# Essays on Economic Determinants of Refugees' Neighborhood Integration

Madhinee Valeyatheepillay

München 2021

# Essays on Economic Determinants of Refugees' Neighborhood Integration

Inaugural-Dissertation

zur Erlangung des Grades

Doctor oeconomiae publicae (Dr. oec. publ.)

an der Ludwig-Maximilians-Universität München

2021

vorgelegt von

Madhinee Valeyatheepillay

| | |
|---|---|
| Referentin: | Prof. Panu Poutvaara, Ph.D. |
| Korreferent: | Prof. Dr. Matz Dahlberg |
| Promotionsabschlussberatung: | 14. Juli 2021 |

Datum der mündlichen Prüfung: 29. Juni 2021

Namen der Berichterstatter: Panu Poutvaara, Matz Dahlberg, Joachim Winter

# Acknowledgments

Acknowledgement and thanks goes to Panu Poutvaara, my supervisor, for letting me be part of his team at the ifo Center for International Institutional Comparisons and Migration Research and allowing me to explore a topic which is highly relevant to society. I am grateful for the invaluable comments Panu provided me on my thesis. I am most grateful to Matz Dahlberg for providing invaluable feedback, hosting me during my time at Uppsala University, providing stimulating discussions on our projects and the unwavering support throughout my doctoral programme. Thank you for being a great co-author who I thoroughly enjoy working with and learning from. I am grateful to Joachim Winter for being an inspiring professor and serving as a third advisor on this thesis.

I also thank my amazing and bright co-author, Cristina Bratu, who always offers helpful advice and put in hours of work to make the last chapter of this thesis possible. Thank you for all the uplifting words during this time. I learn so much from our collaboration. I would like to express my gratitude to Yvonne Giesing for inspiring me to know more on migration economics and refugee integration, and her feedback on this thesis. My appreciation extends to Martina Braun and Till Nikolka for the encouragement and the laughter. I am grateful to my colleagues, Britta Rude, Sarah Reiter, Christa Hainz, Tanja Stitteneder, Carla Rhode, Mirely Kollmannsberger and Clara Albrecht, who have all contributed to supporting me at ifo over the years and providing feedback. I thank the administrative and printing staff at ifo for their support. My special thanks go to Dagmar Bach, my "Bürofreundin", who always encourages me and for caring to put snacks on my desk to keep me healthy. Thank you to Audrey Au Yong Lyn for the office visits, and research as well as career talks. I am grateful for the feedback provided by participants in conferences and seminars.

I express my gratitude to DAAD for making my research stay at Uppsala University possible. I would like to thank all the people at the Institute for Urban and Housing (IBF) at Uppsala University, particularly Che Yuan-Liang for the Stata advice and the laughter, Fredrik Hansson for the econometrics and fun discussions, Henrik Andersson for the advice on the database and Mattias Öhman, Kristoffer Jutvik, Torsten Santavirta as well as Roger Andersson for their fruitful feedback on my research. I am grateful to Kerstin Larsson, Lena Lubenow, Jenny Sundström, Ulrika Wahlberg and Pamela Tipmanoworm for the administrative support and their kindness. Thank you to Göran Rydgren for the encouragement and guidance. My special thanks go to my two corridor friends at IBF, Timothy Blackwell and Alexander Kalyukin, for their economics, political, geography and fun discussions. Thank you to the Economics Department at Uppsala University for welcoming me during my research stay. I would like to thank Zunjuan, Rinni and Lovisa for sharing their office with me. I appreciated the administrative support provided by Johanna Mörk. Thank you to John Östh for the fruitful discussion and providing me help with the Equipop software.

I would like to thank my friends around the world, neighbors, and Tillmann who have been a constant source of support and encouragement during this time. My special thanks go to my family, particularly my "maman", grandmas, aunts and Emma, who always support me in whichever way they can, and for putting some sparkles in my life when I need it.

Munich, March 2021

# Contents

# List of Figures

# List of Tables

# Introduction

## Background

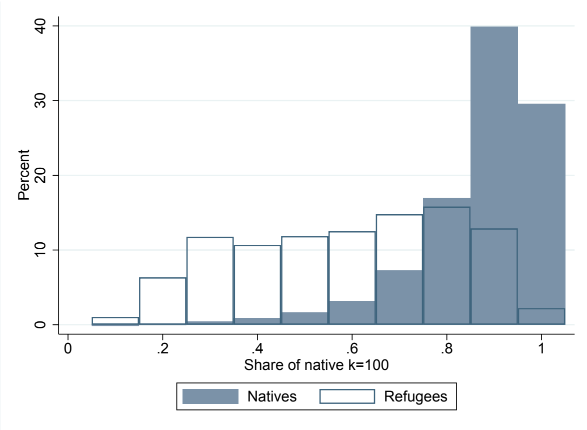
Over 79.5 million individuals were displaced due to conflict, violence, persecution or other human right violations, representing an increase of 38.4 million since 2010.[1] Included in this Figure is about 26 million refugees, with around 50 percent under the age of 18, and 4.2 million asylum seekers. An increasing number of refugees gravitate towards developed countries, such as the United States and European countries, where they are subjected to culture, educational and labor market systems, and in some cases language that differ greatly from their countries of origin. Integrating refugees in the host countries presents major challenges and has become the focus of policy makers. Although integration is relevant for both the refugees and the host countries, existing research on refugee integration remains limited due to scarcity of data. The overarching objective of this thesis is therefore to gain an understanding of the integration outcomes of refugees in the host countries with the view of guiding policy makers. The different chapters focus on the causes and consequences of refugees' residential integration, which is an area of integration that has been less documented in the economics literature to date. This thesis will thus add to the nascent literature on refugees' labor market integration and other dimensions of integration (Aksoy et al., 2020; Battisti et al., 2019; Brell et al., 2020; Brücker et al., 2020; Dahlberg et al., 2020; Fasani et al., 2020) and focus on residential integration on a small geographical scale.

Increasing refugee migration affects the compositions of the host countries' neighborhoods, which in turn raise concerns about whether their integration is affected by living

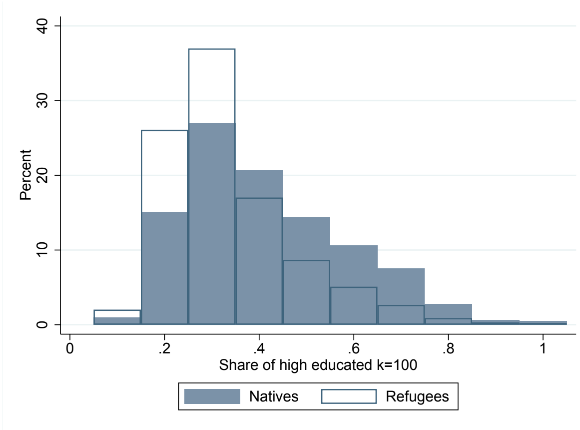
[1] These statistics are obtained from the United Nations High Commissioner for Refugees (2019)

in certain designated neighborhood areas (Malmberg et al., 2018). There is significant variation in the neighborhoods that refugees and natives live in. As shown in Figure 1, the contrast in neighborhoods also holds in Sweden, which is the country studied in this thesis. The figure presents the neighborhood of refugees and natives in terms of share of natives, share of high educated, share of employed and share on welfare living among their 100 closest neighbors in 2017. Most refugees reside in neighborhoods with lower share of natives, high educated, employed and higher share on welfare among their neighbors in comparison to natives. Not only is the variation in neighborhoods noticeable between refugees and natives, but there is also a perceptible variation among the refugees' neighborhoods, particularly in terms of share of high educated and employed among their nearest neighbors. Understanding the variation in refugees' neighborhoods is crucial in combating segregation at the micro level, and examining the consequences of the neighborhood variation can thus promote integration.

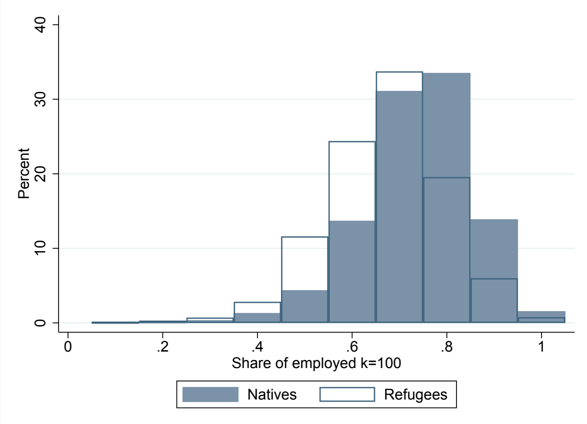
Figure 1: Characteristics of 100 closest neighbors for natives and refugees in 2017



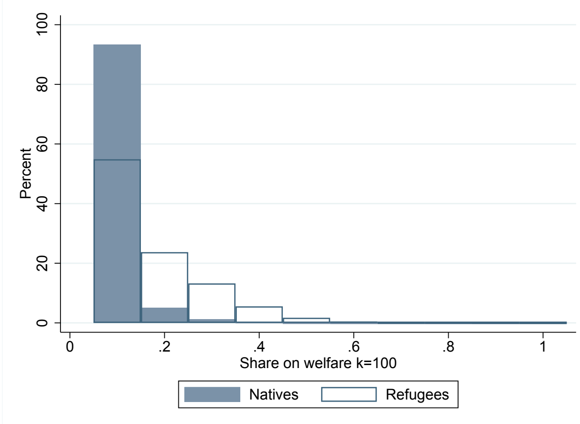

(a) Share of natives




(b) Share of high-educated




(c) Share of employed




(d) Share on welfare

*Notes:* The figure shows the characteristics of the 100 closest neighbors for all refugees and natives above the age of 18 who were residing in Sweden in 2017. Refugees are defined as those possessing a refugee residence permit data. Natives are individuals born in Sweden; high-educated individuals have an education level with at least some tertiary education; employed are defined as those having an earnings; welfare refers to individuals receiving social benefits. Figure 1(d) has a different y-scale compared to the other Figures.
*Source*: Own calculations on data from the GeoSweden database.

## Residential Integration

Understanding the integration of refugees is an issue that is gaining importance in several European countries due to their increase in number. Integration itself is a multidimensional concept. While no consensus on the definition of the concept has been reached, Harder et al. (2018) and Ager and Strang (2008) show that integration includes various dimensions, such as economic, housing and social integration among others. Integration is defined as a two-way process involving the migrants and the host country.

## Introduction

Research from Chetty et al. (2016b) and Sharkey and Faber (2014) show that neighborhoods play a role in individuals' economic, health and educational outcomes. Sharkey and Faber (2014) further demonstrate that residential contexts affect the lives of residents through access to schooling as well as employment opportunities. Interaction with peers and networks in the neighborhoods also affects social integration (Galster, 2008). Furthermore, children's future labor market and schooling outcomes as well as gender gaps in adulthood are shaped by neighborhoods Chetty et al. (2016a,b). Overall, neighbors can form an important part of social networks and diffuse information, knowledge and resources, which could affect labor market and other economic opportunities as well as children's educational outcomes (Chetty et al., 2016b; Ellen and Turner, 1997; Solon et al., 2000). During the last decades, residential segregation has also been a hotly debated topic among policy makers: there is an assumption that immigrants' concentration in neighborhoods leads to the formation of "parallel societies" which are detrimental to their social and labor market integration (Schönwälder, 2007).

Research from Brell et al. (2020) indicates that refugees differ from economic immigrants since they are forced to leave their countries, and therefore cannot select their timing for immigration. They generally arrive to the host country with limited human capital and are expected to learn the host country's language. They are also not allowed to work till their asylum application is processed and it is much harder for a refugee to secure a job. Research shows that they also tend to perform worse in terms of labor market outcomes as compared to other immigrants (Bevelander, 2020; Fasani et al., 2020; Sarvimäki, 2017). Although their labor market integration and recently their multi-dimensional integration outcomes (linguistic, economic, political, social, psychological) have been studied (Aksoy et al., 2020; Dahlberg et al., 2020), their residential integration remains largely undocumented. We may expect that their disadvantage in the labor market may also translate into worse residential integration. Labor immigrants in contrast to refugees have a choice over their residential locations as compared to the refugees who generally have to rely on the local authorities to house them. Therefore, it is imperative to understand how their outcomes are impacted by their residential locations.

## Sweden as a Focus

This thesis is based on data on refugees in Sweden during the past 25 years. The Swedish setting is particularly interesting for the research focus since the country experienced a large inflow of immigrants, especially refugee immigration, in the recent decades (Swedish Migration Agency, 2020). In 2019, there was an estimated 2 million of immigrants living in Sweden, accounting for 19.6 percent of the population (Statistics Sweden, 2020). In 2015, Sweden also had the second highest number of first time asylum applications among the EU28 countries accounting for 16,016 asylum seekers, mostly from Syria, Iraq and Afghanistan, for every million inhabitants (Eurostat, 2016). Important lessons can therefore be drawn from the data in Sweden in relation to refugee migration and this could benefit other European countries who have accepted the challenges of integrating the influx of refugees arriving between 2015 and 2016 during the so-called "refugee crisis".

Sweden has been instrumental in implementing various immigration and integration policies in the EU. Chapters 1 and 2 of this thesis use the refugee placement policy which was applied in 1985 and officially continued until 1994 to provide exogenous variation. The refugee dispersal policy in Sweden is commonly known as the "Sweden-wide" or "All-of-Sweden" strategy (Robinson et al., 2003). This policy meant that refugees' residential preferences were disregarded. It presents the advantage that all refugees, apart from family reunification immigrants, arriving in Sweden were allocated to municipalities through municipality-wise contracts and housed into available apartments in the municipalities upon their arrival. The first and second chapters can therefore assume that exactly which neighborhoods refugees ended up in within the municipalities are exogenous from their point of view. The refugee dispersal policy enables us to evaluate and examine its effect on integration. Overall, the policy allows us to exploit exogenous variation within municipalities as well as provide causal evidence, and surmount the potential issue of refugees self-selecting into neighborhoods.

This thesis has benefited from access to the GeoSweden database, a longitudinal geo-coded micro data collected by Statistics Sweden, which allows analysis over time and enables us to study migration patterns. GeoSweden database is rich, detailed and includes all residents living in the country on the $31^{st}$ of December every year. The database also includes the exact month and year as well as country of origin for each immigrant arriving in Sweden. Swedish administrative data also has the advantage that it contains a variable on the immigration status of the individuals that enables distinction between refugees, labor migrants, family reunification migrants and other migrants. Throughout the chapters of this thesis, refugees are identified through this variable and the sample consists of those who are granted asylum and obtain residence permits in Sweden.

While most countries possess data aggregated to the level of administrative spatial unit, the Swedish administrative data at hand contains detailed coordinates on a 100 by 100 meter grid on where individuals live and work which allows us to study internal mobility patterns. Detailed coordinates are of particular relevance for this thesis as all the chapters empirically study neighborhood integration at various geographical scales, and being in neighborhoods in close proximity may be important for interaction with other neighbors (Van Ham and Tammaru, 2016). The data at hand allows to construct individualized neighborhoods for all individuals living in Sweden.

## k Nearest Neighbor Approach

A feature and contribution that is common throughout the three chapters is the use of the $k$-nearest neighbor approach (Östh, 2014; Turk, Östh, et al., 2017). Our geo-coded data allows us to calculate individualized neighborhoods based on population size, i.e. neighborhoods of different geographical scales. The different chapters of the thesis benefit from contextual neighborhood information on different scales, where scale is computed as counts of nearest neighbors. The construction of small geographical scales allows to investigate potential interaction between individuals, and whether diffusion of information and knowledge takes place in a small neighborhood. The individualized neighbor approach means that the number of individuals in the neighborhood is mostly constant. For each individual, we identify the characteristics of the $k$-nearest neighbors, for instance the share of individuals who are educated, employed or with high earning.

The $k$-nearest neighbor approach presents several advantages in that it constructs neighborhoods with about the same counts of neighbors. This approach also allows small geographical scale of analysis. The small neighborhood sizes can account for residential exposure, i.e., the degree to which immigrants can encounter and have the probability to interact with neighbors with particular characteristics in their neighborhood (Massey and Denton, 1988; Wilkes and Iceland, 2004). A small scale of analysis is important for catching nuances that might otherwise go unnoticed when using predefined bigger neighborhoods, for example, clustering of immigrants inside areas dominated by native Swedes. Furthermore, the literature shows that inter-group contact in small geographic areas will increase trust (Dinesen and Sønderskov, 2018) and it may be expected that contacts at a small scale can enhance relationship and attitude between natives and refugees.

In this thesis, neighborhood integration is defined along both ethnic and socio-economic composition. Chapters 1 and 3 include the share of natives among the refugee' nearest neighbors. From Figure 1, it can be noticed that most natives lived in neighborhoods with 90 and 100 percent of natives among their nearest neighbors while only about 15 percent of refugees lived in high share of natives neighborhoods. Understanding the variation in neighborhoods in terms of share of natives is one of the aims of the first and third chapters since the degree to which a refugee or other migrant is exposed to natives has an impact on learning the native language and other country-specific skills (Cutler and Glaeser, 1997; Edin et al., 2003). The second chapter includes the share of ethnic individuals among the refugees' nearest neighbors. From the socio-economic perspective, chapters 1, 2 and 3 cover share of high educated, share of employed and chapter 3 additionally includes share of high earners and share on welfare. The share of highly educated, employed individuals and high-income earners in the refugees' and other migrants' neighborhood contribute to their access to high-quality social networks through daily, local interactions and transmit knowledge about labor and housing markets in the host country. As shown in Figure 1, refugees are not much represented in the high share of high educated and employed neighborhoods as well as low share on welfare, and it is important to understand the consequences of being located in such neighborhoods.

## Outline of Thesis

Chapters 1 and 2 use the refugee placement policies in order to deal with potential neighborhood selection issues and also to examine the effect of such policies, which are in place in several European countries. These policies were implemented to deal with the issue of refugees clustering geographically, leading to the formation of ethnic enclaves, and leaving major cities with an unequal burden of immigration, higher financial costs and housing shortages (Danzer and Yaman, 2013; Robinson et al., 2003). Several European countries, including Sweden, Germany, Denmark, UK and Ireland, as well as Canada and the US among others, applied refugee placement policies (OECD, 2016). Refugee dispersal policies varied in its implementation from country to country and were not only launched at the national level in certain countries but also at the city level and the neighborhood level. The aim of this policy is to affect refugees' location (Andersson, 2003; Damm, 2005). Dispersal policies may offer benefits in terms of spreading financial costs, opportunities for long-term integration and decreased pressure on housing and social services. Refugee dispersal can therefore lead to several policy implications in terms of regional policies, urban issues, residential segregation, labor market integration, language learning, educational integration and welfare.

Chapter 1 of this thesis aims at closing the gap in how well refugees are integrated into the host country via their small-scale residential integration. It examines the effects of refugees' initial neighborhood characteristics on future residential integration and labor market outcomes by applying the $k$ nearest neighbor approach. We choose small scale neighborhoods, i.e. $k = 100$ nearest neighbors, for our baseline results: this can be understood as the neighbors that the refugees are likely to meet and interact with in their housing blocks. Chapter 1 accounts for the concern that refugees might self-select into neighborhoods by using a Swedish refugee placement policy. Through this policy, we argue that the refugees were exogenously treated with neighbors possessing different characteristics.

The first chapter makes several contributions. Individualized neighborhoods are constructed using the $k$-nearest neighbor approach to investigate small-scale residential integration. Chapter 1 also complements the existing literature by examining other neighbor characteristics than co-ethnics: natives, high educated and employed shares represent larger composition of the population as compared to co-ethnics. A neighborhood quality index is constructed using the different characteristics shares, including share of natives, employed and high educated. This chapter further focuses on initial neighborhood conditions of the refugees upon arrival in the host country. We add to the literature on refugee integration by focusing on an area of integration, i.e. their residential integration, which is less documented.

The results indicate that the higher the quality of the initial neighborhood, defined via share of natives, share highly educated, share employed, and a constructed neighborhood quality index, the higher the future neighborhood quality of the refugees. These results are not driven by stayers. When investigating the effects on earnings, we find weak indications that the quality of the neighborhood affects earnings positively. We can draw a few conclusions from this chapter. The findings suggest that social networks at a small geographic scale can be an important channel in explaining residential and, potentially, labor market integration. The results further indicate that if policy makers want to combat local residential segregation, a well-designed refugee placement policy might achieve this.

Since chapter 1 only indicates a marginal effect of natives, chapter 2 adopts a similar empirical design as the first chapter, and presents evidence on the causal effects of ethnic enclaves on refugees' neighborhood composition of co-ethnics as well as employed co-ethnics, and labor market outcomes on small geographical scales. I investigate whether a potential mechanism through which the effect of ethnic enclaves can impact economic outcomes is through information dissipation within ethnic networks by means of daily local interactions. I account for possible neighborhood selection choices by exploiting a Swedish refugee placement policy in 1990/91.

Chapter 2 contributes to the literature in two aspects. First, this chapter adds to the literature on co-ethnics by examining small geographical scale the effect of ethnic enclaves occurs. The creation of small scale neighborhood sizes allows to investigate where potential interactions between co-ethnics might take place. Second, the data at hand allows to identify exactly who are the refugees.

In a first instance, I show that the initial share of co-ethnic affects the future share of co-ethnic positively, irrespective of the small neighborhood sizes investigated. The findings indicate that the magnitudes of the effects are higher when expanding the neighborhood size. When looking at movers, the results are more similar to the effects in the long run. In the second instance, this paper examines the effects of co-ethnics and employed co-ethnics on labor market outcomes. The results indicate positive and statistically significant effect of ethnic share on earnings and employment 8 years after the refugees' arrival, with the magnitude of the effect being higher at a larger geographical scale. This is in line with the paper by Edin et al. (2003) and Martén et al. (2019) who find a positive effect of co-ethnics on labor market outcomes over time. This result shows that the effect takes some time to kick in. The results point out that interaction occurs on a small geographical scales, and demonstrate that ethnic networks, particularly employed co-ethnics, play an important role for economic success of newly arrived refugees on small geographical scales. Overall, the findings indicate that potential interactions with co-ethnics on small geographical scales are valuable for sharing information to the new refugee arrivals.

The third chapter of this thesis investigates into another determinant of integration, namely age at arrival as it is expected to impact neighbourhood integration through its effects on language acquisition, social networks, and other dimensions of acculturation. This chapter studies residential integration patterns in adulthood for children of refugees who arrive in Sweden before the age of 16 and exploits a siblings design. Using geo-coded information on the residential location of each individual in Sweden, this study takes a novel, data-driven approach in defining neighborhoods and construct individualized $k$-nearest neighborhoods, for $k = 100$ or $k = 1000$.

This chapter adds to the literature in several ways. We study an immigrant category, namely refugees, that is understudied in the literature and focus on their residential integration. Their integration patterns differ from other immigrants and is a worthwhile study to foster their integration in the host country. While previous studies on age at arrival have found that early age at arrival improve school performance, education, and earnings for immigrant children, so far, there has been limited literature on its impact on residential integration of refugees. This chapter fills this gap in the literature by examining refugees' residential integration. Neighborhood is flexibly defined based on the $k$-nearest neighbor approach.

Exploiting a siblings design, the results indicate that, at age 30, refugee children arriving later live in neighborhoods with lower share of natives, high educated, high income earners and higher share of welfare, regardless of the level of $k$. This chapter also provides evidence that refugee children arriving later experience worse labor market outcomes in terms of earnings, lower educational outcomes and likelihood to marry Swedish-born partners at age 30 as compared to children arriving earlier to the host country. Using a decomposition analysis, this study shows that the mean effects of age at immigration on neighborhood integration are only partly explained by economic integration, educational integration and intermarriage. The findings indicate that a large part of the estimated mean age at arrival effects remains unaccounted for, particularly for $k = 100$, which suggests a role for Swedish housing policies, housing discrimination and taste-based preferences in fully explaining the effects of age at arrival.

# Chapter 1

# The Importance of Initial Neighborhood Characteristics for Future Residential Integration and Labor Market Outcomes of Refugees

This chapter is based on joint work with Matz Dahlberg from Uppsala University.

## 1.1 Introduction

The increase in refugee migration throughout Europe has led to significant concerns for their integration in the host countries. Integration has been identified as a multidimensional concept and comprises of various dimensions, including economic, social, education and housing among others (Ager and Strang, 2008; Harder et al., 2018). One important, but mainly neglected measure of how well refugees are integrated into a society is through their small-scale residential integration. When refugees arrive to a new country, does it matter which neighborhood they end up in, in terms of their neighbors' characteristics, for their future residential and labor market outcomes?

Small-scale residential integration is important for a couple of related reasons. First, the composition of individuals in the refugees' immediate neighborhoods matters for the generation of social interactions (for the probability that different groups are exposed to each other), for the formation of networks, and for the transmission of information and knowledge about, e.g., the housing- and labor markets, all of which might be beneficial for the refugee's future outcomes (Ellen and Turner, 1997; Galster, 2008; Johnston and Pattie, 2011; Massey and Denton, 1988; Wilkes and Iceland, 2004). Second, the neighborhood composition matters for the generation of trust. Based on contact theory (Allport et al., 1954), arguments have been raised in the literature that inter-group contact in small geographic areas will increase trust (see, e.g., the discussion in Dinesen and Sønderskov (2018)). As pointed out by Nannestad (2004, 2009), the level of social capital in a local geographic area is important for different forms of integration. Given its importance, we know surprisingly little about the causes and consequences of small-scale residential integration.

In this paper, we will start filling this gap in the literature by studying two questions. First, we will examine what role the socio-economic and demographic compositions of the refugees' *initial* neighborhoods in Sweden have for their future residential integration. In parallel work, Bratu et al. (2021), also using Swedish data, examine what role age at arrival for immigrant children has for small-scale residential integration at age 30. Apart from that paper, we know of no other study examining the determinants of residential integration.[1] Second, we will investigate the effect of neighborhood quality on the refugees' labor market outcomes, in terms of earnings.

We define the quality of a neighborhood through the characteristics of the individuals living in that neighborhood. We use different dimensions which we assume are good proxies for the quality of the social network and the information and knowledge transmitted; share of natives (defined as individuals born in Sweden), share of highly educated, share of employed, and a neighborhood quality index which we construct from the other three variables via factor analysis. The hypothesis is hence that individuals with these characteristics are able to provide useful and high-quality information about Swedish institutional details related to, e.g., the housing- and labor markets and availability of jobs, among others.

The main empirical challenge to deal with is that refugees generally self-select their residential location, making their choice of neighborhood endogenous. To solve this endogeneity problem, we make use of a placement policy that was in effect in Sweden in the early 1990s. In this paper, we argue that exactly which neighborhood within a given municipality that a refugee was placed in was exogenous from the refugee's point of view, implying that the refugees were exogenously treated with different small-scale neighborhood qualities (as defined above). Balancing tests also indicate that the placement into different neighborhoods was exogenous. This enables us to use a research design that allows for a causal interpretation of the results for both research questions.[2]

One reason there is scarce evidence on the determinants of small-scale residential integration is probably due to a lack of fine-grained, geocoded information in most (register-based) databases. In examining our questions, we will use a comprehensive, register-based, database, GeoSweden. There are two aspects that are particularly interesting with this database for this paper. The data at hand contains highly granulated geographic data through coordinate information on a $100 \times 100$ meter level. This allows us to construct individualized neighborhoods for all placed refugees via a $k$-nearest neigh-

[1]Bratu et al. (2021) find that arriving after school-starting age implies that you live in socio-economically weaker neighborhoods at age 30.

[2]The econometric model we use is analogous to the one used in the co-ethnic literature by, e.g., Edin et al. (2003) and Andersson (2020).

bors approach.[3] The main results are presented for the $k = 100$ nearest neighbors, which can be understood as the neighbors that the refugees meet in their housing blocks. This approach enables us to analyze the effects of initial neighborhood characteristics on a small geographical scale, where interactions between neighbors are likely to take place. Furthermore, we are able to exactly identify the *refugees* in the data through a variable that provides information on reason for immigration.

Our results show that the quality of the initial neighborhood, in terms of natives, highly educated, employed, and the neighborhood quality index, leads to residential integration along these dimensions in all years within an eight-year follow-up horizon. As one of the aims of the refugee dispersal policy was to mix natives and immigrants, it is particularly interesting to notice that an increase in the initial share of natives leads to higher share of natives in the future. This result holds for a number of robustness tests, such as different sizes of the individualized neighborhoods ($k = 50$ and $k = 250$), distance restrictions for finding the 100 nearest neighbors, and for a sub-sample of only movers. Our results imply that a placement policy may be effective if the aim is to locally mix refugees with individuals with different characteristics, such as natives, highly educated, and employed. When investigating the second question, we find positive, but weak, evidence that the quality of the refugees' neighborhoods affect earnings.

Our paper makes four main contributions to the literature. The first contribution is the specific focus on *small-scale neighborhood integration.* As mentioned earlier, we know very little about its causes and consequences, probably due to lack of fine-grained geocoded data. Bratu et al. (2021) is the only study we know of that examines a determinant of granular neighborhood integration - age at arrival. Given the granularity of our data, we contribute to the literature on the consequences of neighbors as network. The effects of neighbors on learning and knowledge are acknowledged in the field of geography, sociology and economics. In addition to the process of socialization that occurs through contact with peers in the neighborhood (Galster, 2008), neighbors can also form an important part of social networks and diffuse information, knowledge and resources, which could increase labor market and other economic opportunities (Audretsch and Feldman, 1996;

[3]To calculate the individualized neighborhoods, we use Equipop, a software developed by John Östh; see Östh (2014).

Ellen and Turner, 1997). Using US Census data and examining city block level, Bayer et al. (2008) find that residents in the same block have a high propensity to work together and interpret this positive effect as a social interaction effect, which is mainly operating for individuals of the same socio-demographic characteristics. Further evidence from the US from Schmutte (2015) shows that workers have a higher probability to move to a higher-paying firm if their neighbors are engaged in high-paying firms as well.

Our second contribution is that we have a specific focus on *other traits than co-ethnics.* While the earlier literature has mainly focused on the effects of co-ethnic on immigrants' labor market integration (Andersson et al., 2019; Beaman, 2011; Borjas, 2000; Cutler et al., 2008; Edin et al., 2003; Martén et al., 2019), we investigate the role played by other characteristics of the refugees' neighbors, i.e. share natives, share educated, share employed, and the neighborhood quality index, on residential integration and labor market outcomes. Since co-ethnics remain a small share of all neighbors, on average much smaller than the characteristics we examine, we think it is a valuable contribution to examine the role played by the neighbors with these characteristics.

Our third contribution is that by focusing on the *initial neighborhood conditions* facing refugees upon arrival, we relate to previous research that show that the initial condition in the host country affects the refugees' future economic integration. Aksoy et al. (2020) provide evidence that initial local unemployment have a negative effect on refugees' education, earnings and employment outcomes, and Åslund and Rooth (2007) find that an initial favorable labor market increases earnings in the future.

Our fourth contribution relates to the fact that we focus on *refugees.* Our data contains information on the reason for immigration, implying that among other things, we can exactly identify the individuals that arrive in Sweden as refugees, and hence are affected by the placement policy. Most papers use country of origins to identify refugees as they do not possess information on immigration status (see, e.g., Edin et al. (2003), and Damm (2009)). Refugees remain a group whose integration outcomes are understudied. We add to a nascent literature that studies refugees' labor market- and other forms of integration outcomes (Aksoy et al., 2020; Battisti et al., 2019; Bratu et al., 2021; Dahlberg et al., 2020).

The remainder of the paper is structured as follows. Section 1.2 presents the data as well as the construction of the neighborhood quality index used in the paper. Section 1.3 provides the background on Swedish refugee placement policies and on how we use the dispersal policy to solve neighborhood self-selection issues. In this section, we also presents the descriptive statistics for the refugees in our sample. Section 1.4.1 presents our empirical strategy to examine the effect of initial neighborhood characteristics on future residential integration. Section 1.4.2 reports the results for residential integration and section 1.5 provides the empirical design and the results on future labor market outcomes. Finally, section 1.6 concludes.

## 1.2 Data and Definitions

### 1.2.1 Data source and definition of refugees

The analysis in this paper is based on GeoSweden, a comprehensive database collected on a yearly basis from 1990 until 2014. It covers all residents in Sweden and contains a rich set of background characteristics on all individuals collected from several different registers, including the education, the income and the employment registers.[4] Of importance for this paper is that the data includes very detailed geographical information, given by coordinates that defines $100 \times 100$ meter grids, on where the individuals live and work. In addition, it contains information on exactly when and from which country an individual immigrates to Sweden, emigration information as well as information on migration patterns within Sweden. Specifically, from the annual geocodes, we observe when, from where and to where an individual moves. This makes the database very well suited for examining questions related to immigration, within-country migration, segregation, and integration.

Given the focus on refugees in this paper, it is important that we can correctly identify refugees. A unique aspect of our data is that it contains information on the reason for immigrating to Sweden: we know if an individual come to Sweden to work, to study,

[4]All data is collected and made anonymous by Statistics Sweden, and administered by the Institute for Housing and Urban Research at Uppsala University.

as a refugee, or as a tied family member. Using the refugee information on reason for immigration, we obtain an exact definition of the refugees.[5]

### 1.2.2 Definition of Neighborhood

Using the detailed coordinate information in the data, we construct individualized neighborhoods based on a $k$-nearest neighbor approach. An advantage of that approach for defining neighborhoods is that, since it locates the refugees at the center of their own neighborhoods, it is a good representation of the actual urban context of the refugees. Thus, the resulting neighborhood characteristics is a good representation of the actual urban context surrounding each individual. Another important advantage is that this approach offers a useful way of performing the analysis at a very small scale. A small scale analysis is crucial for catching nuances that might be overlooked when using data on a larger geographical scale, often defined through administratively set borders. One such nuance is small-scale residential segregation. Furthermore, a small scale analysis allows us to observe potential social networks and ties that can be important. Since the nearest neighbors are the individuals that the refugees have a higher likelihood to meet, the nearest neighbors are the ones that are most likely to affect the arriving refugees' integration. As Galster (2008) point out, the behaviors and attitudes of a neighborhood resident can impact his neighbor. The process of socialization occurs through contact with peers in the neighborhoods. Neighbors can thus form an important part of social networks and diffuse information, knowledge and resources, which could increase labor market and other economic opportunities. Using a $k$-nearest neighborhood approach provides better insights into neighborhood contexts and their effects for social integration, and it also allows capturing residential mobility on a smaller scale.

We construct the individualized neighborhoods for each refugee as follows:

1. From our full population registers, and for all years, we identify all coordinates in Sweden (100 $\times$ 100 meter squares) at which at least one individual lives.

[5] Earlier research has typically proxied refugee status with country of origin (see, e.g., Edin et al. (2003), and Damm (2009)).

2. For each of the coordinates identified in the former step, we calculate both the total number of individuals and the total number of individuals with a certain characteristic, such as country of birth, earnings and degree of education, living on that coordinate.[6]

3. Using the information obtained in the former two steps, we construct individualized neighborhoods for all refugees living in Sweden by identifying the characteristics of the $k$ nearest neighbors for each individual (which provides us with the share of individuals among the $k$-nearest neighbors that share a certain characteristic).[7]

Using this approach, we build individualized neighborhoods based on the $k = 100$ nearest neighbors. We argue that this creates individualized neighborhoods of a size small enough for meetings on a daily basis (in the staircase, in the street outside the home, kids to play with around the home/at the local playground for the refugees' children, etc) yet not small enough not to be meaningful (if the number of closest neighbors is made too small, the social network would be too small; in the extreme case it would only be the refugee's own family).[8]

The reason for using the $k = 100$ closest neighbors is hence that we assume that these are people that the refugees are likely to meet and interact with on a daily basis. However, in some less densely populated areas, the algorithm might have to search over quite a long distance to reach the 100 nearest neighbors. In these cases, we would not expect a close and frequent interaction among the "neighbors". As Figure 1.1 shows, in the majority of the cases the algorithm do not have to go far to find the 100 nearest neighbors. Going into details, it turns out that for 49% of the refugees, the 100 nearest neighbors are found within their $100 \times 100$ meter coordinate point (interpreted as if the algorithm has to go 0 meters to find those neighbors). Within a distance of 100 meters, the 100 nearest

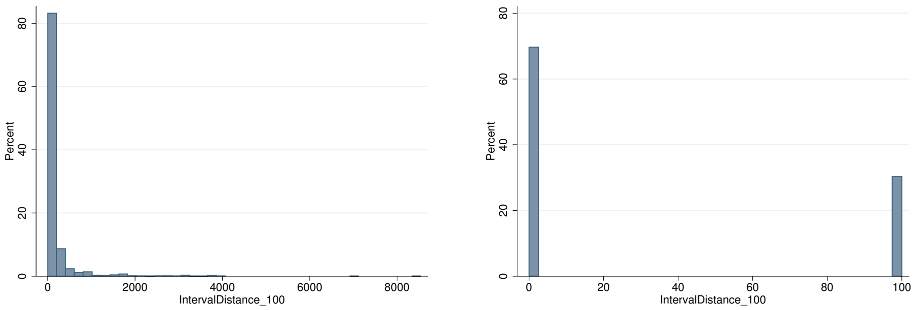
[6]We include all individuals aged 18 and above, including family members in the neighborhoods' individual counts.

[7]For this final step, we use the EquiPop software (Östh, 2014).

[8]In a robustness analysis, we do however check if the results are sensitive to this choice by examining the cases of $k = 50$ and $k = 250$ nearest neighbors. It turns out that the results are not sensitive to the specific choice of these small $k$s.

neighbors are found for 71% of the refugees (see Figure 1.1(b) for the distribution), and allowing a distance of 200 meters the corresponding figure is 84%.

Figure 1.1: Distribution of distance needed for the algorithm to find the refugees' 100 nearest neighbors




(a) All refugees (b) Distance on 100 meters and below

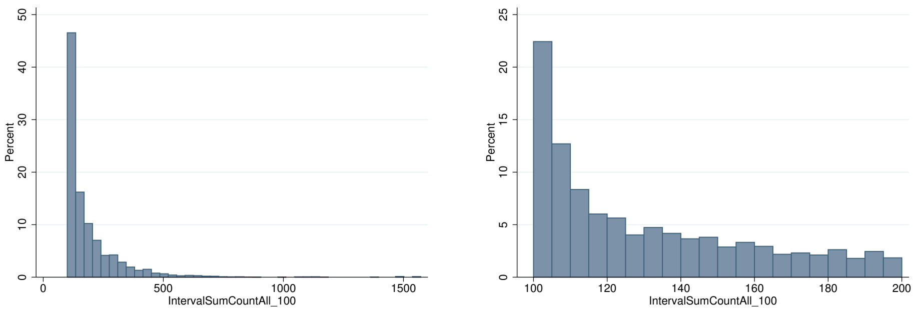
*Notes:* The figure shows the distribution of distance required for the Equipop algorithm to find the refugees' 100 nearest neighbors.
*Source*: Own calculations on data from the GeoSweden database.

It should also be noted that, since we have coordinate points that measure $100 \times 100$ meter, we will not always get exactly the $k$ nearest neighbors. To give an example, assume that we are looking for the $k = 100$ nearest neighbors for a specific individual. If there are more than 100 persons living on the same $100 \times 100$ coordinate point, all those individuals will be used when calculating the shares. Likewise, if there are less than 100 individuals on that point, the algorithm start searching in an adjacent coordinate point. If the sum of all individuals on those two points equals or exceeds 100, all those individuals will be used in the calculations of neighborhood characteristics, and so on. In our case, for $k = 100$, Figure 1.2 shows how many neighbors the algorithm actually identifies when searching for the $k = 100$ nearest neighbors. As can be seen, in the majority of the cases, the algorithm actually finds around 100 close neighbors, but there are cases, in very dense areas, where the algorithm has to pick far more than 100 individuals. For 200 neighbors and less, we cover 71% of the refugees (see Figure 1.2(b) for the distribution of counts).

In the analysis, we will check the sensitivity to both the distance and to the number of closest neighbors found.[9] As will be shown, the results are not sensitive to this.

[9]It shall be noted that finding many neighbors in a close neighborhood is not necessarily an issue in our case. Quite the opposite, actually, if it means that the refugees can benefit from more knowledge and information spillovers.

Figure 1.2: Distribution of actual number of neighbors found when searching for the refugees' $k = 100$ nearest neighbors




(a) All refugees (b) Refugees with 200 counts

*Notes:* The figure shows the distribution of counts. The y-scales are different in Figure (a) and (b).
*Source*: Own calculations on data from the GeoSweden database.

### 1.2.3 Definition of Neighborhood Quality

We characterize the quality of the individualized neighborhoods through three demographic and economic variables; share of natives, share of employed individuals and share of high-educated individuals. Natives are described as individuals born in Sweden regardless of their parents' birth countries. Employed individuals are defined as all those who receive earnings from work at some point during the year. High educated individuals are those who have at least some university education. The rational for using these characteristics is that natives, employed and high-educated all have the potential for providing high-quality networks and high-quality information and knowledge about Swedish institutions, such as the labor market, the housing market and the educational system.

In addition to using these three variables separately, we will also use them to create a single quality index describing each refugee's neighborhood. To construct the index, we will use factor analysis.[10] The estimated neighborhood quality index has a normal

[10]The neighborhood index is constructed by municipality and year. The different shares obtain factor loadings which are the weights and correlations between each share and the factor. Creating a neighborhood quality index allows us to have one measure of neighborhood rather than three different characteristics shares. To be able to use factor analysis, there must be a relationship between the variables. In our case, the correlation between the variables are greater than 0.30 indicating a strong relationship between the variables (Rummel, 1967; Yong, Pearce, et al., 2013). The average factor loadings are shown in Appendix A.1. Moreover, uniqueness, which is the variance not shared with other variables, needs to be low. Uniqueness is less than 0.5 in most of the municipalities and years, which indicates that we can use factor analysis.

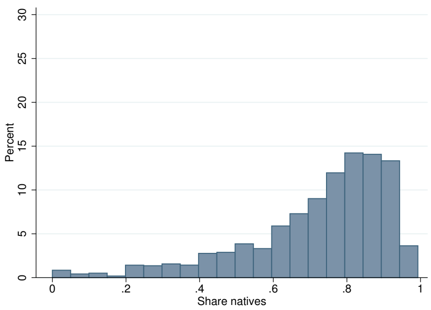
distribution and is centered around a mean of zero. We conduct a Kaiser-Meyer-Olkin Measure of sampling adequacy (KMO) which indicates the proportion of variance in the different characteristics' shares that may be caused by underlying factors. If the KMO measure is greater than 0.5, it indicates that factor analysis is useful. In our case, for about 80 percent of the cases, the KMO is greater than 0.5 for the municipalities. Additionally, we perform Bartlett's test of sphericity to test if the share of natives, share of highly educated and share of employed variables are related. Given that we obtain values of less than 0.05 of the significance level, Bartlett's test indicates that factor analysis is useful in our case.

Figure 1.3 shows the distribution of the refugees' 100 nearest neighbors in their initial arrival year[11] for the three demographic and economic variables and the quality index, respectively. Figure 1.3(a) illustrates that there is quite a large variation in the share of natives among the refugees' $k = 100$ nearest neighbors. In the mean individualized neighborhood of the newly arrived refugees, the natives' share is 0.72; see Table 1.1 for more descriptive statistics. A small percentage of the refugees end up in neighborhoods that have almost no natives.

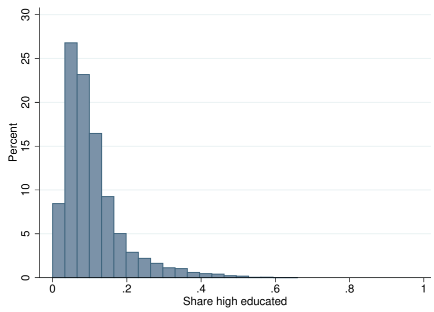
It is apparent in Figure 1.3(b) that most of refugees do not live in neighborhoods with large shares of high-educated individuals (also c.f. Table 1.1). When looking at the share of employed among the refugees' 100 nearest neighbors, we note that the distribution is quite normally distributed, although somewhat skewed to the left. Finally, Figure 1.3(d) presents the distribution of the neighborhood quality index, which mainly varies between -2 and 2 with a mean of zero, and where negative values indicate low-quality neighborhoods and positive values high-quality neighborhoods.

[11] The year of arrival is either 1990 or 1991; see the next section for an explanation for the choice of these years.

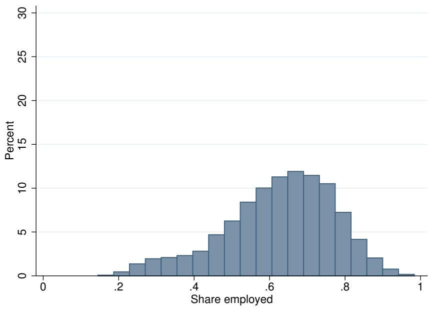
Figure 1.3: Distribution of different characteristics among the refugees' 100 nearest neighbors



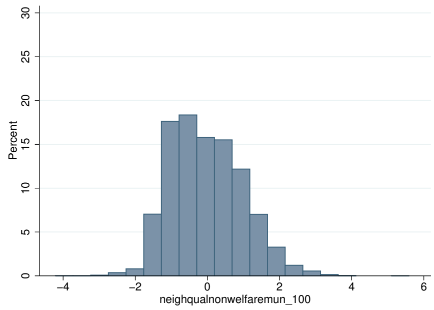

(a) Share natives




(b) Share educated




(c) Share employed




(d) Neighborhood quality index

*Notes:* The figure shows the distribution of initial share of natives, educated individuals, and employed individuals among the refugees' 100 nearest neighbors, and the estimated neighborhood quality index based on the three aforementioned variables, respectively.
*Source*: Own calculations on data from the GeoSweden database.

Table 1.1: Summary statistics: shares of individuals with different characteristics among the refugees' 100 nearest neighbors in the placement year

| | Mean | No. of refugees | p25 | p50 | p75 |
|---|---|---|---|---|---|
| Share of natives | 0.72 | 14051 | 0.63 | 0.78 | 0.86 |
| Share of high educated | 0.11 | 14051 | 0.05 | 0.09 | 0.13 |
| Share of employed | 0.63 | 14051 | 0.54 | 0.64 | 0.73 |
| Quality index | -0.03 | 14024 | -0.81 | -0.13 | 0.67 |

*Notes:* p25 denotes the $25^{th}$ percentile, p50 shows the $50^{th}$ percentile and p75 represents the $75^{th}$ percentile.

*Source:* Own calculations based on data from the GeoSweden database.

## 1.3 Identifying Variation

The methodological problem to solve when estimating the effects of the characteristics of the refugees' initial neighborhood on their future outcomes is related to the refugees self-selecting their residential locations, and hence their neighborhoods. If the refugees self-select into higher-quality neighborhoods based on their own observed and unobserved characteristics, that will most likely lead to biased estimates.

To solve this problem, we will use a refugee placement policy that was in effect in Sweden in the late 1980s and early 1990s. What we argue is that, under the policy, exactly which apartment (and hence initial neighborhood) that a refugee ended up in was exogenous from each placed refugee's point of view. In this section, we will present the basis for our claim of exogeneity. In doing so, we will, first, explain how the placement policy worked and what type of variation in the data we will use for identifying the effects. Second, we will present balancing tests to examine if our research design provides treatment and control groups that are balanced on observable characteristics. Finally, we will present some descriptive statistics on the sample of refugees that will be used in the analysis.

### 1.3.1 The Refugee Placement Policy

The refugee placement policy in Sweden was a two step assignment process. In the first step, refugees were assigned to municipalities via municipality-wise contracts. In the second step, once assigned to a municipality, each refugee was placed in an available apartment at the time of arrival. For this paper, we will not use the variation emanating from the first step. Rather, we will use the within-municipality variation in neighborhood quality emanating from the second step.

The government placed all refugees during those years, with the exception of those arriving on family reunification grounds, i.e. if a refugee had migrated as part of a family member, then he or she would be placed in the same municipality as his or her family. It is therefore very important to be able to separately identify refugees and tied family members. As explained earlier, we are able to do that.

Let us begin by placing the policy in a general context. In recent decades, several countries were faced with the issue that refugees cluster geographically, leading to the formation of ethnic enclaves and leaving major cities with an unequal burden of immigration, higher financial costs and housing shortages (Danzer and Yaman, 2013; Robinson et al., 2003). As a result, several OECD countries applied refugee placement policies to address these concerns.[12] The overall aim of the policy is to affect refugees' location (Andersson, 2003; Damm, 2005) by scattering refugees across counties and municipalities with regards to sufficient housing supply and integration measures in place in those areas.

The refugee dispersal policy in Sweden, commonly known as the "Sweden-wide" or "All-of-Sweden" strategy, was implemented in 1985 and officially continued until 1994 (Robinson et al., 2003). The main motivation of the dispersal policy was to direct refugees away from the metropolitan areas and to aim a balance between urban and rural municipalities.

The Swedish Migration Board (SIV) initially calculated a maximum of 5,000 refugees during 1985 and 60 municipalities in the southern and central parts of Sweden, as well as three northern municipalities, were chosen to participate in the placement policy. The idea was that the refugees should be able to remain in the municipalities in which they were placed and integrate in, e.g., the education and the labor market. However, the actual influx of asylum seekers increased over time and by 1989, 277 out of 284 municipalities took part in the dispersal policy. This implied that housing shortages quickly became apparent in certain municipalities and eventually housing supply was an important determinant in the choice of municipality placement of refugees. The system was under large pressure from 1992 onwards and the placement program might have become less strict after 1991 (Åslund and Rooth, 2007). For this reason, we will use the years 1990 and 1991 in our analysis (as explained above, our data starts in 1990).

Once the refugee had been assigned to a municipality, that municipality was responsible for arranging for an apartment for the refugee. The apartments were typically part of the municipal housing stock.

Sweden has a regulated rental housing market, i.e., rental prices are not determined through the market. Instead of being allocated via the market, rental apartments are allocated via a queuing system. The majority of rental apartments are public (owned by municipal housing companies), and each municipality has its own queue. In addition, the rental market constitutes a quite large share of the housing market and the municipal

[12] See Robinson et al. (2003) for a description of implementation of refugee dispersal policies in different countries; Boswell (2003) for the case of UK and Germany; Bell et al. (2013) for the case of UK; Mayda et al. (2017) and Beaman (2011) for the US; Damm and Vasiljeva (2016) for the case of Denmark. See also OECD (2016) for the implementation in other countries.

housing properties are located in all types of areas within a municipality, ranging from affluent areas to areas that are not so well off.

The system works so that the next apartment that becomes available is given to the first person in the queue who accepts it, and this person then leaves the queue. An increased demand for rental apartments increases the number of queuing individuals, which in turn increases the queuing time to get an offer for an apartment. In many municipalities, the queuing time can be several years.

To get an understanding of how municipalities allocated the placed refugees to specific apartments in the municipalities, we have researched the archives of two of the refugee-receiving municipalities (Uppsala and Västerås municipalities). Municipal documents from the times of the placement indicate that newly arrived refugees were prioritized and were allowed to bypass the municipal queuing system, something that seem to have been more pertinent for refugees with children (*Diarieförda handlingar, centrala staden* 1998)[13], and that they had to accept the first offered apartment offered by the municipality (Västerås).

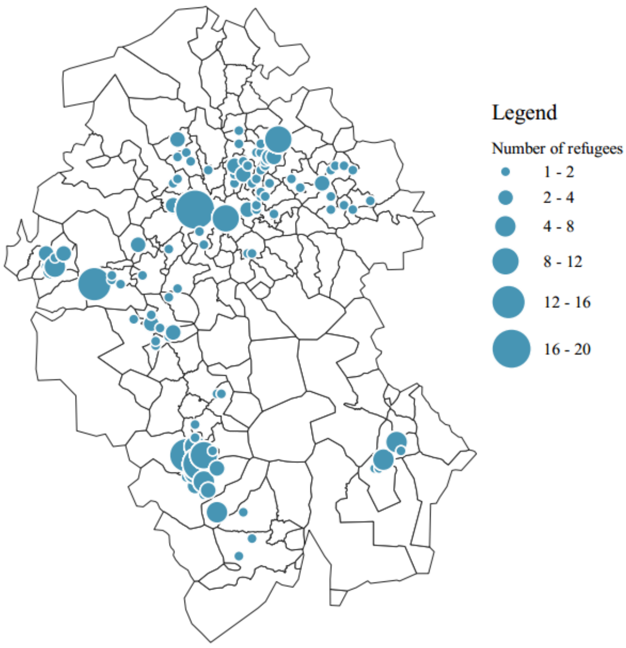
Given these institutional peculiarities, we argue that exactly which apartment within the municipal housing company that first became available after the arrival of the placed refugee must be considered as exogenous from the refugee's point of view. Hence, we assume that exactly which neighborhood a refugee ends up in, in terms of their neighbors' characteristics is exogenous from the individual refugee's point of view.

Apart from being exogenous from the refugees' point of view, we also need enough variation within each municipality to be able to identify any effects. To illustrate how the placed refugees were geographically allocated within a municipality in their initial year in Sweden (after having been granted a residence permit), Figure 1.4 displays a map of the central parts of Uppsala municipality with the residential locations of the refugees arriving in 1990/91 (as measured by the $100 \times 100$ meter coordinate points that we observe in our data). To get some reference point regarding geographic size, we have outlined the smallest administratively set neighborhoods (so called SAMS-areas) in the map. The

[13]For this reason, we will conduct a sensitivity analysis in which we only use refugees with children.

average population in a SAMS-area is approximately 1000. From the map, it is clear that the refugees were not clustered into the same areas, but rather scattered around the city. This shows that the first available apartments in the municipality that became available after the arrival of each refugee were located at quite different locations in the city. In some of these residential locations, the neighbors were high-educated, high-income earners and few received welfare benefits. In other locations, the opposite was true. It is this within-municipality variation that we will use for identifying the effects of the refugees' initial neighborhoods on their future outcomes.

Figure 1.4: Refugee distribution in Uppsala in 1990/91




*Notes:* The figure shows the distribution of refugees in Uppsala during the placement years considered.

*Source:* Own calculations based on data from the GeoSweden database.

### 1.3.2 Balancing Tests

To get a sense of whether the allocation of refugees into different types of neighborhoods *de facto* can be considered as random, we provide balancing tests on the refugees' background charactersitics. Table 1.2 presents the normalized differences, comparing the characteristics of refugees placed in low (below zero) versus high (above zero) quality neighborhoods, as defined via the factor analysis, within each municipality in their initial year in Sweden. The covariates used are age, gender, marital status, having children and country regions. Imbens (2015) shows that large values in the magnitude of 1.00 and above for normalized differences indicate that the groups in low and high native share neighborhoods will be substantially different. The normalized differences depicted in the Table 1.2 are rather small ranging between -0.06 and 0.07. Refugees in the two types of neighborhoods are hence balanced on their background characteristics, indicating that the apartments made available to the refugees on a "first-available-apartment-basis" can be considered as providing an "as-if" random allocation of refugees into different neighborhoods.

When we look at the normalized differences for the individual components in the neighborhood quality index, we see that they are balanced also on these characteristics of the refugees neighbors (see Tables 1.3-1.5).[14]

Table 1.2: Balancing test: neighborhood index ($k = 100$)

| | Low N.Index (N = 6055) | | High N.Index (N = 7996) | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Normalized Difference |
| Age | 33.30 | 6.82 | 33.51 | 6.74 | 0.03 |
| Children | 0.52 | 0.50 | 0.56 | 0.50 | 0.07 |
| Married | 0.30 | 0.46 | 0.28 | 0.45 | -0.04 |
| Female | 0.36 | 0.48 | 0.38 | 0.49 | 0.04 |
| Africa | 0.18 | 0.38 | 0.16 | 0.36 | -0.06 |
| Europe | 0.14 | 0.35 | 0.16 | 0.37 | 0.05 |
| East Asia | 0.08 | 0.28 | 0.09 | 0.28 | 0.00 |
| West Asia | 0.53 | 0.50 | 0.52 | 0.50 | -0.01 |
| Latin America | 0.07 | 0.25 | 0.07 | 0.26 | 0.03 |

*Notes:* The table shows the balancing test for low versus high neighborhood quality index in the initial year for $k = 100$.
*Source:* Own calculations based on data from the GeoSweden database.

[14] Since we conduct a sensitivity analysis for $k = 50$ and $k = 250$, we have also done the balancing tests for these neighborhood sizes. The tests, presented in Appendix Section A.2, indicate that the groups are balanced also for these $k$s.

Table 1.3: Balancing test: native neighbors ($k = 100$)

| | Low Native (N=5713) | | High Native (N=8338) | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Normalized Difference |
| Age | 33.49 | 6.88 | 33.37 | 6.71 | -0.02 |
| Children | 0.54 | 0.50 | 0.55 | 0.50 | 0.02 |
| Married | 0.30 | 0.46 | 0.28 | 0.45 | -0.03 |
| Female | 0.37 | 0.48 | 0.37 | 0.48 | -0.01 |
| Africa | 0.16 | 0.37 | 0.17 | 0.37 | 0.01 |
| Europe | 0.14 | 0.35 | 0.16 | 0.37 | 0.06 |
| East Asia | 0.09 | 0.28 | 0.08 | 0.28 | -0.01 |
| West Asia | 0.54 | 0.50 | 0.52 | 0.50 | -0.05 |
| Latin America | 0.07 | 0.26 | 0.07 | 0.26 | 0.01 |

*Notes:* The table shows the balancing test for low versus high share of natives neighborhoods in the initial year for $k = 100$.
*Source:* Own calculations based on data from the GeoSweden database.

Table 1.4: Balancing test: educated neighbors ($k = 100$)

| | Low Educated (N = 5697) | | High Educated (N = 8354) | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Normalized Difference |
| Age | 33.44 | 6.86 | 33.40 | 6.72 | -0.01 |
| Children | 0.55 | 0.50 | 0.54 | 0.50 | -0.02 |
| Married | 0.29 | 0.45 | 0.29 | 0.46 | 0.02 |
| Female | 0.37 | 0.48 | 0.37 | 0.48 | 0.00 |
| Africa | 0.16 | 0.37 | 0.17 | 0.37 | 0.02 |
| Europe | 0.15 | 0.36 | 0.16 | 0.36 | 0.02 |
| East Asia | 0.09 | 0.28 | 0.08 | 0.28 | -0.01 |
| West Asia | 0.52 | 0.50 | 0.53 | 0.50 | 0.01 |
| Latin America | 0.08 | 0.27 | 0.07 | 0.25 | -0.05 |

*Notes:* The table shows the balancing test for low versus high share of educated neighborhoods in the initial year for $k = 100$.
*Source:* Own calculations based on data from the GeoSweden database.

Table 1.5: Balancing test: employed neighbors (k=100)

| | Low Employed (N = 5953) | | High Employed (N = 8098) | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Normalized Difference |
| Age | 33.15 | 6.73 | 33.62 | 6.81 | 0.07 |
| Children | 0.50 | 0.50 | 0.57 | 0.49 | 0.15 |
| Married | 0.32 | 0.47 | 0.27 | 0.44 | -0.11 |
| Female | 0.36 | 0.48 | 0.38 | 0.49 | 0.05 |
| Africa | 0.18 | 0.39 | 0.15 | 0.36 | -0.09 |
| Europe | 0.14 | 0.35 | 0.16 | 0.37 | 0.05 |
| East Asia | 0.08 | 0.27 | 0.09 | 0.29 | 0.04 |
| West Asia | 0.53 | 0.50 | 0.52 | 0.50 | -0.01 |
| Latin America | 0.07 | 0.25 | 0.07 | 0.26 | 0.03 |

*Notes:* The table shows the balancing test for low versus high share of employed neighborhoods in the initial year for $k = 100$.
*Source:* Own calculations based on data from the GeoSweden database.

### 1.3.3 The Sample of Refugees

Using the information about refugee status in our data, we are able to identify those immigrants arriving in Sweden as refugees in 1990/91. As for all individuals in our data, the register information for the refugees are recorded at the end of each year (which is hence also true for the year in which they obtain their residence permit).

We restrict our sample to consist of refugees who are of working age, 25-55 years old, upon arrival.[15] The refugees arriving in 1990/91 arrived to Sweden from various source countries as seen in Table 1.6 (non-European refugee sending countries) and Table 1.7 (European refugee sending countries). That there are some refugees migrating from European source countries probably indicates that the refugees travelled to Sweden from another country even though they sought asylum from persecution from their birth countries. The refugees could hence have migrated to other OECD-countries before arriving to Sweden. The correlation between birth countries and countries from which the refugees migrate is however 0.88, indicating that there are not many countries for which source country and birth country differ.

[15]The reason for using 55 as the upper age is that we want to have a follow-up horizon that covers several years (we use eight years), which means that all individuals are of working age during this follow-up horizon.

Tables 1.6 and 1.7 present the absolute number and percentage of refugees from each source country. Our sample consists of a total 14,051 refugees, which can be broken down to 6,787 refugees in 1990 and 7,264 refugees in 1991 affected by the placement policy. Table 1.6 shows the refugees who migrated from non-European countries. Overall, the top five refugee sending countries in 1990 and 1991 were Iran, representing nearly 18 percent of our sample, followed by Lebanon with about 14 percent of the sample, closely followed by Ethiopia, Somalia and Iraq. The largest inflow of refugees came from the West Asian region. There is heterogeneity in the number of individuals coming from the refugee sending countries out of the list of 61 source countries. Table 1.7 illustrates refugees coming from Europe, and consists of refugees who had moved to Sweden from a European country. The refugees from European countries form about 20 percent of our sample.

Table 1.6: Sample of non-European refugee sending countries for the considered cohorts

| Refugee sending countries | 1990 | 1991 | Total |
|---|---|---|---|
| Ethiopia | 485 | 381 | 866 |
| | (4.31) | (3.39) | (7.70) |
| Somalia | 253 | 637 | 890 |
| | (2.25) | (5.66) | (7.91) |
| Gambia | 1 | 3 | 4 |
| | (0.01) | (0.03) | (0.04) |
| Tunisia | 26 | 10 | 36 |
| | (0.23) | (0.09) | (0.32) |
| Morocco | 5 | 15 | 20 |
| | (0.04) | (0.13) | (0.18) |
| Uganda | 22 | 42 | 64 |
| | (0.20) | (0.37) | (0.57) |
| Algeria | 16 | 22 | 38 |
| | (0.14) | (0.20) | (0.34) |
| Other Africa | 146 | 219 | 365 |
| | (1.30) | (1.95) | (3.25) |
| Lebanon | 1069 | 844 | 1913 |
| | (9.51) | (7.50) | (17.01) |
| Syria | 139 | 239 | 378 |
| | (1.24) | (2.13) | (3.36) |
| Turkey | 306 | 243 | 549 |
| | (2.72) | (2.16) | (4.88) |
| Iraq | 428 | 591 | 1019 |
| | (3.81) | (5.26) | (9.06) |
| Iran | 1255 | 1210 | 2465 |
| | (11.16) | (10.76) | (21.92) |
| Other West Asia | 83 | 236 | 319 |
| | (0.74) | (2.10) | (2.84) |
| Vietnam | 39 | 36 | 75 |
| | (0.35) | (0.32) | (0.67) |
| Thailand | 72 | 24 | 96 |
| | (0.64) | (0.21) | (0.85) |
| China and Taiwan | 62 | 36 | 98 |
| | (0.55) | (0.32) | (0.87) |
| Phillipines | 6 | 51 | 57 |
| | (0.05) | (0.45) | (0.51) |
| Japan | 1 | 4 | 5 |
| | (0.01) | (0.04) | (0.04) |
| Afghanistan | 39 | 69 | 108 |
| | (0.35) | (0.61) | (0.96) |
| Bangladesh | 65 | 64 | 129 |
| | (0.58) | (0.57) | (1.15) |
| India | 14 | 12 | 26 |
| | (0.12) | (0.11) | (0.23) |
| Pakistan | 72 | 80 | 152 |
| | (0.64) | (0.71) | (1.35) |
| Sri Lanka | 20 | 105 | 125 |
| | (0.18) | (0.93) | (1.11) |
| Other Asian Countries | 192 | 270 | 462 |
| | (1.71) | (2.40) | (4.11) |
| USA | 5 | 2 | 7 |
| | (0.04) | (0.02) | (0.06) |
| Canada | 2 | 4 | 6 |
| | (0.02) | (0.04) | (0.05) |
| Central America | 122 | 92 | 214 |
| | (1.08) | (0.82) | (1.90) |
| Chile | 446 | 45 | 491 |
| | (3.97) | (0.40) | (4.37) |
| Bolivia | 4 | 1 | 5 |
| | (0.04) | (0.01) | (0.04) |
| Peru | 45 | 55 | 100 |
| | (0.40) | (0.49) | (0.89) |
| Brasil | 0 | 2 | 2 |
| | (0.00) | (0.02) | (0.02) |
| Argentina | 18 | 8 | 26 |
| | (0.16) | (0.07) | (0.23) |
| Colombia | 52 | 52 | 104 |
| | (0.46) | (0.46) | (0.92) |
| Other South American Countries | 5 | 16 | 21 |
| | (0.04) | (0.14) | (0.19) |
| Stateless | 7 | 4 | 11 |
| | (0.06) | (0.04) | (0.10) |
| Total | 5522 | 5724 | 11246 |
| | (49.10) | (50.90) | (100.00) |

*Notes:* The table shows the composition of non-European refugees from our sample from the placement policy.
*Source:* Own calculations based on data from the GeoSweden database.

Table 1.7: Sample of European refugee sending countries for the considered cohorts

| Refugee sending countries | 1990 | 1991 | Total |
|---|---|---|---|
| Denmark | 2 | 3 | 5 |
| | (0.07) | (0.11) | (0.18) |
| Finland | 1 | 4 | 5 |
| | (0.04) | (0.14) | (0.18) |
| Norway | 4 | 3 | 7 |
| | (0.14) | (0.11) | (0.25) |
| Iceland | 1 | 0 | 1 |
| | (0.04) | (0.00) | (0.04) |
| Yugoslavia | 291 | 354 | 645 |
| | (10.37) | (12.62) | (22.99) |
| Poland | 157 | 87 | 244 |
| | (5.60) | (3.10) | (8.70) |
| Romania | 122 | 304 | 426 |
| | (4.35) | (10.84) | (15.19) |
| Czech | 56 | 19 | 75 |
| | (2.00) | (0.68) | (2.67) |
| Hungary | 123 | 22 | 145 |
| | (4.39) | (0.78) | (5.17) |
| Greece | 9 | 3 | 12 |
| | (0.32) | (0.11) | (0.43) |
| United Kingdom | 2 | 4 | 6 |
| | (0.07) | (0.14) | (0.21) |
| Ireland | 6 | 9 | 15 |
| | (0.21) | (0.32) | (0.53) |
| Germany | 27 | 27 | 54 |
| | (0.96) | (0.96) | (1.93) |
| France | 3 | 1 | 4 |
| | (0.11) | (0.04) | (0.14) |
| Italy | 7 | 0 | 7 |
| | (0.25) | (0.00) | (0.25) |
| Spain | 3 | 2 | 5 |
| | (0.11) | (0.07) | (0.18) |
| Portugal | 0 | 4 | 4 |
| | (0.00) | (0.14) | (0.14) |
| Netherlands | 10 | 2 | 12 |
| | (0.36) | (0.07) | (0.43) |
| Austria | 4 | 2 | 6 |
| | (0.14) | (0.07) | (0.21) |
| Switzerland | 0 | 1 | 1 |
| | (0.00) | (0.04) | (0.04) |
| Bulgaria | 209 | 277 | 486 |
| | (7.45) | (9.88) | (17.33) |
| Other European Countries | 0 | 4 | 4 |
| | (0.00) | (0.14) | (0.14) |
| Estonia | 0 | 6 | 6 |
| | (0.00) | (0.21) | (0.21) |
| Former Soviet | 228 | 402 | 630 |
| | (8.13) | (14.33) | (22.46) |
| Total | 1265 | 1540 | 2805 |
| | (45.10) | (54.90) | (100.00) |

*Notes:* The table shows the composition of European refugees from our sample from the placement policy.
*Source:* Own calculations based on data from the GeoSweden database.

Table 1.8 shows the descriptive statistics for age, gender, marital status, percentage of refugees with children, refugees obtaining social welfare and low educated and the

Table 1.8: Summary statistics for refugee cohorts considered

| | 1990 | | 1991 | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| Age | 33.35 | (6.73) | 33.48 | (6.83) |
| Female | 0.38 | (0.48) | 0.37 | (0.48) |
| Married | 0.31 | (0.46) | 0.27 | (0.44) |
| Children | 0.56 | (0.50) | 0.53 | (0.50) |
| Deflated income | 15.60 | (87.53) | 14.90 | (85.39) |
| Social welfare | 0.87 | (0.33) | 0.82 | (0.38) |
| Low educated | 0.61 | (0.49) | 1.00 | (0.05) |
| African born | 0.14 | (0.35) | 0.19 | (0.39) |
| Latin American born | 0.10 | (0.31) | 0.04 | (0.20) |
| West Asian born | 0.54 | (0.50) | 0.51 | (0.50) |
| East Asian born | 0.07 | (0.26) | 0.10 | (0.29) |
| European born | 0.14 | (0.34) | 0.17 | (0.37) |
| Observations | 6787 | | 7264 | |

*Notes:* Standard deviations are reported in parentheses. The variables are measured at cohorts' arrival. The variable 'Low educated' comprises of individuals with less than high school education. Income is measured in 100s SEK.

*Source:* Own calculations based on data from the GeoSweden database.

percentage of refugees from different country regions. The average age at migration revolves around 33 years in both the 1990 and 1991 cohort. The majority of refugees are male across the cohorts. Less than half of the refugees are married over the cohorts, ranging from 27 to 31 percent, and nearly half proportion of refugees have children, ranging from 53 to 56 percent. For the education variable (share with less than high school education), there is some heterogeneity over the different cohorts, probably reflecting which countries the majority of the refugees come from in a certain cohort. However, most refugees are low educated. In terms of region of origin, the share of refugees born in West Asia dominates in both 1990 and 1991. The share of Latin American refugees, comprising of Colombians, remains low.

## 1.4 Results: Effects of Initial Neighborhood Quality on Future Neighborhood Quality

To begin with, we are interested in examining the effects of the quality of the refugee's initial neighborhood on the quality of their future neighborhood (i.e., residential integration in terms of socio-economic and demographic characteristics).

These estimates are interesting for a couple of reasons. First, (native) neighborhood integration is an important dimension for policy makers (and was one of the underlying aims of the placement policy). Second, as argued earlier, the quality of the neighborhood one lives in can affect several important outcomes. In the next section, when we examine the placed adults' labor market outcomes, this will serve as the instrument in the first stage to generate an exogenous variation in the quality of the refugees' future neighborhoods, implying that the results here will indicate whether the instrument is relevant.

In the next section, we discuss our empirical specification, and in section 1.4.2, we present our results.

### 1.4.1 Empirical Specification

To estimate the causal effect of initial neighborhood characteristics on future neighborhood characteristics, we estimate yearly cross section equations of the following format:

$$neighborhood_{i,t+z} = \beta_0 + \beta_1 neighborhood_{it} + X_{it} + \delta_a + \epsilon_{it} \tag{1.1}$$

where $neighborhood_{i,t+z}$ represents the quality of the neighborhood in the refugees' individualized neighborhoods after $z$ years, and $neighborhood_{it}$ represents the same shares in the refugees' initial individualized neighborhoods. $X_{it}$ is a set of socio-demographic characteristics and country of origin controls, and $\delta_a$ are municipality fixed effects. This specification is in the style of the first stage estimation in Edin et al. (2003), in which they use the share of co-ethnics in the initial municipality on the share of co-ethnics in the municipality in which the refugees live in $t + 8$. The main differences between our specification and theirs is that we (i) take a more disaggregated approach and look at the characteristics of the refugees very closest neighbors, (ii) focus on other characteristics than co-ethnics, and (iii) look at the time-dynamics (having a follow-up horizon of eight years).[16]

We will estimate equation (1.1) with the quality index as the dependent variable as well as separately for each of the neighbors' characteristic since we think it is important to learn how the refugees integrate in these dimensions, especially in terms of share of natives.

### 1.4.2 Results for Quality of Neighborhood

To examine whether the quality of the refugees' initial neighborhood matters for the quality of their future neighborhood quality, we start out by estimating Equation (1.1) with the estimated index of neighborhood quality as the dependent variable. From the results, presented in Figure 1.5(a) for $k = 100$, there is a clear indication that the quality of the refugees' initial neighborhood matters for the quality of their future neighborhood. Over the 8-year follow-up horizon, there is always a positive effect of the quality of the neighborhood in the initial year (the placement year) on the quality of future neighborhoods. While it drops during the first few years, it stabilizes after approximately five years, indicating that a 1% increase in the initial neighborhood quality leads to an approximately 0.1% higher index of the neighborhood quality in year $t + 5$.

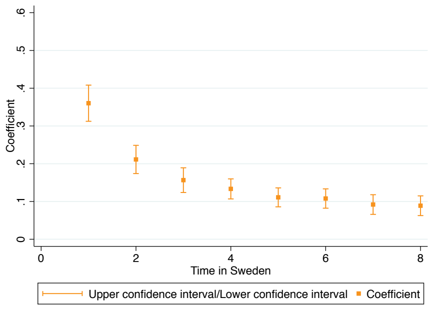
Since measuring the quality of a neighborhood (through the quality of its inhabitants) is a complex phenomenon, it is of interest to look at the separate components used in the construction of the neighborhood quality index. Is it one specific component that drives the results, or do we see similar effects for all the characteristics? From the results, presented in Figures 1.5(b)-(d), it is clear that we see a similar pattern for all three characteristics as for the quality index, even though, in relative terms, the share of natives and the share of high-educated among the $k$-nearest neighbors in the initial neighborhood seem to matter a bit more in the long run compared to the share of employed.

Since an important aim with the placement policy was to combat ethnic segregation (clustering of immigrants/refugees to certain areas), it is very interesting to note the

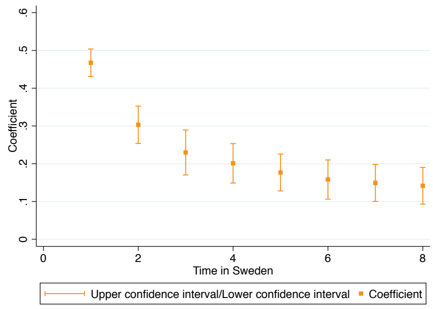
[16] Equation (1.1) is in the same spirit as the first stage in Edin et al. (2003), Åslund et al. (2011), and Andersson (2020).

results for share of natives. A 1% increase in the share of natives among the 100 nearest neighbors in the initial year has the effect that the refugees' will live in a neighborhood with an approximately 0.35% higher share of natives in $t+1$ and more than a 0.1% higher share of natives after eight years (c.f. Figure 1.5(b)). Given that there is an aim for ethnic residential integration, it is clear that the placement policy contains an important aspect: by placing refugees in neighborhoods with more natives, they affect the share of natives among the refugees future neighbors.[17]

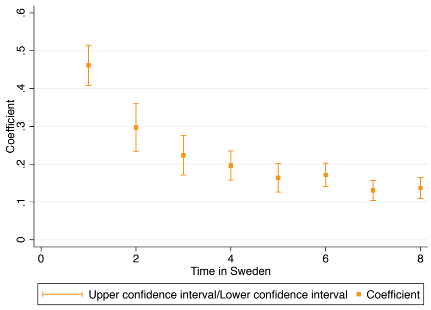
Figure 1.5: Effects of initial neighborhood quality on the quality of future neighborhoods: $k = 100$



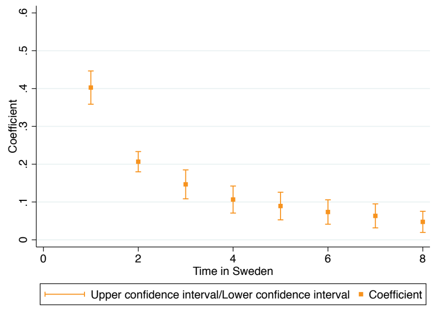

(a) Quality index




(b) Share natives




(c) Share educated




(d) Share employed

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (1.1) and their corresponding 95% confidence intervals.
*Source*: Own calculations on data from the GeoSweden database.

### 1.4.3 Results for Movers

An issue might be that those that stay in the initial (placed) location do not fully reveal their future neighborhood preferences, at least not in the short run (if some of them want

[17]The exact estimates behind Figure 1.5 are presented in Table A.10 - A.13 in the Appendix A.3.

to move but, for some reasons are unable to do so). In addition, if those that do not move to a larger extent were placed in neighborhoods with more advantageous characteristics among their neighbors, there might be a mechanical effect if they do not move.

Movers are defined as any refugee who has a different residential coordinates as their initial one. Table 1.9 shows the summary statistics for movers. There is about 55 percent of refugees that move in $t + 1$ and by $t + 8$, most refugees have moved from their initial residential coordinates.

Table 1.9: Movers

| Duration year | Number of Movers |
|---|---|
| 1 | 7701 |
| | (54.92) |
| 2 | 10345 |
| | (73.77) |
| 3 | 11525 |
| | (82.21) |
| 4 | 12110 |
| | (86.41) |
| 5 | 12492 |
| | (89.11) |
| 6 | 12797 |
| | (91.20) |
| 7 | 12935 |
| | (92.85) |
| 8 | 12894 |
| | (93.77) |

*Notes:* The table shows the absolute number of movers for each year and the percentages are given in parentheses.

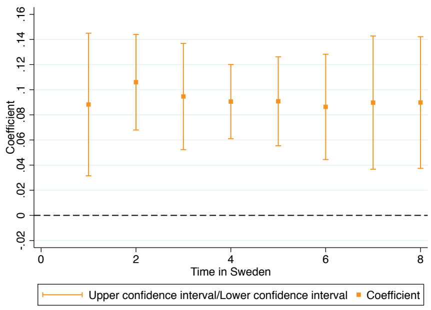
*Source:* Own calculations based on data from the GeoSweden database.

To examine if the results are sensitive to this, we estimate the model for those refugees that have actually moved since the initial placement. These results are presented in Figure 1.6.[18] We see more stable parameter estimates over time, estimates that are more in line with the longer-run effects seen in the baseline analysis. This indicates that it is

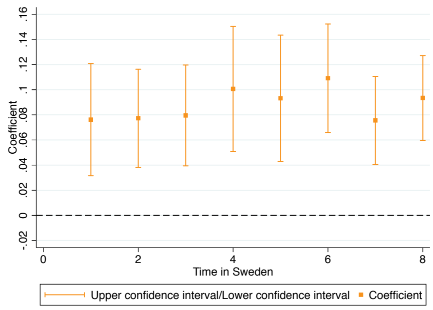
[18] When only looking at movers, we get a smaller sample, namely 7683 movers in t+1, which implies that we get missing values for several municipalities when constructing the neighborhood quality index. Since the small sample yields a lot of uncertainties in the estimates, we do not present the results for the quality index. If there are a small number of refugees in a municipality for a year and they all have the same share of natives employed and high educated, the neighborhood quality index is missing for these refugees.

those that have not yet moved in the first few years after placement that drive the larger initial effects, and that the longer-run results are more in line with the effects found after everybody has found their residential equilibrium.

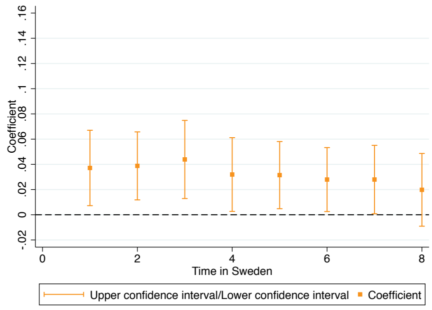
Figure 1.6: Coefficient plots of initial share on future share for movers: $k = 100$




(a) Share natives




(b) Share educated




(c) Share employed

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (1.1) for the movers and their corresponding 95% confidence intervals. The number of observations is 7701 in t+1 and 12894 in t+8.
*Source*: Own calculations on data from the GeoSweden database.

### 1.4.4 Robustness Checks

We have conducted a number of alterations to the baseline model to check the robustness of the results. In this section, we present which alterations we have done and what we have found. The full results are presented in Appendix section A.4.

#### Other Definitions of the Size of the Neighborhood

To examine how sensitive the results are to the choice of neighborhood size $k = 100$, we have re-run the baseline model for $k = 50$ and $k = 250$. It turns out that the exact choice among these three bespoke neighborhood definitions does not matter for the results.

#### Restricting to Refugees Arriving with Children

Since there are indications in municipal documents from the time of the placement policy that refugees with children were prioritized in the municipal housing queue, indicating that those refugees were the first to receive an apartment when one became available for those in the queue, we have estimated the model for parents only. The baseline results are not sensitive to this restriction.

#### Restricting the Distance Needed to Reach the $k = 100$ Closest Neighbors and Restricting the Actual Number of Neighbors Reached when Searching for the $k = 100$ Closest Neighbors

As discussed in section 1.2.2, the algorithm sometimes need to go a long distance to find a refugee's 100 nearest neighbors (c.f. Figure 1.1) and sometimes it will overshoot (getting more neighbors than searched for; c.f. Figure 1.2). To check how sensitive the results are to these specifications, we have, first, restricted the distance to reach the 100 nearest neighbors to be only 100 meters (which cover over 70% of the refugees' neighborhoods) and, second, we have restricted the actual number of neighbors reached when searching for the $k = 100$ closest neighbors to be 200 or less (which also cover over 70% of the refugees' neighborhoods). The baseline results are not sensitive to these alterations.

## 1.5 Results: Effects of Neighborhood Quality on Earnings for Adult Refugees

In the former section, we noticed that the characteristics among the closest neighbors in the refugees' initial neighborhood matter for their future neighborhood composition. In this section, we will examine how the quality of bespoke neighborhoods affects earnings for the placed adults.

In the next section we discuss our empirical specification, and in section 1.5.2, we present our results.

### 1.5.1 Empirical specification

To estimate the causal effect of neighborhood characteristics on earnings, we are interested in estimating an equation of the following format:

$$y_{i,t+z} = \beta_0 + \beta_1 neighborhood_{i,t+z} + X_{i,t+z} + \delta_a + \epsilon_{i,t+z} \tag{1.2}$$

where $y_{i,t+z}$ denotes a refugee's log earnings after $z$ years and $neighborhood_{i,t+z}$ represents the quality of the neighborhood in which the refugee lives in year $t+z$ (as given by share of natives, share high-educated individuals, share employed individuals, and the quality of the neighborhood index in the refugees' individualized neighborhoods, respectively). Since refugees move, and hence self-select, into different neighborhoods in the $z$ years following the initial placement, the methodological problem to solve is the endogeneity of the neighborhood quality in year $t + z$. To solve that problem, we use the quality of the neighborhood in the initial placement year (year $t$) to instrument for the neighborhood quality in year $t + z$ (that is, we estimate Equation 1.1 in a first stage to predict $neighborhood_{i,t+z}$ in a 2SLS analysis). As was clear from the results in section 1.4, the instrument is valid.

Like earlier, $X_{i,t+z}$ is a set of socio-demographic characteristics and country of origin controls[19] and $\delta_a$ are the municipality fixed effects allowing us to control for local amenities, including the local labor market situation. The model is estimated on yearly cross sections, with $z = \{1, ..., Z\}$, to get an understanding of how the estimated effects evolve over time.[20]

Once again, it can be noted that this specification is in the style of the 2SLS estimation in Edin et al. (2003), where they examine how the share of co-ethnics in the refugees' municipalities eight year after arrival affects their labor market outcomes by using the initial

[19] Here we follow Edin et al. (2003) and measure these variables in $t + z$, but we have also measured the $X$-variables in the initial year $t$ to avoid potential problems with bad controls; the results are not sensitive to this.

[20] The identifying variation hence comes from within-municipality variation in neighborhood quality among the refugees.

share of co-ethnics to instrument for the share in $t+8$.[21].The main differences between the specifications are the geographic scale (individualized neighborhoods instead of municipalities), the characteristics analyzed (share natives, educated and employed instead of share co-ethnics), and the follow-up horizon (all eight years after initial placement instead of only year eight).

Even if the instrument is relevant in the sense that it is statistically significant in the first-stage estimation, it might be weak. Weak instruments will lead to biased estimates in the 2SLS estimates. We test for this through the first-stage F-statistic (the results for the F-statistics are presented in Tables A.10, A.11, A.12 and A.13. Stock, Yogo, et al. (2005) proposes a cutoff at 10 for the first-stage F-statistics. In our case, the F-statistics are well above 10 for all characteristics in the early years of the follow-up period. Over time, the F-statistic decreases in size, but stays above 10 in most estimations, the exception being for the share of employed in the later years of the follow-up period (years $t+6$ to $t+8$) where it is slightly below 10. We consider this as reassuring test results.

### 1.5.2 Results

The results for earnings are presented in Figure 1.7.[22] Starting with the neighborhood quality index (see Figure 1.7(a), we can first note that the point estimates are positive for all years, indicating that the quality of the refugees' neighborhoods might have a positive impact on their labor market outcomes, in terms of earnings. For the years covering the period $t+2$ to $t+5$, the point estimates are also statistically significant at least at the five percent significance level. The magnitudes of the point estimates are however not large. To take $t+5$ (which is in year 1995 or 1996, depending on year of arrival) as an example, the point estimate indicates that a ten percent increase in the neighborhood quality index yields 0.05% higher yearly earnings.

[21]This type of 2SLS design has been used in other research as well, see for instance Andersson (2020) in a study on the effect of ethnic enclave on self-employment and labor market outcomes of refugees. Similarly, Åslund et al. (2011) uses the initial residential location of refugees as an instrument to examine the effect of neighborhood characteristics on school performance.

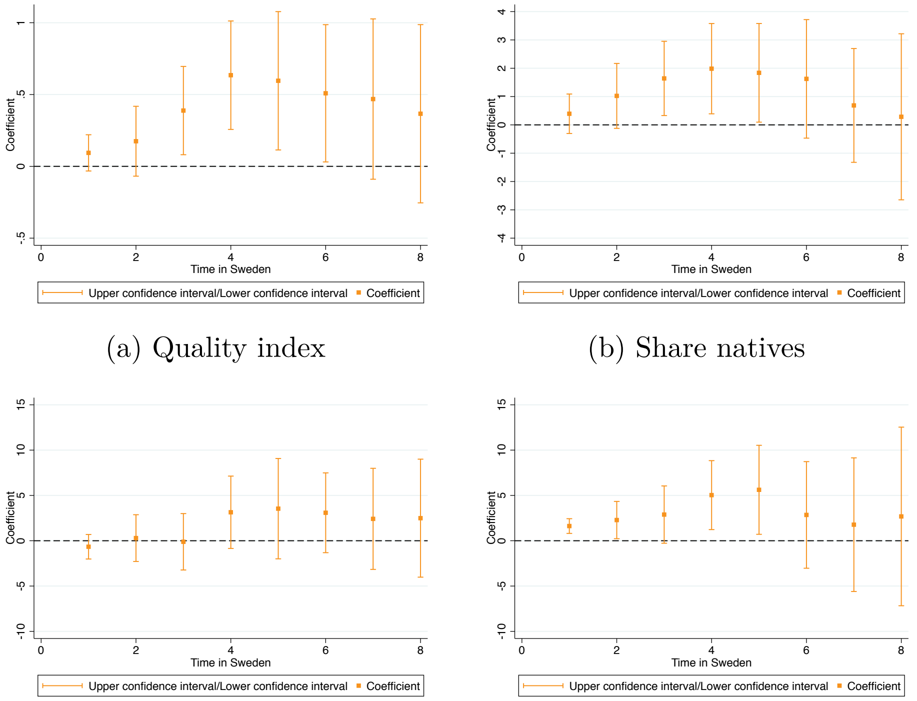
[22]The full results are presented in the tables in Appendix Section A.5.

Turning to the specific components in the neighborhood quality index, the share of natives and the share of employed among the 10 nearest neighbors follow a similar pattern over time as the index. Also in these cases are the estimated magnitudes small. In $t + 5$, the point estimate indicates that a ten percent increase in the share of natives (share of employed) among the 100 closest neighbors yields 0.19% (0.56%) higher yearly earnings.

The characteristic among the closest neighbors that does not seem to matter for earnings is the educational level. The share of high-educated individuals among the 100 nearest neighbors yields statistically as well as economically insignificant results for all years in the follow-up horizon.

We hence find less significant results in the initial years. One interpretation of this finding is that it takes time to form ties at the neighborhood level. A possible interpretation of the insignificant results in the final follow-up years is that in the late 1990s, the Swedish economy was booming, with low unemployment rates. In such a situation, the quality of neighbors might be less important than in hard economic times.

Figure 1.7: Instrumental variable 2SLS: $k = 100$




(a) Quality index

(b) Share natives

(c) Share educated

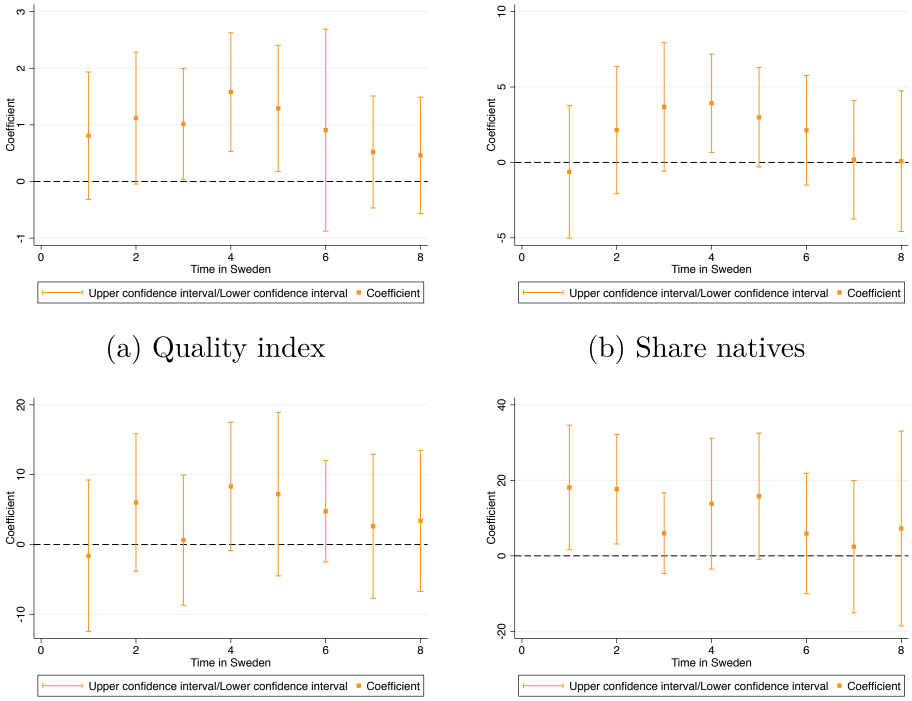
(d) Share employed

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (1.2) and their corresponding 95% confidence intervals. The graphs' y-scales are different for quality index and share of natives.
*Source*: Own calculations on data from the GeoSweden database.

### 1.5.3 Results for Movers

Similar to section 1.4.3, we examine how sensitive the results are to restricting the sample to only movers. From the results, presented in Figure 1.8, it seems like they are fairly similar, both in terms of estimated pattern over time and in terms of magnitudes of the point estimates. One can however note that the most stable results are for the neighborhood quality index and share natives. For share employed, the uncertainty in the point estimates increases.

Figure 1.8: Effects of initial neighborhood quality on future earnings for movers: $k = 100$




(a) Quality index (b) Share natives

(c) Share educated (d) Share employed

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (1.2) for movers and their corresponding 95% confidence intervals. The number of observations is 7701 in t+1 and 12894 in t+8. *Source*: Own calculations on data from the GeoSweden database.

### 1.5.4 Robustness Checks

We have conducted the same type of robustness tests as we did for the effects on future neighborhood quality.[23] The overall conclusion from this robustness analysis is that the results are not specifically sensitive to the alterations to the baseline model that we do. The estimated time pattern is very much the same for all measures of quality index, the most stable results are for the quality index and for share of natives (it is also for these quality measures that we mainly get statistically significant results), and the magnitudes of the point estimates are in all cases small.

[23] All robustness results are presented in Appendix section A.6.

## 1.6 Conclusions

In this paper, we have examined what role the socio-economic and demographic characteristics of the closest neighbors in the initial neighborhood of newly arrived refugees in Sweden play for future outcomes. We use fine-grained coordinate data and examine two specific, but related, questions. First, we examine their role in predicting what type of neighborhood the refugees' will live in in future years. Second, we examine if the characteristics of the neighbors in the refugees' future neighborhoods matter for their labor market outcome, as measured via earnings. We create small neighborhoods using the $k$ nearest neighbor approach, as we hypothesize that the closest neighbors may be a source for interaction and information, and measure the quality of the neighbors as share of natives, highly educated and employed neighbors.

Using the placement policy to solve self-selection into neighborhood issues, and implementing an empirical specification in the style of Edin et al. (2003), this paper reaches two main conclusions. First, we find that the quality of the refugees' initial neighbourhood affect their future neighborhood quality positively throughout an 8-year time horizon, irrespective of the neighbor characteristics looked at, i.e. in terms of share highly educated, share employed, share natives, and a constructed neighborhood index. The results hold for $k = 100$ and $k = 250$, for refugees arriving with children and distance restriction of 100 meters to find the nearest neighbors. Looking at a sample of movers, we notice that the estimates are more stable over years and indicate that stayers in the initial neighborhood drive our results. Second, we find weak indications that the neighborhoods in which the refugees live matter for labor market outcomes (earnings), particularly in the later years as it may take time to form networks. The pattern holds for movers and the results are mostly stable for share of natives and neighborhood quality index.

Overall, our results indicate that networks in the local neighborhood play a role in the residential integration of refugees. This is a conclusion that is in line with the interpretation of the results in Bayer et al. (2008) and Conley and Topa (2002). We can however not rule out that the initial neighborhood might affect the refugees' preferences for having neighbors with certain characteristics. Regardless of the mechanism, it is clear that a placement policy might have an important role to fulfill if policy makers care about residential integration. If the aim of the policy is to combat residential segregation, it looks like it is successful in doing so over time.

Recent evidence show that the neighborhood a child grows up in matters for his or her future outcomes (Chetty et al., 2016a, 2020). Given our results that the quality of the initial neighborhoods matters for the quality of future neighborhoods, the placement of refugees might have important, long-run implications for the refugees' children. Examining what role the quality of the neighborhood has for the refugee children's future educational outcomes is on the top of our research agenda.

We have examined the role of natives among the refugees closest neighbors for the refugees' labor market outcomes. We find signs of positive effects, but where the estimated size of a potential effect is very small. In the co-ethnics literature, where a finding on Swedish data has been that the share of co-ethnics matter for the refugees earnings, they have typically measured the share of co-ethnics at a quite large geographic scale, i.e. municipality. It would be of interest to adopt the approach taken in this paper and estimate the effects of the share of co-ethnics at a more granular scale and see if the results changes. That could say something about mechanisms. Ongoing research looks further into the effect of ethnic enclaves at a smaller geographic scale on residential integration and labor market outcomes.

# Chapter 2

# Local Ethnic Enclaves and Labor Market Outcomes: Small-Scale Geographical Analysis

## 2.1 Introduction

The relevance of ethnic enclaves for labor market outcomes of immigrants has been recognized in existing economic research (Beaman, 2011; Damm, 2009; Edin et al., 2003; Martén et al., 2019). Empirical evidence in Sweden suggests that residing in ethnic enclaves at the municipality level has a positive effect on earnings, particularly for less skilled immigrants (Edin et al., 2003). So far, the mechanisms relating ethnic enclaves to economic success of immigrants remain understudied. A potential mechanism through which the effect of ethnic enclaves can impact economic outcomes is through information dissipation within ethnic networks by means of daily local interactions. Small geographical scales are required to analyze if the effect of ethnic enclaves on future share of co-ethnics and labor market outcomes occur at a small geographical scale where networks and interactions are likely to take place.

Research by Galster (2008) shows that analyzing small-scale residential segregation is crucial as the process of socialization occurs through contact with peers in the neighborhood. Consequently, the behaviors as well as attitudes of a neighborhood resident can impact his neighbor by means of social interaction on a regular basis (Johnston and Pattie, 2011). Neighbors can thus form an important part of social networks and diffuse information, knowledge and resources, which could increase labor market and other economic opportunities (Ellen and Turner, 1997). They can act as an informal hiring network and share information about job opportunities or job trainings (Bertrand et al., 2000) because they are the ones that the refugees are likely to meet on a daily basis. Given that research has shown that the neighbors who are in close proximity matter most for economic outcomes in a more general context (Bayer et al., 2008; Conley and Topa, 2002), it is worth investigating into whether networks and local interactions with co-ethnics and employed co-ethnics operate via small geographical level for future neighborhood compositions and labor market outcomes of refugees.

Using detailed Swedish geocoded data containing coordinates on a 100 by 100 meter grid, and a novel approach, namely the $k$-nearest neighbor approach, this paper presents new evidence on the effect of ethnic enclaves on residential integration and labor market outcomes on small-scale geographical scales for refugees. I define various small neighbourhood sizes, and investigate whether small-scale neighborhood sizes matter through daily interactions and can be crucial for knowledge spillovers for living with co-ethnics as well as employed co-ethnics in the future neighborhood and labor market outcomes. I hypothesize that the share of co-ethnics in the refugees' small neighborhoods can provide newly arrived refugees an informal hiring network and allow them to participate in the labor market. As empirical evidence points out that high quality co-ethnics are the ones who drive labor market integration of incoming refugees (Damm, 2009, 2014), I also characterize the share of employed co-ethnics among the neighbors of the refugees: the closest co-ethnics who are in the labor market can provide crucial information for the refugees.

The main challenge to identifying the causal effect of local ethnic enclaves on integration outcomes is that refugees can self-select into neighborhoods. Therefore, the choice of neighborhood is probably affected by unobserved factors which in turn have an impact on earnings and employment. To address this concern, several studies have used ethnic enclaves' size variation across cities or regions (Borjas, 2000; Cutler et al., 2008). Moreover, a study by Boeri et al. (2015) exploit variation at the residential housing blocks in Italy to deal with the selection problem. The ideal approach to dealing with sorting problem and establishing causal inference in the literature is the use of refugee placement policies[1] as natural experiments (Beaman, 2011; Damm, 2009; Edin et al., 2003). Refugee dispersal policies have the aim of spreading asylum seekers and refugees to residential locations upon their arrival in the host countries and decentralizing immigrants from big cities to other areas.[2] In this paper, I will use the Swedish refugee dispersal policies and exploit exogenous variation within municipalities. I argue that exactly which apartment a refugee resided in was exogenous from his perspective and therefore, the refugee was treated with different co-ethnic and employed co-ethnic neighbors.

[1] In this paper, the terms refugee dispersal policy and refugee placement policy are used interchangeably.

[2] The refugee placement policy has not only been used as an instrument to examine labor market outcomes, but also to analyze educational outcomes, health outcomes (Grönqvist et al., 2012), criminal outcomes (Grönqvist et al., 2015), election outcomes (Dustmann et al., 2016) and attitudes to the welfare state (Dahlberg et al., 2012).

This paper uses administrative data to identify refugees arriving during the placement policy. As the data comprises of detailed geographical coordinates on a 100 by 100 meter grid, it allows the creation of small individualized neighborhoods of various sizes, including $k = 100$, $k = 250$, $k = 500$ and $k = 1000$ nearest neighbors. The $k$-nearest neighbor approach enables small scale analysis, which matters for the arriving refugees' integration. The nearest neighbors are the individuals that the refugees have a higher likelihood to meet. Using small scale neighborhoods enable taking into consideration socialization as well as network patterns and noticing clustering of immigrants of the same ethnicity, which would otherwise be unnoticed at the municipality level. Clustering of co-ethnics and employed co-ethnics in current and future residential locations suggests if networks at a small scale is important. Moreover, clustering of co-ethnics around the newly arrived refugees' who are employed and have an earnings indicates whether information spillovers through interactions with high quality networks.

The baseline results indicate that the effect of initial co-ethnics and employed co-ethnics share in the neighborhoods have positive and statistically significant effects on future co-ethnics and employed co-ethnics in the neighborhoods, regardless of the $k$-nearest neighbor level examined and over different time horizons, i.e. $t + 4$ to $t + 10$, suggesting that meeting co-ethnics in the local area matters. The impact of initial co-ethnics share decreases over time for $k = 100$ and $k = 250$ nearest neighbors. When looking at the effect of co-ethnics and employed co-ethnics on residential integration, I also examine how robust the results are to various estimations, including distance restriction to 500 meters and refugees arriving with children. When looking at movers only, the results are more similar to the effects in the long run, i.e. $t + 10$. The results show that the effect of initial ethnic share is positive and stable to these specifications.

In the next step, I investigate the effect of ethnic share on labor market outcomes from $t + 4$ to $t + 10$, and the results show that the share of co-ethnics and employed co-ethnics have positive and statistically significant effects on earnings and employment from t+8 onwards. This result is in line with Edin et al. (2003) who find a positive impact of ethnic enclaves 8 years after the refugees' arrival, and indicates that it takes time to build networks. The result could also suggest that the recession in Sweden in the 1990s could have also impacted the fact that it takes time for the network effects to impact the refugees' labor market outcomes. The magnitude of this effect is larger at the higher $k$-nearest neighbor level, showing that there is a higher probability of interacting with skilled co-ethnics at a larger geographical scale and these skilled co-ethnics could then disseminate information for the new refugee arrivals. If the quality of the neighborhood is more crucial than just being with co-ethnics, socio-economic residential integration at a small scale geographical level can happen by locating the refugees in neighborhoods with employed neighbors.

This study contributes to the literature on ethnic enclaves in two dimensions. While most research pertaining to the effect of ethnic enclaves on labor market outcomes yield positive effect in terms of fostering economic integration, decreasing job search costs and asymmetric information and participation in ethnic economy[3] (Beaman, 2011; Damm, 2009; Drever and Hoffmeister, 2008; Edin et al., 2003; Martén et al., 2019; Munshi, 2003; Patacchini and Zenou, 2012), the majority of studies to date uses municipality level or city level data due to lack of detailed geographical data (Borjas, 2000; Cutler et al., 2008; Edin et al., 2003). Exploiting exogenous variation within municipalities does not allow to identify if the mechanism of the effect of ethnic enclaves occurs on a small geographical scale and whether information exchanges occur through small scale neighborhoods. Municipalities are also of varying size, with some comprising of higher population density than others. I contribute to the literature by examining small scale neighborhoods since the closest neighbors residing among the refugees matter. A growing body of literature in economics shows that there are substantial labor market benefits from residing close to neighbors who have jobs and that there is a positive effect of social networks on job finding probability, especially for low skilled individuals (Bayer et al., 2008; Calvo-Armengol and Jackson, 2004; Schmutte, 2015; Wahba and Zenou, 2005). Therefore, I create small neighborhood sizes to investigate into potential interactions with co-ethnics.

[3]There are also concerns that living in ethnic enclaves leads to immigrants being segregated from natives, limitation of language assimilation, hinderance of labor market integration as well as upward mobility in their jobs, and lowering wage growth (Battu et al., 2011; Borjas, 2000; Cutler et al., 2008; Danzer and Yaman, 2013; Logan et al., 2003; Xie and Gough, 2011).

Second, this paper benefits from improved data in terms of being able to identify exactly who are refugees, as compared to previous literature on ethnic enclaves. The data used in this paper contains a variable on reasons for immigration from which I can distinguish if an individual immigrates to Sweden as a labor market immigrant, a student, a tied family member or a refugee. Most papers identify refugees through country of birth or origin and thus, lead to oversampling of refugees. Edin et al. (2003) use information on country of birth to identify the refugees and the composition of refugees include those from non-OECD countries, with the exception of Turkey. The inclusion of many countries may also exacerbate the oversampling issue and some refugees would access a smaller network than others. Several other papers use country of origin to compose their refugee sample and only includes the largest refugee sending countries (Damm, 2009, 2014). Although Boeri et al. (2015) use data at a small geographical level,i.e. housing blocks, their sample comprises of both legal and illegal migrants in Italy. In contrast, the current paper benefits from improved data which allows me to exactly identify refugees. This paper thus adds to the literature on refugees' labor market and other forms of integration (Aksoy et al., 2020; Battisti et al., 2019; Bratu et al., 2021; Brell et al., 2020; Dahlberg and Valeyatheepillay, 2021; Dahlberg et al., 2020).

This paper is structured as follows. The next section outlines the data source and presents the $k$ nearest neighbor approach. Section 2.3 introduces the identification strategy and presents balancing tests. Section 2.4 presents the empirical specification to investigate the effect of initial co-ethnics and employed co-ethnics on future residential integration. This Section is followed by Section 2.4.2 and 2.4.3 discussing the results. Section 2.5 presents the effects of co-ethnics and employed co-ethnics on future labor market outcomes and 2.5.2 discusses the results. Finally, Section 2.7 concludes.

## 2.2 Data and Neighborhood Definitions

### 2.2.1 Data Sources and Definition of Refugees

The empirical analysis presented in this paper uses GeoSweden, a rich Swedish administrative data, collected yearly from 1990 to 2014 by Statistics Sweden. The data comprises of population as well as tax registers, and includes all residents living in Sweden at the end of December every year. The data presents an opportunity to examine the effect of ethnic enclaves at a small geographical scale because it contains coordinates on a 100 by 100 meter grid of where all individuals live. The detailed geographical units enable the construction of individualized neighborhood based on the $k$-nearest neighbor approach and get a representation of how many ethnic and employed ethnic neighbors are among the refugees' individualized neighborhoods.

Yet another feature of the GeoSweden database is that it contains a variable on the reason for immigration from which I can identify whether an immigrant comes to Sweden to study, work, as a tied family member or a refugee. Therefore, I can exactly identify the individuals who obtain refugee permits and thus obtain the exact number of refugees in my sample.

### 2.2.2 k Nearest Neighbor Approach

To gain insight into the potential interaction mechanism at work on residing in ethnic enclaves, it is important to construct different small neighborhood scales and examine at which geographical scale ethnic network plays a role for residential and labor market outcomes of the refugees. This research uses the Equipop software to construct individualized neighborhoods based on the population size (Östh, 2014). Compared to using municipalities as area of observations, where the number of individuals varies largely within different municipalities, the $k$-nearest neighbor approach presents the advantage that the number of individuals among the nearest neighbors is rather constant. For the purpose of this research, the share of individuals from a certain ethnicity and the share of employed ethnics living among the refugees' nearest neighbor are constructed for each year from 1990 to 2014.

The steps involved in calculating the individualized neighborhood are as follows: initially, for every year, I locate each 100 by 100 meter coordinate in Sweden where individuals live. Then, I calculate the total population stock and the number of individuals of a given ethnicity residing on that specific coordinate.[4] In this paper, the ethnic stock in

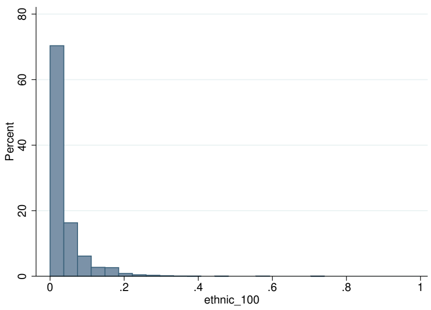
[4] The individual counts of neighbors include all individuals aged 18 and above.

Sweden is calculated because existing members of an ethnic community who resided in Sweden for a certain length of time can disseminate information about the country and how to integrate in the labor market. As Beaman (2011) shows, the vintage of ethnic networks matters for labor market integration of immigrants. The variable ethnic stock is created by totalling the number of individuals born in each of the countries in my sample. In the last step, Equipop calculates the individualized neighborhoods of the refugees by identifying the characteristics of the $k$-nearest neighbors for each refugee. For the purpose of this paper, the share of co-ethnics among the refugees' 100, 250, 500 and 1000 nearest neighbors and share of employed co-ethnics among the refugees' 250, 500 and 1000 nearest neighbors are calculated.

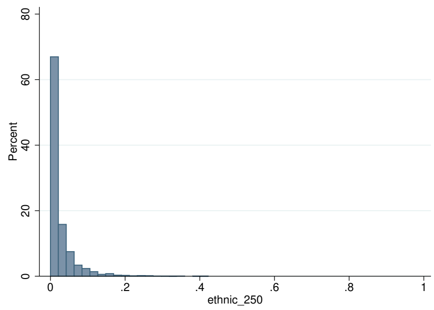
Smaller scale neighborhoods than municipalities allow the assumption that the refugees interact with their neighbors and thus, obtain information that would be valuable for the housing and labor markets. The hypothesis is that ethnic network may operate much more at small individualized neighborhood sizes as ethnic neighbors are more likely to meet each other on a daily basis when they live close by. Spatial proximity can thus matter for information sharing and formation of social network. For instance, Bayer et al. (2008) show that social interactions with individuals living in the same block have an effect on labor market outcomes. Therefore, $k = 100$ is used to capture small scale neighborhood and is indicative of neighbors that the refugees meet in their apartment blocks and recognize as neighbors. Using $k = 250$ would mean interacting with individuals from similar ethnic backgrounds when refugees are with their children at the playground. $k = 500$ nearest neighbor is representative of interacting with individuals from similar ethnic backgrounds when refugees are with their children at the playground. $k = 1000$ would be indicative of ethnic neighbors that the refugees might meet at the local bus stop, train station, at the local shopping center or ethnic clubs.

The analysis distinguishes between 10 different ethnic groups in Sweden; Iranian, Iraqi, Lebanese, Syrian, Ethiopian, Somalian, Vietnamese, Yugoslavian, Turkish and Chilean. Figure 2.1 shows the distribution of the share of co-ethnics in the refugees' initial arrival year (see Table B.1 in Appendix B.1 for detailed summary statistics on the shares). From Figure 2.1, it is apparent that most individuals live among few co-ethnics, regardless of the $k$-nearest neighbor. On average, the ethnic share is rather small and is mostly concentrated at the lower end of the distribution, around 1 to 4 percent, irrespective of the $k$-nearest neighbor. However, there is a variation in this share depending on the area that the refugees reside in and the share can be as high as 83 percent in certain areas for $k = 100$ and 53 percent for $k = 1000$.

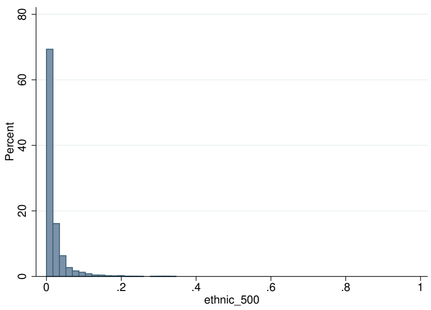
Figure 2.1: Distribution of ethnic share at k nearest neighbor: $k = 100, k = 250, k = 500, k = 1000$



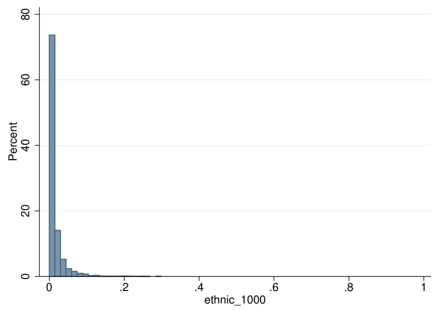

(a) $k = 100$




(b) $k = 250$




(c) $k = 500$




(d) $k = 1000$

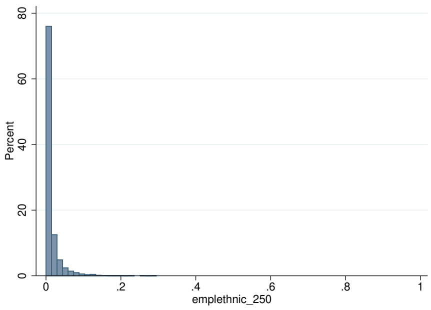
*Notes:* This Figure illustrates the distribution of ethnic share among the $k$ nearest neighbors.
*Source*: Own calculations on data from the GeoSweden database.

Moreover, I define share of employed co-ethnics among the $k$-nearest neighbors because the quality of enclaves matters and employed co-ethnics can provide information and provide job referrals.[5] The distribution of the employed co-ethnics among the nearest neighbors in the refugees' initial arrival year can be found in Figure 2.2 (see Table B.2 in Appendix B.1 for detailed summary statistics on the shares). It can be noticed that,

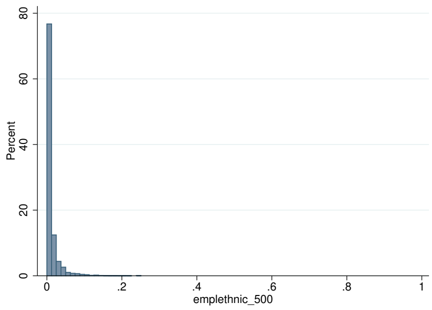
[5]The employed co-ethnics share is not defined for $k = 100$ as the share is too small and most refugees have zero neighbors who are employed co-ethnics at $k = 100$ nearest neighbor level.

on average, there is relatively low employed ethnic share among the refugees' nearest neighbors, irrespective of the $k$-nearest neighbor levels.

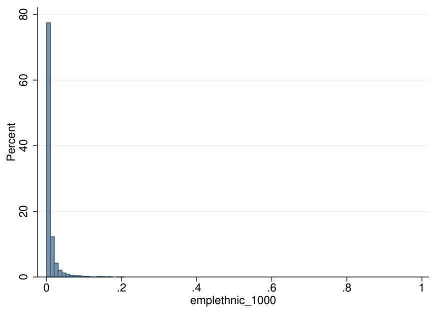
Figure 2.2: Distribution of employed ethnic share at k nearest neighbor: $k = 250, k = 500, k = 1000$




(a) $k = 250$




(b) $k = 500$

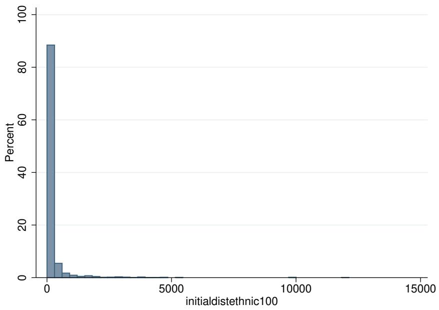



(c) $k = 1000$

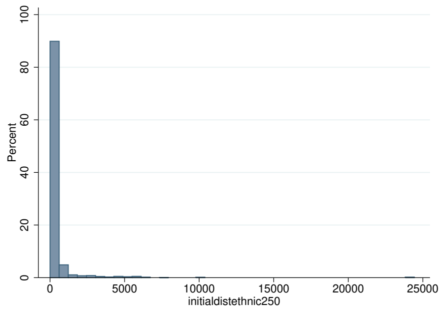
*Notes:* This Figure illustrates the distribution of employed co-ethnics among the $k$ nearest neighbors.
*Source*: Own calculations on data from the GeoSweden database.

If refugees live in areas which are less densely populated, the algorithm for constructing individualized neighborhoods need to search over a long distance to find the nearest neighbors and hence, the co-ethnics residing further away may not be able to interact with the refugees. As shown in Figure 2.3, most refugees live within a short distance to their neighbors. About 93 percent of the refugees live within a 500 meters proximity to their 100 nearest co-ethnic neighbors. At $k = 250$ and $k = 500$ levels, there are about 86 percent and 73 percent co-ethnics living among the refugees' nearest neighbors. There is about 50 percent of refugees living within 500 meters to their $k = 1000$. Therefore, this paper will conduct robustness checks based on the distance.

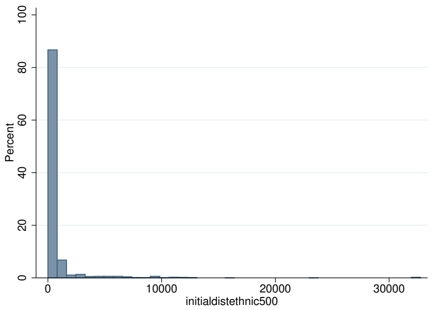
Figure 2.3: Distribution of distance to find the refugees' closest neighbors: $k = 100, k = 250, k = 500, k = 1000$



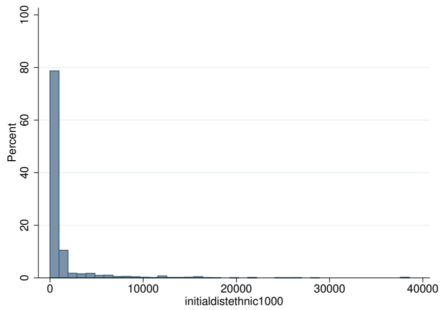

(a) $k = 100$




(b) $k = 250$




(c) $k = 500$




(d) $k = 1000$

*Notes:* This Figure illustrates the distribution of distance needed to find the $k$ nearest neighbors.
*Source*: Own calculations on data from the GeoSweden database.

## 2.3 Identification Strategy

The main methodological challenge when identifying the causal effect of ethnic enclaves on outcomes of immigrants is that refugees could self-select into neighborhoods. Therefore, the choice of neighborhood can be affected by the refugees' observed and unobserved characteristics which in turn will lead to biased estimates for residential and labor market outcomes. Hence, this paper accounts for refugees' self selection into neighborhoods by using the Swedish placement policy. In the subsection 2.3.1, I will outline the Swedish placement policy in more detail and the identifying variation used in this paper. I will then proceed to show balancing tests to investigate if the refugees' allocations were random during the policy.

### 2.3.1 The Refugee Placement Policy

In response to concentration of refugees in metropolitan regions and high asylum applications, Sweden established the refugee placement policy from 1985 to 1994. The aim of the refugee dispersal policy was mainly to spread refugees in various municipalities in the host countries, thus focusing on decentralization from cities and spreading the burden sharing throughout municipalities that present opportunities for integration.[6] Refugee dispersal policies imply that refugees are not free to choose their initial residential locations. In the case of Sweden, the government placed all newly arrived refugees from 1985 to 1994 in municipalities with which the Swedish Migration Board (SIV) had contracts. There was an exception in the policy for family reunification immigrants who could choose their residential locations. Although the policy was in place from 1985 to 1994, the implementation of the placement policy was less strict from 1992 onwards (Åslund and Rooth, 2007). Therefore, this paper only uses refugees arriving 1990 and 1991 to analyze the question at hand.

While the idea of the refugee placement policy was that the refugees would remain and integrate in the municipality they were placed in, the refugees could still move to another location if they found another housing after being placed. Moreover, the refugees still received their social welfare irrespective of moving or staying in placed municipalities. Although Sweden has a regulated housing market and rental apartments are allocated through a queuing system, the refugees were allowed to bypass the queuing system. The bypass of the system was particularly applied if the refugees arrived with their children, but they needed to accept the first apartment provided by the municipality.

The procedures involved during the refugee placement policy were:

1. Once an asylum seeker arrived in Sweden and applied for asylum, he was placed in a refugee center administered by the SIV. Edin et al. (2003) show that there was no correlation between the location of the refugee center and the port of entry.

[6] For more detailed information on the Swedish refugee placement policy, see Andersson (2003) and Dahlberg and Valeyatheepillay (2021).

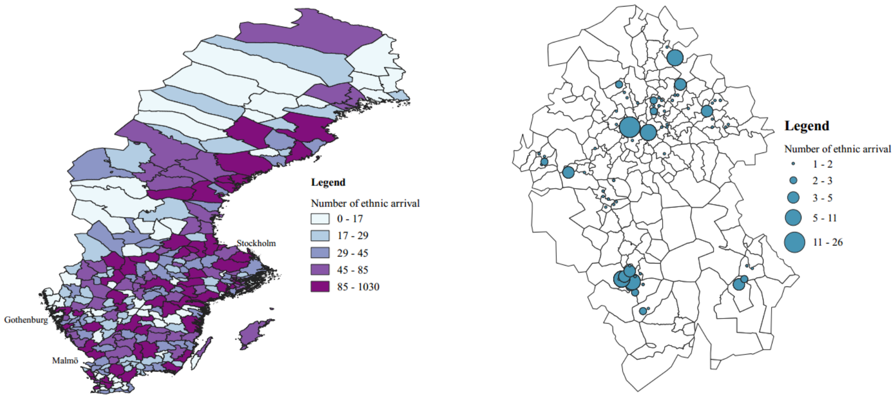
2. Upon obtaining a residence permit, the refugee was placed in one of the contracted municipalities, which comprised of nearly all the 289 Swedish municipalities in 1990 and 1991. The placement officers allocating refugees to municipalities and the refugees had no contact, therefore selection on unobservable characteristics was not likely.

3. Within the municipalities, the refugee was placed in an available accommodation upon arrival.

I will use within municipality variation in neighborhood types and argue that the exact apartment allocated to a refugee by the municipality is exogenous from his or her point of view. Therefore, I assume that exactly which neighborhood a refugee lived in, in terms of ethnic and employed ethnic neighbors' characteristics, is exogenous from the refugee's perspective.

To illustrate where the refugees were placed during the policy, Figure 2.4(a) shows the map of Sweden with the distribution of the absolute number of ethnic arrival from the top 10 refugee source countries in 1990 and 1991. In Figure 2.4(a), the dark purple colors on the map represent higher ethnic arrival in the municipalities and the light colors illustrate lower ethnic arrival. Figure 2.4(a) shows that the refugee placement policy in 1990/91 dispersed refugees throughout Sweden and the refugees are not only placed in the big cities, including Stockholm, Malmö and Gothenburg, but also in the North of Sweden and in the middle of Sweden. As I will require sufficient variation within a municipality to identify any effects, I zoom on a map of Uppsala city center to illustrate the refugee allocation within the municipality center. From Figure 2.4(b), it can be noticed that the refugees were dispersed through different apartments in Uppsala city center and certain coordinates obtained more refugees than others depending on the housing available upon arrival. The dispersal of refugees in different areas of the municipality lends support to the assumption that the characteristics of the refugees' neighbors are exogenous from their individual point of view and that the refugees lived with varying share of co-ethnics.

Figure 2.4: Distribution of ethnic arrival in Sweden and Uppsala in 1990/91




(a) Sweden wide (b) Uppsala city center

*Notes:* Figure 2.4(a) represents the distribution of refugees from the top 10 source countries in 1990/91 in Sweden. Figure 2.4(b) illustrates the distribution of refugees in Uppsala city center.
*Source*: Own calculations on data from the GeoSweden database.

### 2.3.2 Balancing Tests

To identify if the refugee placement policy was random in terms of observable characteristics, this paper also tests whether refugees placed in neighborhoods with more co-ethnics among their nearest neighbors were similar to the ones placed in neighborhoods with few co-ethnics by assessing the normalized differences. Table 2.1 presents the mean as well as standard deviations for the covariates, and the normalized difference[7] by Imbens (2015) for refugees in a high ethnic neighborhoods share among their nearest neighbors versus those placed in low ethnic neighborhoods among the $k = 250$ nearest neighbors. Large values in the magnitude of 1.00 and above for normalized differences show that the refugees in the high versus low co-ethnic neighborhoods are considerably different. The covariates used are age, education, marital status, having children and country regions. The normalized differences between the high and low ethnic neighborhoods are rather modest at a range of -0.38 to 0.23. Most normalized differences are smaller than 0.10, which imply that the refugees in low and high ethnic neighborhoods are balanced, and that the regression analysis will not be as sensitive to specification choices and outliers. The mean age, the average refugee having children and who are married, female

[7]The normalized difference equation is given by $\bar{X}_t$ - $\bar{X}_c$ / $\sqrt{(S^2_{X,t} + S^2_{X,c})/2}$ where $\bar{X}_t$ represents the mean of the high ethnic neighborhood. $\bar{X}_c$ illustrates the low ethnic neighborhood and $S$ is the standard deviations. For more details on the normalized difference, see Imbens (2015).

are mostly similar across both low and high ethnic neighborhoods. Therefore, it can be concluded that refugees in low and high ethnic neighborhoods have relatively similar characteristics, and the placement policy allocated refugees randomly in different neighborhoods. These small normalized differences also apply among the $k = 100, 500, 1000$ nearest neighbors as shown in the Appendix section B.2.

Table 2.1: Balancing test: ethnic neighbors for $k = 250$

| | Low Ethnic Neighborhood N = 4934 | | High Ethnic Neighborhood N = 5642 | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Normalized Difference |
| Age | 33.01 | 6.65 | 33.40 | 6.86 | 0.06 |
| Children | 0.50 | 0.50 | 0.54 | 0.50 | 0.09 |
| Married | 0.29 | 0.46 | 0.31 | 0.46 | 0.03 |
| Female | 0.35 | 0.48 | 0.36 | 0.48 | 0.03 |
| Africa | 0.23 | 0.42 | 0.10 | 0.29 | -0.38 |
| Europe | 0.03 | 0.17 | 0.08 | 0.27 | 0.23 |
| East Asia | 0.06 | 0.23 | 0.06 | 0.24 | 0.01 |
| West Asia | 0.65 | 0.48 | 0.71 | 0.45 | 0.14 |
| Latin America | 0.04 | 0.19 | 0.06 | 0.23 | 0.08 |

*Notes:* The table shows the balancing test for low ethnic neighborhood versus high ethnic neighborhood for $k = 250$ nearest neighbor.
*Source:* Own calculations based on data from the GeoSweden database.

The balancing test for employed ethnic neighborhood for $k = 250$ is presented in Table 2.2. Similarly, the normalized differences are rather small for the low and high employed ethnic neighborhoods. The balancing tests for employed ethnic neighborhoods for $k = 500$ and $k = 1000$ are shown in Appendix B.2 in Tables B.6 and B.7. The tests for the different $k$ levels lend support that the refugee placement policy randomly assigned refugees to neighborhoods.

Table 2.2: Balancing test: employed ethnic neighbors for $k = 250$

| | Low Employed Ethnic N = 4007 | | High Employed Ethnic N = 6569 | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Normalized Difference |
| Age | 33.37 | 6.78 | 33.13 | 6.75 | -0.04 |
| Children | 0.51 | 0.50 | 0.53 | 0.50 | 0.03 |
| Married | 0.30 | 0.46 | 0.30 | 0.46 | 0.01 |
| Female | 0.36 | 0.48 | 0.35 | 0.48 | -0.01 |
| Africa | 0.18 | 0.39 | 0.14 | 0.35 | -0.11 |
| Europe | 0.04 | 0.19 | 0.07 | 0.25 | 0.12 |
| East Asia | 0.05 | 0.21 | 0.07 | 0.25 | 0.08 |
| West Asia | 0.70 | 0.46 | 0.67 | 0.47 | -0.08 |
| Latin America | 0.03 | 0.17 | 0.06 | 0.24 | 0.15 |

*Notes:* The table shows the balancing test for low employed ethnic neighborhood versus high employed ethnic neighborhood for $k = 250$ nearest neighbor.
*Source:* Own calculations based on data from the GeoSweden database.

### 2.3.3 Refugee Sample and Labor Market Outcomes Definitions

I identify the refugees by selecting those who were granted asylum in Sweden and acquired refugee permits. Using the refugee placement policy as an identification strategy and bearing in mind that the data is collected since 1990, this paper considers refugees arriving in 1990 and 1991. Given that the refugees are registered in the data in the year in which they obtain their residence permits, their housing locations subsequent to that year and the residential coordinates of the refugees at the end of the year are observed. As the data does not contain information on self-reported ethnicity, countries of birth are used as a proxy for ethnicity. The analysis considers individuals aged 25 to 55. The reason for carrying out the analysis for 25 year old and older is that this age group is more likely to be on the labor market. The data allows me to follow the refugees over time, and I will observe the refugees over an 7-year horizon as the refugees are still of working age in that time horizon frame.[8]

The sample includes refugees from the top ten countries of origin. By focusing on the top countries, I obtain source countries in which the emigration level is large enough for networks to actually be created in Sweden. I do not consider other countries as there are very small amount of individuals arriving from those countries and those countries, in turn, only have a very small refugee stock living in Sweden. Furthermore, the sample

[8]The sample size differs from year to year due to refugees emigrating to other countries or dying.

includes only single countries because country groupings would lead to measurement error and the individuals from the country groupings may not speak the same language to communicate among themselves. The sample mainly comprises of individuals from non-OECD member countries, with the exception of Turkey.

Table 2.3 presents the absolute number and percentage of refugees from the top ten refugees' birth countries in 1990 and 1991. The country of birth and the sending country may differ for some of the refugees as they can migrate from another country. However, the refugees' birth countries and the refugee sending countries are similar in about 80 percent of the cases. In total, the sample comprises of 10,576 refugees in 1990 and 1991, with about equal absolute number of refugees for the two cohorts. I bundle the refugees in the years 1990 and 1991 together for the analysis. Refugees from Iran consists the largest group in our sample making up about 20 percent of the study population, followed by Iraqis and Lebanese.

Table 2.3: Distribution of refugees' birth countries for the considered cohorts

| Refugees' Birth Countries | 1990 | 1991 | Total |
|---|---|---|---|
| Iran | 1209 | 1185 | 2394 |
| | (11.43) | (11.20) | (22.64) |
| Iraq | 953 | 972 | 1925 |
| | (9.01) | (9.19) | (18.20) |
| Lebanon | 708 | 628 | 1336 |
| | (6.69) | (5.94) | (12.63) |
| Syria | 475 | 536 | 1011 |
| | (4.49) | (5.07) | (9.56) |
| Ethiopia | 402 | 320 | 722 |
| | (3.80) | (3.03) | (6.83) |
| Somalia | 277 | 686 | 963 |
| | (2.62) | (6.49) | (9.11) |
| Vietnam | 271 | 343 | 614 |
| | (2.56) | (3.24) | (5.81) |
| Yugoslavia | 269 | 326 | 595 |
| | (2.54) | (3.08) | (5.63) |
| Turkey | 254 | 261 | 515 |
| | (2.40) | (2.47) | (4.87) |
| Chile | 452 | 49 | 501 |
| | (4.27) | (0.46) | (4.74) |
| Total | 5270 | 5306 | 10576 |
| | (49.83) | (50.17) | (100.00) |

*Note:* The table presents the sample of refugees in the sample. The refugees are identified from the reason for immigration variable in the GeoSweden database.

*Source:* Own calculations from GeoSweden database.

Table 2.4 shows the age, gender, percentage of refugees who are married, having at least one child, employed and on social welfare in the refugees' initial year in Sweden. Additionally, the table illustrates the percentage of individuals with less than high school education, labelled as low educated, and the percentage of refugees living in different country regions. Most of the refugees are young men who have a relatively low educational level. There is also a high percentage of refugees who are not married. 85 percent of the refugees depend on social welfare in the initial year in Sweden. A considerable percentage of refugees originates from the West Asian region in the sample, due to conflicts in countries in that region.

Table 2.4: Descriptive statistics for refugees considered

| Variables | mean | SD |
|---|---|---|
| Age | 33.22 | (6.76) |
| Female | 0.35 | (0.48) |
| Married | 0.30 | (0.46) |
| Children | 0.52 | (0.50) |
| Employed | 0.14 | (0.34) |
| Social welfare | 0.85 | (0.36) |
| Low educated | 0.82 | (0.39) |
| Africa born | 0.16 | (0.37) |
| Latin American born | 0.05 | (0.21) |
| West Asian born | 0.68 | (0.47) |
| East Asian born | 0.06 | (0.23) |
| Eastern European born | 0.06 | (0.23) |
| Observations | 10576 | |

*Note:* Standard deviations are reported in parentheses. The variables are measured at cohorts' arrival. The variable 'Low educated' comprises of individuals with less than high school education.

*Source:* Own calculations from the GeoSweden database.

For the labor market outcomes analysis, the key variables of interest in this paper are earnings and employment. The register data contains the gross yearly income reported to the tax agency by employer. Therefore, the earnings variable represents the gross income of the immigrant, namely the sum of taxable income. Employment is defined as a dummy defined as taking the value of 1 if the individual obtained labor market income and 0 otherwise. As shown in Table 2.4, on average, only 14 percent of the refugees are employed in the initial year.

## 2.4 Results: Effect of Initial Neighborhood on Future Neighborhood

This section analyzes the effects of initial neighborhoods with co-ethnics and employed co-ethnics on future share of co-ethnics and employed co-ethnics in their neighborhoods. The estimation is particularly of interest in order to analyze if networks may lead to ethnic clustering throughout time. The initial share of co-ethnics and employed co-ethnics are used as instruments to provide exogenous variation in the future neighborhood quality. Therefore, the results in this section will demonstrate the relevance of the instrument used. Section 2.4.1 presents the empirical specification and Section 2.4.2 will show the results.

### 2.4.1 Empirical Specification

To examine the effects of co-ethnics and employed co-ethnics in the initial refugees' neighborhoods on future residential composition, I estimate the following yearly equations:

$$ethnic_{i,t+z} = \beta_0 + \beta_1 ethnic_{it} + X_{i,t+z} + \delta_a + \epsilon_{i,t+z} \tag{2.1}$$

where $ethnic_{i,t+z}$ represents the share of ethnic neighbors (employed co-ethnics) in the refugees' individualized neighborhoods after $z$ years for individual $i$. $ethnic_{it}$ is the variable of interest and represents the share of ethnic neighbors (employed co-ethnics) in the refugees' initial individualized neighborhoods in year $t$. $X_{i,t+z}$ is a set of socio-demographic characteristics, including age, age square, gender, marital status, having children, arrival year and country of origin controls.[9] $\delta_a$ denotes the initial placed municipality fixed effects which account for local amenities and differences in labor market. $\epsilon_{i,t+z}$ is the error term. The dependent and independent variables are in continuous shares in this case. I adopt a specification in line with the first stage estimation in Edin et al. (2003). The regression outlined includes initial municipality fixed effects, country of origin fixed effects and immigrant arrival year fixed effects.

The specification adopted in this section in line with the first stage estimation in Edin et al. (2003). The specification outlined, however, differs from Edin et al. (2003) in several respects. While Edin et al. (2003) only examine the effect of co-ethnics in the refugees' initial year on future co-ethnics in $t$+8, I will estimate the effects at different time horizon. I follow the refugees on a 7-year time horizon. Moreover, I estimate the effects of initial share of co-ethnics on future neighbor characteristics in the refugees' residential locations on small geographical scales. The small scale analysis detect if potential interactions with co-ethnics locally lead to residing with more co-ethnics in the future.

### 2.4.2 Results for Ethnic Share

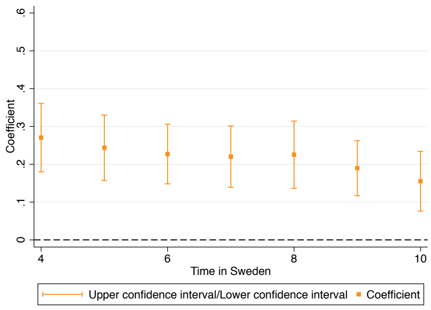
In this Section, I report the results for co-ethnics for the specification from Equation 2.1. Figure 2.5 shows the coefficient plots from estimating the first stage estimations, Equation 2.1, for $t+4$ to $t+10$ for $k=100$, $k=250$, $k=500$ and $k=1000$ nearest neighbors.[10] The choice of the time period is made due to networks taking time to operate. The initial co-ethnic neighborhood matters for future co-ethnic neighborhood throughout the 7-year horizon. The coefficients are statistically significant and positive throughout the years, irrespective of the $k$-nearest neighbor levels. The effect decreases over time but not substantially. Co-ethnics at small-scale neighborhoods matter for being with co-ethnics in the future. Ethnic networks on small geographical scales seem crucial for refugees.

[9] Regressions with $X_{i,t}$ controls have also been considered in order to avoid bad controls and the specifications are robust to those controls.

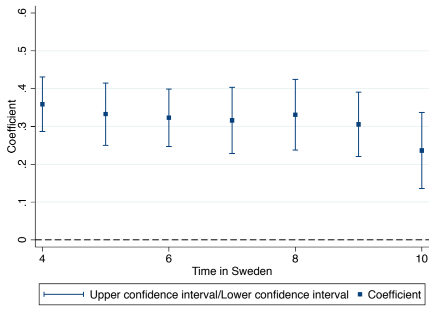
[10] The full results are presented in Table B.8, B.9, B.10 and B.11 in Appendix B.3

Looking at the different $k$-nearest neighbors, we notice that the higher the $k$, the higher the magnitude of the effect of initial ethnic share on future ethnic share. For instance, in Figure 2.5(a), a 1 percent increase in the initial ethnic share leads to almost a 0.27 percent increase in the future share of ethnic in $t + 4$ among the $k = 100$ nearest neighbor. For $k = 250$ nearest neighbors, a 1 percent increase in the initial share of co-ethnics results in a 0.35 percent increase in the future share of co-ethnics. At $k = 500$ nearest neighbors, the future co-ethnics share increase to 0.40 percent in $t + 4$. A 1 percent increase in the initial share of ethnic leads to about 0.44 percent increase in the future ethnic share in $t + 4$ among the $k = 1000$ nearest neighbor. This indicates that residing with co-ethnics on small geographical scales matter for the future neighborhood composition. The result for $k = 1000$ nearest neighbors suggests that it is important to have neighbors of the same ethnicity close enough so that the refugee meet and interact with co-ethnics at the shops and in their local residential area.

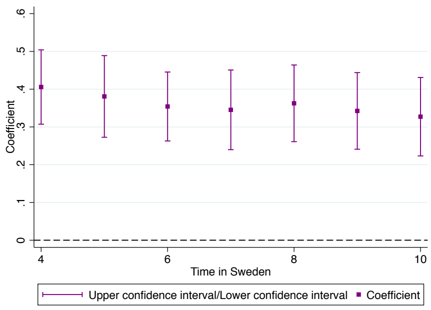
Figure 2.5: Effect of initial co-ethnic share on future co-ethnic share: $k = 100, k = 250, k = 500, k = 1000$



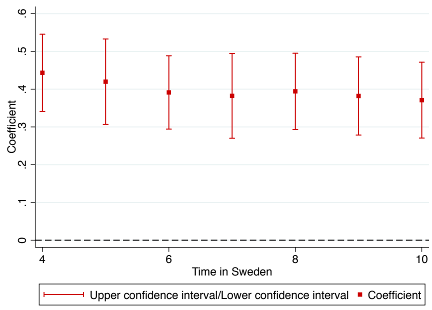

(a) $k = 100$




(b) $k = 250$




(c) $k = 500$




(d) $k = 1000$

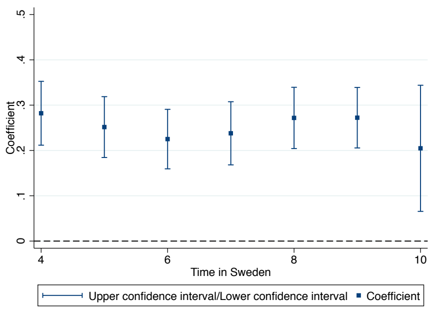
*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.1) and their corresponding 95% confidence intervals.
*Source*: Own calculations on data from the GeoSweden database.

### 2.4.3 Results for Employed Ethnic Share

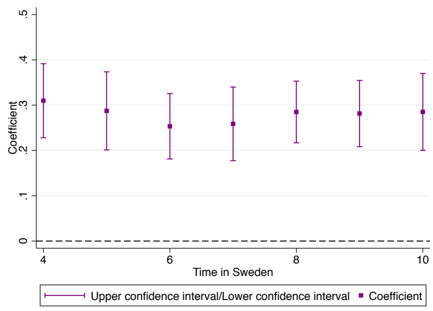
Given that being with employed co-ethnics may be important for integration outcomes, I analyze the effects of initial share of co-ethnics on future small-scale residential compositions. Similar to the share of co-ethnics, initial share of employed co-ethnics has a positive effect for share of employed co-ethnics in the future neighborhood composition, regardless of the $k$ nearest neighbor and the time horizon analyzed. A 1 percent increase in the share of co-ethnics among the $k = 250$ leads to about 0.3 percent increase in the future share of co-ethnics. The effect is rather stable over time, and ranges from 0.35 percent to 0.3 percent for $k = 1000$ nearest neighbors.

This indicates that residential location is also based on socio-economic characteristics. Refugees allocations to more employed co-ethnics in the initial year impact the share of employed co-ethnics among the refugees' future neighbors.

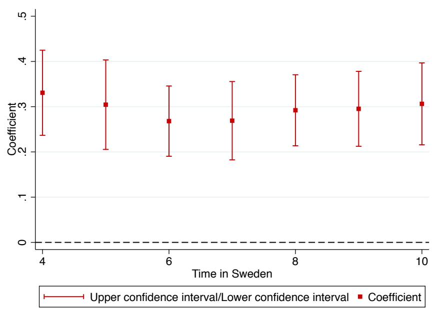
Figure 2.6: Effect of initial employed co-ethnic share on future employed co-ethnic share: $k = 250, k = 500, k = 1000$




(a) $k = 250$




(b) $k = 500$




(c) $k = 1000$

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.1) and their corresponding 95% confidence intervals.
*Source*: Own calculations on data from the GeoSweden database.

### Movers

Although the refugees were initially assigned within municipalities and in specific housing, they could move to another location without restrictions as long as they found alternative housing. Given that stayers in the placed neighborhoods may not disclose their preferences, particularly in the short run, I examine the baseline results for movers. Movers are characterized as those who move from their initial placed neighborhoods and have a different coordinates. About 87 percent of the refugees move from their initial assigned coordinates in $t + 4$ and this percentage increase to reach almost 96 percent in $t + 10$. The results are presented in Figures B.1 and B.2 in Appendix B.4. While the effects of initial co-ethnics and employed co-ethnics are positive and statistically significant as in the baseline results, the effects are similar to the estimates seen in $t + 10$ for all the $k$-nearest neighbors.

### Robustness Checks

I conduct several robustness checks to see if the results hold. I restrict the distance that Equipop finds the nearest neighbors: Equipop can find the nearest neighbors close or far away from the respective refugees depending on whether the refugees reside in a rural or urban area. Therefore, I proceed to carrying out the baseline regression restricting the maximum distance required to reach the nearest neighbors to 500 meters since the algorithm for nearest neighbors can cover a longer distance to reach the nearest neighbors. Restriction of distance to 500 meters means that the refugees meet co-ethnics at the playground or local shops. If immigrants are clustered, the opportunity for interaction is higher and therefore, ethnic capital is more significant. The coefficient plots for this estimation can be seen in Figures B.3 and B.4 in Appendix B.4. The results are robust to these estimations, indicating that ethnic and employed ethnic networks on small geographical scales matter.

As refugee parents were among the first to obtain housing in the municipal housing queue, I estimate Equation 2.1 for refugees arriving in Sweden with their children in Figures B.5 and B.6 in Appendix B.4. The positive effects of initial share of co-ethnics and employed co-ethnics holds for the parents.

## 2.5 Results: Effect of Co-Ethnics on Earnings

To what extent does ethnic network affect the refugees' earnings and employment? In this section, I will investigate the effect of co-ethnics and employed co-ethnics on refugees' labor market outcomes, including earnings and employment. The results will indicate if interactions with co-ethnics and employed co-ethnics is crucial for economic success of refugees.

### 2.5.1 Empirical Specification

To examine the effects of co-ethnics and employed co-ethnics among the nearest neighbors on the refugees' earnings and employment, I estimate the following equation:

$$y_{i,t+z} = \beta_0 + \beta_1 ethnic_{i,t+z} + X_{i,t+z} + \delta_a + \epsilon_{i,t+z} \tag{2.2}$$

where $y_{i,t+z}$ shows the quartic earnings or being employed after $z$ years, respectively. Given that earnings are skewed and the data contains several observations with zero earnings, I create the quartic root of earnings variable as a proxy for log transformations. The quartic root of earnings avoids dropping observations that have earnings equal to zero at the individual level, and it behaves similarly to a logarithmic transformation for positive numbers.[11] Employment is defined as a dummy, taking the value of 0 if the refugee is unemployed and 1 if he is employed. $ethnic_{i,t+z}$ represents the share of co-ethnics (employed co-ethnics) in the refugees' initial individualized neighborhoods. Here again, $X_{i,t+z}$ is a set of socio-demographic characteristics and country of origin controls. $\delta_a$ are the municipality fixed effects allowing us to control for local amenities and unemployment.

The 2SLS estimation is similar to that used in Edin et al. (2003).[12] However, instead of using the share of co-ethnics at the municipality level, I use the share of co-ethnics and employed co-ethnics at smaller geographical scales to allow to investigate into networks

[11] For further details on quartic root, see Ashraf et al. (2015), Brown and Velásquez (2017), Tarozzi et al. (2014), and Tukey (1957)

[12] If we were to use OLS, our results would be biased downward if only low skilled immigrants live in the ethnic enclaves.

and interactions. Moreover, I will investigate the effects at different time horizons rather than the effects in only $t+8$. The instrument used for the 2SLS estimates is the initial share of ethnics or employed ethnics. We use the initial ethnic share as an instrument to solve endogeneity in the neighborhood quality and the exclusion restriction implies that the only effect on labor market outcomes emanate from the current residence location's ethnic share. Given that weak instrument can cause our estimates to be biased, I will present the F statistics in Tables B.8 - B.14 in Appendix B.3. I use the rule of thumb suggested by Stock, Yogo, et al. (2005) and show that the statistics presented compare well to the cutoff of 10.

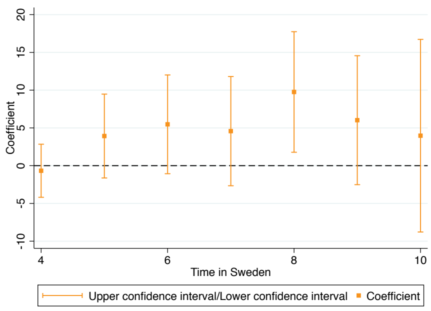
### 2.5.2 Results for Ethnic Share

Figure 2.7 shows the $\beta_1$ coefficients obtained when estimating equation 2.2. The point estimates are positive along long time horizons considered for the different $k$-nearest neighbors. What is noticeable is that being with co-ethnics on small geographical scales has a positive and statistically significant effect on earnings on small geographical scales.[13] This effect holds in $t+8$, which is in line with Edin et al. (2003) who find a positive effect of co-ethnics on earnings at the municipality level 8 years after the refugees' arrival. It can be estimated that a 10 percent increase in the co-ethnic share among the closest neighbors contributes to about 10 percent increase in earnings in t+8, regardless of the $k$ nearest neighbor levels.[14] Networks on a small geographical scale provide better labor market opportunities for the newly arrived refugees. Furthermore, the high magnitude of the results in $t+8$ can be in line with the fact that the refugees arrived in Sweden during the 1990s recession and had to wait until $t+8$ for the economy to be booming for the labor market to recover. Moreover, building networks takes time.

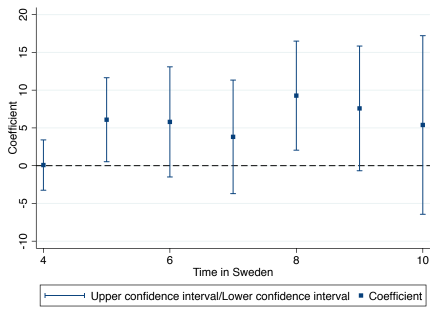
[13]There are some coefficients that are significant at the 10 percent significance level rather than the 5 percent significance level. This is the case for the results for $k = 100$ nearest neighbors in $t+6$. The results for $k = 250$ nearest neighbors are significant if I consider the 10 percent significance level in $t+9$ onwards. The results for $t+5$ for the 500 nearest neighbors is significant at 10 percent as shown in Appendix B.5

[14]The standard errors are rather large and there are more uncertainties in the data over time.

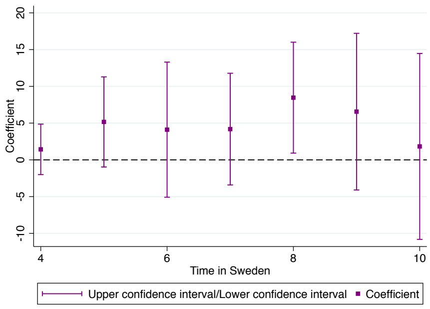
Figure 2.7: Effect of co-ethnic share on future earnings: $k = 100, k = 250, k = 500, k = 1000$



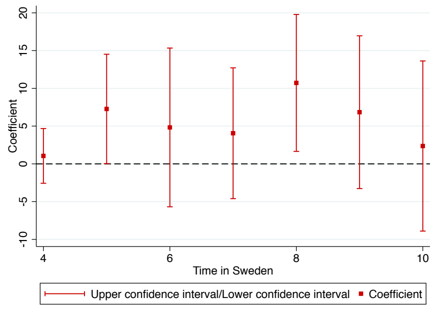

(a) k=100




(b) $k = 250$




(c) $k = 500$




(d) $k = 1000$

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.2) and their corresponding 95% confidence intervals.
*Source*: Own calculations on data from the GeoSweden database.

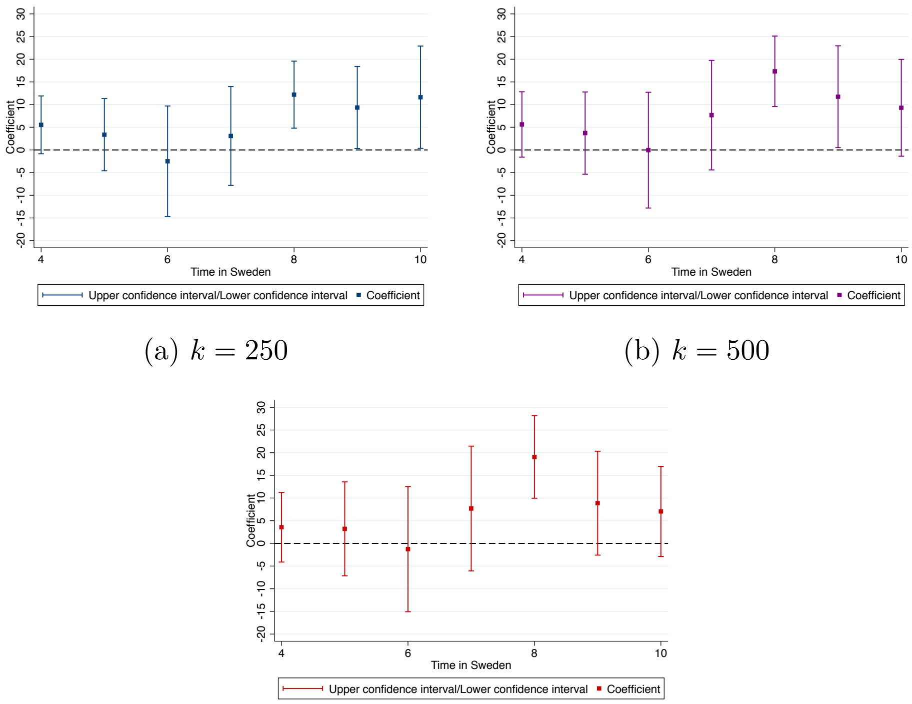
### 2.5.3 Results for Employed Ethnic Share

I expect that the closer the employed ethnic neighbors live to the refugees, the higher the probability of interaction and exchanging information about job opportunities. For instance, Patacchini and Zenou (2012) find that living within one hour travel time to a large number of employed neighbors of the same ethnicity is positively associated with job-finding rates. Turning to the effects of employed co-ethnics on earnings, we can notice that being among employed co-ethnics is important for earnings effects of local ethnic enclaves. I find positive effect in $t+8$ onwards, regardless of the $k$-nearest neighbor levels.[15] Employed co-ethnics in small scale neighborhoods matter for labor market outcomes in

[15]Some of the results are significant if I consider the 10 percent significance level from the year $t + 8$ onwards. This is the case for $t + 9$ and $t + 10$ for $k = 250$ nearest neighbors. The result for earnings is significant at the 10 percent significance level in $t + 9$ for the $k = 500$ nearest neighbors.

the long run. This is in line that social interactions with closest neighbors is important for labor market outcomes, in terms of earnings of the newly arrived refugees.

Figure 2.8: Effect of employed co-ethnic share on future earnings: $k = 250, k = 500, k = 1000$




(a) $k = 250$

(b) $k = 500$

(c) $k = 1000$

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.2) and their corresponding 95% confidence intervals.
*Source*: Own calculations on data from the GeoSweden database.

## 2.6 Results: Effect of Co-Ethnics on Employment

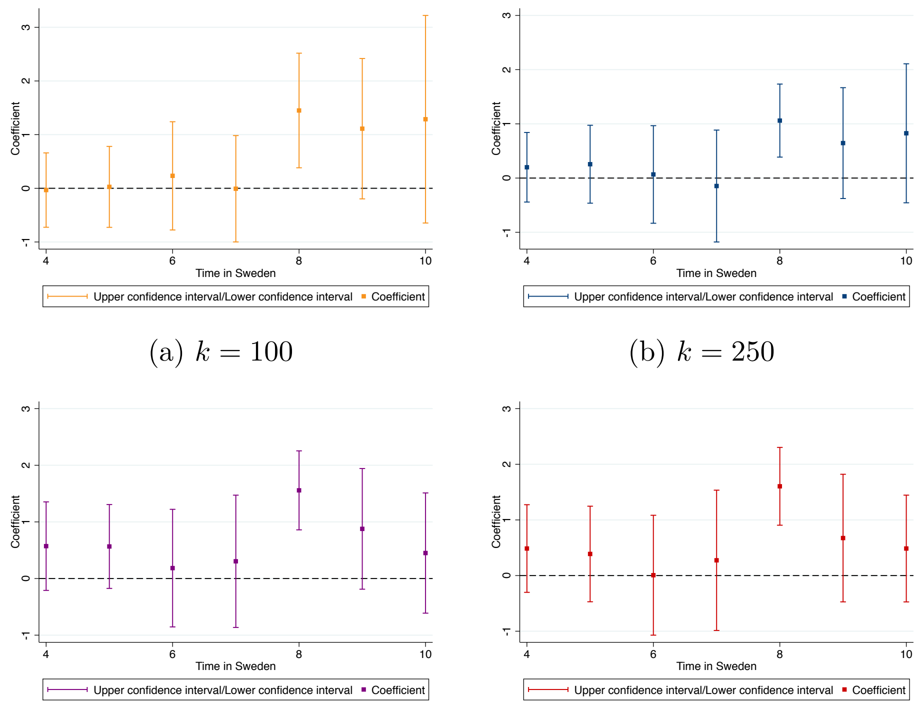
### 2.6.1 Results for Ethnic Share

I examine another labor market outcome, in terms of employment and analyze the effects of share of co-ethnics on probability of employment in each year from $t + 4$ to $t + 10$. Here again, I find that being with co-ethnics in the neighborhood affects employment probability significantly in $t + 8$ onwards, irrespective of the $k$-nearest neighbor level investigated.[16] Being assigned among the nearest neighbors has mostly a positive effect on

[16]The statistical significance of the results holds at the 10 percent level, for t+9 for $k = 100$

probability of employment. This effect is statistically significant, 8 years after arrival. A 10 percent increase in share of co-ethnics among the nearest neighbors increases employment by about 1.5 percentage points for $k = 100$, $k = 250$ and $k = 1000$, and leads to an increase of around 1 percent among the $k = 500$ nearest neighbors in $t + 8$. This results indicate that it may take time to reap the benefits from networks. Although I do not observe how the refugees search for employment, the results demonstrate that networks at a small geographical scale can be crucial for informal hiring or supporting with job search, leading to better job opportunities.

Figure 2.9: Effect of co-ethnic share on future employment: $k = 100, k = 250, k = 500, k = 1000$




(a) $k = 100$ (b) $k = 250$

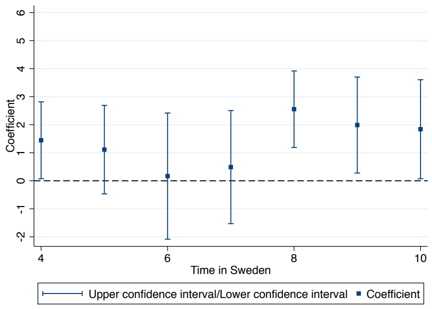
(c) $k = 500$ (d) $k = 1000$

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.2) and their corresponding 95% confidence intervals.
*Source*: Own calculations on data from the GeoSweden database.

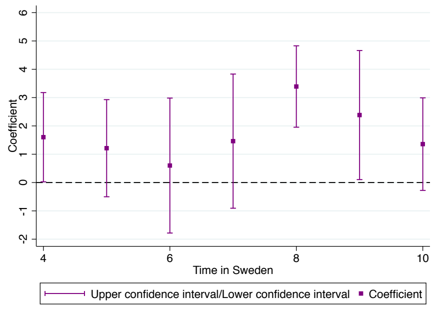
### 2.6.2 Results for Employed Ethnic Share

As the quality of networks can be more important for employment, I investigate the effects of employed co-ethnics on refugees' employment for the $k$-nearest neighbors. The results show that employed co-ethnics impact employment mostly positively in different time horizons, even though the effects are not statistically significant in most years examined. The persistence of the positive and statistically significant effect is seen 8 years onwards after the refugees' arrival. What is noticeable is that these effects are of a higher magnitude as the neighborhood size increases, which indicates that being with employed co-ethnics at a small geographical scale drives successful labor market integration, in terms of employment. It indicates that not only the employed co-ethnics in the refugees' apartment matter but also employed co-ethnics in the local area is important. A 10 percent increase in share of employed co-ethnics among the $k = 250$ nearest neighbors increases employment by about 2.5 percentage points. For $k = 500$ nearest neighbors, the magnitude increases to almost 3.5 percentage points. The importance of employed co-ethnics is more visible at the $k = 1000$ nearest neighbors: a 10 percent increase in share of employed co-ethnics increases employment by about 3.8 percentage points. I find slightly smaller effects of employed co-ethnics on probability of employment in $t + 9$ among the $k = 250$ and $k = 500$ nearest neighbors.

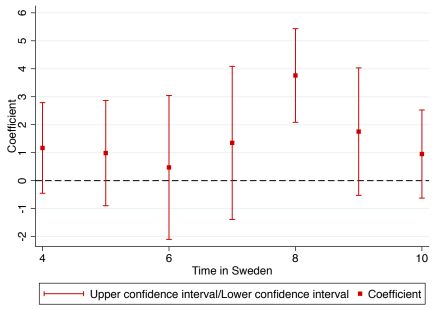
Figure 2.10: Effect of employed co-ethnic share on future employment: k=250, k=500, k=1000




(a) k=250




(b) k=500




(c) k=1000

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.2) and their corresponding 95% confidence intervals.
*Source*: Own calculations on data from the GeoSweden database.

### Robustness Checks

I conduct similar robustness checks as for the effect of co-ethnics and employed co-ethnics in the initial neighborhood for the future neighborhood composition. The results for co-ethnics and employed co-ethnics are shown in Appendix B.6 and B.8. I still find positive results when restricting distance to 500 meters, parents and movers. The results remains positive and statistically significant for distance restriction to 500 meters and movers after a while spent in the host country, i.e. about $t + 8$ years after the refugees' arrival and onwards.

## 2.7 Conclusions

This paper has investigated the effect of co-ethnics and employed co-ethnics on future residential composition and labor market integration at small geographical scales using detailed geographical data. To enable the investigation into potential network mechanism at which the effect occurs, I create neighborhoods of varying sizes of neighbor applying the $k$-nearest neighbor approach. A small scale analysis is chosen since ethnic and employed ethnic neighbors in the close proximity of the refugees can form an important part of their networks and can impact their future neighborhood composition and labor market integration through daily interactions, and dissemination of information.

For the identification strategy, I exploit the Swedish refugee placement policy in 1985 to 1994 as a natural experiment. I only use the year 1990 and 1991 as the treatment years due to data limitations and less strict application of the policy from 1992 to 1994. This paper makes an assumption that exactly which neighborhood the refugee lives in in terms of their ethnic neighbors is exogenous. For the analysis, I use an empirical specification in the style of Edin et al. (2003).

The results show that the initial share of co-ethnics and employed co-ethnics affect the future share of co-ethnics and employed co-ethnics in the neighborhood of the refugees at small geographical scales in different years after the refugees' arrival. This results holds irrespective of the $k$ nearest neighbor level investigated and remains robust to various specifications. The results indicate that interacting with co-ethnics and employed co-ethnics is important for neighborhood composition. I also find that the share of co-ethnics and employed co-ethnics affect positively labor market outcomes, in terms of earnings and employment, in the long run. The quality of the refugees' neighbors are crucial to drive economic success.

Future research could extend this work and apply a similar approach to countries with bigger ethnic enclaves at small geographical scales in order to also be able to examine the effect of ethnic networks and employed ethnic neighbors on labor market integration. There should be further conclusive evidence on potential mechanisms that impact refugees in other countries. Moreover, research could investigate into the optimal enclaves size.

In this paper, I have analyzed refugee placement policy as a determinant of residential integration for adult refugees. It would be interesting to examine another determinant of residential integration and its long run consequences on refugee children.

# Chapter 3

# Age at Arrival and Residential Integration

This chapter is based on joint work with Cristina Bratu from Aalto University and Matz Dahlberg from Uppsala University.

## 3.1 Introduction

European countries are becoming more and more diverse due to large immigration flows, yet smaller geographies, such as neighborhoods, do not always reflect this diversity. This is no less true in Sweden, the country under study in this paper, where the majority of natives live in neighborhoods where 90% of residents are natives; by contrast, only around 20% of refugees live in similarly native-dominant neighborhoods (Figure 3.1).[1] Figure 3.1, while striking, puts together refugees with different characteristics, time spent in the host country and so on. Nevertheless, labor market integration is generally expected to be reflected in residential choices of immigrants: as immigrants earn more, with time spent in the host country, they are more likely to move out of their initial locations, to better residential areas (Massey and Denton, 1985). In turn, arrival at *earlier* ages is particularly beneficial for a host of outcomes, from education and earnings (Alexander and Ward, 2018; Ansala et al., 2019; Böhlmark, 2008; Hermansen, 2017; Lemmermann and Riphahn, 2018)), to health (Berg et al., 2014) and social integration (Åslund et al., 2015). We might therefore expect that children of immigrants are particularly well-placed to make residential choices that reflect their labor market integration, since they spend a considerable amount of time in the host country before making these choices. In this paper, we test if there is empirical support for this hypothesis.

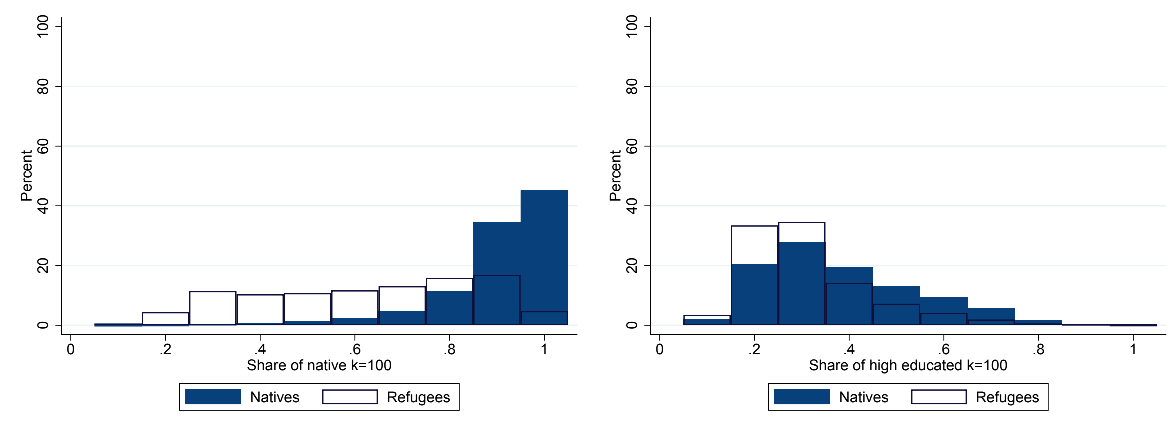
We study whether immigrant children who arrive at earlier ages in Sweden live in better neighborhoods in adulthood. More specifically, we analyze the extent to which age at immigration affects neighborhood composition along two dimensions: i) ethnic composition, measured as the share of natives, defined as individuals born in Sweden, and ii) socio-economic composition, measured via three variables: the share of high-earners, the share of highly-educated individuals, and the share on welfare.[2] We focus on refugees. We hypothesize that the younger refugees are upon arrival, the more time they have to build country-specific knowledge, including language and culture, and to forge social contacts with the native majority, which may affect both their preferences for certain

[1] We clarify how we define neighborhoods later in the introduction. Living in an ethnic enclave need not be detrimental a priori, and as previous research has shown, immigrant enclaves can in fact facilitate labor market integration for newcomers (Damm, 2009, 2014; Edin et al., 2003).

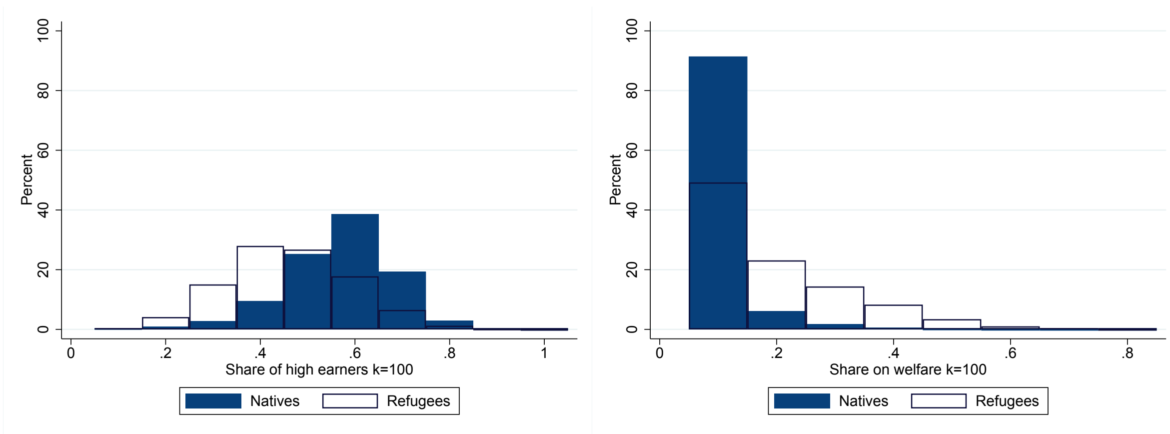
[2] We define these variables more precisely in Section 3.2.

kinds of neighborhoods and their ability to act upon those preferences. We assess to what extent these are likely channels through a decomposition exercise.

Figure 3.1: Characteristics of 100 closest neighbors for natives and refugees in 2014




(a) Share of natives (b) Share of high-educated




(c) Share of high-earners (d) Share on welfare

*Notes:* The figure shows the characteristics of the 100 closest neighbors for all refugees and natives above the age of 18 who were residing in Sweden in 2014. Refugees are defined based on residence permit data. Natives are individuals born in Sweden; high-educated individuals have an education level with at least some tertiary education; high-earners are defined as earning above the median in the municipality; on welfare refers to receipt of social benefits.
*Source*: Own calculations on data from the GeoSweden database.

Our focus on refugees in Sweden is motivated by the following fact, alluded to in the first paragraph: natives and refugees in Sweden live in profoundly different neighborhoods in terms of ethnic and socio-economic composition. Not only are refugees much less likely to live close to natives, they also consistently live in neighborhoods with fewer high-earners, fewer high-educated individuals and a disproportionately larger share of individuals that receive social benefits. These patterns are worrying given that research shows that the type of neighborhoods refugees live in may play an important role in facilitating integration, by influencing economic and social outcomes through networks and fostering social interactions. Evidence shows that neighbors can transmit information, resources and knowledge and influence the behaviors and attitudes of their neighborhood peers (Borjas, 1995; Ellen and Turner, 1997; Johnston and Pattie, 2011; Sampson et al., 2002; Sharkey and Faber, 2014). Moreover, childhood environments shape long-run outcomes of children: children in poor neighborhoods experience worse labor market outcomes in adulthood (Chetty et al., 2016a). If in adulthood, around the time when people start families, refugees live in poor neighborhoods, their children may not be able to do better than their parents, and the refugee-native gap may thus widen over time.

We use administrative data to study refugees born between 1974 and 1984 who arrive in Sweden before the age of 15 and whose outcomes we can observe at age 30. We apply a siblings design to estimate the effect of arriving at different ages relative to a reference group that arrives between the ages of 0 and 3. The within-family analysis enables us to address potential selection bias stemming from the fact that parents with better unobservables may move abroad when their children are younger.[3] We take a data-driven approach in defining neighborhoods. Using geo-coded information on the residential location of each individual in Sweden, - given by 100 $\times$ 100m coordinates - we construct individualized $k$-nearest neighborhoods, for values of $k$ equal to 100 or 1000. This method essentially allows us to identify the characteristics of neighbors at both very granular levels and at more aggregate levels. We provide suggestive evidence for the mechanisms that generate these outcomes by performing a decomposition analysis in the style of Heckman et al. (2013) to analyse how much of the effect of age at arrival on neighborhood integration goes through earnings, education and intermarriage, which is defined as being married to or cohabiting (with children) with a Swedish-born partner.

Our baseline results show that compared to refugee children arriving between age zero and three, refugee children arriving later experience a larger deviation from natives in terms of the composition of their neighbors at age 30. The effects on residential integration both

[3]We note, however, that such issues are likely to be less prevalent in our sample of refugees, who are more likely to move so as to escape violence and conflict, and thus have less control over the timing of their moves.

along the ethnic and socio-economic lines are flat until around school-starting age, when they start declining with each additional year of age since arrival. There are no marked differences between $k = 100$ and $k = 1000$. The effects we find must be considered as large. For example, those that arrive at age 15 live in neighborhoods with a 7 percentage points lower share of natives among their closest neighbors, which amounts to 35 percent of the mean value for the reference group. The corresponding magnitudes for the socio-economic characteristics of their neighbors are approximately 6 percentage points (share high-educated), 7.5 percentage points (share of high-income earners), and 7.5 percentage points (share on welfare).

We next document that age at arrival negatively affects refugees' labor market integration - as measured by income rank and years of education - and the probability of marrying a native. The estimated effects are sizeable. For instance, arriving in Sweden at age 15 rather than at ages 0-3 leads to approximately a 12.5 lower percentile rank in the earnings distribution at age 30, a half a year less of education, and a 28 percentage point lower probability of being married to a native-born partner (conditional on being married).

Finally, we decompose the baseline results in order to assess how much of the effects of age at arrival on residential integration operate through the labor market and education channels and how much through the intermarriage channel. We find that income rank, years of education and intermarriage contribute between about 20 to 40 percent of the variation in neighborhood characteristics. However, a large part of the effects of age at arrival on residential outcomes remains unexplained, particularly for very small neighborhoods ($k = 100$).

Our results are robust to correcting for issues related to variation in population density across areas. We first show, descriptively, that we capture similarly sized neighborhoods within similarly large areas, regardless of area density. We further show that the age at arrival results hold when we weight the regressions to account for population density.

We make several contributions. Due in large part to data limitations in categorizing immigrants by admission category, there are only a few papers that focus on the integration of *refugees*. Since refugees make up a significant proportion of immigrants to Europe, it is an important and highly policy-relevant group to study. While it is a group that is heavily understudied, it is also a group for which there are reasons to believe that their integration process may differ from that of other types of immigrants (see, e.g., the discussion in Brell et al. (2020)). Moreover, the few existing papers primarily focus on labor market integration (see, for example, Battisti et al. (2019) and Fasani et al. (2020) and Dahlberg et al. (2020)). By looking at residential integration, our paper is one of the first in this nascent literature to focus on other forms of integration than the labor market. Integration is a multidimensional process, and understanding how it unfolds along these multiple dimensions is important for developing adequate policy responses (see Harder et al. (2018) for the development of a multidimensional integration index and Aksoy et al. (2020) for an application of that index using German data).

Relating to the age at arrival literature, we bring two main novelties to the table. First, this is the first paper with a specific focus on refugees. The earlier literature has focused on immigrants more generally. Second, we add by having a specific focus on residential integration. The earlier literature has mostly focused on a range of other outcomes, from education and earnings (Alexander and Ward, 2018; Ansala et al., 2019; Böhlmark, 2008; Hermansen, 2017; Lemmermann and Riphahn, 2018), to health (Berg et al., 2014) and social integration (Åslund et al., 2015).

The only earlier paper we know of that has examined the effects of age at arrival on residential integration is Åslund et al. (2015). In their paper, they find that immigrant children arriving at a later age in Sweden have a lower probability of living in the same neighborhood as natives, work with natives, and marry natives. Our paper does however differ in two important ways from Åslund et al. (2015). While we focus on recent cohorts of refugees, Åslund et al. (2015) study the children of earlier cohorts of labor immigrants (mainly immigrating from the other Nordic countries or non-Nordic European countries). Ex ante, it is not clear that the effects should be the same for these two vastly different groups of immigrants. In addition, since we use coordinate-based data, we do not have to rely on administratively defined neighborhoods (as Åslund et al. (2015) do) but can construct individualized neighborhoods.

Our flexible neighbourhood definition is based on a $k$-nearest neighbor approach. This approach presents several advantages: we can create neighborhoods with constant counts of individuals as compared to administrative units. Furthermore, our approach can better capture what refugees identify as their neighborhood, because it puts the refugee at the center of their own neighborhood. Most importantly, we can conduct small scale neighborhood analysis, down to $k = 100$, capturing potential interactions and social networks. To our knowledge, we are the first to look at small-scale neighborhood integration.

The paper is organized as follows: in Section 3.2, we describe the data and elaborate on the $k$-nearest neighbors approach. In Section 3.3, we introduce the empirical specification and discuss potential threats to identification. We present and discuss the results from the baseline estimates and the decomposition analysis in Section 3.4. We conclude in Section 3.5.

## 3.2 Sample Selection and Neighborhood Definition

### 3.2.1 Data

The analysis uses Swedish geo-coded register data from the GeoSweden database, which contains information on all residents in Sweden. The data is collected on a yearly basis from 1990 to 2014 and consists of variables from the population and tax registers. Importantly for our study, it contains information on the country of birth, reason for and year of immigration. It additionally includes detailed geographic information on residential location, given by coordinates on a $100 \times 100$ meter-level.

Our sample consists of refugee children born between 1974 and 1984 and whose age upon arrival in Sweden is between zero and fifteen years old.[4] A child is considered a refugee if they either have at least one parent classified as a refugee or their own permit is a refugee permit. We study residential characteristics at age 30, hence we are implicitly restricting to those immigrants who do not return to their home country before that age. For each child, we link information on their own education level, their income (measured

[4]The age at arrival variable comes primarily from the in-migration register, which is available from 1990 to 2014. For those arriving before 1990, we use a variable from the income register (Louise) that gives the latest year of immigration. We take the value of this variable when the child first enters the Louise register, at age 16. The earliest cohort that we can observe at age 16 is born in 1974, whereas the youngest cohort we can observe at age 30 is born in 1984. Hence, these data restrictions inform our choice of the cohorts under study.

in percentile ranks, relative to everyone in their birth cohort), number of siblings, as well as their parents' education and income rank.[5]

Table 3.1: Summary statistics

| | Mean | Std. dev. | No. of obs. |
|---|---|---|---|
| **Siblings sample** | | | |
| Child percentile income rank | 40.16 | 30.48 | 22,312 |
| Child has college or above | 35.91 | n/a | 22,137 |
| Parent percentile income rank | 12.99 | 16.30 | 22,312 |
| At least one parent with college or above | 31.78 | n/a | 21,242 |
| Average age at arrival | 9.76 | 3.41 | 22,312 |
| | | | |
| **Full sample** | | | |
| Child percentile income rank | 41.34 | 30.77 | 35,535 |
| Child has college or above | 38.69 | n/a | 35,262 |
| Parent percentile income rank | 14.51 | 17.56 | 35,535 |
| At least one parent with college or above | 35.60 | n/a | 33,910 |
| Average age at arrival | 9.84 | 3.46 | 35,535 |

*Notes:* Child percentile income rank refers to the position in the earnings distribution relative to everyone in a given cohort. Parents are ranked relative to all parents with children in a given cohort. The earnings measure captures income from employment and self-employment. College or above is defined as having at least a post-secondary education that takes fewer than 3 years to complete.

Table 3.1 shows summary characteristics for the refugee children in our sample. Since our empirical strategy uses a siblings design, we show how these differ by sample. We see that both the children in the siblings sample and their parents are less likely to have a university degree or above. There are no significant differences in income rank at age 30 in the two groups, and children in both samples arrive, on average, at around age 10.

[5] We measure parents' income rank when the child is between 15 and 19, so as to get at a measure of financial resources available to the child when they were growing up.

### 3.2.2 Constructing Individualized Neighborhoods Using the $k$ Nearest Neighbors Approach

The GeoSweden database collects geographical coordinates given on a $100 \times 100$ meter level on the 31st of December every year. The $100 \times 100$ meter coordinate information in the data allows us to construct individualized neighborhoods of different sizes using the Equipop software developed by Östh (2014).

The procedure for creating individualized neighborhoods is as follows. For each coordinate in the yearly register, we first identify the $k$-nearest neighbors using the Equipop software, which looks for neighbors in the adjacent $100 \times 100$ grids. Similarly, we next identify the neighbors with a particular characteristic among the $k$-nearest. We then take the ratio of these two values so as to obtain the share of neighbors with a certain characteristic among the $k$-nearest neighbors for each coordinate in the yearly register. The $k$-nearest neighborhood approach ensures that individuals residing at the same coordinate obtain the same value for the share of a certain characteristics among their $k$-nearest neighbors.

There are various reasons for using the $k$-nearest neighbor approach. While administrative units are defined differently in different municipalities, the $k$-nearest neighbor approach allows us to construct neighborhoods with almost constant counts of individuals (Östh, 2014). Furthermore, another improvement from this approach is that it can better capture what refugees identify as their neighborhoods, as the refugees are located at the center of their own neighborhoods. Thus, the resulting neighborhood characteristics are good representations of the actual urban context surrounding the individual. Additionally, the $k$-nearest neighbor approach enables the creation of small neighbourhoods. The small scale analysis, down to $k = 100$, used in this paper reveals the individuals that the refugees are most likely to interact with, and is thus required to detect interactions and social networks. This interaction can play an important role for future integration outcomes. According to Galster (2008), the behaviors and attitudes of a neighborhood resident can impact his neighbor. Clustering of refugees with neighbors of certain characteristics also does not go unnoticed with a small scale analysis.

This paper shows results for neighborhoods of two different sizes: 100 and 1000. The different neighborhood sizes allow us to capture the characteristics of individuals that immigrants may encounter and possibly interact with both very locally (such as in the building they live in) and more broadly in the area they live in (at work, in shops etc.).

As described above, the algorithm looks for the closest $k$ neighbors, starting from adjacent grids. Variation in density across areas may pose concerns regarding the kinds of neighborhoods we can capture with this procedure. In high-density areas, on the one hand, it can happen that the adjacent grid contains more than $k$ neighbors. In that case, the algorithm reports all the neighbors that are close. In Figure C.1, we show that for most of our sample, the difference between $k$ and the actual number of neighbors that the algorithm finds is between 0 and 200, for both $k = 100$ and $k = 1000$. In low-density areas, by contrast, the algorithm may have to travel to farther grids in order to reach the desired level of $k$. Figure C.2 shows that this does not seem to be a concern in the case of $k = 100$. As it is expected, slightly larger distances have to be covered in order to reach $k = 1000$ neighbors. Nonetheless, these distances are rarely larger than 400 meters. Together, these figures suggest that we capture similarly sized neighborhoods within similarly large areas, regardless of area density.

We focus on four neighborhood-level characteristics: i) share of natives, where natives are defined as those born in Sweden regardless of their parents' country of birth; ii) share of highly-educated, where high education is defined as having more than high-school education; iii) share of high-earners, that is, those earning above the median in the municipality earnings distribution and iv) share who receive social assistance benefits.[6]

Table 3.2 shows neighborhood characteristics at age 30 for three different subgroups: natives (column 1), the full sample of refugees (column 2), and the siblings sample of refugees. While the neighborhood characteristics at age 30 of the two groups of refugee groups are very similar to each other, there are some clear differences between the native-born individuals and those arriving as refugees. The two groups differ the least in terms of the share of high-educated neighbors, but the native-born individuals have a larger

[6]Note that anyone that receives a non-zero amount of social assistance in a given year is considered to be a welfare recipient.

share of natives, a larger share of high-income earners, and a lower share of individuals on welfare among their neighbors than refugees.

Table 3.2: Outcomes in different groups

| | Natives | Refugees, full sample | Refugees, siblings sample |
|---|---|---|---|
| *k = 100* | | | |
| Share natives | 0.86 | 0.65 | 0.64 |
| Share high-educated | 0.32 | 0.30 | 0.29 |
| Share high-earners | 0.53 | 0.46 | 0.45 |
| Share on welfare | 0.04 | 0.10 | 0.11 |
| *k = 1000* | | | |
| Share natives | 0.85 | 0.66 | 0.66 |
| Share high-educated | 0.32 | 0.30 | 0.30 |
| Share high-earners | 0.51 | 0.45 | 0.45 |
| Share on welfare | 0.04 | 0.09 | 0.10 |
| Observations | 819,420 | 35,535 | 22,312 |

*Notes:* Natives are born in Sweden to Swedish parents. Refugees are born abroad to foreign-born parents and arrive in Sweden between the ages of 0 and 15.

## 3.3 Empirical Strategy

As highlighted by Alexander and Ward (2018), there are two main empirical issues that we have to consider when estimating the effects of age at arrival on neighborhood integration: collinearity and selection bias. In this section, we describe how we address each of these issues in order to get closer to estimating the causal effect of age at arrival on residential integration.

We cannot simultaneously estimate the effect of age at arrival, birth cohort and years spent in Sweden since they are collinear with each other. Therefore, we use natives to identify the birth-cohort neighborhood profile and estimate whether age at arrival influences deviations from this profile. This is accomplished through a two-stage procedure.[7] In the first stage, we use all natives born in the same birth cohorts as the immigrants in our sample and estimate the following equation to identify the birth-cohort neighborhood profile of natives:

$$y_i^{native} = \lambda_b + \varepsilon_i \tag{3.1}$$

where $y_i^{native}$ denotes the natives' neighborhood characteristics and $\lambda_b$ constitute a full set of birth cohort fixed effects.

In the second step, we use our sample of immigrants and examine whether their age at arrival is related to deviations from the native birth-cohort neighborhood profile. This is achieved by regressing immigrants' neighborhood characteristics at age 30 in deviations from the average neighborhood characteristics of natives born in the same birth cohort estimated in equation (3.1), $(y_i - \widehat{\lambda_b})$, on age at arrival in Sweden $(a_i)$ and individual and family characteristics that can be observed in the data $(\mathbf{X}_i)$:[8].

$$y_i - \widehat{\lambda_b} = \alpha + \sum_{a=4}^{15} \beta_a I(a_i = a) + \gamma \mathbf{X}_i + \eta_i \tag{3.2}$$

The second issue we have to address is potential selection bias. The worry is that parents with better unobservables (in terms of, e.g. motivation, parenting skills, and other variables that might be correlated with the outcome variables but that are not observed in the data) may to a larger extent migrate when their children are young. In other words, the controls in equation (3.2) may not capture all child and parent characteristics that drive both earlier arrival in Sweden and later-life outcomes. We therefore estimate a

[7]This procedure has been used earlier in the literature. See, for example, Alexander and Ward (2018), who apply the procedure in an analysis of the effects of age at arrival during the Age of Mass Migration in the United States on labor market outcomes. We adjust this procedure to our setting and estimate birth-cohort instead of life-cycle profiles since all individuals in our sample are observed at the same age.

[8]The reference category pools ages 0-3.

model with family fixed effects that allows us to identify the effect of every additional year of childhood spent in Sweden on later-life outcomes by using within-family differences in age at arrival. The final model is hence given by:

$$y_{ij} - \widehat{\lambda_b} = \alpha + \sum_{a=4}^{15} \beta_a I(a_{ij} = a) + \mu \text{first-born}_{ij} + \theta \text{female}_{ij} + \phi_j + \eta_{ij} \tag{3.3}$$

where $y_{ij}$ is the outcome of child $i$ in family $j$, $a_{ij}$ is the child's age at arrival in Sweden, and $\phi_j$ is the family fixed effect that captures unobserved family characteristics that are common to all siblings in the same family and constant over time. We follow previous literature that highlights the importance of birth order effects and add a dummy for first-born children (Böhlmark, 2008). We additionally control for gender to capture gender differences in the outcomes we consider.

To get a sense of how the baseline category (i.e., those that arrive at age 0-3) is doing relative to their corresponding cohort of natives, Table 3.3 reports the mean of the variable $y_i - \widehat{\lambda_b}$ for that age group. It can first be noted that, on average, those that arrive at age 0-3 have approximately a 20 percentage points lower share of natives among their closest neighbors at age 30 compared to their corresponding native cohort. Even though the two groups have been living in Sweden for more or less their whole life, their close neighborhoods are markedly different in ethnic composition.

For the three socio-economic variables, we see a different picture with almost no, or very small differences, between the two groups. At age 30, those that arrived at age 0-3 have 1 percent more high-educated individuals among their closest neighbors, 4 percent more individuals on welfare, and approximately 4 percent less high-income earners.

In the bottom panel of Table 3.3, we also note that, still compared to their corresponding native cohort, those that arrive early are 10 percentile ranks lower in terms of earned income, they have half a year of less education, they are 9 percent less likely to be married, and they are 48 percent less likely to be married to a native Swede (conditional on being married at age 30).

Table 3.3: Baseline means

| | Baseline mean |
|---|---|
| *Panel A: Residential integration outcomes* | |
| Share natives | |
| $k$ = 100 | -0.20 |
| $k$ = 1000 | -0.19 |
| Share high-educated | |
| $k$ = 100 | 0.01 |
| $k$ = 1000 | 0.01 |
| Share high-earners | |
| $k$ = 100 | -0.05 |
| $k$ = 1000 | -0.04 |
| Share on welfare | |
| $k$ = 100 | 0.04 |
| $k$ = 1000 | 0.04 |
| *Panel B: Other integration outcomes* | |
| Income rank | -10.26 |
| Years of education | -0.57 |
| Marriage | -0.09 |
| Intermarriage | -0.48 |

*Notes:* The baseline means refer to the pooled category of those who arrive between the ages of 0 and 3.

## 3.4 Results

Our results are presented in the following three sections.

In section 3.4.1, we first present the effects of age at immigration on residential integration, which constitute our baseline estimates. In order to examine the extent to which the effects on residential integration work via other integration channels (income, educational attainment, and intermarriage), we first estimate the effects of age at arrival on these three outcomes in section 3.4.2 and then decompose the main effect estimated in section 3.4.1 into the different parts in section 3.4.3.

### 3.4.1 Effects on Residential Integration

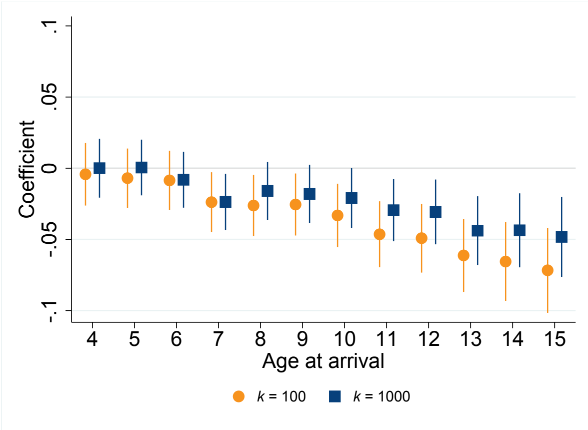
**Residential Integration in terms of Ethnicity**

Figure 3.2a plots the $\beta_a$ coefficients obtained when estimating equation (3.3) with share of natives as the dependent variable. We see from the figure that there is a strong negative relationship between age at arrival and the share of natives among the $k$-nearest neighbors at age 30, for both $k = 100$ and $k = 1000$. The effect is a precisely estimated zero until the age of seven (which roughly corresponds to the school-starting age in Sweden), at which point the effect turns negative.[9] The point estimates at $k = 100$ are slightly more negative than the point estimates at $k = 1000$, implying that refugees have a smaller share of natives among their very closest neighbors.[10] In terms of magnitudes, these coefficients imply that those arriving at age 15 end up in neighborhoods with an approximately five ($k = 1000$) to seven ($k = 100$) percentage point lower share of natives among their neighbors compared to those arriving at ages 0-3 and relative to the corresponding native cohort. These effects amount to 25-35 percent of the mean value for the reference group (c.f. Table 3.3).

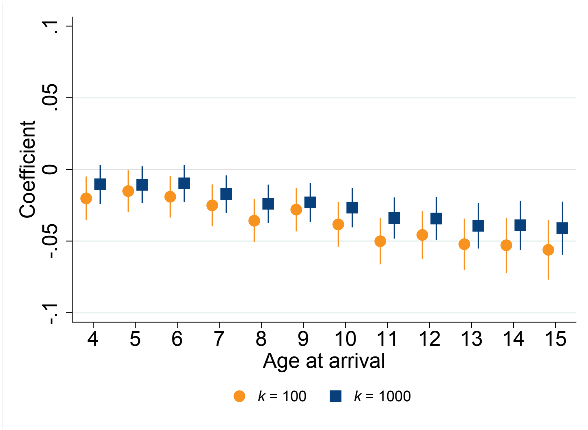
[9]The effect is relative to those arriving at age 0-3 relative to their corresponding native cohort; c.f. equation (3.3).

[10]Our results for the effect of refugees' age at arrival on the share of natives among $k = 1000$ closest neighbors are in line with the results in Åslund et al. (2015) for immigrants in earlier cohorts that typically did not arrive as refugees. They measure residential integration at an administratively-determined unit, the SAMS area, which has on average approximately 1000 inhabitants.

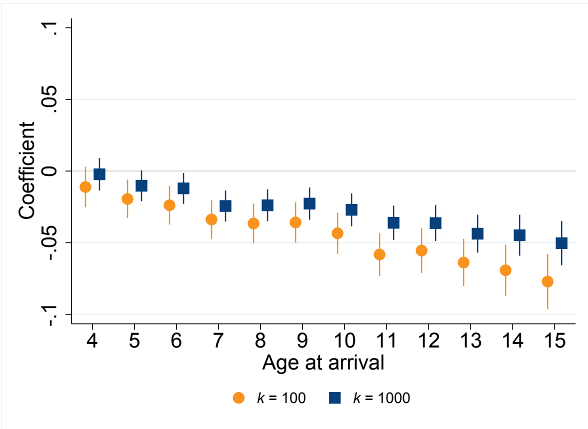
Figure 3.2: Effect of age at arrival on residential integration outcomes



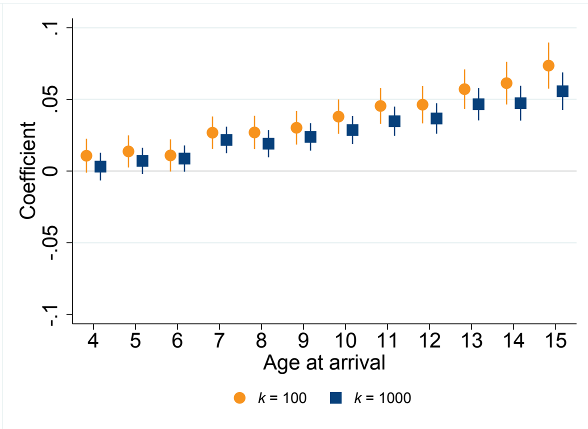

(a) Share of natives




(b) Share of high-educated




(c) Share of high-earners




(d) Share on welfare

*Notes:* The figure shows the $\beta_a$ coefficients obtained when estimating equation (3.3) and their corresponding 95% confidence intervals.
*Source*: Own calculations on data from the GeoSweden database.

### Residential Integration in terms of Socio-economic Characteristics

We focus on three variables when examining the socio-economic composition of the refugee children's neighbors when they reach the age of 30: share highly-educated, share high-income earners, and share on welfare. From the results, presented in Figures 3.2b-3.2d, there are three main conclusions that can be drawn.

First, the older a child is when arriving in Sweden, the more disadvantageous is his or her neighborhood at age 30 (in terms of the neighbors' socio-economic characteristics); there are significantly lower shares of highly educated individuals and high-income earners and a significantly higher share of individuals on welfare compared to the reference category.

For example, being older than 10 years old instead of 0-3 years old when immigrating to Sweden implies an approximately 5 percentage points lower share of highly educated neighbors (c.f. Figure 3.2b), an approximately 6-8 percentage point lower share of high-income earners (c.f. Figure 3.2c), and an approximately 3-6 percentage point higher share of welfare recipients (c.f. Figure 3.2d). Comparing these estimates with the mean values for the reference group (see Table 3.3), we see that the magnitudes of the estimates are sizeable.[11]

Second, as we saw for the share of natives, the effect starts being more pronounced at around school-starting age and, in absolute values, the effects seem to continuously increase in magnitude with each age of arrival.

Third, the effects seem to be fairly similar no matter the size of the neighborhood, even though the effects seem to be somewhat less positive for close neighborhoods ($k = 100$).

### 3.4.2 Effects on Labor Market, Educational, and Marital Integration

The earlier refugee children arrive in a new country, the more time they have to build up country-specific knowledge (e.g. different forms of networks, new language, cultural habits, institutional knowledge). This country-specific knowledge might also affect other forms of (integration) outcomes that, in turn, might affect residential integration. Here we examine the effects on three other important margins: labor market, educational, and social (marital) integration.

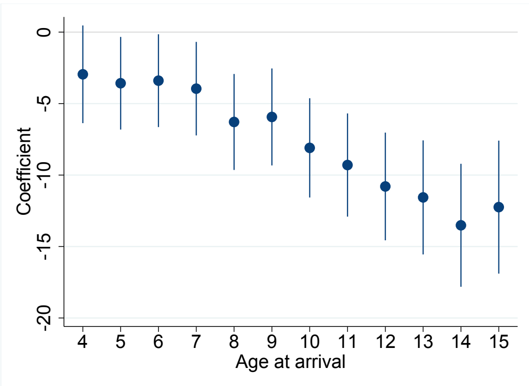
Earlier research on the effects of age at immigration has found that, for immigrants in general, the earlier they arrive, the better they do on the labor market, the higher is their educational achievement, and the more they marry over ethnic lines. From Figures 3.3a-3.3d, we see that this is also true for refugee immigrants. For instance, arriving in Sweden at age 15 instead of at age 0-3 implies that refugees have, on average, approximately a 12.5 lower percentile rank in the earnings distribution at age 30, a half a year less of

[11] In terms of the socio-economic characteristics examined in this paper, those that arrive at age 0-3 live in neighborhoods that are very similar to their native counterparts. This group has a 1 percentage point higher (5 percentage points lower/4 percentage point higher) share of high-educated (high-earners/on welfare) neighbor than the natives.

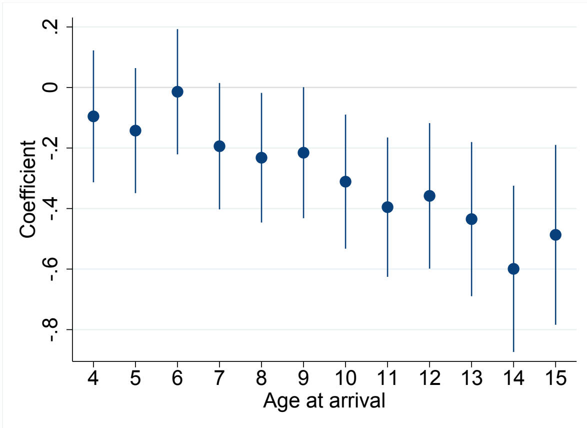
education, and a 28 percentage point lower probability of being married to a native-born partner (conditional on being married at age 30; overall, they are more likely to be married at age 30). Relating the point estimates to the baseline means (see Table 3.3), the effects are very large.

Given that age at arrival matters for labor market, education, and intermarriage outcomes, it is of interest to examine how much of the baseline estimates of age at arrival on residential integration can be explained by these three intermediate channels. We turn to this in the next section.

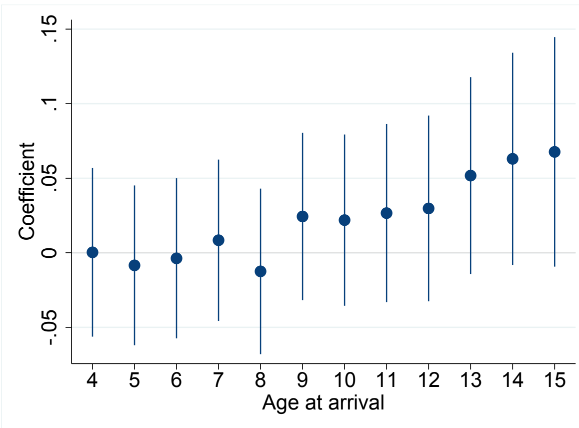
Figure 3.3: Effect of age at arrival on other integration outcomes



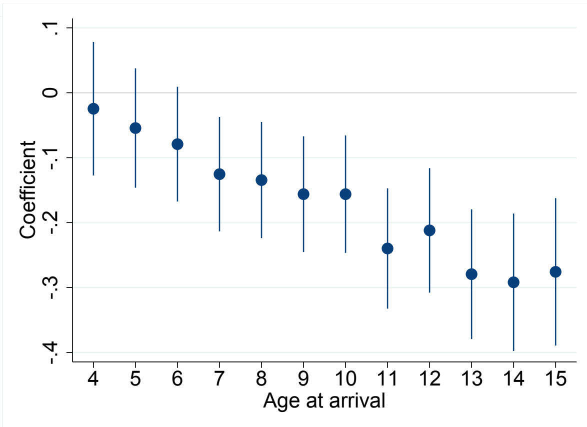

(a) Income rank




(b) Years of education




(c) Married or cohabiting




(d) Married to native, conditional on being married

*Notes:* The figure shows the $\beta_a$ coefficients obtained when estimating equation (3.3) and their corresponding 95% confidence intervals.
*Source*: Own calculations on data from the GeoSweden database.

### 3.4.3 Decomposing the Main Effect on Residential Integration

We conduct a decomposition of the main effects in the style of Heckman et al. (2013), in which we decompose the effects of age at immigration on neighborhood integration via economic integration, educational integration and intermarriage. To be able to interpret this as a causal effect from the mediators, we need to make strong assumptions. The first assumption is that all unobserved factors should be uncorrelated with both age at arrival and the mediators and orthogonal to the link between the mediators and neighborhood integration. For this reason, we rather think of this method as a descriptive tool to better understand our results.

Since the estimated effects observed in Figures 3.2a-3.2d are fairly linear, we have chosen to estimate equation (3.3) with age of the child entering linearly in the decomposition exercise (that is, we decompose a linear effect of age at arrival). The reason for this choice is in terms of clarity; instead of presenting a decomposition analysis for each and every age coefficient estimated in Figures 3.2a-3.2d, we present an overall decomposition analysis.

The decomposition is conducted in three steps:

1. We first estimate equation (3.3) with a linear age variable and with the variables income rank, years of education and intermarriage as additional covariates, and save the coefficients on these three additional variables and the main effect of age. These coefficients are in columns (1)-(4) in Table C.1.
2. We then estimate equation (3.3) with a linear age at arrival variable, separately for each of the variables income rank, years of education and intermarriage as outcome variables. We save the coefficient on the age variable from each of these regressions (columns (5)-(7) in Table C.1).
3. Finally, we calculate the contribution of each of the three "channel" variables. This is done by multiplying the coefficient on each variable as estimated in the first step with the respective coefficient on age as estimated in the second step. This means that we weight the contribution of each variable to the main outcome by the effect

of age on that variable. These estimated contributions can be found in columns (8)-(10) of Table C.1.

The total effect is equal to the main effect of age plus the contributions considered, and the shares are equal to each contribution divided by the total effect. These shares are presented in Table 3.4.[12]

Table 3.4: Decomposition

| | Income rank | Years of education | Intermarriage | Residual |
|---|---|---|---|---|
| *Panel A: k = 100* | | | | |
| Share natives | 0.0822 | 0.0234 | 0.2872 | 0.6073 |
| Share high-educated | 0.0695 | 0.0890 | 0.2606 | 0.5809 |
| Share high-earners | 0.0987 | 0.0276 | 0.2307 | 0.6431 |
| Share on welfare | 0.0673 | 0.0222 | 0.1267 | 0.7838 |
| *Panel B: k = 1000* | | | | |
| Share natives | 0.1062 | 0.0291 | 0.4015 | 0.4632 |
| Share high-educated | 0.0642 | 0.0950 | 0.3065 | 0.5342 |
| Share high-earners | 0.0750 | 0.0312 | 0.2305 | 0.6633 |
| Share on welfare | 0.0657 | 0.0220 | 0.1588 | 0.7535 |

*Notes:* The table presents the decomposition analysis for the married sample. The estimates to construct this table can be seen in Table C.1 from the Appendix.

The overall impression from the results is that there is a large part of the variation in the baseline effect of age at immigration on neighborhood integration that is still unexplained even after accounting for potential effects going through the three mediators. If we look at $k = 100$, we see that the unexplained variation varies from just below 60 percent (for

[12]The decomposition presented in Table 3.4 is based on those individuals that had married at age 30. The reason for this is that we want to decompose the main affects into all three intermediate channels. However, it can be noted that when we use the full sample and decompose the baseline effects into the labor market and education channels, we get shares for these intermediate channels that are very similar to those in Table 3.4, see Table C.2 and the corresponding Table C.3 with the estimates obtained at steps 1-3 in the decomposition exercise.

share high-educated among the $k = 100$ nearest neighbors) to 78 percent (for share on welfare). Of the three mediators, the largest part of the baseline estimates are accounted for by intermarriage and the smallest part by years of education. If we compare the results for $k = 100$ and $k = 1000$, we note that there is a larger unexplained variation in the share of natives for $k = 100$. For the socio-economic variables, the unexplained variation is more similar over neighborhoods of different sizes. It is however worth stressing that the estimated shares presented in Table 3.4 can probably not be given a causal interpretation, so they should be interpreted with this in mind.

## 3.5 Conclusions

The aim of this paper was to examine if, and to what extent, refugee children who arrive at earlier ages in Sweden live in better neighborhoods in adulthood. We reach three conclusions. In our baseline results, we find that those that arrive at younger ages in Sweden (and in particular before school-starting age) are more geographically integrated at age 30: among their very closest neighbors, they have larger shares of natives, highly educated and high-earning individuals and a lower share of individuals on welfare (compared to their older siblings and once we account for time-invariant, unobserved family characteristics). This indicates that a longer exposure to the host country from an earlier age results in better residential integration outcomes in adulthood (in terms of close neighbors' ethnicity and socio-economic composition).

A long exposure to the host country might, however, also affect other margins, such as labor market and education outcomes, as well as marriage patterns which, in turn, might affect the refugees' choice of residential area at age 30. Examining this, we find that the younger the refugees are when they arrive, the more they earn, the more educated they are and the more likely they are to marry Swedish-born partners by the age of 30.

When examining how large a share of the baseline results is explained by the three intermediate channels, our results indicate that they explain some but far from all of the mean age at arrival effects estimated in the baseline analysis. The unexplained variation, that is, the variation left after accounting for intermediate effects via the labor market, education, and intermarriage channels, is for almost all characteristics and neighborhood sizes larger than 50 percent, and when looking at the characteristics among the 100 nearest neighbors, it varies between 60 and 80 percent.

How can we understand this large unexplained variation in residential integration? What can affect residential integration that does not work via the three intermediate channels examined in this paper? As we see it, there are at least three potential candidates. First, there can be a taste-based explanation that works independently from the three channels. Arriving at different ages can, for example, have differential effects on preferences for certain types of neighbors to have contact with. That this might be a possible explanation is indicated by the mean values for those that arrive between ages 0 and 3 (c.f. Table 3.3 ); at age 30, those individuals live in neighborhoods that are more or less identical to their corresponding cohort of native-born individuals in terms of socio-economic characteristics, but with markedly fewer individuals born in Sweden. One interpretation of this is that they have preferences for interacting with neighbors that are similar to them, both in terms of socio-economic characteristics and in terms of country of birth.

Second, even if they do well in the labor market, and can afford to live in any neighborhood that matches their preferences, they may not be able to realize those choices in the presence of discrimination in the housing market. Ethnic-based housing market discrimination could explain the pattern observed in Table 3.3.

Third, even those that arrive late and do not manage as well in the labor market may enter more affluent neighborhoods due to the way Swedish housing policies are designed. Tenure mix policies, where the aim is to build different forms of housing tenures in the same neighborhood (e.g. owner-occupied as well as rentals), are intended to promote social mix. In addition, the Swedish rental system in such that rents are not market-determined and individuals are placed in municipality-specific queues for rental apartments, whereby available apartments are offered to the person that has spent the longest in the queue. Since the municipality-owned companies have their properties in all types of areas, affluent as well as less affluent, a given individual can end up in areas with affluent neighbors, independent of his or her own income. We think these types of housing policies have the potential to explain a large part of the unexplained variation observed in the estimates. Examining these three types of explanations would be an important topic for future research.

# Appendices

# Appendix A

# Appendix to Chapter 1

## A.1 Average Factor Loadings by Year

The factor analysis has been conducted separately for each municipality and each year. To give a sense of the factor loadings obtained for the three variables used in the analysis, Table A.1 presents the yearly means over the municipalities.

Table A.1: Average factor loadings by year

| Year | Share of natives | Share of high educated | Share of employed |
|---|---|---|---|
| **1990** | 0.6227 | 0.7094 | 0.8214 |
| **1991** | 0.7475 | 0.7366 | 0.865 |
| **1992** | 0.7086 | 0.7508 | 0.8574 |
| **1993** | 0.7467 | 0.7467 | 0.8802 |
| **1994** | 0.7795 | 0.6429 | 0.8993 |
| **1995** | 0.784 | 0.6605 | 0.8963 |
| **1996** | 0.7732 | 0.6862 | 0.8964 |
| **1997** | 0.7793 | 0.7111 | 0.8961 |
| **1998** | 0.7818 | 0.7277 | 0.8918 |
| **1999** | 0.7655 | 0.7335 | 0.8853 |

*Notes:* This table shows the average factor loadings for the three different neighborhood characteristics for each year.
*Source:* Own calculations based on data from the GeoSweden database.

## A.2 Balancing Tests for *k=50* and *k=250*

Tables 1.2-A.9 provides balancing tests for individualized neighborhoods based on $k = 50$ and $k = 250$.

Table A.2: Balancing test: neighborhood index ($k = 50$)

| | Low N.Index (N = 6039) | | High N.Index (N = 8012) | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Normalized Difference |
| Age | 33.39 | 6.85 | 33.44 | 6.72 | 0.01 |
| Children | 0.52 | 0.50 | 0.56 | 0.50 | 0.09 |
| Married | 0.30 | 0.46 | 0.28 | 0.45 | -0.04 |
| Female | 0.36 | 0.48 | 0.38 | 0.49 | 0.04 |
| Africa | 0.18 | 0.39 | 0.15 | 0.36 | -0.09 |
| Europe | 0.15 | 0.35 | 0.16 | 0.37 | 0.04 |
| East Asia | 0.08 | 0.27 | 0.09 | 0.28 | 0.03 |
| West Asia | 0.52 | 0.50 | 0.53 | 0.50 | 0.02 |
| Latin America | 0.07 | 0.25 | 0.07 | 0.26 | 0.02 |

*Notes:* The table shows the balancing test for low versus high neighborhood quality index in the initial year for $k = 50$.
*Source:* Own calculations based on data from the GeoSweden database.

Table A.3: Balancing test: native neighbors ($k = 50$)

| | Low Native (N=5841) | | High Native (N=8210) | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Normalized Difference |
| Age | 33.51 | 6.91 | 33.35 | 6.68 | -0.02 |
| Children | 0.53 | 0.50 | 0.55 | 0.50 | 0.05 |
| Married | 0.30 | 0.46 | 0.28 | 0.45 | -0.04 |
| Female | 0.37 | 0.48 | 0.37 | 0.48 | 0.01 |
| Africa | 0.17 | 0.37 | 0.17 | 0.37 | 0.00 |
| Europe | 0.15 | 0.36 | 0.16 | 0.36 | 0.02 |
| East Asia | 0.09 | 0.28 | 0.08 | 0.28 | 0.00 |
| West Asia | 0.53 | 0.50 | 0.52 | 0.50 | -0.03 |
| Latin America | 0.07 | 0.25 | 0.07 | 0.26 | 0.03 |

*Notes:* The table shows the balancing test for low versus high share of natives neighborhoods in the initial year for $k = 50$.
*Source:* Own calculations based on data from the GeoSweden database.

Table A.4: Balancing test: educated neighbors ($k = 50$)

| | Low Educated (N = 5792) | | High Educated (N = 8259) | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Normalized Difference |
| Age | 33.36 | 6.87 | 33.46 | 6.71 | 0.01 |
| Children | 0.53 | 0.5 | 0.55 | 0.5 | 0.03 |
| Married | 0.29 | 0.46 | 0.29 | 0.45 | -0.01 |
| Female | 0.36 | 0.48 | 0.38 | 0.48 | 0.02 |
| Africa | 0.17 | 0.37 | 0.16 | 0.37 | -0.01 |
| Europe | 0.15 | 0.35 | 0.16 | 0.36 | 0.03 |
| East Asia | 0.09 | 0.28 | 0.08 | 0.28 | 0.00 |
| West Asia | 0.53 | 0.50 | 0.52 | 0.50 | 0.00 |
| Latin America | 0.07 | 0.26 | 0.07 | 0.25 | -0.02 |

*Notes:* The table shows the balancing test for low versus high share of educated neighborhoods in the initial year for $k = 50$.
*Source:* Own calculations based on data from the GeoSweden database.

Table A.5: Balancing test: employed neighbors ($k = 50$)

| | Low Employed (N = 6078) | | High Employed (N = 7973) | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Normalized Difference |
| Age | 33.17 | 6.73 | 33.61 | 6.80 | 0.07 |
| Children | 0.50 | 0.50 | 0.58 | 0.49 | 0.17 |
| Married | 0.32 | 0.47 | 0.27 | 0.44 | -0.12 |
| Female | 0.36 | 0.48 | 0.38 | 0.49 | 0.05 |
| Africa | 0.19 | 0.39 | 0.15 | 0.36 | -0.10 |
| Europe | 0.15 | 0.36 | 0.16 | 0.36 | 0.02 |
| East Asia | 0.08 | 0.27 | 0.09 | 0.28 | 0.03 |
| West Asia | 0.52 | 0.50 | 0.53 | 0.50 | 0.02 |
| Latin America | 0.07 | 0.25 | 0.07 | 0.26 | 0.03 |

*Notes:* The table shows the balancing test for low versus high share of employed neighborhoods in the initial year for $k = 50$.
*Source:* Own calculations based on data from the GeoSweden database.

Table A.6: Balancing test: neighborhood index ($k = 250$)

| | Low N.Index (N = 5909) | | High N.Index (N = 8142) | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Normalized Difference |
| Age | 33.32 | 6.84 | 33.49 | 6.73 | 0.03 |
| Children | 0.53 | 0.50 | 0.55 | 0.50 | 0.04 |
| Married | 0.30 | 0.46 | 0.28 | 0.45 | -0.04 |
| Female | 0.37 | 0.48 | 0.38 | 0.48 | 0.02 |
| Africa | 0.16 | 0.37 | 0.17 | 0.38 | 0.03 |
| Europe | 0.14 | 0.34 | 0.16 | 0.37 | 0.08 |
| East Asia | 0.09 | 0.28 | 0.08 | 0.28 | -0.01 |
| West Asia | 0.54 | 0.50 | 0.51 | 0.50 | -0.06 |
| Latin America | 0.07 | 0.26 | 0.07 | 0.25 | -0.02 |

*Notes:* The table shows the balancing test for low versus high neighborhood quality index in the initial year for $k = 250$.
*Source:* Own calculations based on data from the GeoSweden database.

Table A.7: Balancing test: native neighbors ($k = 250$)

| | Low Native (N=5667) | | High Native (N=8384) | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Normalized Difference |
| Age | 33.60 | 6.85 | 33.30 | 6.72 | -0.04 |
| Children | 0.55 | 0.50 | 0.54 | 0.50 | -0.02 |
| Married | 0.29 | 0.46 | 0.29 | 0.45 | -0.01 |
| Female | 0.38 | 0.48 | 0.37 | 0.48 | -0.02 |
| Africa | 0.15 | 0.36 | 0.18 | 0.38 | 0.07 |
| Europe | 0.14 | 0.35 | 0.16 | 0.37 | 0.06 |
| East Asia | 0.09 | 0.28 | 0.08 | 0.28 | -0.01 |
| West Asia | 0.54 | 0.50 | 0.51 | 0.50 | -0.06 |
| Latin America | 0.08 | 0.27 | 0.07 | 0.25 | -0.04 |

*Notes:* The table shows the balancing test for low versus high share of natives neighborhoods in the initial year for $k = 250$.
*Source:* Own calculations based on data from the GeoSweden database.

Table A.8: Balancing test: employed neighbors ($k = 250$)

| | Low Employed (N = 5791) | | High Employed (N = 8260) | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Normalized Difference |
| Age | 33.28 | 6.80 | 33.51 | 6.76 | 0.03 |
| Children | 0.52 | 0.50 | 0.56 | 0.50 | 0.07 |
| Married | 0.31 | 0.46 | 0.28 | 0.45 | -0.06 |
| Female | 0.37 | 0.48 | 0.38 | 0.48 | 0.02 |
| Africa | 0.17 | 0.38 | 0.16 | 0.37 | -0.03 |
| Europe | 0.14 | 0.35 | 0.16 | 0.37 | 0.07 |
| East Asia | 0.08 | 0.27 | 0.09 | 0.28 | 0.02 |
| West Asia | 0.53 | 0.50 | 0.52 | 0.50 | -0.03 |
| Latin America | 0.07 | 0.26 | 0.07 | 0.25 | -0.02 |

*Notes:* The table shows the balancing test for low versus high share of employed neighborhoods in the initial year for $k = 250$.
*Source:* Own calculations based on data from the GeoSweden database.

Table A.9: Balancing test: educated neighbors ($k = 250$)

| | Low Educated (N = 5624) | | High Educated (N = 8127) | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Normalized Difference |
| Age | 33.40 | 6.85 | 33.43 | 6.72 | 0.00 |
| Children | 0.55 | 0.50 | 0.54 | 0.50 | -0.02 |
| Married | 0.29 | 0.45 | 0.29 | 0.45 | 0.01 |
| Female | 0.37 | 0.48 | 0.37 | 0.48 | 0.00 |
| Africa | 0.16 | 0.37 | 0.17 | 0.38 | 0.03 |
| Europe | 0.15 | 0.36 | 0.16 | 0.36 | 0.01 |
| East Asia | 0.09 | 0.28 | 0.08 | 0.28 | -0.01 |
| West Asia | 0.53 | 0.50 | 0.52 | 0.50 | -0.01 |
| Latin America | 0.08 | 0.26 | 0.07 | 0.25 | -0.03 |

*Notes:* The table shows the balancing test for low versus high share of educated neighborhoods in the initial year for $k = 250$.
*Source:* Own calculations based on data from the GeoSweden database.

## A.3 Full Results for Neighborhood Integration

Tables A.10-A.13 presents the detailed results for the effects of initial neighborhood quality on future neighborhood quality (corresponding to Figure 1.5; c.f. Equation (1.1)).

Table A.10: Effect of initial neighborhood quality index on future neighborhood quality index, $k = 100$

| VARIABLES | (1) Neighborhood t+1 | (2) t+2 | (3) t+3 | (4) t+4 | (5) t+5 | (6) t+6 | (7) t+7 | (8) t+8 |
|---|---|---|---|---|---|---|---|---|
| Neighborhood t | 0.360*** | 0.211*** | 0.157*** | 0.133*** | 0.111*** | 0.108*** | 0.0919*** | 0.0888*** |
| | (0.0243) | (0.0190) | (0.0166) | (0.0135) | (0.0128) | (0.0130) | (0.0132) | (0.0133) |
| Age | yes | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Observations | 13,989 | 13,991 | 13,987 | 13,976 | 13,980 | 13,984 | 13,873 | 13,696 |
| $R^2_{neighqual}$ | 0.171 | 0.095 | 0.093 | 0.096 | 0.098 | 0.109 | 0.120 | 0.122 |
| $F-stats$ | 47.95 | 24.44 | 16.39 | 20.28 | 15.17 | 14.19 | 11.78 | 12.72 |
| No. Municipalities | 283 | 283 | 283 | 283 | 283 | 283 | 283 | 283 |

*Notes:* All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses. Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

Table A.11: Effect of initial share of native on future share of native, $k = 100$

| VARIABLES | (1) Share of native t+1 | (2) t+2 | (3) t+3 | (4) t+4 | (5) t+5 | (6) t+6 | (7) t+7 | (8) t+8 |
|---|---|---|---|---|---|---|---|---|
| Share of native t | 0.467*** | 0.302*** | 0.228*** | 0.199*** | 0.176*** | 0.157*** | 0.148*** | 0.140*** |
| | (0.0280) | (0.0318) | (0.0350) | (0.0265) | (0.0282) | (0.0302) | (0.0342) | (0.0333) |
| Age | yes | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Observations | 14,021 | 14,023 | 14,019 | 14,014 | 14,019 | 14,031 | 13,931 | 13,751 |
| $R^2_{native}$ | 0.362 | 0.230 | 0.178 | 0.160 | 0.159 | 0.161 | 0.162 | 0.170 |
| $F-stats$ | 52.11 | 24.83 | 16.15 | 18.74 | 14.12 | 10.24 | 10.13 | 10.90 |
| No. Municipalities | 283 | 283 | 283 | 283 | 283 | 283 | 283 | 283 |

*Notes:* All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses. Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

Table A.12: Effect of initial share of high educated on future share of high educated, $k = 100$

| VARIABLES | (1) Share of high educated t+1 | (2) t+2 | (3) t+3 | (4) t+4 | (5) t+5 | (6) t+6 | (7) t+7 | (8) t+8 |
|---|---|---|---|---|---|---|---|---|
| Share of high educated t | 0.461*** | 0.297*** | 0.224*** | 0.197*** | 0.165*** | 0.172*** | 0.131*** | 0.137*** |
| | (0.0140) | (0.0154) | (0.0156) | (0.0156) | (0.0155) | (0.0156) | (0.0155) | (0.0152) |
| Age | yes | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Observations | 14,021 | 14,023 | 14,019 | 14,014 | 14,019 | 14,031 | 13,931 | 13,751 |
| $R^2_{edu}$ | 0.387 | 0.272 | 0.248 | 0.213 | 0.200 | 0.193 | 0.191 | 0.189 |
| $F-stats$ | 35.20 | 24.98 | 13.22 | 11.85 | 10.97 | 10.56 | 10.38 | 10.16 |
| No. Municipalities | 283 | 283 | 283 | 283 | 283 | 283 | 283 | 283 |

*Notes:* All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses. Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

Table A.13: Effect of initial share of employed on future share of employed, $k = 100$

| VARIABLES | (1) Share of employed t+1 | (2) t+2 | (3) t+3 | (4) t+4 | (5) t+5 | (6) t+6 | (7) t+7 | (8) t+8 |
|---|---|---|---|---|---|---|---|---|
| Share of employed t | 0.402*** | 0.206*** | 0.145*** | 0.106*** | 0.0885*** | 0.0724*** | 0.0618*** | 0.0460*** |
| | (0.0260) | (0.0168) | (0.0192) | (0.0175) | (0.0161) | (0.0139) | (0.0159) | (0.0158) |
| Age | yes | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Observations | 14,021 | 14,023 | 14,019 | 14,014 | 14,019 | 14,031 | 13,931 | 13,751 |
| $R^2_{empl}$ | 0.344 | 0.248 | 0.198 | 0.178 | 0.176 | 0.181 | 0.185 | 0.187 |
| $F-stats$ | 52.10 | 28.89 | 11.88 | 10.99 | 10.64 | 9.98 | 8.81 | 8.79 |
| No. Municipalities | 283 | 283 | 283 | 283 | 283 | 283 | 283 | 283 |

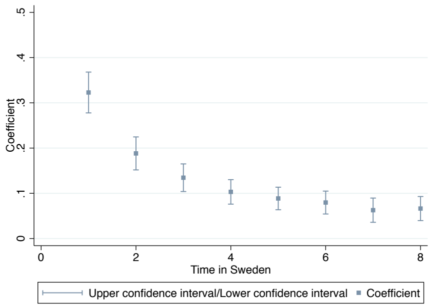
*Notes:* All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses. Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

# A.4 Robustness Checks for Neighborhood Integration

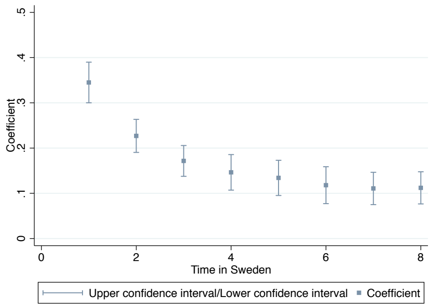
In this section, we present the results from the robustness analysis conducted on the model in Equation (1.1); c.f. section 1.4.2.

## A.4.1 Other Definitions of the Size of the Neighborhood

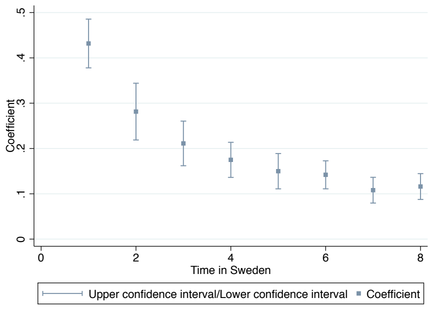
Figure A.1: Effects of initial neighborhood quality on the quality of future neighborhoods: $k = 50$



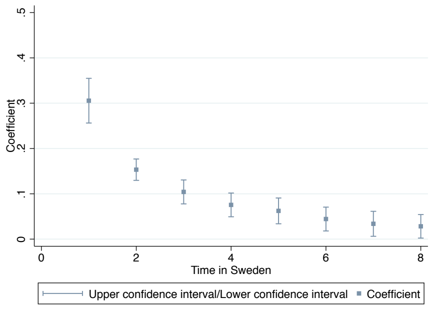

(a) Quality index




(b) Share natives



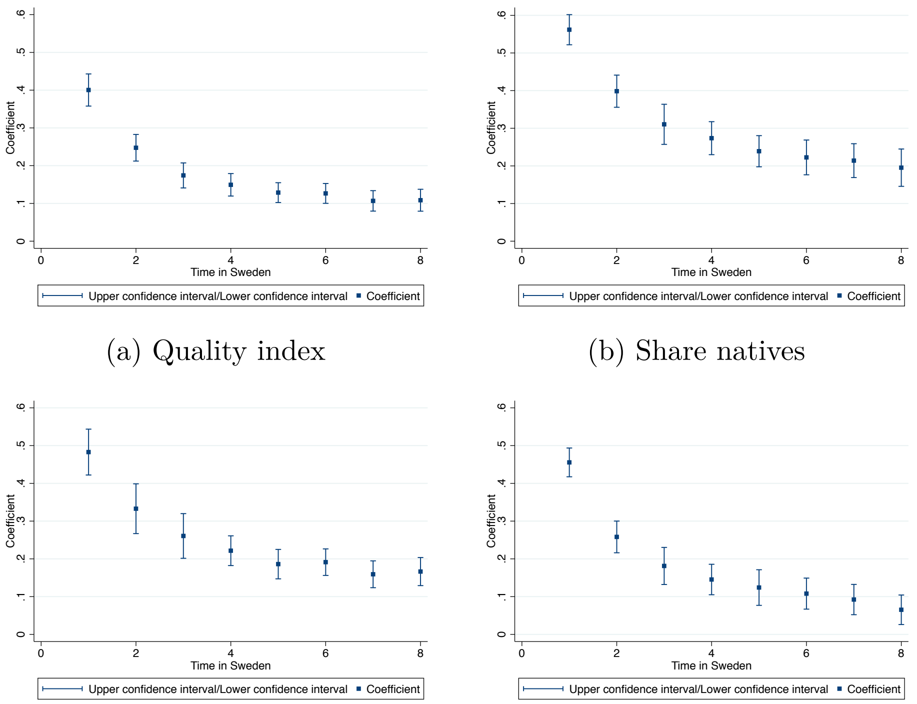

(c) Share educated




(d) Share employed

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (1.1) for $k = 50$ and their corresponding 95% confidence intervals.
*Source*: Own calculations on data from the GeoSweden database.

Figure A.2: Effects of initial neighborhood quality on the quality of future neighborhoods: $k = 250$



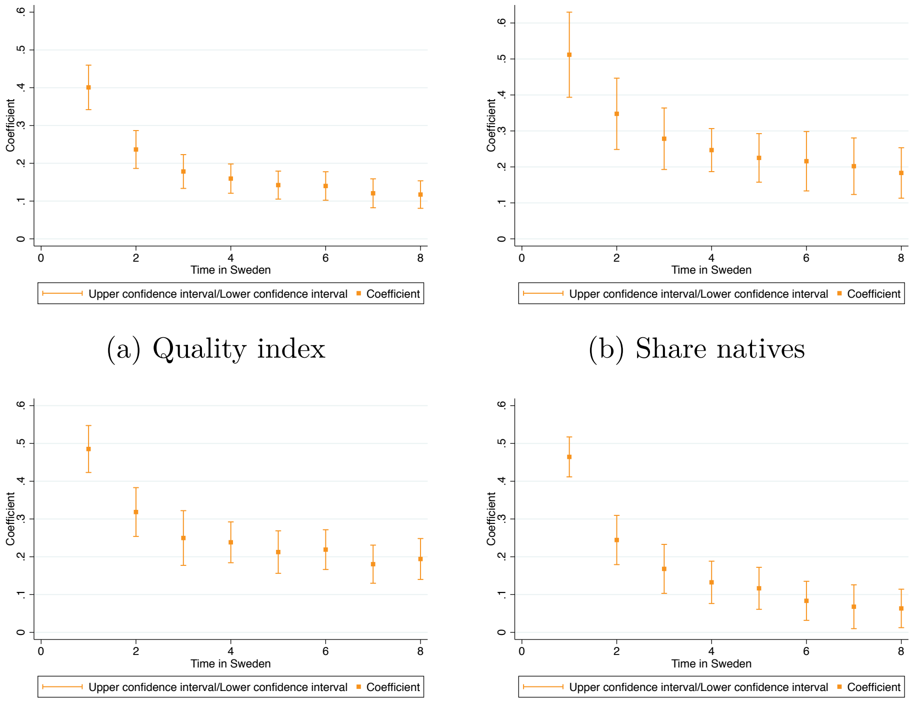

(a) Quality index (b) Share natives

(c) Share educated (d) Share employed

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (1.1) for $k = 250$ and their corresponding 95% confidence intervals.
*Source*: Own calculations on data from the GeoSweden database.

## A.4.2 Restricting to Refugees Arriving with Children

Figure A.3: Effects of initial neighborhood quality on the quality of future neighborhoods for parents only: $k = 100$




(a) Quality index

(b) Share natives

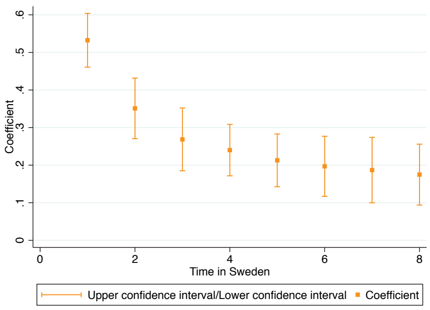
(c) Share educated

(d) Share employed

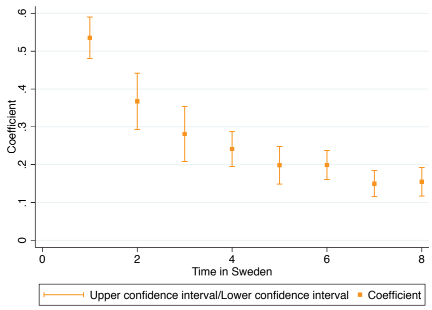
*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (1.1) for parents and their corresponding 95% confidence intervals. The number of observations is 7609 in t+1 and 7503 in t+8.
*Source*: Own calculations on data from the GeoSweden database.

### A.4.3 Restricting the Distance Needed to Reach the *k=100* Closest Neighbors and Restricting the Actual Number of Neighbors Reached when Searching for the *k=100* Closest Neighbors

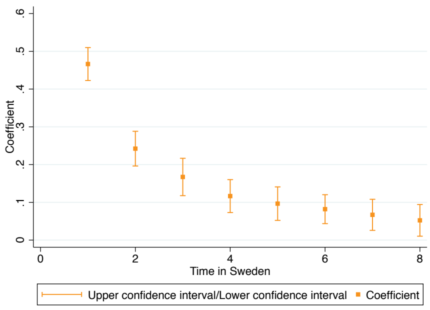
Figure A.4: First stage regressions restricting distance 100 meters: $k = 100$




(a) Share natives



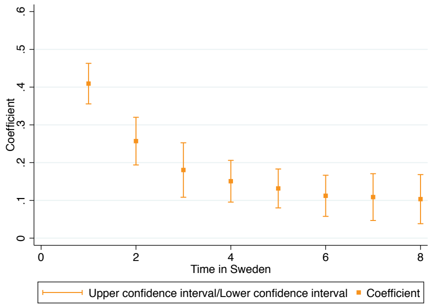

(b) Share educated



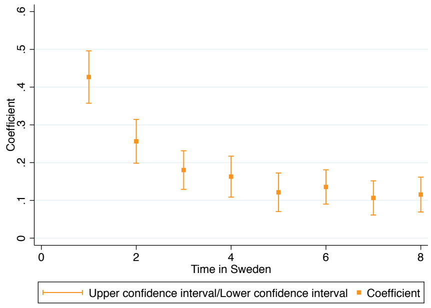

(c) Share employed

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (1.1) for distance restriction to 100 meters and their corresponding 95% confidence intervals. The number of observations is 9974 in t+1 and 9741 in t+8.
*Source*: Own calculations on data from the GeoSweden database.

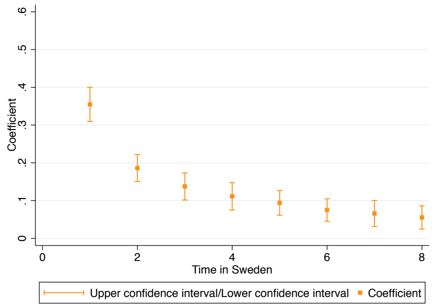
Figure A.5: First stage regressions restricting neighbor counts to 200: $k = 100$




(a) Share natives




(b) Share educated




(c) Share employed

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (1.1) for population counts restriction to 200 individuals and their corresponding 95% confidence intervals.
*Source*: Own calculations on data from the GeoSweden database.

## A.5 Full Results for Earnings

Tables A.14-A.17 presents the detailed results for the effects of initial neighborhood quality on future neighborhood quality (corresponding to Figure 1.7; c.f. Equation (1.2)).

Table A.14: Effect of neighborhood quality index on future earnings, 2SLS regressions: $k = 100$

| VARIABLES | (1) ln earnings t+1 | (2) t+2 | (3) t+3 | (4) t+4 | (5) t+5 | (6) t+6 | (7) t+7 | (8) t+8 |
|---|---|---|---|---|---|---|---|---|
| Neighborhood | 0.0938 | 0.175 | 0.388** | 0.635*** | 0.596** | 0.509** | 0.468* | 0.366 |
| | (0.0641) | (0.124) | (0.156) | (0.192) | (0.245) | (0.243) | (0.284) | (0.315) |
| Age | yes | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Observations | 13,989 | 13,991 | 13,987 | 13,976 | 13,980 | 13,984 | 13,873 | 13,696 |
| $R^2_{neighqual}$ | 0.221 | 0.167 | 0.135 | 0.120 | 0.131 | 0.143 | 0.140 | 0.146 |
| No. Municipalities | 283 | 283 | 283 | 283 | 283 | 283 | 283 | 283 |

*Notes:* All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses. Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

Table A.15: Effect of native share on future earnings, 2SLS regressions: $k = 100$

| VARIABLES | (1) ln earnings t+1 | (2) t+2 | (3) t+3 | (4) t+4 | (5) t+5 | (6) t+6 | (7) t+7 | (8) t+8 |
|---|---|---|---|---|---|---|---|---|
| Share of native | 0.372 | 1.012 | 1.622** | 1.974** | 1.876** | 1.672 | 0.774 | 0.459 |
| | (0.419) | (0.732) | (0.737) | (0.849) | (0.933) | (1.365) | (1.509) | (1.754) |
| Age | yes | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Observations | 14,021 | 14,023 | 14,019 | 14,014 | 14,019 | 14,031 | 13,931 | 13,751 |
| $R^2_{native}$ | 0.222 | 0.171 | 0.144 | 0.140 | 0.142 | 0.146 | 0.143 | 0.146 |
| No. Municipalities | 283 | 283 | 283 | 283 | 283 | 283 | 283 | 283 |

*Notes:* All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses. Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

Table A.16: Effect of share of high educated on future earnings, 2SLS regressions: $k = 100$

| VARIABLES | (1) ln earnings t+1 | (2) t+2 | (3) t+3 | (4) t+4 | (5) t+5 | (6) t+6 | (7) t+7 | (8) t+8 |
|---|---|---|---|---|---|---|---|---|
| Share of high educated | -0.663 | 0.292 | -0.115 | 3.141 | 3.540 | 3.086 | 2.413 | 2.490 |
| | (0.687) | (1.310) | (1.581) | (2.028) | (2.813) | (2.237) | (2.833) | (3.307) |
| Age | yes | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Observations | 14,021 | 14,023 | 14,019 | 14,014 | 14,019 | 14,031 | 13,931 | 13,751 |
| $R^2_{edu}$ | 0.220 | 0.165 | 0.136 | 0.130 | 0.131 | 0.138 | 0.136 | 0.140 |
| No. Municipalities | 283 | 283 | 283 | 283 | 283 | 283 | 283 | 283 |

*Notes:* All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses. Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

Table A.17: Effect of share of employed on future earnings, 2SLS regressions: $k = 100$

| VARIABLES | (1) ln earnings t+1 | (2) t+2 | (3) t+3 | (4) t+4 | (5) t+5 | (6) t+6 | (7) t+7 | (8) t+8 |
|---|---|---|---|---|---|---|---|---|
| Share of employed | 1.621*** | 2.282** | 2.886* | 5.035*** | 5.619** | 2.848 | 1.773 | 2.681 |
| | (0.412) | (1.044) | (1.607) | (1.933) | (2.495) | (2.989) | (3.746) | (5.006) |
| Age | yes | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes | yes |
| Observations | 14,021 | 14,023 | 14,019 | 14,014 | 14,019 | 14,031 | 13,931 | 13,751 |
| $R^2_{empl}$ | 0.224 | 0.174 | 0.147 | 0.137 | 0.135 | 0.158 | 0.152 | 0.160 |
| No. Municipalities | 283 | 283 | 283 | 283 | 283 | 283 | 283 | 283 |

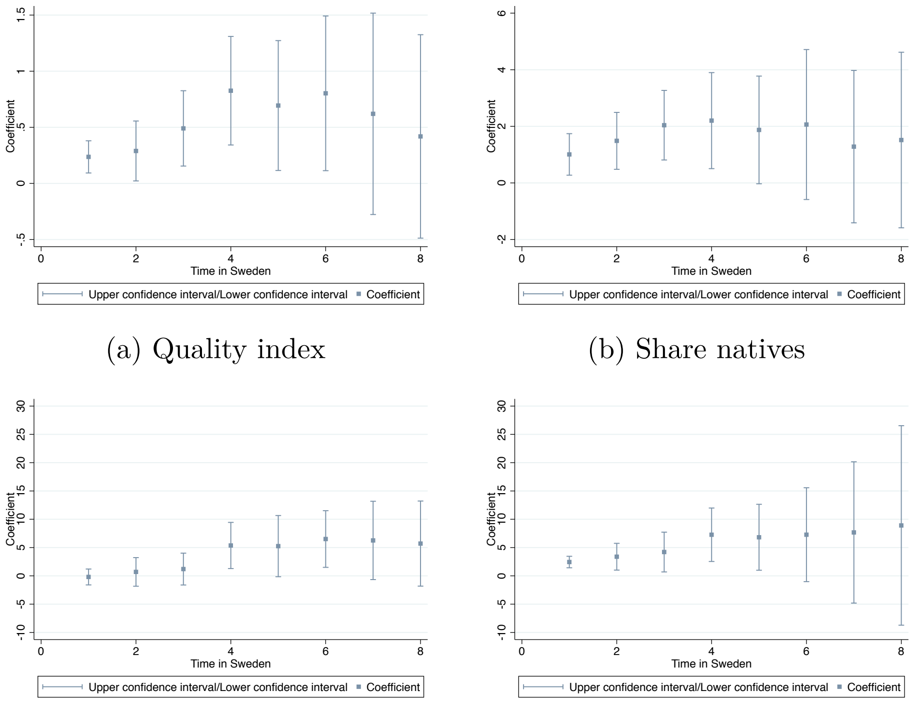
*Notes:* All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses. Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

## A.6 Robustness Checks for Earnings

In this section, we present the results from the robustness analysis conducted on the model in Equation (1.2).

### A.6.1 Other Definitions of the Size of the Neighborhood

Figure A.6: Instrumental variable 2SLS: $k = 50$




(a) Quality index

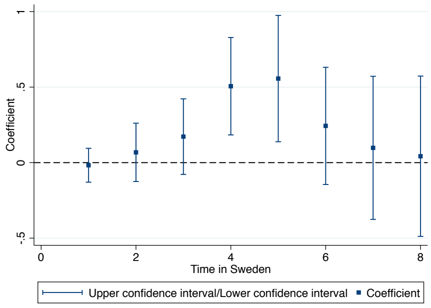
(b) Share natives

(c) Share educated

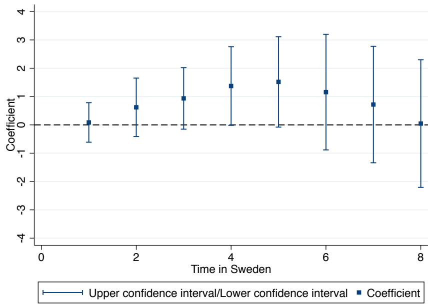
(d) Share employed

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (1.2) and their corresponding 95% confidence intervals for $k = 50$. The y-scales for the figures for quality index and share natives are different to those for share educated and employed.
*Source*: Own calculations on data from the GeoSweden database.

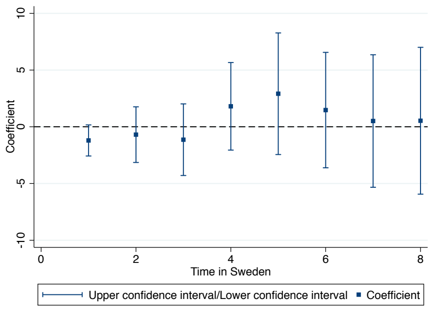
Figure A.7: Instrumental variable 2SLS: $k = 250$



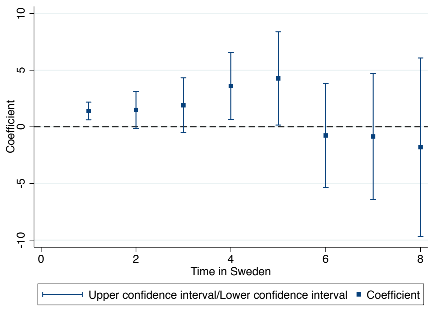

(a) Quality index




(b) Share natives




(c) Share educated

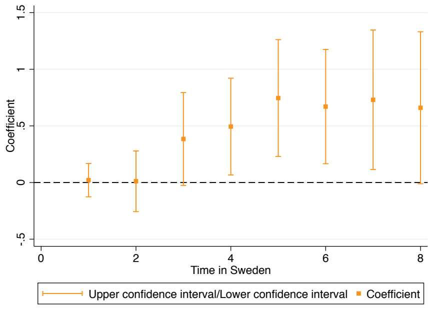



(d) Share employed

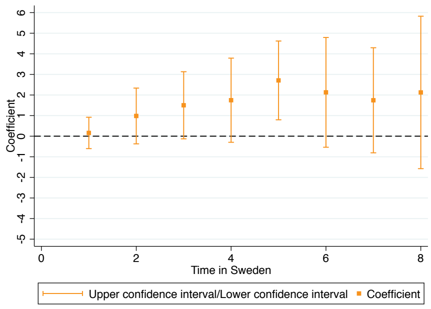
*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (1.2) and their corresponding 95% confidence intervals for $k = 250$. The y-scales for the figures for quality index and share natives are different to those for share educated and employed.
*Source*: Own calculations on data from the GeoSweden database.

### A.6.2 Restricting to Refugees Arriving with Children

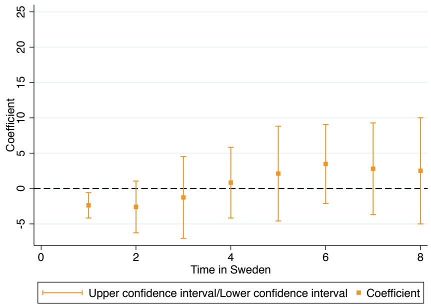
Figure A.8: Parents, 2SLS regressions: $k = 100$



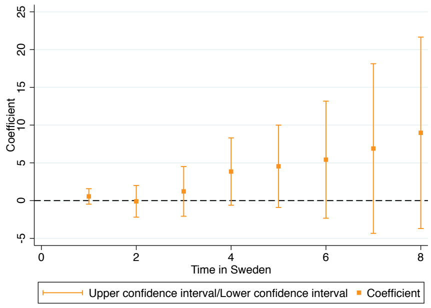

(a) Quality index




(b) Share natives




(c) Share educated

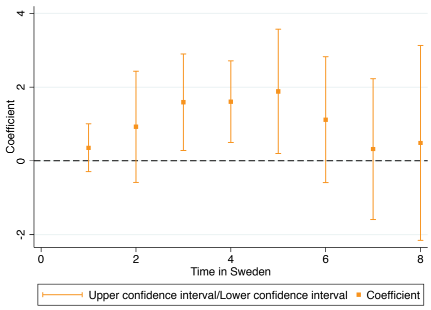



(d) Share employed

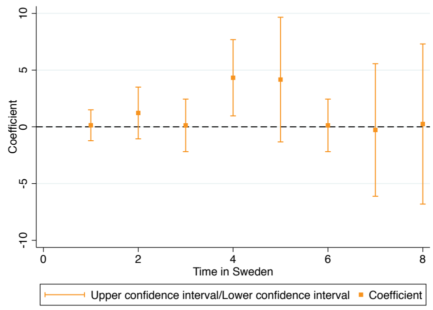
*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (1.2) for parents and their corresponding 95% confidence intervals. The y-scales for the figures for quality index and share natives are different from share educated and employed.
*Source*: Own calculations on data from the GeoSweden database.

### A.6.3 Restricting the Distance Needed to Reach the *k=100* Closest Neighbors and Restricting the Actual Number of Neighbors Reached when Searching for the *k=100* Closest Neighbors

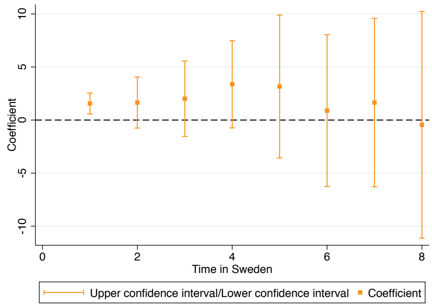
Figure A.9: Restrict distance to 100 meters, 2SLS regressions: $k = 100$




(a) Share natives



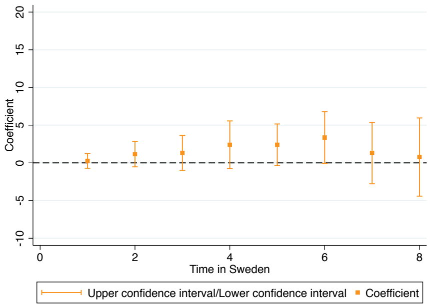

(b) Share educated



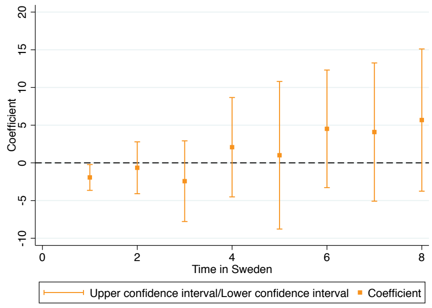

(c) Share employed

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (1.2) for distance restriction to 100 meters and their corresponding 95% confidence intervals. The number of observations is 9974 in t+1 and 9741 in t+8.
*Source*: Own calculations on data from the GeoSweden database.

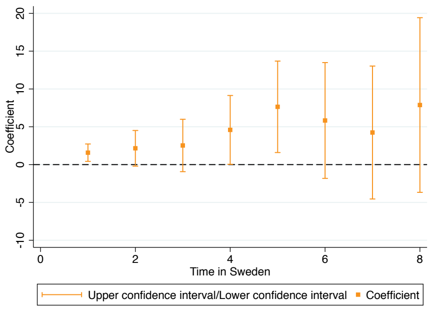
Figure A.10: Restrict neighbor counts to 200, 2SLS regressions: $k = 100$




(a) Share natives




(b) Share educated




(c) Share employed

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (1.2) for neighbor count restriction to 200 individuals and their corresponding 95% confidence intervals.
*Source*: Own calculations on data from the GeoSweden database.

# Appendix B

# Appendix to Chapter 2

## B.1 Additional Descriptive Statistics

Table B.1: Descriptive statistics for shares of co-ethnics

| | Mean | No.of refugees | p25 | p50 | p75 |
|---|---|---|---|---|---|
| Share of co-ethnics $k = 50$ | 0.04 | 10576 | 0.00 | 0.02 | 0.06 |
| Share of co-ethnics $k = 100$ | 0.03 | 10576 | 0.00 | 0.02 | 0.05 |
| Share of co-ethnics $k = 250$ | 0.02 | 10576 | 0.00 | 0.01 | 0.03 |
| Share of co-ethnics $k = 500$ | 0.02 | 10576 | 0.00 | 0.01 | 0.02 |
| Share of co-ethnics $k = 1000$ | 0.01 | 10576 | 0.00 | 0.01 | 0.02 |

*Note:* The table presents summary statistics on shares of co-ethnics in the initial year. p25 denotes the $25^{th}$ percentile, p50 shows the $50^{th}$ percentile and p75 represents the $75^{th}$ percentile.

*Source:* Own calculations from the GeoSweden database.

Table B.2: Descriptive statistics for shares of employed co-ethnics

| | Mean | SD | No.of refugees |
|---|---|---|---|
| Share of employed co-ethnics $k = 250$ | 0.01 | 0.02 | 10576 |
| Share of employed co-ethnics $k = 500$ | 0.01 | 0.02 | 10576 |
| Share of employed co-ethnics $k = 1000$ | 0.01 | 0.02 | 10576 |

*Note:* The table presents summary statistics on shares of employed co-ethnics in the initial year. p25 denotes the $25^{th}$ percentile, p50 shows the $50^{th}$ percentile and p75 represents the $75^{th}$ percentile.

*Source:* Own calculations from the GeoSweden database.

## B.2 Balancing Tests

Table B.3: Balancing test: ethnic neighbors for $k = 100$

| | Low Ethnic Neighborhood N = 4860 | | High Ethnic Neighborhood N = 5716 | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Normalized Difference |
| Age | 32.87 | 6.60 | 33.51 | 6.89 | 0.09 |
| Children | 0.50 | 0.50 | 0.54 | 0.50 | 0.08 |
| Married | 0.30 | 0.46 | 0.31 | 0.46 | 0.02 |
| Female | 0.34 | 0.47 | 0.37 | 0.48 | 0.05 |
| Africa | 0.23 | 0.42 | 0.10 | 0.30 | -0.35 |
| Europe | 0.03 | 0.16 | 0.08 | 0.27 | 0.23 |
| East Asia | 0.06 | 0.23 | 0.06 | 0.23 | 0.00 |
| West Asia | 0.65 | 0.48 | 0.71 | 0.46 | 0.13 |
| Latin America | 0.04 | 0.20 | 0.05 | 0.23 | 0.07 |

*Notes:* The table shows the balancing test for low versus high ethnic neighborhood in the initial year for $k = 100$.
*Source:* Own calculations based on data from the GeoSweden database.

Table B.4: Balancing test: ethnic neighbors for $k = 500$

| | Low Ethnic Neighborhood N = 5011 | | High Ethnic Neighborhood N = 5565 | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Normalized Difference |
| Age | 33.00 | 6.69 | 33.41 | 6.83 | 0.06 |
| Children | 0.49 | 0.50 | 0.55 | 0.50 | 0.12 |
| Married | 0.30 | 0.46 | 0.30 | 0.46 | 0.02 |
| Female | 0.34 | 0.47 | 0.37 | 0.48 | 0.05 |
| Africa | 0.23 | 0.42 | 0.09 | 0.29 | -0.39 |
| Europe | 0.03 | 0.17 | 0.08 | 0.27 | 0.23 |
| East Asia | 0.06 | 0.24 | 0.05 | 0.23 | -0.03 |
| West Asia | 0.64 | 0.48 | 0.71 | 0.45 | 0.15 |
| Latin America | 0.03 | 0.18 | 0.06 | 0.24 | 0.13 |

*Notes:* The table shows the balancing test for low versus high ethnic neighborhood in the initial year for $k = 500$.
*Source:* Own calculations based on data from the GeoSweden database.

Table B.5: Balancing test: ethnic neighbors for $k = 1000$

| | Low Ethnic Neighborhood N = 4955 | | High Ethnic Neighborhood N = 5621 | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Normalized Difference |
| Age | 33.07 | 6.68 | 33.35 | 6.83 | 0.04 |
| Children | 0.48 | 0.50 | 0.56 | 0.50 | 0.16 |
| Married | 0.30 | 0.46 | 0.30 | 0.46 | -0.01 |
| Female | 0.34 | 0.47 | 0.37 | 0.48 | 0.06 |
| Africa | 0.24 | 0.42 | 0.09 | 0.29 | -0.40 |
| Europe | 0.03 | 0.17 | 0.08 | 0.27 | 0.23 |
| East Asia | 0.06 | 0.23 | 0.06 | 0.24 | 0.01 |
| West Asia | 0.65 | 0.48 | 0.71 | 0.46 | 0.13 |
| Latin America | 0.03 | 0.17 | 0.06 | 0.24 | 0.14 |

*Notes:* The table shows the balancing test for low versus high ethnic neighborhood in the initial year for $k = 1000$.
*Source:* Own calculations based on data from the GeoSweden database.

**Employed Co-ethnics**

Table B.6: Balancing test: employed ethnic neighbors $k = 500$

| | Low Employed Ethnic N = 4267 | | High Employed Ethnic N = 6309 | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Normalized Difference |
| Age | 33.31 | 6.80 | 33.16 | 6.74 | -0.02 |
| Children | 0.50 | 0.50 | 0.53 | 0.50 | 0.06 |
| Married | 0.30 | 0.46 | 0.31 | 0.46 | 0.02 |
| Female | 0.35 | 0.48 | 0.36 | 0.48 | 0.01 |
| Africa | 0.20 | 0.40 | 0.13 | 0.34 | -0.19 |
| Europe | 0.03 | 0.18 | 0.07 | 0.26 | 0.18 |
| East Asia | 0.06 | 0.23 | 0.06 | 0.24 | 0.02 |
| West Asia | 0.68 | 0.47 | 0.68 | 0.47 | 0.00 |
| Latin America | 0.03 | 0.17 | 0.06 | 0.24 | 0.15 |

*Notes:* The table shows the balancing test for low versus high employed ethnic neighborhood in the initial year for $k = 500$.
*Source:* Own calculations based on data from the GeoSweden database.

Table B.7: Balancing test: employed ethnic neighbors $k = 1000$

| | Low Employed Ethnic N = 4438 | | High Employed Ethnic N = 6138 | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Normalized Difference |
| Age | 33.30 | 6.81 | 33.16 | 6.73 | -0.02 |
| Children | 0.49 | 0.50 | 0.54 | 0.50 | 0.10 |
| Married | 0.30 | 0.46 | 0.30 | 0.46 | 0.01 |
| Female | 0.35 | 0.48 | 0.36 | 0.48 | 0.03 |
| Africa | 0.21 | 0.41 | 0.12 | 0.33 | -0.24 |
| Europe | 0.03 | 0.18 | 0.07 | 0.26 | 0.19 |
| East Asia | 0.05 | 0.23 | 0.06 | 0.24 | 0.02 |
| West Asia | 0.68 | 0.47 | 0.68 | 0.47 | 0.00 |
| Latin America | 0.02 | 0.16 | 0.06 | 0.24 | 0.19 |

*Notes:* The table shows the balancing test for low versus high employed ethnic neighborhood in the initial year for $k = 1000$.
*Source:* Own calculations based on data from the GeoSweden database.

## B.3 First Stage Results for Neighborhood Composition

Table B.8: Effect of initial co-ethnic share on future co-ethnic share: $k = 100$

| VARIABLES | (1)<br>Ethnic t+4 | (2)<br>t+5 | (3)<br>t+6 | (4)<br>t+7 | (5)<br>t+8 | (6)<br>t+9 | (7)<br>t+10 |
|---|---|---|---|---|---|---|---|
| Ethnic t | 0.269*** | 0.242*** | 0.217*** | 0.214*** | 0.223*** | 0.187*** | 0.153*** |
| | (0.0574) | (0.0469) | (0.0450) | (0.0425) | (0.0484) | (0.0371) | (0.0465) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| $R^2$ | 0.245 | 0.244 | 0.219 | 0.221 | 0.213 | 0.214 | 0.217 |
| F stats | 15.89 | 15.87 | 13.85 | 14.20 | 13.91 | 13.86 | 13.98 |
| Observations | 10,543 | 10,549 | 10,314 | 10,200 | 10,345 | 10,245 | 10,166 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

*Notes:* All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses. Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

Table B.9: Effect of initial co-ethnic share on future co-ethnic share: $k = 250$

| VARIABLES | (1) Ethnic t+4 | (2) t+5 | (3) t+6 | (4) t+7 | (5) t+8 | (6) t+9 | (7) t+10 |
|---|---|---|---|---|---|---|---|
| Ethnic t | 0.357*** | 0.330*** | 0.313*** | 0.308*** | 0.328*** | 0.302*** | 0.233*** |
| | (0.0452) | (0.0441) | (0.0436) | (0.0431) | (0.0497) | (0.0405) | (0.0595) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| Observations | 10,543 | 10,314 | 9,914 | 10,200 | 10,345 | 10,245 | 10,166 |
| $R^2$ | 0.273 | 0.270 | 0.245 | 0.238 | 0.231 | 0.228 | 0.227 |
| F stats | 18.62 | 18.43 | 16.23 | 16.11 | 15.90 | 15.64 | 15.51 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

*Notes:* All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses. Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

Table B.10: Effect of initial co-ethnic share on future co-ethnic share: $k = 500$

| VARIABLES | (1)<br>Ethnic t+4 | (2)<br>t+5 | (3)<br>t+6 | (4)<br>t+7 | (5)<br>t+8 | (6)<br>t+9 | (7)<br>t+10 |
|---|---|---|---|---|---|---|---|
| Ethnic t | 0.404*** | 0.378*** | 0.348*** | 0.337*** | 0.360*** | 0.338*** | 0.324*** |
| | (0.0537) | (0.0558) | (0.0457) | (0.0490) | (0.0506) | (0.0451) | (0.0532) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| Observations | 10,543 | 10,549 | 10,314 | 10,200 | 10,345 | 10,245 | 10,166 |
| $R^2$ | 0.274 | 0.268 | 0.249 | 0.235 | 0.229 | 0.230 | 0.236 |
| F stats | 18.67 | 18.30 | 16.44 | 15.97 | 15.78 | 15.73 | 15.98 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

*Notes:* All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses. Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

Table B.11: Effect of initial co-ethnic share on future co-ethnic share: $k = 1000$

| VARIABLES | (1) Ethnic t+4 | (2) t+5 | (3) t+6 | (4) t+7 | (5) t+8 | (6) t+9 | (7) t+10 |
|---|---|---|---|---|---|---|---|
| Ethnic t | 0.442*** | 0.417*** | 0.391*** | 0.372*** | 0.391*** | 0.378*** | 0.368*** |
| | (0.0547) | (0.0573) | (0.0423) | (0.0479) | (0.0449) | (0.0414) | (0.0487) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| Observations | 10,543 | 10,549 | 10,314 | 10,200 | 10,345 | 10,245 | 10,166 |
| $R^2$ | 0.273 | 0.256 | 0.251 | 0.239 | 0.233 | 0.238 | 0.242 |
| F stats | 18.62 | 17.55 | 16.55 | 16.19 | 15.98 | 16.20 | 16.30 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

*Notes:* All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses. Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

### Employed Co-ethnics

Table B.12: Effect of initial employed co-ethnic share on future employed co-ethnic share: $k = 250$

| VARIABLES | (1) Employed Ethnic t+4 | (2) t+5 | (3) t+6 | (4) t+7 | (5) t+8 | (6) t+9 | (7) t+10 |
|---|---|---|---|---|---|---|---|
| Employed Ethnic t | 0.282*** | 0.252*** | 0.225*** | 0.238*** | 0.272*** | 0.272*** | 0.205*** |
| | (0.0358) | (0.0341) | (0.0333) | (0.0354) | (0.0343) | (0.0339) | (0.0707) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| Observations | 10,543 | 10,549 | 10,314 | 10,200 | 10,345 | 10,245 | 10,166 |
| $R^2$ | 0.273 | 0.256 | 0.251 | 0.239 | 0.233 | 0.238 | 0.242 |
| F stats | 17.93 | 18.72 | 15.69 | 15.72 | 15.17 | 13.88 | 12.50 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

*Notes:* All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses. Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

Table B.13: Effect of initial employed co-ethnic share on future employed co-ethnic share: $k = 500$

| VARIABLES | (1) Employed Ethnic t+4 | (2) t+5 | (3) t+6 | (4) t+7 | (5) t+8 | (6) t+9 | (7) t+10 |
|---|---|---|---|---|---|---|---|
| Employed Ethnic t | 0.310*** | 0.287*** | 0.253*** | 0.259*** | 0.285*** | 0.281*** | 0.285*** |
| | (0.0415) | (0.0438) | (0.0366) | (0.0412) | (0.0345) | (0.0371) | (0.0431) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| Observations | 10,543 | 10,549 | 10,314 | 10,200 | 10,345 | 10,245 | 10,166 |
| $R^2$ | 0.273 | 0.256 | 0.251 | 0.239 | 0.233 | 0.238 | 0.242 |
| F stats | 18.21 | 19.14 | 15.25 | 15.90 | 15.34 | 14.05 | 13.38 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

*Notes:* All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses. Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

Table B.14: Effect of initial employed co-ethnic share on future employed co-ethnic share: $k = 1000$

| VARIABLES | (1) Employed Ethnic t+4 | (2) t+5 | (3) t+6 | (4) t+7 | (5) t+8 | (6) t+9 | (7) t+10 |
|---|---|---|---|---|---|---|---|
| Employed Ethnic t | 0.331*** | 0.304*** | 0.268*** | 0.269*** | 0.292*** | 0.295*** | 0.306*** |
| | (0.0478) | (0.0502) | (0.0394) | (0.0440) | (0.0399) | (0.0421) | (0.0460) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| Observations | 10,543 | 10,549 | 10,314 | 10,200 | 10,345 | 10,245 | 10,166 |
| $R^2$ | 0.273 | 0.256 | 0.251 | 0.239 | 0.233 | 0.238 | 0.242 |
| F stats | 18.17 | 17.16 | 14.97 | 14.51 | 13.89 | 12.76 | 12.09 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

*Notes:* All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses. Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

## B.4 Robustness Checks for Neighborhood Composition

### Movers

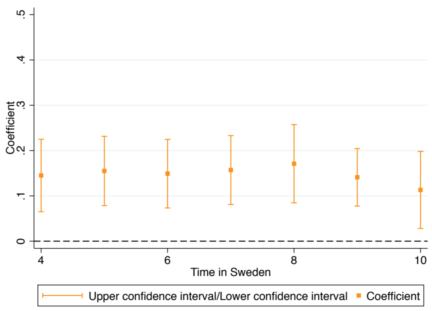
Table B.15: Movers

| Duration year | Number of Movers |
|---|---|
| 4 | 9222 |
| | (87.47) |
| 5 | 9486 |
| | (89.90) |
| 6 | 9540 |
| | (92.49) |
| 7 | 9543 |
| | (93.56) |
| 8 | 9624 |
| | (93.03) |
| 9 | 9631 |
| | (94.00) |
| 10 | 9690 |
| | (95.31) |

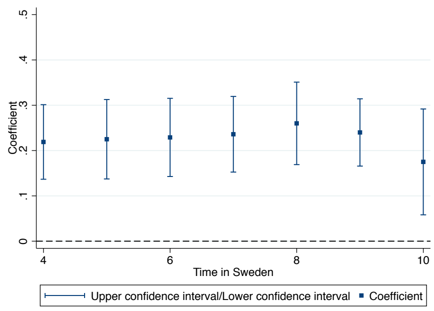
*Note:* The table shows the absolute number of movers for each year and the percentages are given in parentheses.

*Source:* Own calculations based on data from the GeoSweden database.

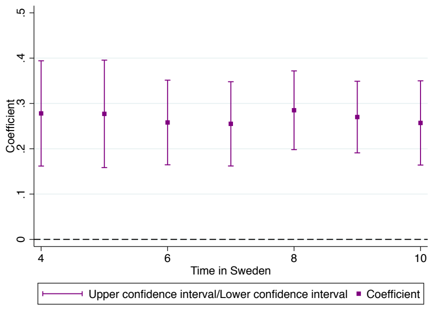
Figure B.1: Movers: effect of initial co-ethnic share on future ethnic share, first stage regressions: $k = 100, k = 250, k = 500, k = 1000$



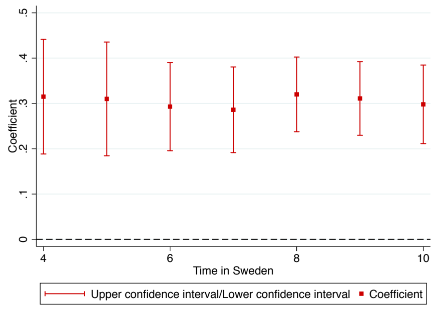

(a) $k = 100$




(b) $k = 250$



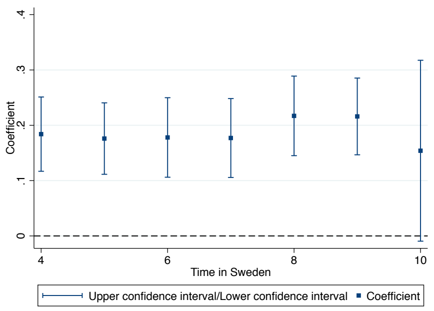

(c) $k = 500$



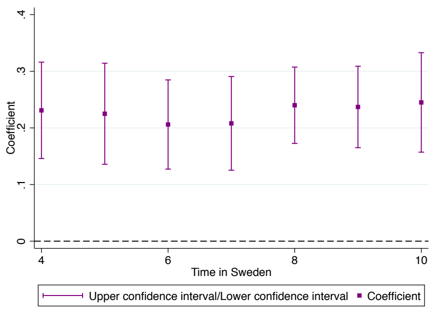

(d) $k = 1000$

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.1) for movers and their corresponding 95% confidence intervals. The number of observations is 9222 in t+4 and 9690 in t+10. *Source*: Own calculations on data from the GeoSweden database.

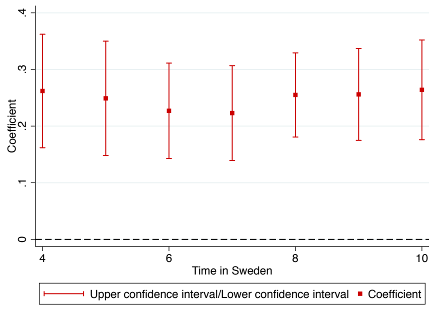
Figure B.2: Movers: effect of initial employed co-ethnic share on future employed ethnic share, first stage regressions: $k = 250, k = 500, k = 1000$




(a) $k = 250$




(b) $k = 500$

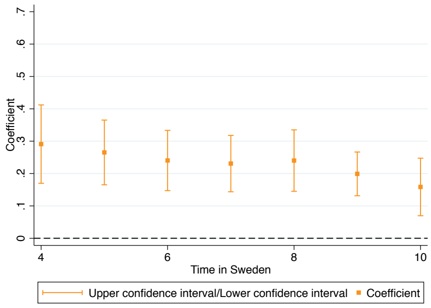



(c) $k = 1000$

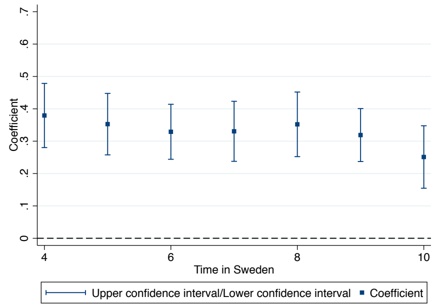
*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.1) for movers and their corresponding 95% confidence intervals. The number of observations is 9222 in t+4 and 9690 in t+10. *Source*: Own calculations on data from the GeoSweden database.

## Distance Restriction to 500 meters

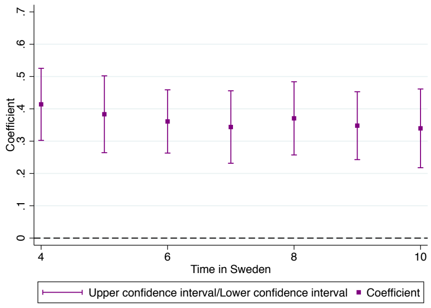
Figure B.3: Distance: effect of initial co-ethnic share on future ethnic share, first stage regressions: $k = 100, k = 250, k = 500, k = 1000$



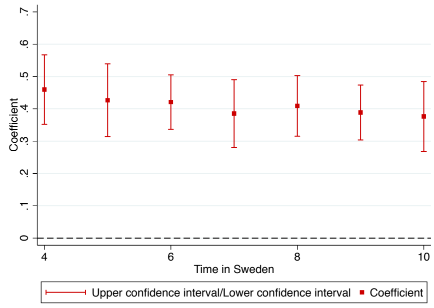

(a) $k = 100$




(b) $k = 250$



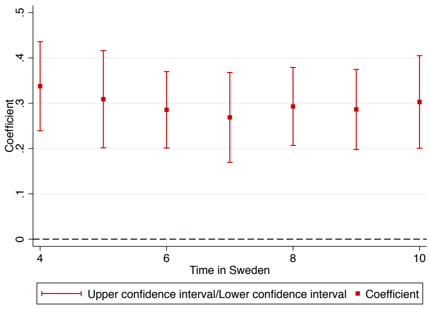

(c) $k = 500$



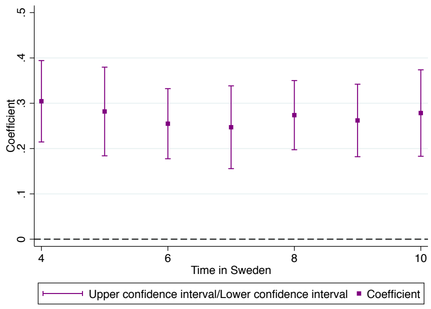

(d) $k = 1000$

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.1) for distance less than 500 meters and their corresponding 95% confidence intervals. The number of observations is 9136 in t+4 and 8805 in t+10 for $k = 250$. For $k = 500$, the number of observations is 7721 in t+4 and 7446 in t+10. The number of observations varies from 5308 in t+4 to 5118 in t+10 for $k = 1000$.
*Source*: Own calculations on data from the GeoSweden database.

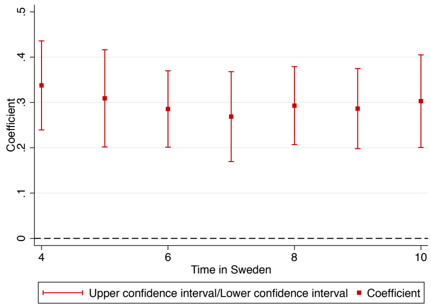
Figure B.4: Distance: effect of initial employed co-ethnic share on future employed ethnic share, first stage regressions: $k = 250, k = 500, k = 1000$




(a) $k = 250$




(b) $k = 500$

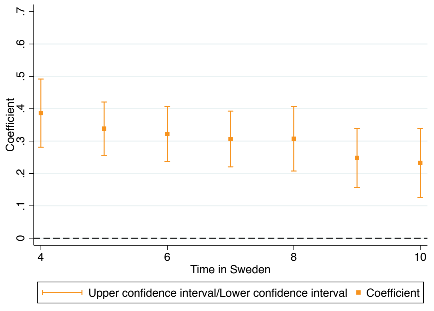



(c) $k = 1000$

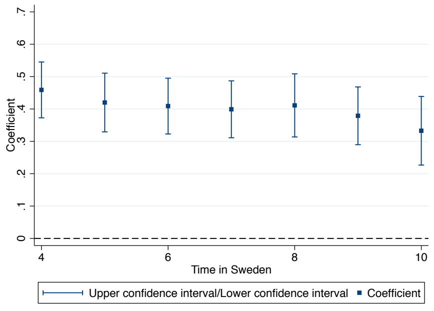
*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.1) for distance less than 500 meters and their corresponding 95% confidence intervals. The number of observations is 9136 in t+4 and 8805 in t+10 for $k = 250$. For $k = 500$, the number of observations is 7721 in t+4 and 7446 in t+10. The number of observations varies from 5308 in t+4 to 5118 in t+10 for $k = 1000$.
*Source*: Own calculations on data from the GeoSweden database.

## Parents

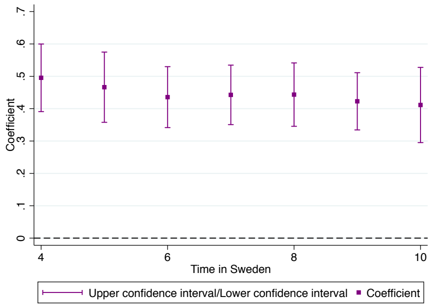
Figure B.5: Parents: effect of initial co-ethnic share on future ethnic share, first stage regressions: $k = 100, k = 250, k = 500, k = 1000$



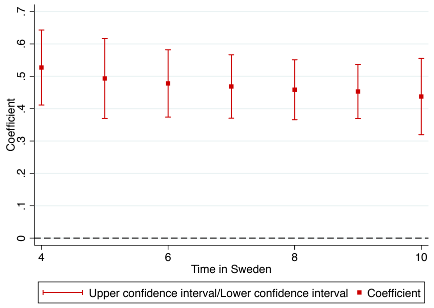

(a) $k = 100$




(b) $k = 250$




(c) $k = 500$

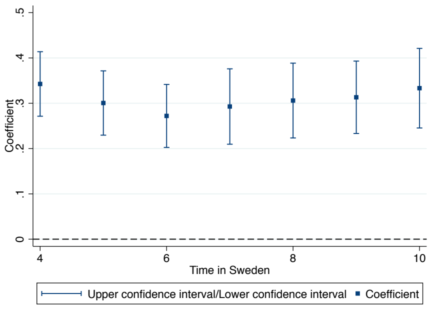



(d) $k = 1000$

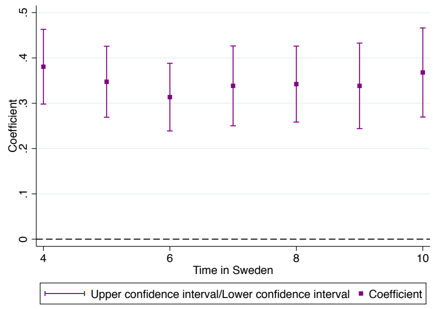
*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.1) for parents and their corresponding 95% confidence intervals. The number of observations is 5501 in t+4 and 5367 in t+10.

*Source*: Own calculations on data from the GeoSweden database.

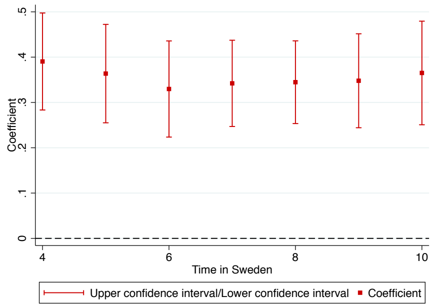
Figure B.6: Parents: effect of initial employed co-ethnic share on future employed ethnic share, first stage regressions: $k = 250, k = 500, k = 1000$




(a) $k = 250$




(b) $k = 500$




(c) $k = 1000$

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.1) for parents and their corresponding 95% confidence intervals. The number of observations is 5501 in t+4 and 5367 in t+10.
*Source*: Own calculations on data from the GeoSweden database.

## B.5 Full Results for Earnings

Table B.16: Effect of co-ethnic share on earnings: $k = 100$

| VARIABLES | (1) Earnings t+4 | (2) t+5 | (3) t+6 | (4) t+7 | (5) t+8 | (6) t+9 | (7) t+10 |
|---|---|---|---|---|---|---|---|
| Ethnic | -0.675 | 3.923 | 5.475* | 4.573 | 9.759** | 6.022 | 3.971 |
| | (1.785) | (2.819) | (3.321) | (3.672) | (4.053) | (4.335) | (6.475) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| Observations | 10,543 | 10,549 | 10,314 | 10,200 | 10,345 | 10,245 | 10,166 |
| $R^2$ | 0.113 | 0.101 | 0.087 | 0.103 | 0.057 | 0.116 | 0.144 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

*Notes:* The table presents the 2SLS estimates of ethnic share on earnings at different $k = 100$ nearest neighbor levels. All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses.Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

Table B.17: Effect of co-ethnic share on earnings: $k = 250$

| VARIABLES | (1)<br>Earnings t+4 | (2)<br>t+5 | (3)<br>t+6 | (4)<br>t+7 | (5)<br>t+8 | (6)<br>t+9 | (7)<br>t+10 |
|---|---|---|---|---|---|---|---|
| Ethnic | 0.0800 | 6.085** | 5.795 | 3.815 | 9.273** | 7.583* | 5.388 |
| | (1.692) | (2.825) | (3.703) | (3.818) | (3.667) | (4.192) | (6.004) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| Observations | 10,543 | 10,549 | 10,314 | 10,200 | 10,345 | 10,245 | 10,166 |
| $R^2$ | 0.112 | 0.090 | 0.089 | 0.110 | 0.076 | 0.109 | 0.139 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

*Notes:* The table presents the 2SLS estimates of ethnic share on earnings at different $k = 250$ nearest neighbor levels. All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses.Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

Table B.18: Effect of co-ethnic share on earnings: $k = 500$

| VARIABLES | (1) Earnings t+4 | (2) t+5 | (3) t+6 | (4) t+7 | (5) t+8 | (6) t+9 | (7) t+10 |
|---|---|---|---|---|---|---|---|
| Ethnic | 1.428 | 5.161* | 4.106 | 4.185 | 8.462** | 6.560 | 1.827 |
| | (1.745) | (3.113) | (4.668) | (3.856) | (3.831) | (5.409) | (6.421) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| Observations | 10,543 | 10,549 | 10,314 | 10,200 | 10,345 | 10,245 | 10,166 |
| $R^2$ | 0.110 | 0.099 | 0.100 | 0.109 | 0.093 | 0.120 | 0.155 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

*Notes:* The table presents the 2SLS estimates of ethnic share on earnings at different $k = 500$ nearest neighbor levels. All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses.Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

Table B.19: Effect of co-ethnic share on earnings: $k = 1000$

| VARIABLES | (1) Earnings t+4 | (2) t+5 | (3) t+6 | (4) t+7 | (5) t+8 | (6) t+9 | (7) t+10 |
|---|---|---|---|---|---|---|---|
| Ethnic | 1.052 | 7.271** | 4.826 | 4.057 | 10.72** | 6.845 | 2.362 |
| | (1.840) | (3.679) | (5.336) | (4.398) | (4.603) | (5.136) | (5.720) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| Observations | 10,543 | 10,549 | 10,314 | 10,200 | 10,345 | 10,245 | 10,166 |
| $R^2$ | 0.110 | 0.090 | 0.098 | 0.110 | 0.085 | 0.122 | 0.154 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

*Notes:* The table presents the 2SLS estimates of ethnic share on earnings at different $k = 1000$ nearest neighbor levels. All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses.Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.

*Source:* Own calculations based on data from the GeoSweden database.

### Employed Co-Ethnics

Table B.20: Effect of employed co-ethnic share on earnings: $k = 250$

| VARIABLES | (1) Earnings t+4 | (2) t+5 | (3) t+6 | (4) t+7 | (5) t+8 | (6) t+9 | (7) t+10 |
|---|---|---|---|---|---|---|---|
| Employed Ethnic | 5.257 | 1.806 | -4.109 | 1.826 | 10.58*** | 8.356* | 11.02* |
| | (3.258) | (4.043) | (6.549) | (5.739) | (3.977) | (4.861) | (6.129) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| Observations | 10,543 | 10,549 | 10,314 | 10,200 | 10,345 | 10,245 | 10,166 |
| $R^2$ | 0.120 | 0.118 | 0.102 | 0.127 | 0.140 | 0.153 | 0.158 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

*Notes:* The table presents the 2SLS estimates of employed ethnic share on earnings for $k = 250$ nearest neighbor levels. All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses.Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

Table B.21: Effect of employed co-ethnic share on earnings: $k = 500$

| VARIABLES | (1) Earnings t+4 | (2) t+5 | (3) t+6 | (4) t+7 | (5) t+8 | (6) t+9 | (7) t+10 |
|---|---|---|---|---|---|---|---|
| Employed Ethnic | 5.421 | 2.303 | -1.525 | 6.246 | 15.31*** | 10.53* | 8.862 |
| | (3.766) | (4.678) | (6.685) | (6.331) | (4.097) | (5.787) | (5.777) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| Observations | 10,543 | 10,549 | 10,314 | 10,200 | 10,345 | 10,245 | 10,166 |
| $R^2$ | 0.117 | 0.117 | 0.111 | 0.128 | 0.131 | 0.148 | 0.161 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

*Notes:* The table presents the 2SLS estimates of employed ethnic share on earnings for $k = 500$ nearest neighbor levels. All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses.Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

Table B.22: Effect of employed co-ethnic share on earnings: $k = 1000$

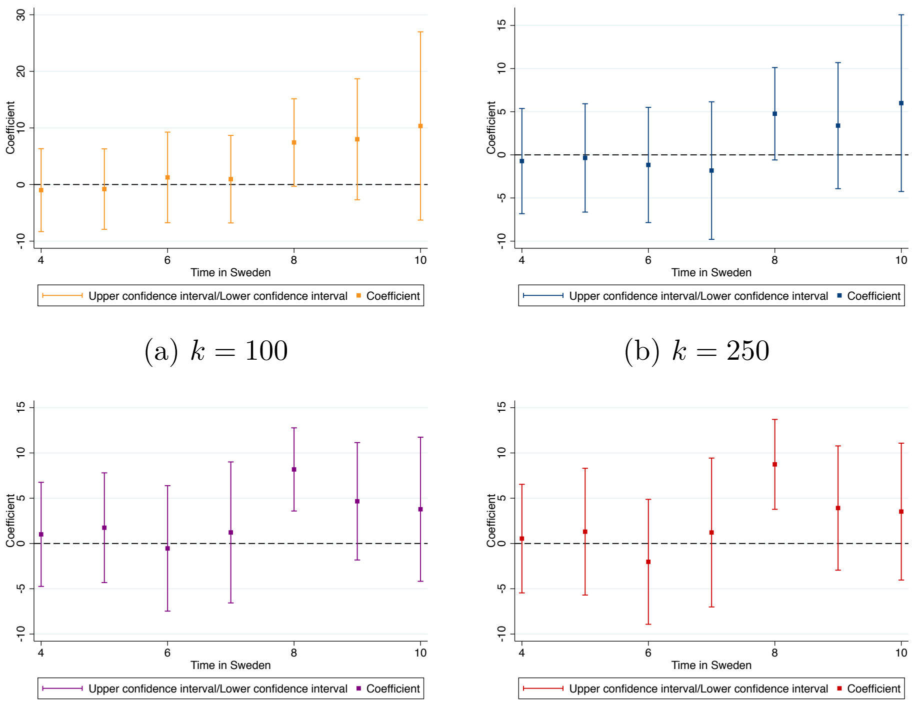
| VARIABLES | (1) Earnings t+4 | (2) t+5 | (3) t+6 | (4) t+7 | (5) t+8 | (6) t+9 | (7) t+10 |
|---|---|---|---|---|---|---|---|
| Employed Ethnic | 3.503 | 2.042 | -2.514 | 6.392 | 17.20*** | 7.835 | 6.884 |
| | (4.005) | (5.264) | (7.153) | (7.132) | (4.471) | (5.797) | (5.294) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| Observations | 10,543 | 10,549 | 10,314 | 10,200 | 10,345 | 10,245 | 10,166 |
| $R^2$ | 0.114 | 0.115 | 0.111 | 0.125 | 0.130 | 0.149 | 0.162 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

*Notes:* The table presents the 2SLS estimates of employed ethnic share on earnings for $k = 1000$ nearest neighbor levels. All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses.Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

## B.6 Robustness Checks for Earnings

### Movers

Figure B.7: Movers: effect of initial co-ethnic share on earnings, 2SLS regressions: $k = 100, k = 250, k = 500, k = 1000$

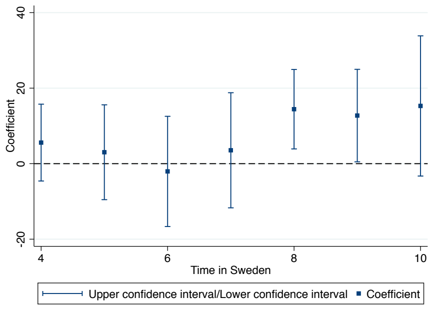



(a) $k = 100$ (b) $k = 250$

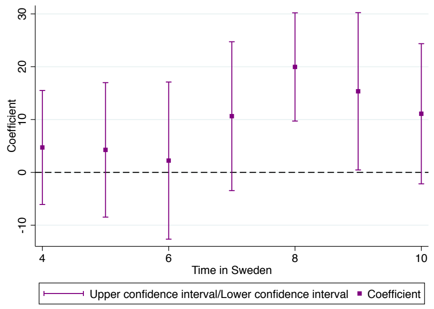
(c) $k = 500$ (d) $k = 1000$

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.2) for movers and their corresponding 95% confidence intervals. The number of observations is 9222 in t+4 and 9690 in t+10.
*Source*: Own calculations on data from the GeoSweden database.

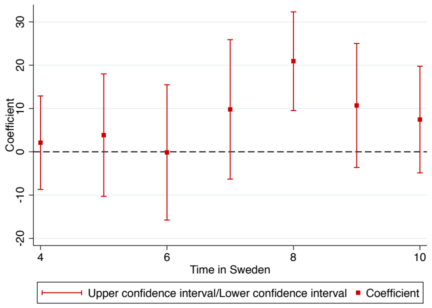
Figure B.8: Movers: effect of initial employed co-ethnic share on earnings, 2SLS regressions: $k = 250, k = 500, k = 1000$




(a) $k = 250$




(b) $k = 500$

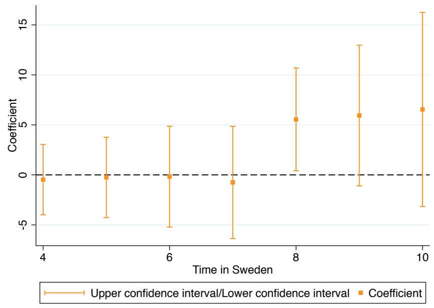



(c) $k = 1000$

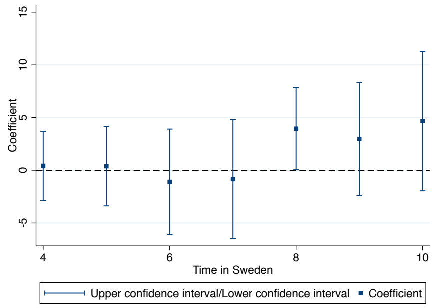
*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.2) for movers and their corresponding 95% confidence intervals. The number of observations is 9222 in t+4 and 9690 in t+10. *Source*: Own calculations on data from the GeoSweden database.

## Distance Restriction to 500 meters

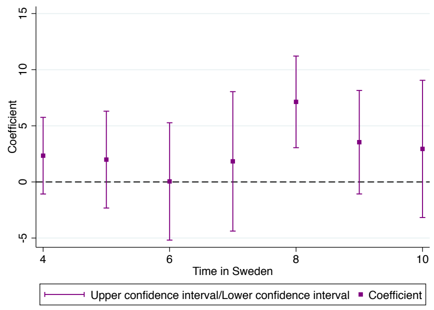
Figure B.9: Distance: effect of initial co-ethnic share on earnings, 2SLS regressions: $k = 100, k = 250, k = 500, k = 1000$



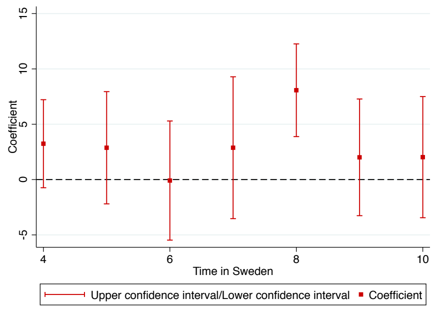

(a) $k = 100$




(b) $k = 250$



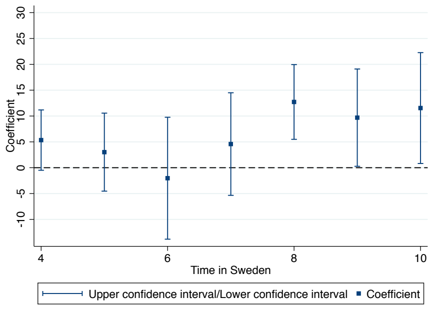

(c) $k = 500$



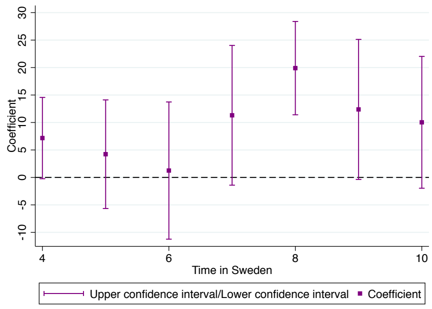

(d) $k = 1000$

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.2) for distance less than 500 meters and their corresponding 95% confidence intervals. The number of observations is 9136 in t+4 and 8805 in t+10 for $k = 250$. For $k = 500$, the number of observations is 7721 in t+4 and 7446 in t+10. The number of observations varies from 5308 in t+4 to 5118 in t+10 for $k = 1000$.
*Source*: Own calculations on data from the GeoSweden database.

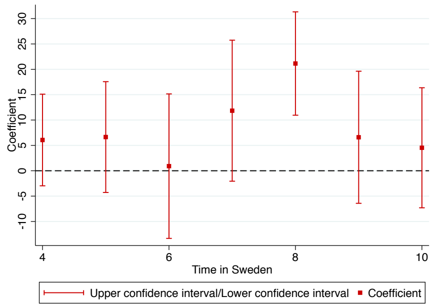
Figure B.10: Distance: effect of initial employed co-ethnic share on earnings, 2SLS regressions: $k = 250, k = 500, k = 1000$




(a) $k = 250$




(b) $k = 500$

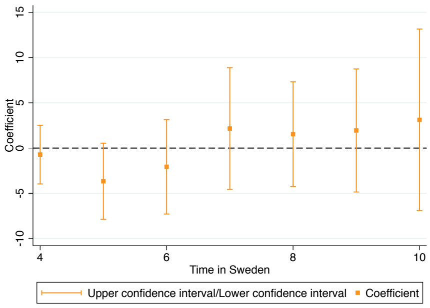



(c) $k = 1000$

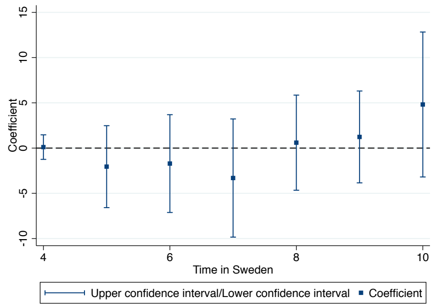
*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.2) for distance less than 500 meters and their corresponding 95% confidence intervals. The number of observations is 9136 in t+4 and 8805 in t+10 for $k = 250$. For $k = 500$, the number of observations is 7721 in t+4 and 7446 in t+10. The number of observations varies from 5308 in t+4 to 5118 in t+10 for $k = 1000$.
*Source*: Own calculations on data from the GeoSweden database.

## Parents

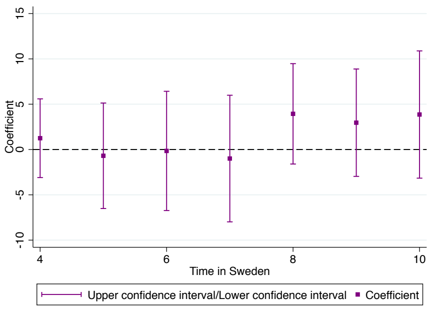
Figure B.11: Parents: effect of initial co-ethnic share on earnings, 2SLS regressions: $k = 100, k = 250, k = 500, k = 1000$



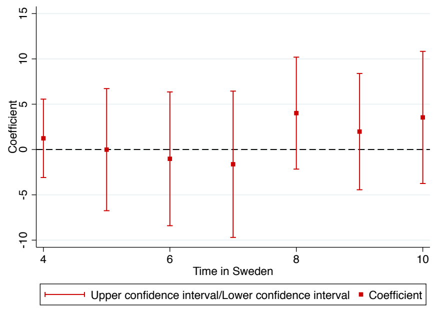

(a) $k = 100$




(b) $k = 250$



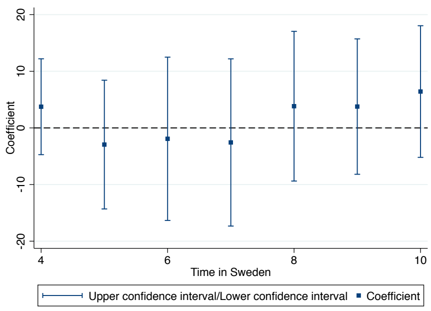

(c) $k = 500$



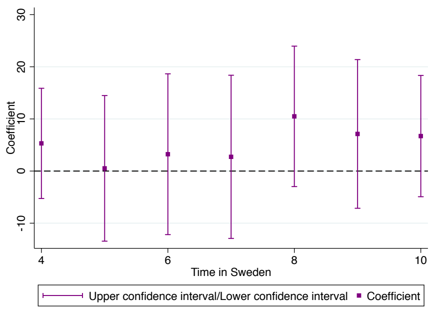

(d) $k = 1000$

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.2) for parents and their corresponding 95% confidence intervals. The number of observations is 5501 in t+4 and 5367 in t+10.
*Source*: Own calculations on data from the GeoSweden database.

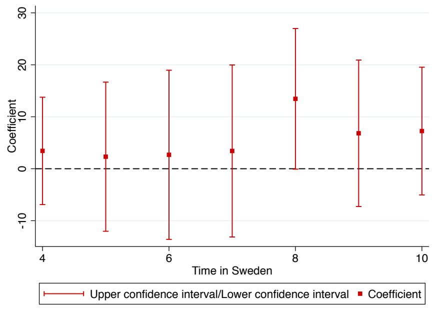
Figure B.12: Parents: effect of initial employed co-ethnic share on earnings, 2SLS regressions: $k = 250, k = 500, k = 1000$




(a) $k = 250$




(b) $k = 500$




(c) $k = 1000$

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.2) for parents and their corresponding 95% confidence intervals. The number of observations is 5501 in t+4 and 5367 in t+10.

*Source*: Own calculations on data from the GeoSweden database.

## B.7 Full Results for Employment

Table B.23: Effect of co-ethnic share on employment: $k = 100$

| VARIABLES | (1) Employment t+4 | (2) t+5 | (3) t+6 | (4) t+7 | (5) t+8 | (6) t+9 | (7) t+10 |
|---|---|---|---|---|---|---|---|
| Ethnic | -0.0330 | 0.0264 | 0.232 | -0.00828 | 1.450*** | 1.111* | 1.287 |
| | (0.352) | (0.383) | (0.512) | (0.503) | (0.542) | (0.664) | (0.982) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| Observations | 10,543 | 10,549 | 10,314 | 10,200 | 10,345 | 10,245 | 10,166 |
| $R^2$ | 0.105 | 0.107 | 0.098 | 0.110 | 0.066 | 0.103 | 0.099 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

*Notes:* The table presents the 2SLS estimates of ethnic share on employment for $k = 100$ nearest neighbor levels. All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses.Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

Table B.24: Effect of co-ethnic share on employment: $k = 250$

| VARIABLES | (1) Employment t+4 | (2) t+5 | (3) t+6 | (4) t+7 | (5) t+8 | (6) t+9 | (7) t+10 |
|---|---|---|---|---|---|---|---|
| Ethnic | 0.199 | 0.255 | 0.0663 | -0.147 | 1.060*** | 0.645 | 0.826 |
| | (0.326) | (0.365) | (0.457) | (0.524) | (0.342) | (0.519) | (0.651) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| Observations | 10,543 | 10,549 | 10,314 | 10,200 | 10,345 | 10,245 | 10,166 |
| $R^2$ | 0.103 | 0.105 | 0.099 | 0.110 | 0.094 | 0.126 | 0.129 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

*Notes:* The table presents the 2SLS estimates of ethnic share on employment for $k = 250$ nearest neighbor levels. All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses.Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

Table B.25: Effect of co-ethnic share on employment: $k = 500$

| VARIABLES | (1) Employment t+4 | (2) t+5 | (3) t+6 | (4) t+7 | (5) t+8 | (6) t+9 | (7) t+10 |
|---|---|---|---|---|---|---|---|
| Ethnic | 0.572 | 0.566 | 0.184 | 0.304 | 1.557*** | 0.877 | 0.450 |
| | (0.397) | (0.376) | (0.527) | (0.593) | (0.354) | (0.541) | (0.539) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| Observations | 10,543 | 10,549 | 10,314 | 10,200 | 10,345 | 10,245 | 10,166 |
| $R^2$ | 0.098 | 0.101 | 0.098 | 0.106 | 0.079 | 0.120 | 0.140 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

*Notes:* The table presents the 2SLS estimates of ethnic share on employment for $k = 500$ nearest neighbor levels. All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses.Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

Table B.26: Effect of co-ethnic share on employment: $k = 1000$

| VARIABLES | (1) Employment t+4 | (2) t+5 | (3) t+6 | (4) t+7 | (5) t+8 | (6) t+9 | (7) t+10 |
|---|---|---|---|---|---|---|---|
| Ethnic | 0.486 | 0.388 | 0.00620 | 0.274 | 1.604*** | 0.674 | 0.486 |
| | (0.400) | (0.436) | (0.547) | (0.640) | (0.355) | (0.582) | (0.487) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| Observations | 10,543 | 10,549 | 10,314 | 10,200 | 10,345 | 10,245 | 10,166 |
| $R^2$ | 0.100 | 0.103 | 0.100 | 0.107 | 0.086 | 0.127 | 0.140 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

*Notes:* The table presents the 2SLS estimates of ethnic share on employment for $k = 1000$ nearest neighbor levels. All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses.Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.
*Source:* Own calculations based on data from the GeoSweden database.

### Employed Co-ethnics

Table B.27: Effect of employed co-ethnic share on employment: $k = 250$

| VARIABLES | (1) Employment t+4 | (2) t+5 | (3) t+6 | (4) t+7 | (5) t+8 | (6) t+9 | (7) t+10 |
|---|---|---|---|---|---|---|---|
| Employed Ethnic | 1.445** | 1.112 | 0.166 | 0.487 | 2.555*** | 1.988** | 1.840** |
| | (0.697) | (0.802) | (1.143) | (1.024) | (0.693) | (0.870) | (0.897) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| Observations | 10,543 | 10,549 | 10,314 | 10,200 | 10,345 | 10,245 | 10,166 |
| $R^2$ | 0.111 | 0.113 | 0.101 | 0.114 | 0.116 | 0.136 | 0.142 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

*Notes:* The table presents the 2SLS estimates of employed ethnic share on employment for $k = 250$ nearest neighbor levels. All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses.Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.

*Source:* Own calculations based on data from the GeoSweden database.

Table B.28: Effect of employed co-ethnic share on employment: $k = 500$

| VARIABLES | (1) Employment t+4 | (2) t+5 | (3) t+6 | (4) t+7 | (5) t+8 | (6) t+9 | (7) t+10 |
|---|---|---|---|---|---|---|---|
| Employed Ethnic | 1.602** | 1.213 | 0.600 | 1.462 | 3.391*** | 2.383** | 1.357 |
| | (0.800) | (0.871) | (1.210) | (1.203) | (0.730) | (1.157) | (0.830) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| Observations | 10,543 | 10,549 | 10,314 | 10,200 | 10,345 | 10,245 | 10,166 |
| $R^2$ | 0.108 | 0.111 | 0.102 | 0.113 | 0.106 | 0.131 | 0.146 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

*Notes:* The table presents the 2SLS estimates of employed ethnic share on employment for $k$500 nearest neighbor levels. All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses.Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.

*Source:* Own calculations based on data from the GeoSweden database.

Table B.29: Effect of employed co-ethnic share on employment: $k = 1000$

| VARIABLES | (1) Employment t+4 | (2) t+5 | (3) t+6 | (4) t+7 | (5) t+8 | (6) t+9 | (7) t+10 |
|---|---|---|---|---|---|---|---|
| Employed Ethnic | 1.169 | 0.986 | 0.472 | 1.350 | 3.760*** | 1.753 | 0.951 |
| | (0.823) | (0.956) | (1.306) | (1.390) | (0.850) | (1.157) | (0.800) |
| Age | yes | yes | yes | yes | yes | yes | yes |
| $Age^2$ | yes | yes | yes | yes | yes | yes | yes |
| Female | yes | yes | yes | yes | yes | yes | yes |
| Married | yes | yes | yes | yes | yes | yes | yes |
| Female*Married | yes | yes | yes | yes | yes | yes | yes |
| Children | yes | yes | yes | yes | yes | yes | yes |
| Female*Children | yes | yes | yes | yes | yes | yes | yes |
| Education dummy | yes | yes | yes | yes | yes | yes | yes |
| Immigration year FE | yes | yes | yes | yes | yes | yes | yes |
| Country of origin FE | yes | yes | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | yes | yes | yes | yes |
| Observations | 10,543 | 10,549 | 10,314 | 10,200 | 10,345 | 10,245 | 10,166 |
| $R^2$ | 0.105 | 0.108 | 0.101 | 0.111 | 0.108 | 0.136 | 0.147 |
| No. Municipalities | 268 | 265 | 255 | 253 | 247 | 247 | 250 |

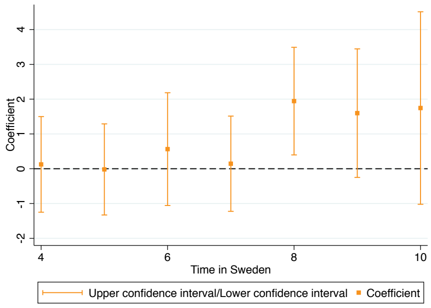
*Notes:* The table presents the 2SLS estimates of employed ethnic share on employment for $k = 1000$ nearest neighbor levels. All specifications are estimated with immigration year, country of origin and municipality fixed effects. Standard errors are in parentheses.Coefficients with * are statistically significant at the 10 percent level, those with ** at the 5 percent level, those with *** at the 1 percent level.

*Source:* Own calculations based on data from the GeoSweden database.

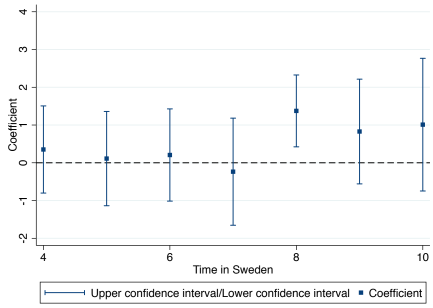
## B.8 Robustness Checks for Employment

### Movers

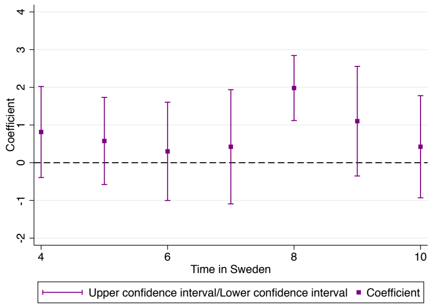
Figure B.13: Movers: effect of initial co-ethnic share on employment, 2SLS regressions: $k = 100, k = 250, k = 500, k = 1000$



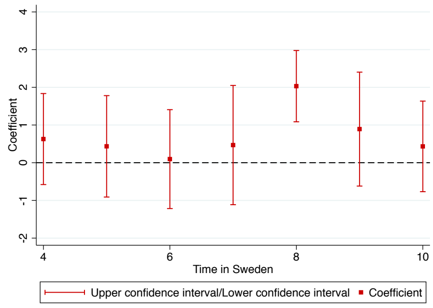

(a) $k = 100$




(b) $k = 250$



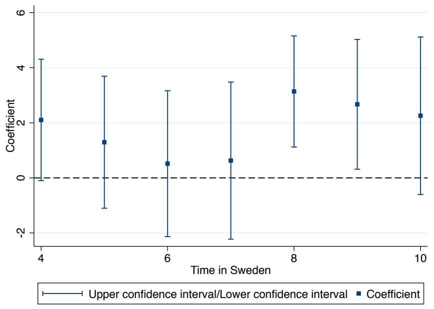

(c) $k = 500$



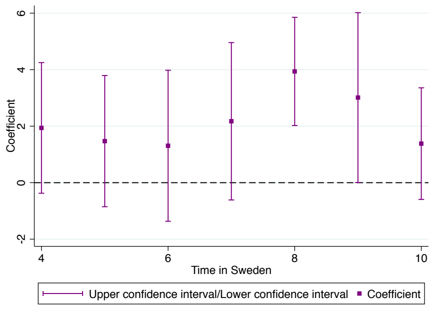

(d) $k = 1000$

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.2) for movers and their corresponding 95% confidence intervals. The number of observations is 9222 in t+4 and 9690 in t+10.
*Source*: Own calculations on data from the GeoSweden database.

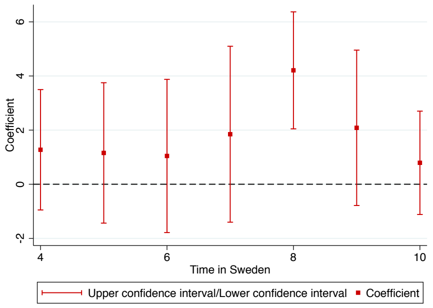
Figure B.14: Movers: effect of initial employed co-ethnic share on employment, 2SLS regressions: $k = 250, k = 500, k = 1000$




(a) $k = 250$




(b) $k = 500$

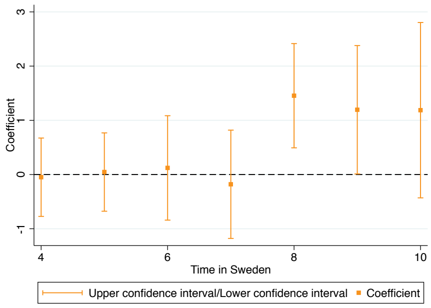



(c) $k = 1000$

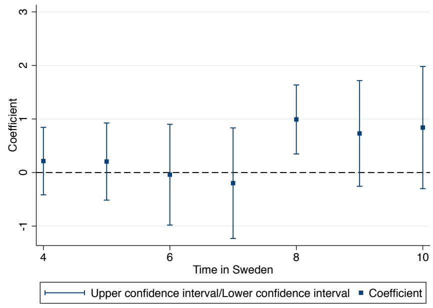
*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.2) for movers and their corresponding 95% confidence intervals. The number of observations is 9222 in t+4 and 9690 in t+10. *Source*: Own calculations on data from the GeoSweden database.

### Distance Restriction to 500 meters

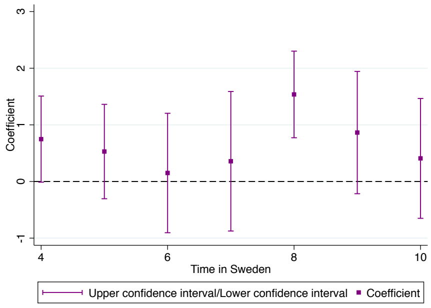
Figure B.15: Distance: effect of initial co-ethnic share on employment, 2SLS regressions: $k = 100, k = 250, k = 500, k = 1000$



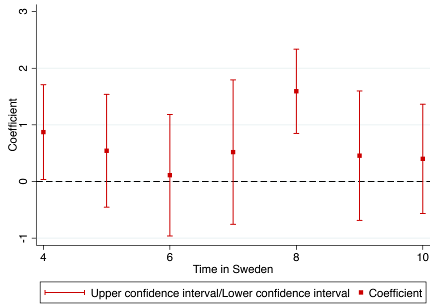

(a) $k = 100$




(b) $k = 250$



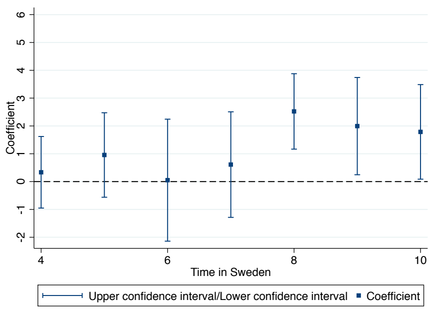

(c) $k = 500$



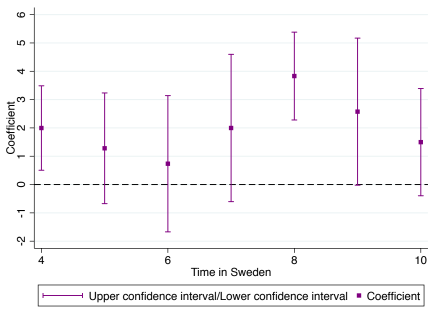

(d) $k = 1000$

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.2) for distance less than 500 meters and their corresponding 95% confidence intervals. The number of observations is 9136 in t+4 and 8805 in t+10 for $k = 250$. For $k = 500$, the number of observations is 7721 in t+4 and 7446 in t+10. The number of observations varies from 5308 in t+4 to 5118 in t+10 for $k = 1000$.
*Source*: Own calculations on data from the GeoSweden database.

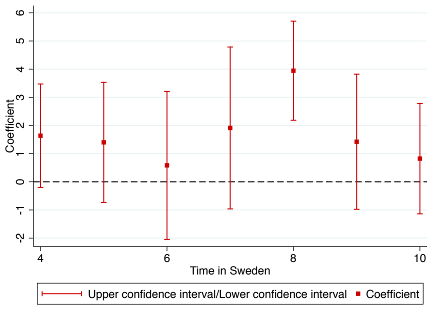
Figure B.16: Distance: effect of initial employed co-ethnic share on employment, 2SLS regressions: $k = 250, k = 500, k = 1000$




(a) $k = 250$




(b) $k = 500$

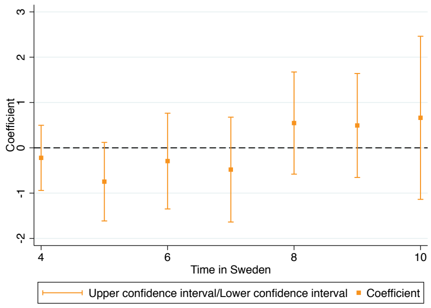



(c) $k = 1000$

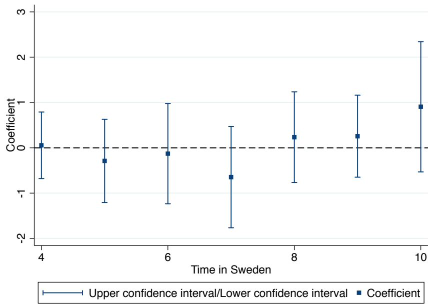
*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.2) for distance less than 500 meters and their corresponding 95% confidence intervals. The number of observations is 9136 in t+4 and 8805 in t+10 for $k = 250$. For $k = 500$, the number of observations is 7721 in t+4 and 7446 in t+10. The number of observations varies from 5308 in t+4 to 5118 in t+10 for $k = 1000$.
*Source*: Own calculations on data from the GeoSweden database.

## Parents

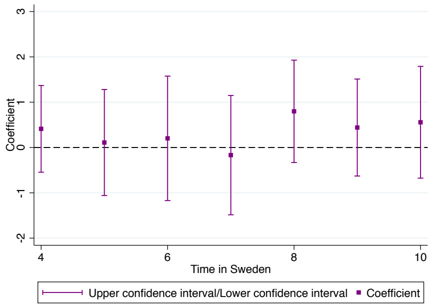
Figure B.17: Parents: effect of initial co-ethnic share on employment, 2SLS regressions: $k = 100, k = 250, k = 500, k = 1000$



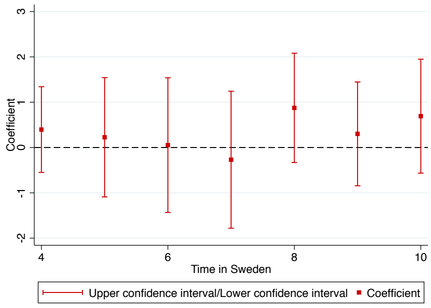

(a) $k = 100$




(b) $k = 250$



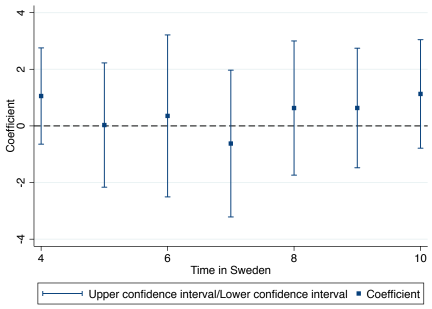

(c) $k = 500$



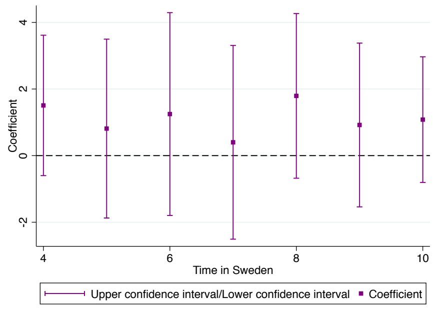

(d) $k = 1000$

*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.2) for parents and their corresponding 95% confidence intervals. The number of observations is 5501 in t+4 and 5367 in t+10.
*Source*: Own calculations on data from the GeoSweden database.

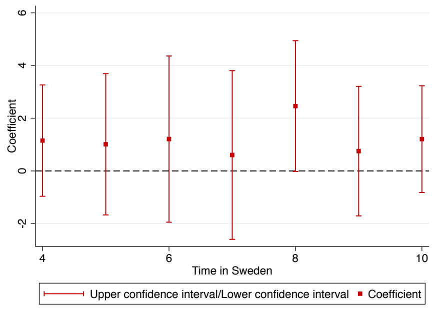
Figure B.18: Parents: effect of initial employed co-ethnic share on employment, 2SLS regressions: $k = 250, k = 500, k = 1000$




(a) $k = 250$




(b) $k = 500$




(c) $k = 1000$

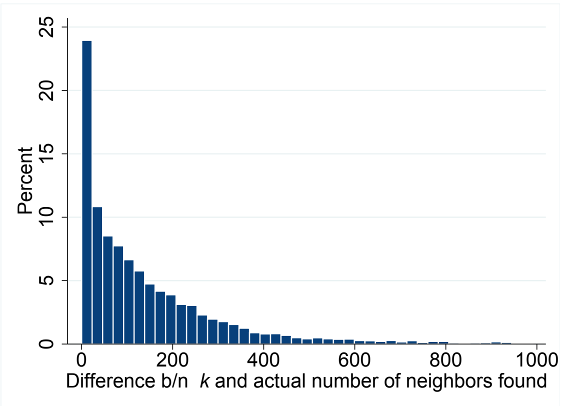
*Notes:* The figure shows the $\beta_1$ coefficients obtained when estimating equation (2.2) for parents and their corresponding 95% confidence intervals. The number of observations is 5501 in t+4 and 5367 in t+10.
*Source*: Own calculations on data from the GeoSweden database.

# Appendix C

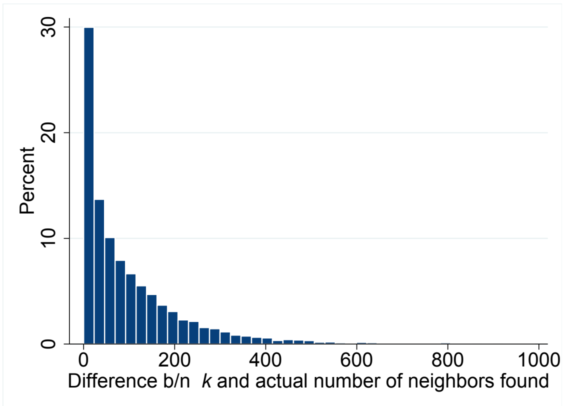
# Appendix to Chapter 3

## C.1 Checks for $k$-nearest Algorithm

Figure C.1: Difference between $k$ and observed $k$



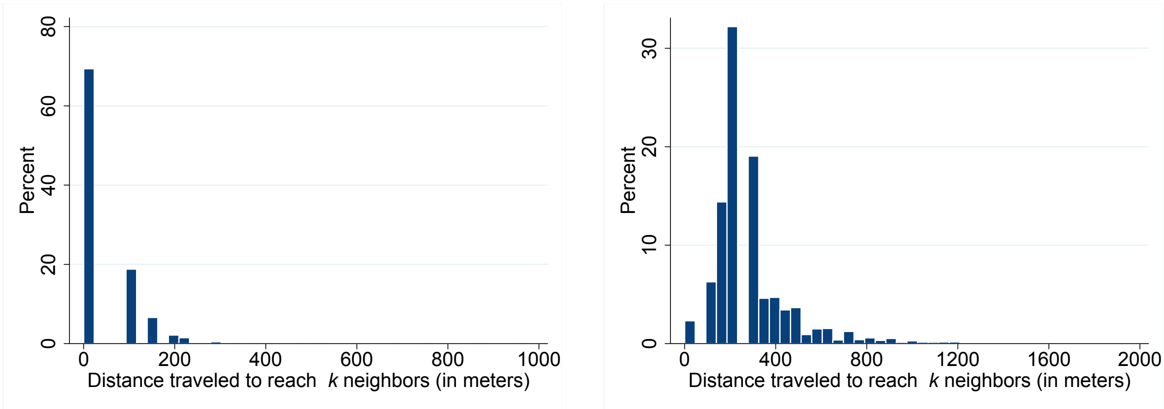

(a) $k = 100$




(b) $k = 1000$

*Notes:* The x-axis is cropped in order to make the graphs easier to read.

Figure C.2: Distance traveled by EquiPop software to find $k$ neighbors

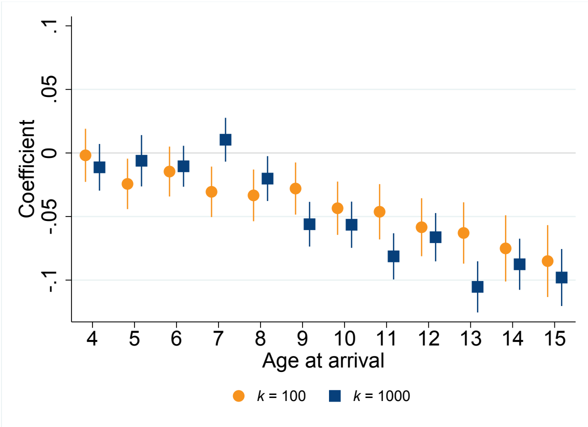



(a) $k = 100$ (b) $k = 1000$

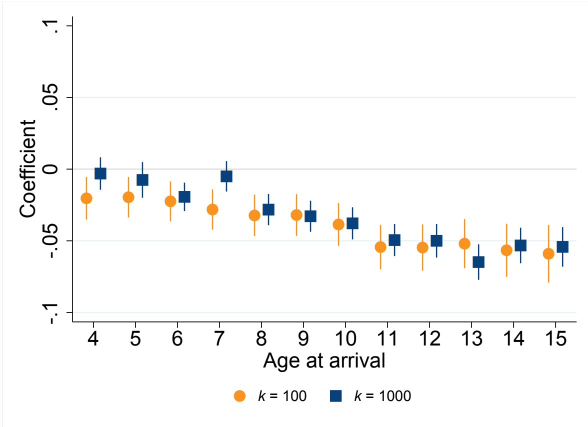
*Notes:* The x-axis is cropped in order to make the graphs easier to read.

## C.2 Results from Weighted Regressions

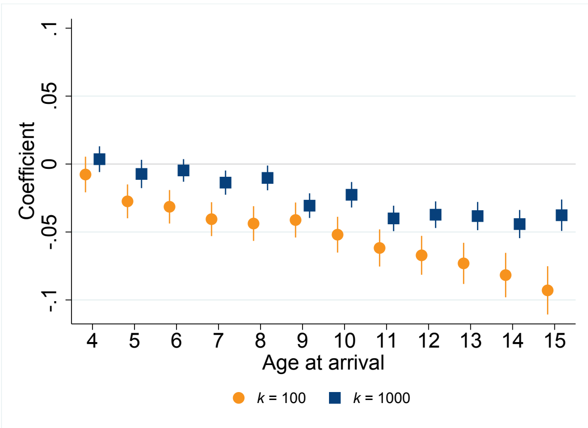
Figure C.3: Effect of age at arrival on residential integration outcomes



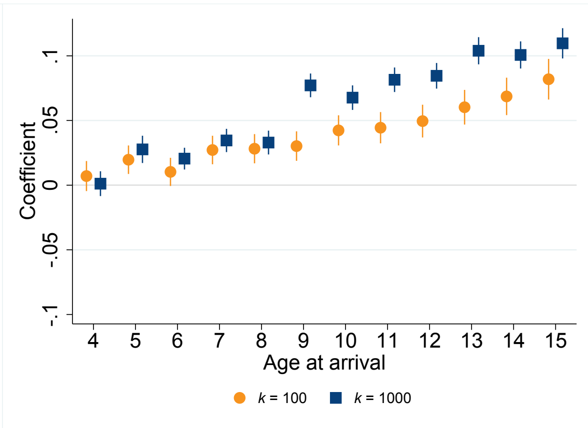

(a) Share of natives




(b) Share of high-educated




(c) Share of high-earners




(d) Share on welfare

*Notes:* The figure shows the coefficients from the weighted regressions.
*Source*: Own calculations on data from the GeoSweden database.

## C.3 Decomposition

### C.3.1 Married Sample

Table C.1: Decomposition: steps to obtain shares; married sample

| | Coefficients from augmented eq. (1) | | | | Effect of age on channels | | | Contributions | | | | Shares | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) |
| | Age | I | ED | IM | I | ED | IM | I | ED | IM | T | I | ED | IM | R |
| | | | | | | | | (2) × (5) | (3) × (6) | (4) × (7) | (1) + (8) +(9) + (10) | (8)/(11) | (9)/(11) | (10)/(11) | (1)/(11) |
| *Panel A: k = 100* | | | | | | | | | | | | | | | |
| Share natives | -0.0051 | 0.0007 | 0.0049 | 0.0945 | -0.9235 | -0.0401 | -0.0254 | -0.0007 | -0.0002 | -0.0024 | -0.0084 | 0.0822 | 0.0234 | 0.2872 | 0.6073 |
| Share high-educated | -0.0021 | 0.0003 | 0.0082 | 0.0378 | -0.9235 | -0.0401 | -0.0254 | -0.0003 | -0.0003 | -0.0010 | -0.0037 | 0.0695 | 0.0890 | 0.2606 | 0.5809 |
| Share high-earners | -0.0032 | 0.0005 | 0.0034 | 0.0453 | -0.9235 | -0.0401 | -0.0254 | -0.0005 | -0.0001 | -0.0012 | -0.0050 | 0.0987 | 0.0276 | 0.2307 | 0.6431 |
| Share on welfare | 0.0053 | -0.0005 | -0.0037 | -0.0335 | -0.9235 | -0.0401 | -0.0254 | 0.0005 | 0.0001 | 0.0009 | 0.0067 | 0.0673 | 0.0222 | 0.1267 | 0.7838 |
| *Panel A: k = 1000* | | | | | | | | | | | | | | | |
| Share natives | -0.0023 | 0.0006 | 0.0037 | 0.0799 | -0.9235 | -0.0401 | -0.0254 | -0.0005 | -0.0001 | -0.0020 | -0.0051 | 0.1062 | 0.0291 | 0.4015 | 0.4632 |
| Share high-educated | -0.0015 | 0.0002 | 0.0065 | 0.0332 | -0.9235 | -0.0401 | -0.0254 | -0.0002 | -0.0003 | -0.0008 | -0.0028 | 0.0642 | 0.0950 | 0.3065 | 0.5342 |
| Share high-earners | -0.0028 | 0.0003 | 0.0033 | 0.0389 | -0.9235 | -0.0401 | -0.0254 | -0.0003 | -0.0001 | -0.0010 | -0.0043 | 0.0750 | 0.0312 | 0.2305 | 0.6633 |
| Share on welfare | 0.0037 | -0.0004 | -0.0027 | -0.0308 | -0.9235 | -0.0401 | -0.0254 | 0.0003 | 0.0001 | 0.0008 | 0.0049 | 0.0657 | 0.0220 | 0.1588 | 0.7535 |

*Notes:* The table shows the decomposition analysis for the married sample. The variable I denotes the income rank. ED stands for education level. IM represents intermarriage. Columns (1-3) show the regressions using equation (1). Columns (4-5) shows the estimated effect of age on income rank and years of education. T denotes the total effect and R represents the residual.

### C.3.2 Full Sample

Table C.2: Decomposition

| | Income rank | Years of education | Residual |
|---|---|---|---|
| *Panel A: k = 100* | | | |
| Share natives | 0.1003 | 0.0350 | 0.8647 |
| Share high-educated | 0.0947 | 0.1077 | 0.7977 |
| Share high-earners | 0.1128 | 0.0307 | 0.8565 |
| Share on welfare | 0.0790 | 0.0295 | 0.8915 |
| *Panel B: k = 1000* | | | |
| Share natives | 0.1165 | 0.0439 | 0.8396 |
| Share high-educated | 0.0887 | 0.1080 | 0.8033 |
| Share high-earners | 0.1035 | 0.0326 | 0.8640 |
| Share on welfare | 0.0684 | 0.0274 | 0.9042 |

*Notes:* The table presents the decomposition analysis for the full sample. The estimates to construct this table can be seen in Table C.3.

Table C.3: Decomposition: steps to obtain shares; full sample

| | Coefficients from augmented eq. (1) | | | Effect of age on channels | | Contributions | | | Shares | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | Age | I | ED | I | ED | I | ED | T | I | ED | R |
| | | | | | | (2) × (4) | (3) × (5) | (1) + (6) + (7) | (6)/(8) | (7)/(8) | (1)/(8) |
| *Panel A: $k$ = 100* | | | | | | | | | | | |
| Share natives | -0.0055 | 0.0006 | 0.0048 | -1.0858 | -0.0461 | -0.0006 | -0.0002 | -0.0063 | 0.1003 | 0.0350 | 0.8647 |
| Share high-educated | -0.0034 | 0.0004 | 0.0100 | -1.0858 | -0.0461 | -0.0004 | -0.0005 | -0.0043 | 0.0947 | 0.1077 | 0.7977 |
| Share high-earners | -0.0048 | 0.0006 | 0.0038 | -1.0858 | -0.0461 | -0.0006 | -0.0002 | -0.0057 | 0.1128 | 0.0307 | 0.8565 |
| Share on welfare | 0.0050 | -0.0004 | -0.0036 | -1.0858 | -0.0461 | 0.0004 | 0.0002 | 0.0056 | 0.0790 | 0.0295 | 0.0295 |
| *Panel A: $k$ = 1000* | | | | | | | | | | | |
| Share natives | -0.0035 | 0.0004 | 0.0039 | -1.0858 | -0.0461 | -0.0005 | -0.0002 | -0.0041 | 0.1165 | 0.0439 | 0.8396 |
| Share high-educated | -0.0027 | 0.0003 | 0.0080 | -1.0858 | -0.0461 | -0.0003 | -0.0004 | -0.0034 | 0.0887 | 0.1080 | 0.8033 |
| Share high-earners | -0.0034 | 0.0004 | 0.0027 | -1.0858 | -0.0461 | -0.0004 | -0.0001 | -0.0039 | 0.1035 | 0.0326 | 0.8640 |
| Share on welfare | 0.0041 | -0.0003 | -0.0027 | -1.0858 | -0.0461 | 0.0003 | 0.0001 | 0.0046 | 0.0684 | 0.0274 | 0.9042 |

*Notes:* The table shows the decomposition analysis for the full sample. The variable I denotes the income rank. ED stands for years of education. Columns (1-3) show the regressions using equation (1). Columns (4-5) shows the estimated effect of age on income rank and years of education. T denotes the total effect and R represents the residual.

# Bibliography


Ager, Alastair and Alison Strang (2008). “Understanding Integration: A Conceptual Framework.” *Journal of Refugee Studies* 21 (2), pp. 166–191.

Aksoy, Cevat Giray, Panu Poutvaara, and Felicitas Schikora (2020). “First Time Around: Local Conditions and Multi-dimensional Integration of Refugees”. *IZA Discussion Paper Series* No. 13914.

Alexander, Rohan and Zachary Ward (2018). “Age at Arrival and Assimilation during the Age of Mass Migration”. *The Journal of Economic History* 78 (3), pp. 904–937.

Allport, Gordon Willard, Kenneth Clark, and Thomas Pettigrew (1954). “The Nature of Prejudice”.

Andersson, Henrik (2020). “Ethnic Enclaves, Self-employment, and the Economic Performance of Refugees: Evidence from a Swedish Dispersal Policy”. *International Migration Review*, p. 0197918320912195.

Andersson, Roger (2003). *Settlement Dispersal of Immigrants and Refugees in Europe: Policy and Outcomes*. Vancouver Centre of Excellence Vancouver.

Andersson, Roger, Sako Musterd, and George Galster (2019). “Port-of-Entry Neighborhood and Its Effects on the Economic Success of Refugees in Sweden”. *International Migration Review* 53 (3), pp. 671–705.

Ansala, Laura, Ulla Hämäläinen, and Matti Sarvimäki (2019). “Age at Arrival, Parents and Neighborhoods: Understanding the Educational Attainment of Immigrants’ Children”. *Journal of Economic Geography*.

Ashraf, Nava et al. (2015). “Savings in Transnational Households: A Field Experiment among Migrants from El Salvador”. *Review of Economics and Statistics* 97 (2), pp. 332–351.

Åslund, Olof, Anders Böhlmark, and Oskar Nordström Skans (2015). "Childhood and Family Experiences and the Social Integration of Young Migrants". *Labour Economics* 35, pp. 135–144.

Åslund, Olof and Dan-Olof Rooth (2007). "Do When and Where Matter? Initial Labour Market Conditions and Immigrant Earnings." *The Economic Journal* 117 (518), pp. 422–448.

Åslund, Olof et al. (2011). "Peers, Neighborhoods, and Immigrant Student Achievement: Evidence from a Placement Policy." *American Economic Journal: Applied Economics* 3 (2), pp. 67–95.

Audretsch, David B and Maryann P Feldman (1996). "R&D Spillovers and the Geography of Innovation and Production". *American Economic Review* 86 (3), pp. 630–640.

Battisti, Michele, Yvonne Giesing, and Nadzeya Laurentsyeva (2019). "Can Job Search Assistance Improve the Labour Market Integration of Refugees? Evidence from a Field Experiment". *Labour Economics* 61, p. 101745.

Battu, Harminder, Paul Seaman, and Yves Zenou (2011). "Job Contact Networks and the Ethnic Minorities". *Labour Economics* 18 (1), pp. 48–56.

Bayer, Patrick, Stephen L Ross, and Giorgio Topa (2008). "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes". *Journal of Political Economy* 116 (6), pp. 1150–1196.

Beaman, Lori A (2011). "Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the US." *The Review of Economic Studies* 79 (1), pp. 128–161.

Bell, Brian, Francesco Fasani, and Stephen Machin (2013). "Crime and Immigration: Evidence from Large Immigrant Waves." *Review of Economics and Statistics* 21 (3), pp. 1278–1290.

Berg, Gerard J Van den et al. (2014). "Critical Periods during Childhood and Adolescence". *Journal of the European Economic Association* 12 (6), pp. 1521–1557.

Bertrand, Marianne, Erzo FP Luttmer, and Sendhil Mullainathan (2000). "Network Effects and Welfare Cultures". *The Quarterly Journal of Economics* 115 (3), pp. 1019–1055.

Bevelander, Pieter (2020). "Integrating Refugees into Labor Markets". *Iza World of Labor.*

Boeri, Tito et al. (2015). "Immigration, Housing Discrimination and Employment". *The Economic Journal* 125 (586), F82–F114.

Böhlmark, Anders (2008). "Age at Immigration and School Performance: A Siblings Analysis using Swedish Register Data". *Labour Economics* 15 (6), pp. 1366–1387.

Borjas, George (1995). "Ethnicity, Neighborhoods, and Human-Capital Externalities". *American Economic Review* 85 (3), pp. 365–90.

Borjas, George J (2000). "Ethnic Enclaves and Assimilation". *Swedish Economic Policy Review* 7 (2), pp. 89–122.

Boswell, Christina (2003). "Burden-sharing in the European Union: Lessons from the German and UK Experience." *Journal of Refugee Studies* 16 (3), pp. 316–335.

Bratu, Cristina, Matz Dahlberg, and Madhinee Valeyatheepillay (2021). "Age at Arrival and Residential Integration".

Brell, Courtney, Christian Dustmann, and Ian Preston (2020). "The Labor Market Integration of Refugee Migrants in High-Income Countries". *Journal of Economic Perspectives* 34 (1), pp. 94–121. DOI: 10.1257/jep.34.1.94.

Brown, Ryan and Andrea Velásquez (2017). "The Effect of Violent Crime on the Human Capital Accumulation of Young Adults". *Journal of Development Economics* 127, pp. 1–12.

Brücker, Herbert, Yuliya Kosyakova, and Eric Schuß (2020). *Fünf Jahre seit der Fluchtmigration 2015: Integration in Arbeitsmarkt und Bildungssystem macht weitere Fortschritte.* Tech. rep. IAB-Kurzbericht.

Calvo-Armengol, Antoni and Matthew O Jackson (2004). "The Effects of Social Networks on Employment and Inequality". *American Economic Review* 94 (3), pp. 426–454.

Chetty, Raj, Nathaniel Hendren, and Lawrence F Katz (2016a). "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment". *American Economic Review* 106 (4), pp. 855–902.

Chetty, Raj et al. (2016b). "Childhood Environment and Gender Gaps in Adulthood". *American Economic Review* 106 (5), pp. 282–88.

Chetty, Raj et al. (2020). "Race and Economic Opportunity in the United States: An Intergenerational Perspective". *The Quarterly Journal of Economics* 135 (2), pp. 711–783.

Conley, Timothy G and Giorgio Topa (2002). “Socio-economic Distance and Spatial Patterns in Unemployment”. *Journal of Applied Econometrics* 17 (4), pp. 303–327.

Cutler, David M and Edward L Glaeser (1997). “Are Ghettos Good or Bad?” *The Quarterly Journal of Economics* 112 (3), pp. 827–872.

Cutler, David M, Edward L Glaeser, and Jacob L Vigdor (2008). “When are Ghettos Bad? Lessons from Immigrant Segregation in the United States.” *Journal of Urban Economics* 63 (3), pp. 759–774.

Dahlberg, Matz, Karin Edmark, and Heléne Lundqvist (2012). “Ethnic Diversity and Preferences for Redistribution.” *Journal of Political Economy* 120 (1), pp. 41–76.

Dahlberg, Matz and Madhinee Valeyatheepillay (2021). “The Importance of Initial Neighborhood Characteristics for Future Residential Integration and Labor Market Outcomes of Refugees”.

Dahlberg, Matz et al. (2020). “Labor Market Integration of Low-Educated Refugees: RCT Evidence from an Ambitious Integration Program in Sweden”. *IFAU WP* 21.

Damm, Anna Piil (2005). “The Danish Dispersal Policy on Refugee Immigrants 1986-1998: A Natural Experiment?” *Department of Economics, Aarhus School of Business Aarhus*.

– (2009). “Ethnic Enclaves and Immigrant Labor Market Outcomes: Quasi-experimental Evidence.” *Journal of Labor Economics* 27 (2), pp. 281–314.

– (2014). “Neighborhood Quality and Labor Market Outcomes: Evidence from Quasi-random Neighborhood Assignment of Immigrants.” *Journal of Urban Economics* 79, pp. 139–166.

Damm, Anna Piil and Kristine Vasiljeva (2016). “The Danish Experience with Refugee Influx in the 1980s and 1990s.” *Ieb Report*. Institut d’Economia de Barcelona, pp. 7–9.

Danzer, Alexander M and Firat Yaman (2013). “Do Ethnic Enclaves Impede Immigrants’ Integration? Evidence from a Quasi-experimental Social-interaction Approach.” *Review of International Economics* 21 (2), pp. 311–325.

*Diarieförda handlingar, centrala staden* (1998). Socialförvaltningen, Uppsala, Sweden, Dnr 1988:98.

Dinesen, Peter and Kim Sønderskov (Jan. 2018). "Ethnic Diversity and Social Trust. A Critical Review of the Literature and Suggestions for a Research Agenda". ISBN: 9780190274801. DOI: 10.1093/oxfordhb/9780190274801.013.13.

Drever, Anita I and Onno Hoffmeister (2008). "Immigrants and Social Networks in a Job-Scarce Environment: The Case of Germany". *International Migration Review* 42 (2), pp. 425–448.

Dustmann, Christian, Kristine Vasiljeva, and Anna Piil Damm (2016). "Refugee Migration and Electoral Outcomes." *CReAM Discussion Paper* 19, p. 16.

Edin, Per-Anders, Peter Fredriksson, and Olof Åslund (2003). "Ethnic Enclaves and the Economic Success of Immigrants: Evidence from a Natural Experiment." *The Quarterly Journal of Economics* 118 (1), pp. 329–357.

Ellen, Ingrid Gould and Margery Austin Turner (1997). "Does Neighborhood Matter? Assessing Recent Evidence." *Housing Policy Debate* 8 (4), pp. 833–866.

Eurostat (2016). "Record Number of Over 1.2million First Time Asylum Seekers Registered in 2015". *Eurostat News Release* 4.

Fasani, Francesco, Tommaso Frattini, and Luigi Minale (2020). "(The Struggle for) Refugee Integration into the Labour Market: Evidence from Europe". *CEPR Discussion Paper* 12718.

Galster, George C (2008). "Quantifying the Effect of Neighbourhood on Individuals: Challenges, Alternative Approaches, and Promising Directions." *Schmollers jahrbuch* 128 (1), pp. 7–48.

Grönqvist, Hans, Per Johansson, and Susan Niknami (2012). "Income Inequality and Health: Lessons from a Refugee Residential Assignment Program". *Journal of Health Economics* 31 (4), pp. 617–629.

Grönqvist, Hans, Susan Niknami, and Per Olof Robling (2015). "Childhood Exposure to Segregation and Long-Run Criminal Involvement: Evidence from the "Whole of Sweden" Strategy". *The Swedish Institute for Social Research (SOFI), Stockholm University.*

Harder, Niklas et al. (2018). "Multidimensional Measure of Immigrant Integration". *Proceedings of the National Academy of Sciences* 115 (45), pp. 11483–11488.

Heckman, James, Rodrigo Pinto, and Peter Savelyev (2013). "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes". *American Economic Review* 103 (6), pp. 2052–86.

Hermansen, Are Skeie (2017). "Age at Arrival and Life Chances among Childhood Immigrants". *Demography* 54 (1), pp. 201–229.

Imbens, Guido W (2015). "Matching Methods in Practice: Three Examples". *Journal of Human Resources* 50 (2), pp. 373–419.

Johnston, Ron and Charles Pattie (2011). "Social networks, Geography and Neighbourhood Effects". *The SAGE Handbook of Social Network Analysis. SAGE Publications Ltd, London*, pp. 301–311.

Lemmermann, Dominique and Regina T Riphahn (2018). "The Causal Effect of Age at Migration on Youth Educational Attainment". *Economics of Education Review* 63, pp. 78–99.

Logan, John R, Richard D Alba, and Brian J Stults (2003). "Enclaves and Entrepreneurs: Assessing the Payoff for Immigrants and Minorities". *International Migration Review* 37 (2), pp. 344–388.

Malmberg, Bo et al. (2018). "Residential Segregation of European and Non-European Migrants in Sweden: 1990–2012". *European Journal of Population* 34 (2), pp. 169–193.

Martén, Linna, Jens Hainmueller, and Dominik Hangartner (2019). "Ethnic Networks can Foster the Economic Integration of Refugees". *Proceedings of the National Academy of Sciences* 116 (33), pp. 16280–16285.

Massey, Douglas S and Nancy A Denton (1985). "Spatial Assimilation as a Socioeconomic Outcome." *American Sociological Review*, pp. 94–106.

– (1988). "The Dimensions of Residential Segregation". *Social Forces* 67 (2), pp. 281–315.

Mayda, Anna Maria et al. (2017). "The Labor Market Impact of Refugees: Evidence from the US Resettlement Program." *US Department of State Office of the Chief Economist Working Paper* (2017-04).

Munshi, Kaivan (2003). "Networks in the Modern Economy: Mexican Migrants in the US Labor Market". *The Quarterly Journal of Economics* 118 (2), pp. 549–599.

Nannestad, Peter (2004). "A Game Real Actors Won't Play? Integration of Ethnic Minorities in Denmark as a Collective Action Dilemma". *The International Migration Review* 38 (1), pp. 287–308. ISSN: 01979183, 17477379.

– (2009). "Making Integration Work". *Handbook of Social Capital: The Troika of Sociology, Political Science and Economics*. Cheltenham, UK: Edward Elgar Publishing. ISBN: 9781845423230.

OECD (2016). *Making Integration Work: Refugees and Others in Need of Protection.* OECD Publishing.

Östh, John (2014). *Introducing the EquiPop Software: An Application for the Calculation of k-nearest Neighbour Contexts/Neighbourhoods.*

Patacchini, Eleonora and Yves Zenou (2012). "Ethnic Networks and Employment Outcomes". *Regional Science and Urban Economics* 42 (6), pp. 938–949.

Robinson, Vaughan, Roger Andersson, and Sako Musterd (2003). *Spreading the 'Burden'?: A Review of Policies to Disperse Asylum Seekers and Refugees.* Policy Press.

Rummel, Rudolph J (1967). "Understanding Factor Analysis". *Journal of Conflict Resolution* 11 (4), pp. 444–480.

Sampson, Robert J, Jeffrey D Morenoff, and Thomas Gannon-Rowley (2002). "Assessing "Neighborhood Effects": Social Processes and New Directions in Research". *Annual Review of Sociology* 28 (1), pp. 443–478.

Sarvimäki, Matti (2017). "Labor Market Integration of Refugees in Finland". *VATT Research Reports* 185.

Schmutte, Ian M (2015). "Job Referral Networks and the Determination of Earnings in Local Labor Markets". *Journal of Labor Economics* 33 (1), pp. 1–32.

Schönwälder, Karen (2007). *Residential segregation and the integration of immigrants: Britain, the Netherlands and Sweden.* Tech. rep. WZB Discussion Paper.

Sharkey, Patrick and Jacob W Faber (2014). "Where, When, Why, and for Whom Do Residential Contexts Matter? Moving Away from the Dichotomous Understanding of Neighborhood Effects". *Annual Review of Sociology* 40, pp. 559–579.

Solon, Gary, Marianne E Page, and Greg J Duncan (2000). "Correlations between Neighboring Children in their Subsequent Educational Attainment". *Review of Economics and Statistics* 82 (3), pp. 383–392.

Statistics Sweden (2020). *Summary of Population Statistics 1960 – 2019*. available at: `https://www.scb.se/en/finding-statistics/statistics-by-subject-area/population/population-composition/population-statistics/pong/tables-and-graphs/yearly-statistics--the-whole-country/summary-of-population-statistics/`. accessed 9 February 2021.

Stock, James H, Motohiro Yogo, et al. (2005). "Testing for weak instruments in linear IV regression". *Identification and Inference for Econometric Models: Essays in honor of Thomas Rothenberg* 80 (4.2), p. 1.

Swedish Migration Agency (2020). *History*. available at: `https://www.migrationsverket.se/English/About-the-Migration-Agency/Migration-to-Sweden/History.html/`. accessed 9 February 2021.

Tarozzi, Alessandro et al. (2014). "Micro-loans, Insecticide-treated Bednets, and Malaria: Evidence from a Randomized Controlled Trial in Orissa, India". *American Economic Review* 104 (7), pp. 1909–41.

Tukey, John W (1957). "On the comparative anatomy of transformations". *The Annals of Mathematical Statistics*, pp. 602–632.

Turk, Umut, John Östh, et al. (2017). "Inequality of Opportunity in Sweden: A Spatial Perspective". *University Of Verona Department of Economics Working Paper* 9.

United Nations High Commissioner for Refugees (2019). *Global Trends Forced Displacement in 2019*. available at: https://www.unhcr.org/globaltrends2019/.

Van Ham, Maarten and Tiit Tammaru (2016). "New Perspectives on Ethnic Segregation over Time and Space. A Domains Approach". *Urban Geography* 37 (7), pp. 953–962.

Wahba, Jackline and Yves Zenou (2005). "Density, Social Networks and Job Search Methods: Theory and Application to Egypt". *Journal of Development Economics* 78 (2), pp. 443–473.

Wilkes, Rima and John Iceland (2004). "Hypersegregation in the Twenty-first Century". *Demography* 41 (1), pp. 23–36.

Xie, Yu and Margaret Gough (2011). "Ethnic Enclaves and the Earnings of Immigrants". *Demography* 48 (4), pp. 1293–1315.

Yong, An Gie, Sean Pearce, et al. (2013). “A beginner’s guide to factor analysis: Focusing on exploratory factor analysis”. *Tutorials in Quantitative Methods for Psychology* 9 (2), pp. 79–94.