this association is in effect during the first time period, but becomes too small during the second to be significantly different from zero. As discussed in the section on in-to-out-migration ratio, this is likely due to the unemployment rate being considerably higher from 1995 to 2005. It's also worth nothing that this association is only significant for women.

The predictor with the smallest association coefficient with in-migration is the share of foreigners. A one percent increase in the share of foreigners is associated with a 0.03 percent increase in in-migration for the whole time period. Looking at the two time periods separately shows that the association only came into effect from 2006 to 2015. As discussed in the previous chapter, this can probably be explained by the greatly increased number of foreigners in Finland during the second time period. The size of the association is very small, and it is likely there was simply not enough foreigners during the first period for their share of the population to show as having a significant association with in-migration.

6.3 Out-Migration

Next, I'll look at the results analysing municipal out-migration, to get a better understanding of which municipal utility factors are particularly associated with the population of a municipality moving away. The results of the regressions for out-migration can be seen in tables 5, 6 and 7. Table 5 shows the results for all in-migrants, whereas tables 6 and 7 show the results for men and women, respectively. As with in-migration and in-to-out-migration ratio, the model used in this part of the analysis is log-log in nature, and the results should thus be interpreted roughly as percent changes in out-migration per one percent change in the predictors.

Out-Migration: Both Sexes

Predictors	1995-2015				1995-2005				1996-2015 (difference to 1995-2005)			
	Estimate	*Std. Error*	*p-value*		*Estimate*	*Std. Error*	*p-value*		*Estimate*	*Std. Error*	*p-value*	
Average household income per consumption unit	0.06	0.06	0.3353		0.13	0.07	0.0680	.	-0.13	0.07	0.0637	.
Higher education rate	0.07	0.05	0.1106		0.01	0.05	0.8982		0.01	0.03	0.6080	
Municipal tax rate	0.14	0.08	0.0637	.	0.14	0.10	0.1800		**0.25**	0.11	0.0238	*
Share of foreigners	**0.09**	0.01	<0.001	***	0.03	0.02	0.0732	.	**0.06**	0.01	<0.001	***
Share of working age adults	**1.60**	0.10	<0.001	***	**1.56**	0.12	<0.001	***	**-0.49**	0.11	<0.001	***
Unemployment rate	**0.14**	0.01	<0.001	***	**0.15**	0.02	<0.001	***	**-0.04**	0.01	0.0014	**
Population	**0.81**	0.03	<0.001	***	**0.88**	0.04	<0.001	***	0.00	0.00	0.5627	
R-Squared / Adj. R-Squared	0.23915 / 0.19725				0.25168 / 0.20956							

Table 5

Out-Migration: Men

Predictors	1995-2015				1995-2005				1996-2015 (difference to 1995-2005)			
	Estimate	*Std. Error*	*p-value*		*Estimate*	*Std. Error*	*p-value*		*Estimate*	*Std. Error*	*p-value*	
Average household income per consumption unit	0.02	0.07	0.736		0.11	0.08	0.1776		-0.15	0.08	0.0639	.
Higher education rate	**0.14**	0.05	0.009	**	0.07	0.06	0.2541		0.03	0.03	0.3129	
Municipal tax rate	0.16	0.09	0.068		0.17	0.11	0.1293		**0.25**	0.12	0.0401	*
Share of foreigners	**0.11**	0.01	<0.001	***	0.03	0.02	0.1050		**0.08**	0.01	<0.001	***
Share of working age adults	**1.63**	0.11	<0.001	***	**1.65**	0.13	<0.001	***	**-0.66**	0.13	<0.001	***
Unemployment rate	**0.12**	0.01	<0.001	***	**0.12**	0.02	<0.001	***	**-0.04**	0.02	0.0043	**
Population	**0.74**	0.03	<0.001	***	**0.80**	0.04	<0.001	***	0.00	0.01	0.8721	
R-Squared / Adj. R-Squared	0.18208 / 0.13704				0.19657 / 0.15135							

Table 6

Out-Migration: Women

Predictors	1995-2015				1995-2005				1996-2015 (difference to 1995-2005)			
	Estimate	*Std. Error*	*p-value*		*Estimate*	*Std. Error*	*p-value*		*Estimate*	*Std. Error*	*p-value*	
Average household income per consumption unit	0.09	0.07	0.1996		0.16	0.08	0.0607	.	-0.12	0.08	0.1604	
Higher education rate	0.03	0.05	0.6290		-0.02	0.06	0.7749		-0.03	0.03	0.4201	
Municipal tax rate	0.17	0.09	0.0589	.	0.14	0.12	0.2471		0.24	0.13	0.0611	.
Share of foreigners	**0.07**	0.01	<0.001	***	0.02	0.02	0.2145		**0.04**	0.01	0.0038	**
Share of working age adults	**1.42**	0.12	<0.001	***	**1.30**	0.14	<0.001	***	-0.25	0.13	0.0595	.
Unemployment rate	**0.16**	0.01	<0.001	***	**0.18**	0.02	<0.001	***	**-0.05**	0.02	0.0035	**
Population	**0.89**	0.03	<0.001	***	**0.97**	0.04	<0.001	***	0.00	0.01	0.8362	
R-Squared / Adj. R-Squared	0.20456 / 0.16076				0.21118 / 0.16678							

Table 7

The ability of the model to predict out-migration is significantly better than its ability to predict in-to-out-migration ratio or in-migration. The model without the period interaction variable reaches an r-squared value of 23.9 percent and an adjusted r-squared value of 19.7 percent. The model with the period interaction variable reaches an r-squared value of 25.2

percent and an adjusted r-squared value of 21.0 percent. The predictors in the model can explain a whole fifth of the variation in out-migration between municipalities. However, when the model without the interaction variable is run without population as a predictor, the r-squared value drops to 11.4 percent and the adjuster r-squared value to 6.5 percent. The model's predictive ability regarding out-migration is still slightly higher than regarding in-migration, but nowhere near as high as with the population variable included. This makes sense as since the population acts as a pool of potential out-migrants, naturally an increase in population is good at predicting an increase in out-migration.

The predictor with the highest association with out-migration is the share of working age adults. A one percent increase in the share of working age adults is associated with a 1.6 percent increase in out-migration. This association is larger than the one between this predictor and in-migration. Looking at the two time period separately also shows that the association has grown weaker in the second time period compared to the first. These results are consistent with the observations made about the association between the share of working age adults and in-to-out-migration ratio, for which the negative association is strong during the first time period when the association between the share of working age adults and out-migration is stronger than the one between the share of working age adults and in-migration. In the second time period the association grows stronger for in-migration and weaker for out-migration, and thus the negative association with in-to-out-migration disappears.

A possible narrative to explain this observation could simply be the following: from 1995 to 2005 there were many working age adults compared to 2006-2015 in smaller municipalities with weaker prospects. These adults were moving into municipalities with better prospects, and naturally municipalities with more potential migrants experienced more out-migration. From 2006 to 2015 a lot of the potential migration had already happened, and there were now relatively more working age adults in municipalities with better prospects than those with weaker ones compared to the earlier time period. However, since migration still occurred mostly from municipalities with weak prospects to municipalities with good ones, the association with in-migration had increased and the association with out-migration decreased.

What's also interesting is that while the change in the association between the share of working age adults and in-migration from the first time period to the second was significant only for women, the change in the association with this predictor and out-migration is significant only for men. So over time, the share of working age adults has become a stronger predictor for women moving in and a weaker predictor for men moving out.

The predictor with the second highest association with out-migration is population. As already discussed, it makes sense for a larger pool of potential migrants to be associated with the number of migrants. A one percent increase in population is associated with a 0.81 percent increase in out-migration for the whole time period. There seems to be no significant difference in the size of this association between 1995-2005 and 2006-2015. The most interesting fact about this coefficient is that it is less than one. This indicates that while higher population is associated with higher out-migration, the share of population to migrate diminishes as the size of the population grows larger.

The third predictor associated with out-migration is the unemployment rate. A one percent increase in the unemployment rate is associated with a 0.14 percent increase in out-migration. Looking at the two time periods separately shows that this association is higher from 1995 to 2005 (0.15) and diminishes slightly during the second time period. This is similar to what happened with unemployment and in-migration, except this time the coefficient remains large enough during the second period to still be significantly different from zero. The fact that unemployment's association with both in-migration and out-migration grows weaker in the second time period supports the theory that as unemployment was lower from 2006 to 2015 compared to 1995 to 2005, it became less relevant in migratory decision making. It is also noteworthy that as with in-migration, the association between unemployment and out-migration is larger for women than it is for men.

The predictor with the lowest yet still significant association with out-migration is the share of foreigners. A one percent increase in the share of foreigners is associated with a 0.09 percent increase in out-migration for the whole time period. Looking at the two time periods separately shows that this association is very weak during the first time period (0.03) and only marginally significant, but grows larger and more clearly significant during the second time period (0.09). This is probably due to there simply being very few foreigners in Finland during the first time period compared to the second. It is also noteworthy that the association coefficient is larger between the share of foreigners and out-migration than it is between the share of foreigners and in-migration. This explains the negative association between the share of foreigners and the in-to-out-migration ratio.

Municipal tax rate has no significant association with out-migration for the whole time period of 1995-2015, but when looking at the two time periods separately, we can see that the association turns significant during the second time period of 2006-2015. During this period, a one percent increase in municipal tax rate is associated with a 0.39 percent increase in out-migration.

Finally, it is of note that the higher education rate appears to be positively associated with out-migration for men from 1995 to 2015. This association does not show for women or both sexes combined. A one percent increase in the higher education rate is associated with a 0.14 percent increase in out-migration. Looking at the two time periods separately shows that this association is too small during the first time period to be significantly different from zero. The association grows during the second time period, but the change is also too small to reach significance. This is probably due to the sample size available for each time period getting smaller as we move from a one period of 21 years to two periods of around 10 years. Nevertheless, the association hints that highly educated men might be on average more migratory than women and their less educated counterparts.

6.4 Municipality-to-Municipality Migration and Distance

In the last section of this chapter I'll look at the results of the modified gravity model used to ascertain how distance is associated with migration between two municipalities. I ran two separate specifications of the model, one with all the observations and one with observations of zero municipality-to-municipality migration removed from the data. As with the models looking at in- and out-migration, the models used in this section are log-log in nature and thus the results should be interpreted roughly as percent changes in the dependent variable per one percent increase in the distance between municipalities. The results are presented in table 8.

Municipality to Municipality Migration

Predictors	**1995-2015**				**1995-2005**				**1996-2015 (difference to 1995-2005)**			
	Estimate	*Std. Error*	*p-value*		*Estimate*	*Std. Error*	*p-value*		*Estimate*	*Std. Error*	*p-value*	
With all observations												
Distance												
Without Controls	**-0.5060**	0.0006	<0.001	***	**-0.5085**	0.0006	<0.001	***	**0.0353**	0.0006	<0.001	***
R-Squared / Adj. R-Squared	0.2395 / 0.2393				0.2410 / 0.2408							
With Controls	**-0.5018**	0.0006	<0.001	***	**-0.5044**	0.0006	<0.001	***	**0.0358**	0.0006	<0.001	***
R-Squared / Adj. R-Squared	0.2372 / 0.2370				0.2387 / 0.2385							
With zero-observations removed												
Distance												
Without Controls	**-0.7888**	0.0014	<0.001	***	**-0.7888**	0.0014	<0.001	***	**0.0234**	0.0015	<0.001	***
R-Squared / Adj. R-Squared	0.4311 / 0.4302				0.4314 / 0.4305							
With Controls	**-0.7898**	0.0014	<0.001	***	**-0.7899**	0.0014	<0.001	***	**0.0242**	0.0015	<0.001	***
R-Squared / Adj. R-Squared	0.4329 / 0.4319				0.4332 / 0.4323							

Table 8

First off, looking at the r-squared and adjusted r-squared values shows that distance alone is a very strong predictor of migration flows. Almost a quarter of the variation in

municipality-to-municipality migration flows is accounted for by distance. When the observations with zero migration are left out, this number climbs to over 40 percent. A one percent increase in the distance between municipalities is associated with an approximately 0.506 percent decrease in migration between those municipalities. This negative association is hardly surprising. What is somewhat surprising is how consistent the size of this association stays with and without the control variables or the period interaction variable in the mix. Adding the control variables reduces the association by 0.042 for the whole time period.

Looking at the two time periods separately show that the association has grown very slightly but still significantly weaker over time. This is as expected, as travel time between different municipalities has grown shorter from the first period to the second due to, for example, improved road networks. It might be interesting to run the same analysis using the actual travel time between municipalities as a proxy for distance instead of straight line distance to see whether this changes things.

The associations I find between distance and municipality-to-municipality migration are somewhat smaller than have been found in previous studies on the subject. As mentioned in chapter 3, Poot et al. (2016) found a one percent increase in straight line distance to be associated with a 0.62 – 0.75 percent decrease in migration flows over multiple periods of five years. A likely reason for the stronger estimates gained in their study is the five-year time periods of scrutiny used instead of year-to-year migration. The longer time period ensures there are very little zero-migration pairs present. In this thesis, I analyze migration flows on a yearly basis, which leads to many smaller and more remote municipality-pairs not having any migration happen between them. The large amount of non-migration lowers the estimates gained for the predictor coefficients. Indeed, as can be seen in table 8, when all the observations with zero migration are removed from the data, the coefficient jumps to approximately 0.79, which is close to the upper estimates gained by Poot et al.

7. Conclusions

My aim in this thesis was to determine to what extent it is possible to predict municipal net migration, in-migration and out-migration using municipality characteristics. The characteristics looked at were average household income per consumption unit, higher education rate, municipal tax rate, the share of foreigners, the share of working age adults, unemployment rate and population. My results show these variables are quite insufficient to predict net migration, as the model utilized could explain only approximately 1.2 percent of variation in the in-to-out-migration ratio between municipalities. The variables fared somewhat better at predicting in- and out-migration, with the model explaining roughly 6 percent of the variation in in-migration and roughly 21 percent of the variation in out-migration. The variable explaining most of the variation in out-migration was population, as removing this variable from the model lowered its predictive ability regarding out-migration to a level similar to that of in-migration.

I found average household income to have a very marginally significant negative association with net migration. This was somewhat surprising, as the variable is strongly positively correlated with the in-to-out-migration ratio. Likely income is strongly correlated with some other variable that in turn has a positive association with net migration, such as high education rate. One possible explanation for the negative association between average household income and net migration could be the collective bargaining system affecting numerous areas of business in Finland. If wages are agreed upon collectively, a worker can't expect to gain a better salary by moving to another region, unless they are willing to change occupations as well. As the costs of living are higher in areas with high average income, it makes sense for high average household income to be negatively associated with net migration.

I found high education rate to be positively associated with the in-to-out-migration ratio. This association exists due to high education rate being positively associated with in-migration. There was no association between the variable and out-migration. The association of this variable with in-migration has also grown smaller over the time, for women in particular. Perhaps the overall increase in education level in Finland and the reduced gap in average education between the most and least educated municipalities has caused education to become less important for migratory decision making.

I found municipal tax rate to be negatively associated with the in-to-out-migration ratio. The association was the highest of all the predictors in the model, which is not that surprising, as municipal tax rate is strongly negatively associated with in-migration for the whole time period, and positively associated with out-migration from 2006 to 2015. The association with in-migration is also somewhat stronger for men than for women. There was not much difference between sizes of the coefficients for men and women when it comes to out-migration, but the significance of the coefficients turned out to be larger for men.

I found the share of foreigners to be positively associated with both in- and-out-migration for the whole time period. Looking at the two time periods separately showed that these association only became significant from 2006 to 2015. This is probably due to there simply being a lot more foreigners in Finland during that time than the earlier time period of 1995-2005. The association between the share of foreigners and out-migration turned out to be bigger than the one between share of foreigners and in-migration, and so the variable has a negative association with the in-to-out-migration ratio.

I found the share of working age adults to be strongly positively associated with both in- and out-migration for the whole time period. The association with out-migration was stronger however, which is why the variable was negatively associated with the in-to-out-migration ratio. Looking at the two time periods separately showed that over time the share of working age adults became more strongly associated with in-migration for women and less strongly associated with out-migration for men. This led to the association between the share of working age adults and the in-to-out-migration ratio turning positive during the second time period of 2006-2015.

I found the unemployment rate to be negatively associated with in-migration and positively associated with out-migration, and thus naturally it was negatively associated with the in-to-out-migration ratio. The associations were stronger for women than for men, for whom the association with in-migration was not significant at all. The associations also turned out to grow weaker from the first time period to the second. This could be due to the overall level of unemployment reducing in Finland, which might have led to unemployment becoming less of an issue in migration decisions.

I found population to be highly positively associated with both in- and out-migration. The association with out-migration was stronger however, which is why population was negatively associated with the in-to-out-migration ratio. This feels somewhat counterintuitive, as the municipalities drawing the most migrants in Finland, like Helsinki or Tampere, all have large populations. Still, what this finding indicates is that should we have

two municipalities similar in all respects, and should one of them gain an increase in population, the increase in the amount of in-migrants would not be enough to off-set the amount of new out-migrants brought on by the increase in the size of the potential migrant pool.

Finally, I found distance to be a significant determinant of migration between two municipalities, explaining almost a quarter of the variation. If zero- migration observations are left out of the data, this number climbs to over forty percent. This also increases the size of the coefficient to a level comparable to what has been found in some other migration studies, which is understandable since these studies often use five-year periods as the periods during which migration happens, and thus have very little if any zero-migration observations in their data.

To conclude, municipalities that experienced a lot of in-migration from 1996 to 2015 had on average higher levels of education, low municipal tax rates, low unemployment and high population. There was also a larger than normal share of foreigners and working age adults in these municipalities. Municipalities that experienced a lot of out-migration had on average higher unemployment. There was also a larger than normal share of foreigners and working age adults in these municipalities, likely due to these groups being more mobile than other demographic groups. From 1995 to 2005, this resulted in the share of working age adults being actually negatively associated with net migration. This has changed over time however, and in the future municipalities with a high share of working age adults can be expected to have high net migration. Likewise, it appears the role of unemployment and education in migratory decision-making has lessened over time. This is likely due to unemployment in general having lowered significantly and the average level of education having risen significantly over the last two decades in Finland.

8. References

Alonso, W. (1978), *A theory of movements*, in Systems approach to human settlements, edited by N. Hansen.

Alonso, W. (1986), *Systemic and log-linear models: From here to there, then to now, and this to that*, Discussion paper 86-10, Center for Population Studies, Harvard University

Davies S. et al (2001), *A Conditional Logit Approach to U.S. State-to-State Migration*, Journal of Regional Science, Vol. 41, No. 2, pp. 337-360

Greenwood M. (1969), *An Analysis of the Determinants of Geographic Mobility in the United States*, The Review of Economics and Statistics, Vol. 51, No. 2, pp. 189-194

Greenwood M. & Hunt G. (1984), *Econometrically accounting for identities and restrictions in models of interregional migration*, Regional Science and Urban Economics 14, pp.113-28.

Greenwood M. & Hunt G. (1984), *Migration and interregional employment redistribution in the United States*. American Economic Review 74, pp 957-969

Greenwood M. & Hunt G. (1985), *Econometrically accounting for identities and restrictions in models of interregional migration: Further thoughts*, Regional Science and Urban Economics 15, pp 605-614

Greenwood M. & Hunt G. (2003), *The Early History of Migration Research*, International Regional Science Review, Vol. 26, No. 1, pp. 3-37

Leisure, J. & Lewis, R. (1977), *Internal Migration in the USSR: 1897-1926*, Demography, Vol. 4, No. 2, pp. 479-496

Liang Z. & White, M (1996)., *Internal Migration in China, 1950-1988*, Demography, Vol. 33, No. 3, pp. 375-384

Liu Y. & Shen J. (2014), *Spacial patterns and determinants of skilled internal migration in China, 2000-2005*, Papers in Regional Science, Vol. 94, Iss. 4, pp. 749-771

Makower, J. et al (1938), *Studies in Mobility of Labour: A Tentative Statistical Measure*, Oxford Economic Papers I, pp. 83-123

McFadden, D. (1978), *Modelling the choice of residential location*, In Spatial interaction theory and planning models, edited by A. Karlquist, L. Lundquist, F. Snickars, and J. W. Weibull, Amsterdam

McFadden, D. (1981), *Econometric models of probabilistic choice*, In Structural analysis of discrete data with econometric applications, edited by C. F. Manski and D. McFadden , Cambridge

Piras, R. (2017, first published in 2016), *A long-run analysis of push and pull factors of internal migration in Italy. Estimation of a gravity model with human capital using homogenous and heterogenous approaches*, Papers in Regional Science, Vol. 96, Iss. 3, pp. 571-602

Poot J. et al (2016, *The Gravity Model of Migration: The Successful Comeback of an Ageing Superstar in Regional Science*, The Institute for the Study of Labor (IZA) Discussion Paper No. 10329

Ravenstein, E.G. (1885), *The laws of migration*, Journal of the Royal Statistical Society 48, pp 167-235.

Ravenstein, E.G. (1889), *The laws of migration*, Journal of the Royal Statistical Society 52, pp. 241-305

Richard A. Easterlin (1965), *Long Swings in U.S. Demographic and Economic Data*, Demography, II, p. 490-507

Schwartz, A. (1968), *Migration and Life Span Earnings in the U.S.* Unpublished Ph.D. dissertation, University of Chicago

Schultz, T. P. (1982), *Lifetime migration within educational strata in Venezuela: Estimates of a logistic model*, Economic Development and Cultural Change 30, pp. 559-593.

Thomas S. (1963), *Internal Migration in Sweden: A Recent Study*, Population Index, Vol. 29, No. 2, pp. 125-129

Young E.C. (1924), *The Movement of the Farm Population (Bulletin426)*, Ithaca: New York Agricultural Experiment Station