Henrik Andersson

Immigration and the Neighborhood
ECONOMICS AT UPPSALA UNIVERSITY

The Department of Economics at Uppsala University has a long history. The first chair in Economics in the Nordic countries was instituted at Uppsala University in 1741.

The main focus of research at the department has varied over the years but has typically been oriented towards policy-relevant applied economics, including both theoretical and empirical studies. The currently most active areas of research can be grouped into six categories:

* Labour economics
* Public economics
* Macroeconomics
* Microeconometrics
* Environmental economics
* Housing and urban economics

Additional information about research in progress and published reports is given in our project catalogue. The catalogue can be ordered directly from the Department of Economics.
Essay 1 (with Kristoffer Jutvik): This paper uses quasi-experimental evidence to understand how changes in migration policy affect the number of asylum seekers. We look specifically at a sudden, regulatory change in the Swedish reception of Syrian asylum seekers. The change took place in September 2013, and implied that all Syrian asylum seekers would be granted permanent, instead of temporary residence permits. Using high frequency data and an interrupted time series set-up, we study the extent to which this change caused more Syrian citizens to apply for asylum in Sweden, and how the change affected the distribution of asylum seekers in Europe. Results show that the change in policy almost doubled the number of asylum seekers from Syria within 2013, with a significant jump in numbers already within the first week after the implementation of the policy. While this also decreased the share of asylum seekers to other large recipient countries (Germany), the effects were highly temporary.

Essay 2: In this paper I estimate the causal effect of ethnic enclaves on the probability of self-employment. To account for neighborhood selection I make use of a refugee dispersal program. Results indicate that larger ethnic enclaves, measured as the share of self-employed coethnics in the municipality immigrants first arrive into, affect the probability of self-employment positively, while the share of all other coethnics has a negative effect. Results however also indicate that there is a long term economic penalty to being placed with a larger share of self-employed coethnics, an effect which is partly mediated through the choice of self-employment.

Essay 3 (with Heléne Berg and Matz Dahlberg): In this paper we investigate the migration behavior of the native population following foreign (refugee) immigration, with a particular focus on examining whether there is any support for an ethnically based migration response. If ethnicity is the mechanism driving the change in natives' migration behavior, our maintained hypothesis is that native-born individuals who are more ethnically similar to arriving refugees should not change their migration behavior to the same extent as native-born individuals with native-born parents (who are ethnically quite different from refugees). Using rich geo-coded register data from Sweden, spanning over 20 consecutive years, we account for possible endogeneity problems with an improved so-called "shift-share" instrumental variable approach; in particular, our strategy combines policy-induced initial immigrant settlements with exogenous contemporaneous immigration as captured by refugee shocks. We find no evidence of neither native flight nor native avoidance when studying the full population. We do, however, find native flight among individuals who are expected to be more mobile, and within this group, we find that all natives, irrespective of their parents' foreign background, react similarly to increased immigration. Our results therefore indicate that preferences for ethnically homogeneous neighborhoods may not be the dominant channel inducing flight. Instead our estimates indicate that immigration leads to more socio-economically segregated neighborhoods. This conclusion can have important implications for the ethnically based tipping point literature.

Essay 4 (with Matz Dahlberg): In this paper we examine the short-run housing market effects of refugee immigration to Sweden. Given that Sweden is a major refugee receiving country, it constitutes an interesting and important case to study. To deal with the endogeneity resulting from the refugees' location choices, we use an econometric specification that includes neighborhood fixed effects and an instrumental variable that is based on a historical settlement pattern mainly determined by a refugee placement policy. We find that refugee immigration to small neighborhoods has no average effect on changes in housing prices in that neighborhood. We find a positive effect on increased housing supply, measured as the number of objects on sale. The zero effect of immigration on housing prices stands in contrast to the negative results found in earlier studies. We hypothesize that the reason is due to different preferences for homogeneity in Sweden, and/or to institutional features in the Swedish rental sector.
Acknowledgements

The making of this thesis has been a collaborative effort. I have benefited from plenty of help from creative and intelligent people, and given the sheer number of individuals who, knowingly or unknowingly, influenced the content, I won’t be able to mention all. Here is, however, my attempt to thank some of the most important contributors.

First off, I thank my supervisors, Matz Dahlberg and Per Engström. Matz, your influence cannot be overestimated. It was your suggestion to study white flight that led me into the route of migration, which is now the theme of this thesis. You’re encouragement, drive and genuine interest in the topic as well as your PhD students, make you a great mentor. It is fair to say that my thesis would have been both of much lower quality, as well as about something completely different, had I ended up with a different supervisor. Per, thank you for helping me think about theory, mechanisms and what my results really mean; in short, thank you for making me think about economics. I sincerely hope to be able to continue working with both of you in the future.

Beyond the help of my supervisors, Matti Sarvimäki and Karin Edmark provided insightful and truly helpful comments at my licentiate and final seminar. In both of these occasions, I am also very thankful to comments from P-A Edin. Further important was the cooperation with my co-author Hélène Berg, who in many ways functioned as an extra supervisor during the formative second year as a PhD student. In addition, a long list of people, including Susanne Urban, Cristina Bratu, Tim Blackwell, Peter Fredriksson, Che-Yuan Liang, Mattias Engdahl, Jon Fiva, Alicia Adsera, Gideon Goerdt, Florian Morath and Albert Saiz, have at different stages and in different forms given valuable comments on the texts that make up this thesis. To this list of helping hands, I must also add a countless number of participants at seminars and workshops.

The majority of my time as a PhD-student was spent at the Institute for Housing and Urban Research (IBF). IBF is a great interdisciplinary environment, and I’m thankful for perspectives from sociology, geography and political science. I am especially grateful to Kristoffer Jutvik, for long morning discussions, a sharp political eye, and for your positive take on life. Special thanks also to Göran Rydén and Kerstin Larsson, for
teaching me elegant words as well as always being good coffee-company, to Ann-Sofie Wigg Bodin for all the help with job-market references, and of course to all my fellow PhD-students, without whom everyday work would be considerably bleaker. IBF is a unique workplace, and it will be hard to replace.

The PhD program also included a number of (for me) fairly technical courses at the Department of Economics. Admittedly, this wasn’t always a fun and inspiring experience, but it was made a whole lot better given the company of my fantastic cohort: Maria, Olle, Paula, Dagmar, Lucas, Aino-Maija and Franklin. It goes without saying that you are all brilliant academics, and I have benefited greatly from our conversations, our studying and your general advice. Most importantly however, thank you for being able to talk about other things than economics, for seeing the world outside of academia, and of course, for all the pancakes.

At the Department of Economics, I also owe lots of gratitude to the administrative staff, with a special thanks to Katarina Grönvall. Oskar Nordström Skans and Luca Repetto provided excellent advice through the job market process, during which I also enjoyed the company of Sebastian Escobar.

In addition to the time in Uppsala, I also spent six months as a visitor at UCLA. This would not have been possible without Leah Boustan, who was kind enough to invite me as a guest researcher. For making my time in California to an experience of a lifetime, I am also grateful to Marva and Cody Shearer, the best hosts anyone could ask for.

Outside academia, I have been fortunate to have friends who helped me sustain my social life. Thanks to Stefan, Dick, Per, Gustav, Fredrik and Robert, for being great travel companions, and (at least seemingly) actually interested in what I do. I am also thankful to Caroline, Johan, Jenny, Eric, Adam, Sandra, Henrik and Fanny, for all the laughs and conversations.

My final thanks goes to my family: My brother in law Christian for help with everything from guitar to occasional accommodation, my niece Ellie and nephew Oscar, for giving me a new perspective on life, my siblings, Linus and Louise, for being my single greatest source of inspiration, and to my parents, Kristina and Ante, for programming me with the self-confidence, which made me believe I could do this in the first place. Last, to Emelie, for being my outlier. For being colorful, unique, and so wonderfully weird.

Svartbäcken, Uppsala, April 2018
Henrik Andersson
Introduction

The topic of international migration has grown in importance, taking a more prominent role in political discussions as well as in academic debates over the past decades. This holds true globally, with an overall increase in the number of migrants worldwide\(^1\), as well as for Sweden, the case in study within this thesis.

Sweden has seen a particularly interesting development with regards to international migration over the past decades. Figure 1 shows the development of foreign born individuals, with a registered address in Sweden from 1950-2017. As can be seen, especially since the late 1980’s, the number has increased greatly. Given low or sometimes zero population increase in natives, the increase in absolute numbers also represent a parallel increase in the share of foreign born. In 1950, less than 3 percent of the population were born in another country, which can be compared to more than 18 percent today.

\(^1\)According to the UN International Migration Report (2017), the estimated number of global migrants grew from 173 million in 2000, to 220 million in 2010, to 258 million in 2017.
Figure 1. Number and composition of foreign born in Sweden 1950-2014.

Notes: Y-axis in units of thousands.
Source: Statistics Sweden.

Figure 1 also reflects a change in the composition of the foreign born population. During the 50’s, immigration to Sweden mainly came from neighboring countries. Labor market immigration from especially Yugoslavia, but also from other Eastern European countries, Greece and Italy increased the share of the non-Nordic European population during the 60’s and 70’s, only to be replaced largely by refugee and tied family migration in the 80’s and onward. Today a clear majority of the foreign born population was born in countries outside of Europe.²

These changes, in numbers and composition of the population, could hold important implications socio-economically, geographically and politically. The change in the population also provides the main motivation for the focus on immigration to Sweden in this thesis: Something profound has happened, and it should be studied.

²Swedish population statistics, including changes over time in numbers and composition, is found at https://www.scb.se/en/finding-statistics/statistics-by-subject-area/population/population-composition/population-statistics/.
Besides the obvious relevance of this study for our understanding of matters in Sweden, there are a number of reasons to believe that a focus on the Swedish case is beneficial from a larger perspective.

First, the increase in migrants and the foreign born population has over a fairly short historical time period made Sweden into a high immigration country, especially as compared to other Western nations.\textsuperscript{3} Research focusing on causes and consequences of international migration naturally benefits from studying a case with a substantial amount of migrants.

Second, over the last thirty years, Sweden has experimented with different reception, integration and immigration policies. These are interesting study objects in themselves, but can also provide grounds for helpful variation to be exploited by econometricians, seeking to establish causal relations. Two of these policies, the placement strategy of 1985-1994, and the change in assessments of Syrian asylum seekers in 2013, are used in this thesis.

Third, immigration and the policies mentioned could not have been studied, had it not been for the data, which is unique and of very high quality. All four papers in one way or another use features from GeoSweden, an individual full population database, spanning from 1990 to 2014.\textsuperscript{4} The data has, among other things, information from income, education, property and migration registries, providing longitudinal knowledge over many important dimensions, such as country of birth, time of immigration, place of residence and yearly socio-economic information. As I will discuss in the papers, this sometimes allows me to study mechanisms and stories that other papers may not have been able to do.

With this background in mind, I now move on to the content of the thesis.

The migration decision: Asylum seekers and the choice of destination

Of the research questions related to migration, the essential one is of course the question of migration itself, including the reason for migrating as well as the choice of destination. Already in 1885, the geographer E.G Ravenstein used aggregate population statistics for Ireland, Scotland and England, to lay forward his “laws of migration”, which at the time

\textsuperscript{3}As an example, within Europe, only Switzerland has a larger foreign born population. See \url{www.pewglobal.org/interactives/migration-tables/} for a global comparison.

\textsuperscript{4}GeoSweden is administered by the Institute for Housing and Urban Research in Uppsala, and is made anonymous by Statistics Sweden.
included for example the importance of distance (Ravenstein noted that most migrants travel only short distances) and the gravitating force of industrial centers in attracting labor market migrants (Ravenstein, 1885).

While Ravenstein never used the term himself, his work is sometimes categorized as a basis for the theoretical research labeled “push and pull” theories.⁵ Common in these models is a functional approach, where migration is explained as a result of weighting costs of migrating against the benefits at the destination. An early, highly influential paper is Sjaastad (1962), who characterized all migration as a function of the private returns to migrating.⁶ Sjaastad described the process as one in which the prospective migrant compares the private, economic returns to real income, with the financial and psychological costs of leaving ones home. Sjaastad’s characterization of migration remains heavily influential, and much modern work on migration by economists rely on or explicitly study private returns to migration.⁷

Arguably some migratory behavior, for example refugee migration, happen under sudden and highly threatening situations, and cannot as easily be modeled as a cost-benefit decision. An alternative, modern example is the model by Moore and Shellman (2007), who theorize that refugee migration is a two stage process. In the first stage, individuals are assumed to decide whether or not they wish to leave their home. In the second stage, given such a decision has been made, a destination is chosen. Such a hypothesis arguably better captures some of the decisions taken under situations where the prime target is to run from something, rather than to something else.

Other researchers have criticized the push and pull theories for putting too much emphasis on the individual. Some of these theories instead highlight migration as a household decision, where families jointly decide on sending certain members of the family for migration. This potentially provides the family with both better income as well as diversifies the household risks (Stark and Bloom, 1985). Another group of researchers highlight the importance of networks. After the first migrants leave for a certain destination, the costs of continued migration along the same path drop. Networks, or “chains”, have for example been used to explain the strong increase in migration from Mexico to the United States

---

⁵Several of the theories under this umbrella are often also labeled as neo-classical (Castles et al., 2014).
⁶Contemporary to Sjaastad was Lee (1966), who separated the conditions influencing the choice of migrating into four factors, including characteristics at origin and destination, intervening obstacles and personal matters.
⁷See for example (Borjas, 1999; Boustan, 2017).
Massey and Garcia, 1987). Last, a number of institutional theories have downplayed the choice of the individual all together. A prominent example is Piore (1979), who instead sees international migration as a part of the global system, where structural, permanent demand of labor in developed nations constantly leads to inflow of cheap labor.

As noted by Massey et al. (1993), there is no reason to believe that any of the above theories would work exclusively in relation to one another. Networks might be highly important in certain cases, while institutional and socio-economic contexts can put restrictions on both individual decisions based on private returns as well as household decisions with an explicit risk dimension. The importance of different factors should instead be concluded from empirical investigations.

The first chapter of the thesis, “Do asylum seekers respond to policy changes? Evidence from a Swedish-Syrian case” (co-written with Kristoffer Jutvik), is a quantitative case study looking at the effect of pull factors on where asylum seekers go. More specifically, we use a sudden change in Swedish migration policy, which made residence permits for Syrian asylum seekers permanent rather than temporary, and study the effect on the number of asylum seekers from Syria to Sweden. To further shed light on the story we look into the distribution of asylum seekers in Europe, and the characteristics of those coming.

A now fairly long empirical discussion within political science, sociology and economics has taken an interest in the destination choice of refugees. While it seems fairly uncontroversial that push factors in the origin country is the crucial key to understanding why the number of asylum seekers vary over time, the reason why some countries receive much larger shares than others are not always as easily explained. Some factors are fixed, and outside the scope of any policy-maker, such as a common language, colonial ties or short geographical distance between the source and destination. Asylum policies are, however, clearly within the scope of national decision making, and the efficiency of such policies are still under heavy debate. Whether or not policies have any effect, is related to the theoretical discussion outlined above, that is, essentially, what enters the decision margin of migrants in general and asylum seekers in particular? The question is also related to a debate within political science, where researchers disagree on the reach and de facto power of national states in the current, increasingly international era (Sassen, 1996; Freeman, 1998).

Most of the quantitative studies in the field are based on yearly panel data, where changes in policies in year $t$ are correlated with changes in asylum flows the same or following years. While such analyses provide
great insight, there is a lack of papers which look at high frequency data, with changes in months or weeks, and which can estimate causal effects. Our approach bridges this gap, by estimating the weekly change in asylum seekers as a response to a sudden policy change in residence permits for Syrians. The sudden change combined with the research design allows us to consider the causal effect.

The study provides several insights. First, the policy change did effect the number of asylum seekers, and fairly fast. The first two weeks following the policy change saw a doubling of the weekly number of asylum seekers. Also, while the effect was substantial, it also seems to have been rather temporary. Already within a year, the great share of asylum seekers from Syria again started opting for Germany. Last, studying the characteristics of those arriving, a clear conclusion is that the share of individuals arriving alone, without another family member, decrease. This supports the idea of migration being an important household, rather than an individual, decision.

Settlement: Ethnic enclaves and economic activity

Having studied one dimension of the destination choices of migrants, I move on to the settlement. How settlement takes place and develops is naturally, just as the migration decision itself, tightly connected to the policies of the receiving state. Castles et al. (2014) write:

At one extreme, openness to settlement, granting of citizenship and gradual acceptance of cultural diversity may allow the formation of ethnic communities, which are seen as part of a multicultural society. At the other extreme, denial of the reality of settlement refusal of citizenship and rights to settlers, and rejection of cultural diversity may lead to formation of ethnic minorities, whose presence is widely regarded as undesirable.

Noteworthy is that “ethnicity” is a debated concept in itself. A simple, broad way of looking at it is a sense of group belonging, based on common markers such as values, culture, language and origin. In this thesis I am restricted to the Swedish register data, which provides many benefits overall, but in terms of ethnicity is limited. Since I know the origin, proxyd either by birth or emigration country, this will function as the defining feature of ethnicity.

Regardless if neighborhoods develop along the logic of ethnic minorities or communities (or anything in-between), it is an empirically documented observation that immigrants tend to spatially cluster along
Loosely defined, geographical spaces or neighborhoods with a high concentration of coethnics are sometimes labeled as “ethnic enclaves”.

How does living close to coethnics, or in geographical spaces with a disproportional presence of coethnics, affect important economic and social indicators? On the one hand, there are reasons to believe negative aspects might come from such segregation, stemming from distance to natives, who can provide essential skills in host-country human capital, including language, and to some extent social codes. On the other hand, the networks running through ethnic enclaves might be important for finding jobs through informal channels.

An empirical literature, primarily in economics, deals with causal effects of living in an ethnic enclave. Prior papers have used different research designs to find the effect of ethnic enclaves on labor market outcomes (Munshi, 2003; Edin et al., 2003; Bayer et al., 2008; Damm, 2009; Beaman, 2012), welfare uptake (Bertrand et al., 2000; Åslund and Fredriksson, 2009) and industry specialization (Kerr and Mandorff, 2016).

In chapter II, titled “Ethnic enclaves, self-employment and the economic performance of refugees”, I add to the literature on causal effects of ethnic enclaves on economic outcomes, by asking if living close to individuals born in the same country increases the probability that an individual will become self-employed.

The paper uses the refugee dispersal policy in place in Sweden between 1985-1994, which meant that refugees were not allowed themselves to decide where to live in the country, but were placed by the immigration board in contracted municipalities. This research design has been used before, and is a well-established method to construct exogenous variation in the geographical sorting of immigrants. I look at the effect of number of coethnics, or skilled coethnics, in the municipality of arrival, on the probability of self-employment several years after arrival.

The results support an increased probability of self-employment, primarily stemming from being placed with coethnics who themselves run a business. A larger concentration of coethnics, measured as the pure number of coethnics, coethnics with capital or higher income do not give any increase in the probability of self-employment. The results therefore support a mechanism, where quality information or knowledge is

---

8See for example the first stage of the regressions developed in Chapter III and IV in this thesis. In these I show that we can predict the neighborhoods new refugees locate in based on the locational pattern of earlier waves of immigrants from the same origin countries.

9See for example Edin et al. (2003); Åslund and Fredriksson (2009).
spread within the enclaves. Individuals who have themselves run a business, have know-how on the process of businesses, which in turn can be transferred to new immigrants who can become self-employed. This mechanism stand in contrast to the sometimes suggested niche market, where individuals become self-employed to sell to a niche, ethnic market. Last, while the probability for self-employment increases, this does not imply a one-dimensional success story. Many self-employed show poor economic outcomes, with income trajectories over the following twenty years being on average lower than those who became employed on the formal market.

Effects on the receiving economy: Native flight and housing prices

The last two chapters of the thesis deal with the related issues of the effect of migration on the destination country. Such effects, if any, will of course be dependent on a number of features, including the time frame, institutional and policy context, and perhaps most importantly the composition and characteristics of those arriving.

The effects also depend on the geographical scale chosen. One branch of the literature have for example taken an interest in the nation-wide effect on fiscal outcomes.\textsuperscript{10} In my case, I focus on neighborhood effects, since it will be on this geographical level that day to day encounters take place, making it an interesting arena for economic studies. A battery of empirical papers have studied the effect of immigration on a number of host country features, for example wages (Dustmann et al., 2016, 2017), preferences for redistribution (Luttmer, 2001; Dahlberg et al., 2012), trust (Alesina and La Ferrara, 2002), participation in social activities (Alesina and La Ferrara, 2000), charitable giving (Andreoni et al., 2016), collective action (Vigdor, 2004) and the size and mix of publicly provided goods and services (Alesina et al., 1999, 2000).

I study two outcomes on neighborhood level: House prices and domestic migration by natives. Chapter III, “Migrating natives and foreign immigration: Is there a preference for ethnic residential homogeneity?” , which is co-authored with Matz Dahlberg and Heléne Berg, studies the effect of immigration on native domestic migration behavior. Assuming natives (defined as being born in Sweden) have some preference for living with other natives, increased immigration to a neighborhood can lead to natives moving out of the area (native flight) or natives avoiding

\textsuperscript{10}For a Swedish example see Ruist (2015).
that very neighborhood, when they move (native avoidance). Similar mechanisms have been suggested in several countries, including the UK and the US.\footnote{See for example Saiz and Wachter (2011) and the Sá (2014).}

It is noteworthy that most previous research treat immigration as one concept, estimating effects regardless of reason for migration or origin country. In this paper, we focus on the specific group of refugees, which is a highly debated, and in numbers substantial group to focus on.

Our estimates for the full population in Sweden suggest neither native flight nor avoidance. However, when looking specifically at those owning a home, we detect some native flight. Even more importantly, this flight seems to be happening for different groups of natives, stemming from different parental backgrounds. Both natives with parents born in Sweden and parents born outside of OECD, react to increased immigration by moving out.

There are a couple of essential take-aways from this: First, native flight need not be a reaction along ethnic lines. If we assume that natives with parents born in the same countries as contemporary arriving refugees, are ethnically closer to the same refugees, compared to natives with both parents in Sweden, flight responses due to ethnically motivated preferences would differ for these native groups. As mentioned, our results show that both of these native groups practice flight to the same extent. Our results are hence indications of a socioeconomic story, where individuals may instead be reacting to other traits of immigrants, such as a lower socioeconomic status. Second, the fact that home owners but not renters react, suggests it is important to take mobility constraints into account. Immigrants, and refugees in particular, largely live in rental apartments’ their initial years in Sweden. Since the rental stock is inelastic, a year with unusually large immigration therefore also means that unusually few rental apartments are available. Many natives who live in rental apartments may not have the economic muscles, and therefore not the option, to buy, and so, assuming most natives are not prepared to move a long distance, unusually few natives move from one rental apartment to another, the year following large scale in-migration. This constraint does not apply to those who own their apartment or house, since they can sell, and move to another place.

The last chapter of the thesis, called “Refugee immigration and the housing market”, is written together with Matz Dahlberg, and studies the effect of refugee immigration into a neighborhood on housing prices. The outcome is a natural follow-up from the previous study on native
flight and avoidance, since population dynamics are tightly linked to the formation of housing prices.

Given native flight and/or avoidance, we would theoretically expect negative price developments from increased immigration. If people move out, and especially if socio-economically stronger households move out, as a response to immigration, we would expect both fewer bidders, as well as more bidders to have relatively less capital.

There are, however, counter-acting mechanisms. This can happen in the form of demand from immigrants themselves: more immigrants equals more people, which implies more bidders and thereby higher prices. Increased demand could also come from natives, if natives substitute rentals for owning a home, as a response to immigration.

Given theoretical mechanisms predicting both negative as well as positive effects, it becomes an empirical question to evaluate the overall impact. Previous studies from Italy (Accetturo et al., 2014), the US (Saiz and Wachter, 2011) and the UK (Sá, 2014) have found that the negative side caused by native flight and avoidance (or a reaction to a perceived drop in local amenities) tends to outweigh the positive demand effects, causing prices to drop.

Our results are insignificant and close to zero, indicating that, in Sweden, there is no average effect from increased immigration on housing prices. The results are robust to a number of sensitivity checks and alternative specifications. Our results hence stand in contrast to the earlier findings, where prices have been seen to drop due to increased immigration. We see two possible reasons for this. First, as described in chapter III, in Sweden, refugees largely occupy rental apartments, meaning that there are unusually few rental apartments available on the market, the year after the immigration flow. Assuming natives still wish to live in the same neighborhoods that immigrants move into, they will be forced to buy, rather than rent. More individuals opting for buying rather than renting will indirectly raise prices. If this “crowding out” effect is pronounced due to different institutional reasons in the Swedish case, it could explain the results. The other, more straightforward explanation, is that the native Swedes have weaker preferences for living in homogeneous neighborhoods, as compared to the populations in earlier research.
References


I. Do asylum seekers respond to policy changes? Evidence from the Swedish-Syrian case

With Kristoffer Jutvik

Acknowledgments: The authors are grateful to Karin Edmark, P-A Edin, Matz Dahlberg, Gunnar Myrberg, Cristina Bratu, Per Engström, Torsten Santavirta, Tim Blackwell, Darrel Robinson and seminar participants at the Migration Agency in Sweden, the Institute for Housing and Urban Research (Uppsala University) and the Department of Economics (Uppsala University), for helpful suggestions and comments.
1 Introduction

The relevance of “responsibility-sharing” in the field of migrant reception policy has become increasingly central from a European perspective. On the one hand, in a system where asylum seekers to some extent select host-destinations after entering the EU, it has been suggested that the national state has no or less control of its own borders (Soysal, 1994; Sassen, 1996, 1998). On the other hand, it is apparent how a number of countries develop comparatively restrictive policy structures and attitudes towards migrants and asylum seekers, for example using visa requirements and border checks, which may in turn affect migration flows to specific destinations (Böcker and Havinga, 1998; Freeman, 2004; Castles et al., 2014). The effectiveness of such policies are further heavily dependent on, besides the overarching European political migration infrastructure, what migrants actually know about countries they go to, and how much such knowledge enters their decision making process.

A now fairly large literature with contributions from political science, economics and sociology, has sought to understand if refugees respond to host country policies, and to what extent. Until this point, a large proportion of previous studies can be divided into two broad approaches: One quantitative branch has tried to estimate cross-country differences in the number of asylum seekers as a function of both source and destination country institutions and policies. The other, qualitative approach, has interviewed migrants to better understand in what way and to what extent host country policies matter for the decision on where to go. Although both provide valuable insights, it is essential to further understand how national policies cause responses in asylum flows, and how fast these responses might take place.

In this paper we therefore take a different approach. Using a sudden change in Swedish residence permit policy with regards to Syrian citizens, we estimate the effect of the change in policy on the number of Syrian asylum seekers going to Sweden. The sudden change and high frequency data allow us to evaluate the effect causally already within weeks of the policy change. The possibility to estimate immediate causal impacts of migration policies is a distinct new feature we believe this paper brings to the literature.

Briefly, the policy change, which was implemented in September 2013 (RCI 14/2013, 2013), meant that all asylum seekers from Syria were granted permanent instead of temporary residence permits. Compared

---

1See Neumayer (2004); Thielemann (2006); Moore and Shellman (2007); Hatton (2009, 2016); Keogh (2013); Brekke et al. (2017).

2See Collyer (2005); Gilbert and Koser (2006); Crawley (2010).
to other European states, the change made Swedish assessments of Syrian asylum applications unique and liberal and the new directives were reported in international media instantly after implementation. In order to evaluate the effect of this change, we use an interrupted time series approach to estimate the change in number of asylum applications in Sweden. We further consider the distribution of asylum seekers among the two major European recipients, Germany and Sweden. Also, besides providing permanent shelter, the policy change implied that one could seek family reunification immediately after a residence permit was granted. Since this policy change has different implications for primarily people with and without a family, we also use detailed Swedish register data to study the change in characteristics among the accepted refugees.

The paper adds several novelties to the current state of the literature. Firstly, the high frequency of the data allows us to study effects already within a week after the policy reform. The high frequency will be instructive on how fast information on asylum policy changes travels as well as enters the decision margin of those affected. Also, the current literature most often uses yearly data, which can be problematic given the seasonal patterns of migration. Secondly, using a sudden change in policy in combination with an interrupted time series approach, we can estimate the causal effect of a policy change on the number of asylum seekers. Current quantitative papers struggle with providing causal effects, since all cross-country comparisons suffer more or less from endogeneity issues. Thirdly, this study suggests how a policy change affects the composition of asylum seekers or, in other words, if increasingly beneficial migration policies attract specific categories of migrants after implementation.

In a broader sense, this study adds not only to the academic debate on determinants of migration flows, but it is also relevant from a policy perspective, suggesting how even medium alterations of national policy can have major effects on migrant inflows. Such knowledge may be of utmost importance in the process of planning and implementing reception policies in terms of capacities on both the national and supranational levels of government.

The paper provides three main empirical results. First, the results clearly indicate that Syrians reacted to the policy change; the weekly number of asylum seekers to Sweden almost doubled within 2013. This change also affected the distribution within Europe, with the share applying in Germany dropping as the numbers in Sweden increased. The policy also had a substantial effect rather fast. Already within a couple

\[ \text{See (Brekke et al., 2017) for a discussion on identification.} \]
of weeks, the number of asylum seekers had greatly increased. Sec-
ond, while the effect was seen on a European level, already within a
year the trends were reversed, with Germany receiving the majority
of applications among the two countries. A combination between long
waiting periods in Sweden, a parallel decrease in Germany, as well as
a German labor market reform, making the labor market for refugees
accessible faster, are all plausible explanations for this change. Third
and last, the share of households arriving together, as well as children
and women, dropped, as a response to the policy. This is most consis-
tent with a story, which argues that due to the new possibility of family
reunification, households send one individual who can later apply for
reunification, rather than attempting the often risky trip all together.

The study has the following structure: The next section gives an
overview of previous research, after which we discuss the policy change
as well as other institutional details in the Swedish-Syrian case. Section
4 describes the empirical approach, section 5 gives the results and section
6 concludes.

2 Push and pull factors

There is a long line of research with the specific aim to understand the
international flows of asylum seekers. Within this line of research, nega-
tive aspects in countries of origin or positive aspects of host-destinations,
so-called push and pull factors, have largely been used as prevalent ex-
planations to observed migration patterns. Simply put, push factors
are generally defined as any structural reasons causing individuals to
move out of a place, while pull factors represent any factors attracting
individuals to a destination (Zimmermann, 1996).

A common method for evaluating the importance of push and pull
factors is country paired panel data. The method consists of pairing
host and source countries, and regress the in-between country flow of
asylum seekers or refugees, on the economic, social and political values
both in the source and host country. Generally, the results from these
studies tend to support the idea that push factors outweigh pull factors
as explanation for the great flows of asylum (Hatton, 2009, 2016). As far
as pull factors go, some are fixed or historically dependent and clearly
outside the scope of any policy-maker, such as geographical distance,
a common language, whether the source country is a former colony or
not, or the size of the refugee stock already present at the destination
(Neumayer, 2004; Thielemann, 2006). There is also some evidence that
economic factors, such as unemployment (Thielemann, 2006), or income level (Moore and Shellman, 2007; Keogh, 2013) at destination matters.

These and other studies have also addressed the effectiveness of migration policy as a pull or deterrence measure. A common method has been to use yearly data, and study the effect of recognition rates on asylum flows. Papers using this method have tended to find small but significant effects, suggesting that lower recognition rates are correlated with smaller asylum flows in the same or later years (Neumayer, 2004; Keogh, 2013). A more detailed focus on different types of policies is found in Thielemann (2006), and later Hatton (2009), who provides a taxonomy of three different kinds of migration policy: policies related to entry, the application process and welfare. The first of these entails any policy that deals with border control, and physical entrance into the country. Policies on the process of application refer to rules regarding the status of a refugee, most often empirically captured by the already mentioned recognition rate. Welfare policies include the generosity of the welfare system, conditional on having a residence permit. Results so far point in somewhat different directions. While Hatton (2009, 2016) find deterrent effects from stricter border control and asylum processes, but not from welfare changes, Brekke et al. (2017) and Thielemann (2006) find access policies to have only a minor effect, while stricter welfare policies can have a clear deterrence effect. As noted by Brekke et al. (2017) the results are potentially, partially influenced by endogeneity.

An underlying assumption in many of the quantitative studies above is that the asylum seekers have fairly extensive knowledge about the potential destination countries. This assumption has however been critiqued and questioned by a number of studies. For instance, Böcker and Havinga (1998) argue that the choice of destination for asylum seekers is unconscious. In that sense, the destination is, in most cases, random rather than decided on rational grounds based on a comparison between advantages and disadvantages between contexts. There is a growing body of qualitative studies that provide similar conclusions (Gilbert

---

4 The recognition rate is generally defined as the share of all applicants who are granted a residence permit.

5 To see how endogeneity might cause issues, consider for example an increase in asylum flows from country $i$ to country $j$ in $t$. Due to the increased pressure, country $j$ changes policies, perhaps lowering recognition rates, or adopting stricter access policy measures. Two opposing effects will now occur: First, a policy effect, which potentially decreases the number of applicants. Second, a network effect, where the higher number of applicants from country $i$ in period $t$ makes it easier for other applicants from country $i$ to come in period $t + 1$. A simple model regressing the number of asylum seekers on policy changes might therefore underestimate the effect of the policy.
and Koser, 2006; Crawley, 2010; Mayblin, 2016). Of these Gilbert and Koser (2006) adds an interesting contribution by performing a large set of interviews with asylum seekers regarding their prior knowledge about policy and practice before arriving to the United Kingdom. The study holds that only a few of the respondents arrived with significant knowledge about the destination whilst the rest knew very little. The authors highlight the fact that most of the migrants left their country in a rush and were given false information on their journey, which, combined with low education and limited networks in the UK, implies limited knowledge on the country of destination.\(^6\)

There are also a few studies that, like our study, focus on one country. Among these, Holzer et al. (2000) provides an interesting contribution, investigating the effects of increasingly restrictive migration policies on the inflow of asylum seekers into Switzerland. Although the authors find that restrictive policies reduce the inflow of migrants, the prevalence of push factors is emphasized. The decrease in inflow is mainly explained by fewer applications from culturally and geographically distant countries whilst the inflow from conflict areas, in this case former Yugoslavia, remained stable over the time period despite the policy changes.

In summary, even though previous studies seem rather unanimous about the importance of push factors, as well as certain (often fixed) pull factors, the effects of migration policy changes remain disputed. It is also evident that there is a lack of research using high frequency data, and with an empirical framework that can capture causality.

3 The policy and the Syrian case

Before going into the methodological setup, this section describes the developments and directives provided by the Swedish Migration Agency (SMA) more in detail.

3.1 Developments and directives in the conflict in Syria

Starting in late 2011, as the Syrian conflict rose in intensity and scope, the SMA crafted a number of internal documents containing guidelines and descriptions of recent developments in the conflict referred to as

\(^6\)It should be noted that even if refugees themselves are not aware of destination country particulars, policy changes could still affect where refugees go, if they use smugglers. In fact, Gilbert and Koser (2006) note that the majority of the respondents in their sample used smugglers in some form.
RCI (Instructions from the General Counsel).\textsuperscript{7} These documents serve as guidelines in decision-making and rely on information from internal as well as international sources such as the United Nations High Commissioner for Refugees (UNHCR). Starting in December 2011 the General Counsel of the SMA stipulated that the conflict in Syria was not considered to be an armed but a severe conflict (RCI 32/2011, 2011). This meant that the SMA did not consider the Syrian conflict \textit{in itself} to be of such magnitude that asylum applications generally should be granted. A few months later, in January 2012, the General Counsel decided to hold all expulsions of Syrian citizens (RCI 1/2012, 2012) as the conflict had become more intense. These directives were followed by similar recommendations by UNHCR in March 2012.

Later, in June 2012, given reports of escalation in the conflict, another document (RCI 14/2012, 2012) stipulated that there was a situation of \textit{general violence in the country}. The SMA stated that even if reports from independent sources obstructed a uniform and detailed overview of the situation, the conflict was to be considered as \textit{extremely severe}. In other words, the judgment by the SMA at this time was that any person, merely by being situated on Syrian territory, risked such treatment as stipulated in the Swedish Aliens Act.\textsuperscript{8} This is a central statement as it meant that \textit{all} Syrian applicants should be granted asylum and therefore a residence permit. Of these, approximately 7 out of 10 applicants were granted \textit{temporary} residence permits, allowing three years of residence. The remaining, approximately 3 out of 10 applicants, were granted permanent residence permits. These directives were in many regards in harmony with immigration policies in the rest of Europe, even though the length of the Swedish permit was longer.

\textsuperscript{7}In Swedish: “Rättscchefens instruktioner”.

\textsuperscript{8}There are two categories of protection in the Swedish Aliens Act (that are relevant for this study), \textit{refugees} and \textit{persons in need of subsidiary protection}. In accordance with the Geneva Convention, a person is a \textit{refugee} if he or she has a well-founded fear of persecution due to race, nationality, religious or political beliefs, gender, sexual orientation or affiliation to a particular social group. \textit{Subsidiary protection status}, refers to persons being at risk of death penalty, being subjected to corporal punishment, torture or other inhumane or degrading treatment or, as civilians, risk injury or death due to an armed conflict. Persons that fulfill the prerequisite in any of the above categories have the right to obtain temporary or permanent residence permits. Normally, persons in need of protection obtain permanent residence permits. However, if the permit is granted due to an armed conflict, temporary permits may be granted given the scale, intensity and durability of the conflict (RCI 14/2012, 2012; RCI 14/2013, 2013).
3.2 The policy change

About a year later, in September 2013, the SMA made a new evaluation of the conflict in Syria (RCI 14/2013, 2013). The General Counsel stated that the conflict was still imprinted by a situation of general violence as described in the former RCI from 2012. However, at this time the SMA had reliable information about the severity of the conflict, confirming the atrocities. It was further stated that the conflict was in a deadlock position, in which both sides believed near victory was possible and that the number of actors participating in the conflict had increased. Given the situation of general violence and increased complexity of the conflict, the General Counsel made the judgment that the unrest in Syria would go on for an extensive period of time and stated that all Syrian asylum seekers should be granted permanent residence permits. For the individual applicant the change meant, besides gaining a secure residence status, the possibility to apply for family reunification. Holders of a temporary residence permit did not have this option.

*Figure 1.* The share of permanent residence permits among all granted residence permits (among Syrian applications)

\[\text{Note: Share of permanent residence permits among all granted residence permits. Data concerns distribution among Syrian applications only. Note that the recognition rate was 100\% before and after the policy change.}\]

*Source: Swedish Migration Agency.*
A few important details and implications should be noted regarding the policy change. First, the decision was made by the SMA, and was hence not a party-political decision. This is important since the decision was not the result of a long parliamentary debate, or political discussion, making it practically impossible to react to the change in directives before the actual implementation. Second, the change was de facto implemented. Before September 2013 approximately 3/10 Syrian residence permits were permanent, which changed to 10/10 after the reform. A detailed view of this development is seen in Figure 1. Third, from a European perspective the changed directives stood out as unique and liberal. To the best of our knowledge, there was no parallel change in migration policy from other countries in Europe. Fourth, the policy change was immediately reported in media both nationally (Swedish Radio, Dagens Nyheter, and Expressen) and in international media such as BBC and Al-Jazeera. Essentially then, information about the reform was readily available for anyone seeking it, already the next morning. The impact and significance of the policy change is suggested in Figure 2 which visualize the relative search interest for the Arabic translation of “Sweden” and “Germany” on Google over 2013. The frequency of searches for “Sweden” increases with 83 %-units over the two weeks following the policy change while the relative search interest for “Germany” stays at low levels. The search interest for “Sweden” remains high before it returns to its pre-change level at the end of the year.

3.3 The route: Journey from Syria to Europe

At the end of 2014, Syria had become the world’s major origin country of refugees with 8 million Syrian citizens internally displaced and more than 4 million refugees outside the country. Although most Syrians flee to the neighboring countries of Turkey, Lebanon, Jordan and Iraq (UNHCR, 2016), many attempt to travel further into the European Union with the aim to apply for asylum. The accumulating concentration of migrants at the border of the EU has put significant pressure on the main entry points of Italy, Greece and Hungary. In total, there were 770

---

9Swedish Radio is the public service radio broadcasting company in Sweden. Dagens Nyheter and Expressen are two of the largest daily newspapers in the country.

10Similarly, the search interest for “Sweden” on Google increased in Lebanon and Turkey, harboring large populations of Syrian refugees, in the week of implementation as well. We have investigated the search interest for a number of terms, such as “residence permit”, “asylum Sweden”, “migration policy”, “Swedish Migration Board”, but due to insufficient data coverage we have not been able to see any changes in search interest over the time period.
Figure 2. Relative search interest for Arabic translation of “Sweden” and “Germany” on Google

Note: The figure displays the relative search frequency in relation to the highest point in the graph. The value of 100 marks the highest interest.

Source: Google Trends.
000 irregular border crossings into the EU reported between 2010 and 2015. Recent reports indicate that the lion’s share of asylum seekers only transit through the above countries on their journey towards the northern part of the EU of which Germany (largest absolute migration from Syria) and Sweden (largest per capita migration from Syria) have been the most popular destinations for Syrians (ESPON, 2015).

In order to enter the EU, asylum seekers from Syria rely on irregular border-crossings using mainly two routes; the Central Mediterranean route and the Eastern Mediterranean route. First, the Central Mediterranean route goes from Libya to Malta and Italy. Between the years of 2010 - 2015 there were over 50 000 irregular border-crossings by Syrian citizens reported on this route. Syrian migrants arriving to Italy were normally asked for fingerprints by the authorities. However, as stated by respondents in interviews, many left before a claim for asylum was initiated with the aim to travel further and hand in their application in other countries (Jörum, 2015).\footnote{\textit{Italy is a part of the Dublin Regulation, which states that the country in which asylum is first applied for, is also the country that is responsible for rejecting or accepting the application. An individual who has turned in an application in country A can hence not restart a new application process in country B.}} The journey to Sweden from Italy is described as being rather uncomplicated and made with help of public transportation (Jörum, 2015).

The second route; the Eastern Mediterranean route, goes through Turkey - harboring almost 2 million Syrian refugees - to Greece. This route has over 60 000 irregular border-crossings by Syrian citizens between the years of 2010 and 2015 (ESPON, 2015). The Syrian migrants in Greece describe that they left fingerprints knowing that it wouldn’t change their possibilities to apply for asylum in other European countries (Greece is since 2011 exempted from the Dublin regulation). The route from Greece to Sweden is described as more complex compared to the route from Italy, going through Macedonia and Serbia, before entering EU via Hungary with the help of human smugglers (Jörum, 2015).

Ideally, we would like a time frame, estimating the average, approximate time it would take for a Syrian asylum seeker to go to Sweden. However, how fast an individual can react to the policy will depend on which route is chosen, and where the individual is located when the policy is instigated. Particularly, we expect individuals already in Italy or close to Italy on the central Mediterranean route, to be able to react fairly quickly, perhaps even within a week. For those still in Syria, it is reasonable to expect much longer time frames.
4 Empirical method and data

We now turn to the empirical set-up of our study. To estimate the causal effect of the policy change on the number of Syrian asylum seekers to Sweden, we make use of an interrupted time series design. The design uses two key features:

First, we must have a clearly defined point in time of intervention. For our baseline estimate, which uses weekly data, this point in time will be the 36th week of 2013. We also do monthly estimates, and use the month of September as intervention point. Both points in time are reasonable, given that the change took place early in the 36th week, which was also the first week of September.

Second, we need data on the outcome (in our case number of asylum seekers from Syria) before and after the intervention. For this, we have collected data from SMA on weekly numbers of Syrian asylum applications over the period 2010-2016. For validity checks, we also have applications from the largest other source countries for asylum. The data will be used to create an underlying time trend of number of asylum seekers. Let $T$ be the number of weeks passed since the start of the study period, and $\text{reform}_t$ a dummy variable that takes the value 1 for all observations after the intervention. Our outcome is $\# \text{asylumseekers}_t$, measuring the weekly number of asylum seekers from Syria to Sweden. This leaves us with the following:

$$\# \text{asylumseekers}_t = \beta_0 + \beta_1 T + \beta_2 \text{reform}_t + \beta_3 \text{reform}_t \times T + \epsilon_t \; \; (1)$$

In equation 1, $\beta_1$ represents the effect of the underlying time trend, while $\beta_2$ gives the level effect of the reform. $\beta_3$ is an interaction term capturing any change in the trend after the reform. The effect is hence measured comparing the situation before and after the reform, given the underlying linear trend, as well as the change in the trend after the reform, in the number of asylum seekers. In our baseline estimates, we will compare weekly effects over the year 2013.

4.1 Empirical considerations

Given the sudden change in policy, and the high frequency of the data, the interrupted time series is a convincing way to capture the causal effect of the policy change on the number of asylum seekers from Syria.

---

12 For a good introduction to the method, see Bernal et al. (2017).
13 For information on other data-sources used, we refer to the Appendix, and section B.
While it is possible that other time-confounders could affect the estimation results, we argue that these are unlikely to be an issue in our setting.

First, the increase of Syrians into Sweden should be specific to Syrians, and not a general increase of asylum applications from all source countries to Sweden. As a robustness check, we therefore add a control group, looking at the change in Syrian applications compared to the general increase of applications from other large sending countries. Specifically, we take out the top four asylum countries besides Syria, which during this time were Somalia, Afghanistan, Iraq and Albania. The average number of applications from these countries then form a control group, which is seen in the multiple group model in Equation 2.\footnote{See Linden (2015) for a full description on the multiple group interrupted time series model.}

\[ \text{asylumseekers}_t = \beta_0^M + \beta_1^M T + \beta_2^M \text{reform}_t + \beta_3^M \text{reform}_t * T + \beta_4 Z + \beta_5 Z * T + \beta_6 Z * \text{reform}_t + \beta_7 Z * \text{reform}_t * T + \varepsilon_t \] (2)

To the already described terms, we add Z; a dummy taking the value 1 for the treatment group, Syria, and the value 0, for the control group (Iraq, Afghanistan, Somalia and Albania). $\beta_4$ gives the difference in the intercept of Syrian asylum seekers and the control countries, prior to the intervention. The difference in the trend, or the slope, is given by $\beta_5$, and $\beta_6$ shows the difference in the effect of the reform between Syria and the control countries. Last, $\beta_7$ shows the difference between the slope after the reform. In this case $\beta_6$ is the main coefficient of interest, providing a difference-in-differences style estimator. As will be clear from the graphical illustration of this model in Figure 1, the number of asylum seekers from the control group countries neither increased nor decreased over the intervention line.

Second, we want to make sure that an increase in the number of Syrian asylum seekers does not reflect a general trend in Europe. If substitute countries to Sweden (such as Germany or Denmark) changed legislations or altered the general strictness of their migration policy, it can have spillover effects on Sweden. The control group approach in Equation 2 provides a good test for this as well, since if large general asylum policy changes occurred in substitute countries, we would expect the number of control group country asylum seekers to Sweden to increase or decrease as a response. To further account for this, we make use of
the DEMIG POLICY database, an extensive policy database covering migration related regulatory changes in 45, mostly high income countries between 1945 and 2014. The database covers 6,000 changes, including the one studied in this paper.\textsuperscript{15} Looking at all changes made during 2013, there were no other large scale changes such as the Swedish one, which targeted asylum seekers.\textsuperscript{16}

Third, we need the severity of the conflict in Syria to remain at a similar level. Should the conflict become more (or less) severe over time, any increase (decrease) in asylum applications could reflect the increased (decreased) intensity of the conflict rather than the liberalization of Swedish policies. It is hard to find weekly or monthly estimates of the severity of the Syrian conflict. However, as noted in the previous section, the change in policy did not occur as a response to an increase in the intensity of the conflict, but rather as a response to better, more reliable information, as well as durability. Also, and most importantly, we will further look at the distributional change in Europe, studying the change in the share of asylum seekers who opts for Sweden, rather than the main alternative country: Germany. For these estimates the severity of the conflict should not matter.

Furthermore, it is known that migration is cyclical over the year. In general fewer people migrate over the winter than the summer. Should the reform happen at the same time as a “high season”, estimates can be inflated. In our case the problem is likely the opposite. Migration tends to rise during spring and summer, while the Swedish reform took place in the beginning of the fall. In the Appendix we include regressions where seasonality of migration is addressed.\textsuperscript{17}

Last, as always when performing time series analysis, autocorrelation might cause problems. We deal with this in the simplest way possible, by performing all our estimates with Newey West standard errors.

\textsuperscript{15}For more information on this database, see De Haas et al. (2015).
\textsuperscript{16}An exemption is a German ad-hoc resettlement program. However, this program was small, launched already during the spring, and did not change the situation for Syrians fleeing to Germany.
\textsuperscript{17}Specifically, to address this issue, we use data for all Syrian applications from 2010 to 2015. Based on this period, we put a weight on each month, using so called fourier series. Fourier series is a concept in trigonometry, where each observation is weighed on the interval of (-1,1). If part of the estimate is due to seasonal changes, this should be picked up by using this model. We present regression results including fourier terms in the Appendix.
4.2 More on interpretation

Besides the identification issues mentioned above, the interpretation of the coefficient $\beta_2$ in Equation 1 and $\beta_6$ in Equation 2 depends on how we interpret the change in trends after and before the reform.

A possible effect of the policy is that the number of asylum seekers greatly increase in the following weeks after the policy change, and then drop over the remain of 2013 (in fact this is what we see in the results). Consider for example a group of Syrians already in refugee camps in Italy or possibly Greece. These could react on a short time horizon, should they have the information to act on. For individuals who are still in Syria, however, the time frame from decision to actual asylum application in Sweden might be several months. Note that this would create a steep increase shortly just after the implementation, only to be followed by a downward drop in number of asylum applications. Now, focusing on Equation 1, given a downward post-policy trend during 2013, the interaction term $\beta_3$ will produce a counterfactual that assumes the trend after the new policy is implemented, would have been downward sloping also in absence of the policy change. We will therefore get a higher estimate including the interaction term, than without it. We run models both with and without interaction effects, and make our conclusions based on our most conservative estimates.

5 Results

The following section provides the results of the paper. First we provide the baseline estimates, showing a clear effect in the weekly number of asylum seekers to Sweden. We continue with studying the distributional European dimension and end with a discussion on the changing characteristics of those seeking asylum.

5.1 Baseline estimates - did the policy increase asylum applications to Sweden?

We start by plotting the weekly amount of asylum seekers against the weeks of the year 2013 (Figure 3). The plot further includes two predicted trends on both sides of the intervention line (time of policy change). As seen from the figure, there is a clear jump in the number of asylum seekers at the intervention line. The jump peaks in the next week (second week of September), after which the trend starts slowly moving downward again. Simply by studying the graphical evidence, we
Figure 3. Plotting the number of asylum seekers from Syria to Sweden weekly (including control group countries)

Note: The number of asylum applications to Sweden from Syrian citizens, per week. Results starting from the first week of 2013 until the first week of 2014. The control group consists of the average number of asylum seekers from the 4 largest sending countries besides Syria: Iraq, Afghanistan, Somalia and Albania. The predicted lines are based on the model specified in Equation 2.

Source: Swedish Migration Agency.
can hence conclude that there is a substantial effect already within the first week.

Moving on to the estimations, a single group interrupted time series regression (Equation 1) including an interaction term gives an estimate of 435 more Syrian per week. Given that the weekly average of Syrian asylum seekers before the reform (during 2013) was around 200 individuals, the effect is very large. Note, however, that this estimate uses an interaction term of the time trend, estimating the effect of the reform conditional on the downward slope seen on the right hand side of the intervention line in Figure 3. The estimate of 435 Syrians per week should therefore be seen as an upper bound estimate. We might instead want to measure the effect, conditional on a linear trend over the full period. This can easily be done by estimating the model without the interaction term. Doing this gets an estimate of 330 Syrians more per week, which arguably still is a very large estimate. Conducting a back of the envelope calculation, the estimate implies more than 5000 additional Syrian asylum seekers arriving over a 16 week period, who would not have arrived without the policy change. Regression tables are found in the Appendix, in section A.\textsuperscript{18}

Furthermore, to make sure that this time period was not one with a general increase (or decrease) of asylum seekers to Sweden, or with general changes in Europe, we wish to compare our estimates to the number of asylum seekers from other large source countries. Using Equation 2, we add a plot of a control group, consisting of the average number of asylum seekers from Somalia, Iraq, Albania and Afghanistan. These countries represent the four largest source countries during the period, besides Syria. This plot clearly shows a strong stability over the entire period, with no apparent jump over the intervention line in 2013. Since very little happened, interrupted time series regressions based on the multiple group design in Equation 2, practically replicate the results of the baseline regression. These results are also found in the Appendix.

5.2 The European context

Having shown that the policy change affected the number of asylum applications from Syria, a couple of follow-up questions appear. First, we would like to know how the change effected the distribution of asylum seekers in Europe. Did all Syrians coming to Europe now opt for Sweden? Or did the larger flow of asylum seekers go elsewhere?

\textsuperscript{18}We have also estimated monthly estimates over a 2 year period. These give strong positive effects, and are available in the Appendix.
To look into this we use Eurostat data, and compare the share of asylum applications into Europe. Doing this reveals that Germany and Sweden cover the majority of the applications from Syria over the period, accounting for more than 60% of the total number of applications in the EU. Thus, we focus on these two countries in our comparison.

In Figure 4, we plot the share of Syrian applications on the European level for Sweden and Germany from the start of 2013 to the end of 2014. We use monthly estimates, since Eurostat lacks data on weekly level.\(^{19}\) As can be seen, the share of applications are relatively stable and at an equal level between the countries, at about 30% each, in the months leading up to the policy change.

\textit{Figure 4.} Share of European asylum seekers going to Germany or Sweden

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{Share of Syrian asylum seekers in Europe}
\end{figure}

\textit{Note:} Monthly share of Syrians seeking asylum in Europe, going to Germany or Sweden. \textit{Source:} Eurostat.

In September 2013, and the following months, the trends change drastically when, reflecting our baseline results, migration flows turn to Sweden over Germany. The trend lasts for about four months before it, again, harmonizes with the German level and then drops in the summer of 2014. This exercise suggests that the Swedish change of policy

\(^{19}\text{We start at 2013 for illustrative purposes. The number of asylum seekers was much fewer in the beginning of 2012, making the measure of shares much less reliable with a much higher variance.}\)
had an intense and direct effect on the European distribution of Syrian
asylum seekers, however, the effect was rather limited in a longer time
frame. It seems the effect on the European distribution of Syrian asy-
lum seekers was significant in the short run, as an increasing number of
refugees came to Sweden that otherwise would have applied for asylum
in Germany. The swift change in the shares between the two countries
suggests that these were mainly asylum migrants who had already left
Syria, with a relatively short distance to travel to Germany or Sweden.

After the initial change in distribution, the effect decreases and Ger-
many becomes the major host-destination for Syrians. This is a puzzling
trend, given that Sweden keeps the permanent residence permits through
the entire period. Providing a complete and exact answer to why the
trends diverged is outside the scope of this study. However, there are
some developments that may help us understand it better.

First, as a consequence of the sudden inflow of Syrian asylum seek-
ers, the effectiveness of the Swedish bureaucracy, handling applications
for asylum, gradually dropped compared to their German counterpart.
On average, the SMA processed one Syrian claim 54 days faster than
their German authorities in 2013. The relative difference between the
countries gradually increased in favor of Germany over 2014, before it
sky-rocketed in 2015, when one claim was processed 167 days faster in
the German system.20 Second, while there were no major changes in
German residence permit policy during this period (see DEMIG policy
database), the country did liberalize the access to the labor market.
In May of 2014, the Bundestag decided that asylum seekers would be
allowed to work already within 3 months, instead of the previous 9.21
This option already existed in Sweden, but combined with a larger low
wage sector in Germany, the changed legislation could be part of what
explains the change in asylum flows. Thirdly, during the time after the
policy change there was an ongoing and increasingly central political
and public discussion in Sweden regarding capacities, strategies and lo-
cal migrant reception that may have affected migrants that otherwise
would have come to Sweden. The signals sent out from the political
debate, combined with relative increases in Swedish waiting periods,
during which no family reunification was possible, and the labor market
reforms in Germany, are all plausible reasons for the trend reversal.

20Note that this refers to the time of residence permit. Hence, someone given a res-
idence permit in 2015, on average had waited 167 days longer in Sweden than in
Germany.
21See https://www.bundesregierung.de/Content/EN/Artikel/2014/04/
5.3 Who came?

Another follow-up question is whether the policy changed the characteristics of those coming. The policy created grounds for a more long-term future in Sweden for the arriving migrants. Did this policy change attract a certain group over others? Specifically this question is of interest since the policy made family reunification available for many more.

Ideally, to answer this question, we would like to compare key values of those arriving just before the policy change to those arriving just after. For this we make use of a very detailed database of individual register data, called GeoSweden. The data-set holds anonymous individual level information on all residents with a registered address in Sweden between 1990 and 2014. It includes information on demographics, labor market status, education, the reason for immigration, the month of residence permit and much more. It can however not identify the date for the actual asylum application, which means we have to approximate a treatment and a control group. To construct two groups to compare, we do the following: First we select all who Syrian refugees who received their residence permit during 2013, up until September, which is when the policy change took place. This group could therefore not have selected Sweden based on the new, more liberal rules. Second, data from the SMA shows that 98% of the Syrian asylum seekers who got their residence permits in March 2014 applied for asylum after the policy change. Thus, a clear majority of Syrians receiving their permits in March 2014 selected Sweden after the new asylum rules were instated. We therefore consider the Syrian-born individuals who got their residence permits in March 2014 and onwards as treated by the policy. The database ends in December 2014; hence we cannot include any cases after this.

In Table 1 we compare some basic individual demographic characteristics of the full population of those coming. As can be seen, there is a big jump in the share of men. The pre-reform share of 58 percent has now increased to 68 percent, which in absolute numbers represent an increase of around 4000 men. Also, while the mean age doesn’t change significantly, comparisons between age-groups reveal that younger adults come in greater numbers, while the share of children decreases. A third, fairly big change, is the big drop in the share of individuals arriving with a household member. Before the policy change, more than 60 percent of the individuals arriving were part of a household arriving together, which drops to a minority of 44 percent after the reform.

These numbers can be complemented with a comparison using only the adult population (individuals older than 18). For this, we see a

\footnote{A household member in this case includes a spouse or a child/parent.}
Table 1. *Individual characteristic-differences between asylum seekers arriving before and after the reform*

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (Before Reform)</td>
<td>Mean (After Reform)</td>
<td>Difference</td>
</tr>
<tr>
<td>FULL POPULATION</td>
<td>N=7,640</td>
<td>N=14,172</td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>0.58 (0.49)</td>
<td>0.68 (0.47)</td>
<td>0.105***</td>
</tr>
<tr>
<td>Age</td>
<td>28.28 (17.95)</td>
<td>28.61 (15.87)</td>
<td>0.336</td>
</tr>
<tr>
<td>Child (under 18)</td>
<td>0.32 (0.47)</td>
<td>0.27 (0.44)</td>
<td>-0.053***</td>
</tr>
<tr>
<td>Young (18-35)</td>
<td>0.34 (0.47)</td>
<td>0.39 (0.49)</td>
<td>0.048***</td>
</tr>
<tr>
<td>Middle-age (35-60)</td>
<td>0.28 (0.45)</td>
<td>0.31 (0.46)</td>
<td>0.033***</td>
</tr>
<tr>
<td>Arrived with Family</td>
<td>0.62 (0.49)</td>
<td>0.44 (0.50)</td>
<td>-0.176***</td>
</tr>
<tr>
<td>ADULTS</td>
<td>N=5,174</td>
<td>N=10,351</td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>0.61 (0.49)</td>
<td>0.72 (0.45)</td>
<td>0.118***</td>
</tr>
<tr>
<td>Age</td>
<td>37.62 (13.88)</td>
<td>35.73 (12.06)</td>
<td>-1.883***</td>
</tr>
<tr>
<td>Young (18-35)</td>
<td>0.50 (0.50)</td>
<td>0.53 (0.50)</td>
<td>0.028***</td>
</tr>
<tr>
<td>Middle-age (35-60)</td>
<td>0.41 (0.49)</td>
<td>0.43 (0.49)</td>
<td>0.015*</td>
</tr>
<tr>
<td>Married</td>
<td>0.59 (0.49)</td>
<td>0.61 (0.49)</td>
<td>0.024***</td>
</tr>
<tr>
<td>Men (Married)</td>
<td>0.53 (0.50)</td>
<td>0.68 (0.47)</td>
<td>0.157***</td>
</tr>
<tr>
<td>With Children in Swe.</td>
<td>0.40 (0.49)</td>
<td>0.29 (0.45)</td>
<td>-0.114***</td>
</tr>
</tbody>
</table>

Notes: Individual characteristics for asylum seekers arriving from Syria who got a residence permit in Sweden January-September 2013 (column 1), and residence permit between March 2014 and December 2014 (column 2). Upper panel shows the full population, with mean numbers for Men: share who are men, Age: mean age, Child, Young and Middle-age, which represent the share in the different age groups and Arrived with Family, which shows the share that arrived to Sweden with at least one household member. The lower panel shows result only for adults (>18). T-tests on the mean difference is found in column (3), where *** = p<0.01.

similar pattern, where more men and fewer individuals with children go to Sweden. Note that the latter does not mean that the men are not
fathers, only that they do not have children in Sweden. A fairly likely story for these changes in characteristics is the possibility for family reunification. Before the reform, an asylum seeker coming from Syria in most cases received temporary residence permit without the possibility of family reunification. Many families may in this case have chosen to go together, or not at all. The journey as an asylum seeker to Europe is known to be risky, and it is likely even more risky for children. Therefore, given family reunification as an option, it is possible to send only one family member, who can thereafter apply for family reunification in Sweden. Our results in Table 1 show a picture consistent with this story, and that it was mostly the men who came before the rest of the family.

6 Conclusion

In this study we have set out to estimate the effect of a policy change on the number of Syrian asylum seekers coming to Sweden in order to learn more about how asylum seekers respond to policy changes. To this point, previous studies have largely been divided into two categories of quantitative and qualitative research, whereas the former points to the importance of varying push and pull factors and the latter often argues that asylum seekers, to a large extent, are unaware of the characteristics of the host destinations they end up in. Although these studies provide valuable insights, it is still essential to learn more about how national policy changes cause migration flows into specific nations and how fast these changes affect the distribution of migrants. In order to remedy this knowledge gap, we have estimated the effect of a sudden Swedish policy change in September 2013. The change meant that all Syrian asylum seekers would be granted permanent rather than temporary residence permits. By using an interrupted time series design, we study the effect of the change on the number of Syrian asylum seekers to Sweden.

Significantly, our main conclusion is that the policy change had a clear and direct effect on the number of Syrian asylum seekers coming to Sweden. The weekly data gives an estimate of at least 330 more Syrians per week within 2013, after the policy change, with a significant jump already within the first week after the policy was declared. Given that the average weekly number of asylum seekers before the reform was around 200, the effect should be seen as substantial.

This effect is mirrored in our analysis of the European distribution of Syrian asylum seekers which suggests that Sweden became the major host-destination in Europe for Syrian asylum seekers instantly after
the policy change and in the following months. Interestingly, we see a comparable decline in the inflow of migrants to the other main European destination for Syrians asylum seekers, Germany, as a direct consequence of the policy change suggesting that a large number of refugees selected Sweden over Germany at the time. The results do hence strongly support a view that many Syrian asylum seekers were informed about policy changes. However, after a few months, the migration flows settle and Germany becomes the main destination for refugees. The declining effect of the policy change is indeed an interesting puzzle for future studies to study more in detail. Without giving an encompassing answer to this puzzle, we suggest that the decreasing share of asylum seekers coming to Sweden may have been caused by restraints put on the Swedish bureaucracy after the policy change, tripling the time for handling applications made by Syrian citizens, from signals sent from Swedish politicians and the media, and possibly from a German labor market reform, decreasing the waiting time before asylum seekers could enter the labor market.

Lastly, we conclude that the policy change had some interesting effects on the composition of Syrian asylum seekers coming to Sweden. We conclude that the share of households arriving together decreased as a response to the policy. Given the nature of the reform, enabling family reunification, these results are logical as one individual (most often men) can take the risky trip alone, and when in Sweden apply for permits for their family members at a later stage. These results contrast somewhat to Hatton (2009), who finds clear policy effects from border polices and changing policies on application, but not from changes in welfare policy. The results are further in line with theories putting emphasis on the household as the decision unit (Massey and Garcia, 1987; Massey et al., 1993).

Given our conclusion, that asylum seekers do react to policy changes, the generalizability of these results is central to discuss further. One important aspect of this is how we should categorize the investigated policy in terms of impact. Should it be considered as having higher or lower impact compared to other similar policy changes? In this regard, we assume that there are a few circumstances that cause the effect to be in the upper bound. Firstly, at the time of the policy change Syrian refugees were the largest displaced group in the world. There was also a large group of Syrians already settled in Sweden which may have increased the attractiveness of the country. Furthermore, the relative income of Syrian citizens was relatively high, compared to other conflict areas, which may have made Syrians relatively more mobile. Taken together these aspects suggest that, although there is a clear effect of the
policy change, the effect reported here must be considered to be in the upper bound.

The conclusions made in this paper points to several interesting follow-up studies. First, with a fairly strong agreement among researchers on the importance of push factors as well as fixed pull factors (geographical distance, language, colonial ties) in understanding the flows of asylum seekers, the effects of policy is still a debated arena. Specific studies tracking specific changes in singular countries are, we believe, a fruitful way forward. Also, results in this study has suggested clear, but temporary effects. Future research explaining which policies that might create temporary, and which could create more long term effects, are of great interest. Last, in a broader perspective the study highlights the inter-dependence between states where even minor changes in one nation may have significant implications in others. Future studies will therefore do well to study spill-over effects.
References


# Appendices

## A Regression Tables

### Table A.1. Interrupted time series model. Estimating the number of asylum seekers per week to Sweden in 2013 as function of September policy change

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Basic</th>
<th>Interaction term</th>
<th>Adjusted Seasonality*</th>
<th>Adj. Seasonality + Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weeks</td>
<td>0.00603</td>
<td>1.622</td>
<td>3.494</td>
<td>1.557*</td>
</tr>
<tr>
<td></td>
<td>(1.284)</td>
<td>(1.046)</td>
<td>(3.382)</td>
<td>(0.934)</td>
</tr>
<tr>
<td>Reform = 1</td>
<td>351.6***</td>
<td>435.6***</td>
<td>351.0***</td>
<td>423.6***</td>
</tr>
<tr>
<td></td>
<td>(47.49)</td>
<td>(40.71)</td>
<td>(66.33)</td>
<td>(39.278)</td>
</tr>
<tr>
<td>Week * Reform</td>
<td>-15.76***</td>
<td>-19.99*</td>
<td>-15.27***</td>
<td>-15.27***</td>
</tr>
<tr>
<td></td>
<td>(2.955)</td>
<td>(11.36)</td>
<td>(3.00)</td>
<td>(3.00)</td>
</tr>
<tr>
<td>Control groups</td>
<td>0.721</td>
<td></td>
<td></td>
<td>(1.115)</td>
</tr>
</tbody>
</table>

Observations: 52
Interaction term: NO
Fourier term: NO
Control groups: NO

Notes: OLS regressions of the number of Syrian asylum seekers to Sweden per week, in 2013, regressed on a dummy representing the change in policy, in the first week of September 2013 (Reform = 1), number of weeks into 2013 (Weeks) and an interaction term. Column (2) estimate the model in Equation 1, with Weeks = β₁, Reform = 1= β₂ and Week * Reform = β₃. Column (1) shows the same model excluding β₃. The third column further adds fourier terms to adjust for seasonality. Column (4) estimates the model from Equation 2, where Control groups is equivalent to β₂ M (the reform effect for the four largest non-Syrian asylum countries at the time). *** p<0.01, ** p<0.05, * p<0.1. Newey West Standard errors adjusted for (1) lag in parentheses.

* The adjustment for seasonality is done using fourier terms. These are created in three steps: First, we use monthly data. Second, we treat the year as a clock, giving each month a degree value from 30 (January) to 360 (December). Third, we take out fourier terms, which is done using the formula: \( \cos(\text{degree}_t \times \frac{\pi}{180}) \) and \( \sin(\text{degree}_t \times \frac{\pi}{180}) \), where \( \text{degree}_t \) represents the degree value allocated to each month (\( t \)). The formula gives two variables in which each month is given a value on the interval of [-1,1].
Table A.2. Interrupted time series model. Estimating the number of asylum seekers per month to Sweden 2012-2014 as function of September policy change

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1) Basic</th>
<th>(2) Interaction term</th>
<th>(3) Adjusted Seasonality</th>
<th>(4) Adj. Seasonality + Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Months</td>
<td>53.35***</td>
<td>44.09***</td>
<td>30.04**</td>
<td>48.99***</td>
</tr>
<tr>
<td></td>
<td>(16.16)</td>
<td>(12.12)</td>
<td>(14.53)</td>
<td>(10.34)</td>
</tr>
<tr>
<td>Reform = 1</td>
<td>810.1**</td>
<td>771.51*</td>
<td>850.1***</td>
<td>765.6**</td>
</tr>
<tr>
<td></td>
<td>(356.2)</td>
<td>(436.5)</td>
<td>(287.0)</td>
<td>(331.4)</td>
</tr>
<tr>
<td>Months * Reform</td>
<td>27.38</td>
<td>34.15</td>
<td>15.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(46.57)</td>
<td>(27.69)</td>
<td>(39.14)</td>
<td></td>
</tr>
<tr>
<td>Control groups</td>
<td>23.31</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Observations 36 36 36 36

Interaction term NO YES YES YES
Fourier term NO NO YES YES
Control groups NO NO NO YES

Notes: *** p<0.01, ** p<0.05, * p<0.1. Newey west standard errors adjusted for (1) lag. See Table A.1 for more information on specifications.

B Data sources

The study uses several data sources. The Swedish Migration Agency (SMA) has information on the monthly, weekly and (although not analyzed here) daily number of asylum seekers by citizenship. The SMA also hold data on the recognition rate for asylum seekers as well as the number of recognized individuals in need of protection, who received either temporary or permanent residence permit over time. In addition to that, the SMA holds individual data regarding date of application, date of decision, and type of decision for all Syrian applicants.

For the number of asylum seekers from Syria to Europe, we used EuroStat. Google trend data was used to measure the relative search frequency for “Sweden”, “Germany” and other relevant terms during 2013.

To be able to study the change in individual characteristics on those arriving before and after the reform, we used GeoSweden. GeoSweden is a data-base that holds information on all residents with a registered address in Sweden 1990-2014. The data is based on tax-registries, collected by Statistics Sweden, but administered by the Institute for Housing and Urban Research at Uppsala University. The data includes information on a variety of characteristics, such as labor market status, place of stay, income, immigration and demographic status.
II. Ethnic enclaves, self-employment and the economic performance of refugees

Acknowledgments: The author is grateful to Matz Dahlberg, Per Engström, Matti Sarvimäki, Per-Anders Edin, Alícia Aderà, Susanne Urban, Gideon Goerd, Olof Åslund, Che-Yuan Liang, Peter Fredriksson and seminar participants at Uppsala University, SAE of 2017 in Barcelona, the 2017 IIPF conference in Tokyo, Norface Workshop in Hague, March 2017 and Migration Workshop at IBF, August 2017 for helpful comments and discussions.
1 Introduction

The segregation of foreign born residents presents an interesting trade-off from a welfare point of view. On the one hand, social and physical distance to natives decrease access to essential host country skills, but on the other hand, residential concentration of coethnics (ethnic enclaves) can foster networking and thereby employment opportunities. Understanding the relationship between enclaves and economic activity is therefore important in order to assess the impact of residential segregation.

A now fairly large literature has therefore sought to use various natural experiments to understand and identify the effect of ethnic enclaves on different socioeconomic outcomes, such as employment and income (Munshi, 2003; Edin et al., 2003; Bayer et al., 2008; Damm, 2009; Beaman, 2012), welfare uptake (Bertrand et al., 2000; Åslund and Fredriksson, 2009) and industry specialization (Kerr and Mandorff, 2016). In this paper I attempt to further shed light on the relation between ethnic enclaves and economic outcomes, by estimating the causal effect of residential concentration of coethnics on the probability of self-employment. Self-employment has particular importance; partly because self-employment rates tend to be higher for foreign born than for native born, but also because previous research has shown a tendency of immigrant business owners to hire other coethnics (Åslund et al., 2014). While previous research has taken an interest in the relationship between the size of an enclave and the probability of self-employment, this paper is, to the best of my knowledge, the first one using a natural experiment to provide a causal estimate.

As a simple way of characterizing self-employment and ethnic enclaves, I consider two broad channels. Firstly, self-employment can be a function of the quantity of coethnics. A larger number of coethnics could imply more networking, but also a larger potential market. Assuming that people with a specific ethnicity have some common preferences, a large number of residents who belong to the same ethnic group open possibilities through so called “ethnic markets”. Entrepreneurs can sell to these niche markets, which, in turn, also provide employment (Light, 1972). Secondly, qualified coethnics can provide know-how, skills, in-

---

1 See for example: “Immigrant’s self-employment and entrepreneurship activities” (in “The missing Entrepreneurs 2017”).
2 Two closely related paper are Eliasson (2014), who uses a similar identification strategy to investigate the specific channel of how coethnic bankers affect self-employment probabilities, and Andersson et al. (2017), who study the association between ethnic enclaves and self-employment among Middle-Eastern immigrants in very small neighborhoods (1 km²).
formation, contacts and possibly capital, which are all useful elements to set up a business. Plausibly, only someone with knowledge about the process of business will be able to guide others embarking on a self-employment venture. In addition to providing causal estimates, a second important target of this study is therefore to dig deeper into the mechanisms, including separating the quantitative and qualitative channels within the same empirical framework. I argue that it is the access to coethnics with some relevant qualities (preferably that they themselves run a business), and the skills, institutional knowledge and contacts they provide, which is the key component of the ethnic enclave.

To study the question at hand I use high quality Swedish register data, which includes rich individual information on all permanent residents in Sweden. The data allows me to investigate the probability of self-employment as a function of source country and neighborhood variables, as well as individual characteristics. Anyone with a taxable business income\(^3\) is defined as self-employed, and the size of ethnic enclaves is measured primarily through the self-employed coethnics in the municipality or all other coethnics, both as a share of the municipal population. The variation in the first case is an attempt to capture the quality channel, while the second primarily provides an approximation of the size of a potential niche market.\(^4\)

Since I study the effect of local characteristics on individuals, endogenous geographical sorting is an issue. A newly arrived migrant seeking to start a business, could opt for a place with suitable characteristics for the business in mind. If characteristics of the place drive both self-employment tendencies and the settlement behavior of coethnics, a simple linear regression will be biased. As a way of addressing this endogeneity concern I use a Swedish dispersal policy in place between 1985 and 1994. The policy allowed the government to place all newly arrived refugees in contracted municipalities, which during the years investigated included almost all of the Swedish municipalities. Since individual preferences of the arriving refugees in general were ignored (Borevi and Myrberg, 2010), the policy effectively took away the selection problem, by not allowing for the individuals themselves to decide

\(^3\)To be exact, the definition requires the income to be “active” as opposed to “passive”, which are taxation concepts affecting liability. As I will discuss further in section 4, the separation is in fact not particularly important, since almost no one in my sample opts for passive income.

\(^4\)It should be noted that two individuals are referred to as “coethnics” if they are born in the same country. This definition is used because of Swedish register data, which has information on country of birth, however, not on ethnicity. Birth country is hence the best available proxy for ethnicity. I further extend the definition in the results section, approximating ethnicity by language spoken in the country of birth.
where to move. The set-up of the empirical estimation will be to regress an indicator of self-employment within five years after arrival in Sweden, on ethnic enclave information in the municipality of arrival, which will be the result of the dispersal policy. The preferred specification will further include both municipality and country by cohort fixed effects.\(^5\)

As already noted, the focus on self-employment as an outcome first and foremost complements the literature on the causal effect of ethnic enclaves or networks on varying economic outcomes. Second, the paper adds causal evidence to a group of papers demonstrating associative evidence regarding the size of the enclave and the probability of self-employment. Positive effects are found in the U.S. (Borjas, 1986; Lofstrom, 2002; Fairlie and Woodruff, 2005), Sweden (Andersson and Hammarstedt, 2015) and Australia (Le, 2000). On the negative side, Clark and Drinkwater (2002, 2010) find worse employment and self-employment outcomes from enclave size in Britain and Yuengert (1995) finds no support for the enclave hypotheses in the US.\(^6\)

Third, on a more general level, the paper is connected to the broader literature on the determinants of self-employment.\(^7\)

All in all there are three main points to stress as added values of this study. First, despite many books and papers written on this topic, there is still an evident lack of papers with credible identification methods. By using an arguably exogenous sorting of immigrants, this paper fills a part of that gap, improving the literature methodologically. Second, the paper adds to our theoretical understanding of self-employment.

\(^5\)Note that using the refugee placement policy as a way to get exogenous sorting from the point of view of the arriving refugee is an established method used in several studies. See for example (Edin et al., 2003; Åslund and Fredriksson, 2009; Eliasson, 2014).

\(^6\)Related is also Kalnins and Chung (2006), who find longer survival rates for Gujarati Indian-owned hotels, when more hotels in the vicinity are owned by coethnics. Sociological studies of Cubans in Miami are further found in Wilson and Portes (1980) and Portes and Bach (1985). For a larger review, see: Aldrich and Waldinger (1990) and more recently Fairlie and Lofstrom (2015). Here it also deserves to be mentioned that there is a large documented country heterogeneity in self-employment among different ethnic groups. For example, Fairlie and Lofstrom (2015) note that while 23.1 percent of Korean immigrants in the US are business owners, only 5.1 percent of migrants from the Philippines are registered as self-employed individuals. Similar heterogeneities exist in other countries, such as Britain (Clark and Drinkwater, 2010) and Sweden (Andersson and Hammarstedt, 2015). There is no strong a priori reason for this pattern, and differences in ethnic enclaves across groups can serve as an explanatory factor. Other possible explanations include human capital (Lofstrom and Wang, 2009), home country business experience (Akee et al., 2013), labor market discrimination (Constant and Zimmerman, 2006) and access to capital (Eliasson, 2014).

\(^7\)See Simoes et al. (2016) for a review of determinants of self-employment.
processes. As noted, there are different possible mechanisms through which ethnic clusters might cause self-employment activity. The richness of the data allows me to compare and explore different mechanisms in detail, including the separation of the treatment variable based on self-employed coethnics or all other coethnics. Third, further assessing the economic impact of self-employment, I study the performance of the businesses, specifically asking whether some of the economic negative effects of ethnic segregation can be balanced by business networks and the entrepreneurial possibilities stemming from enclaves.

The baseline estimates show a significant positive effect of the municipal share of self-employed coethnics, in the municipality of arrival, on the probability of self-employment within five years. In the preferred specification, a standard deviation increase in the share of coethnics with business income increases the probability of self-employment with around 2 percentage points. Given that only around 4.5 percent of the sample have any business income within the first five years, this is not a negligible effect. The quantitative estimates, looking at all other coethnics, are mostly negative, with the interpretation that a larger amount of coethnics in general causes a higher tendency for non-self-employment activity. These results are robust to a number of different lag specifications, fixed effects, covariates, functional forms, interaction effects and alternative definitions of the explanatory and dependent variable. The estimations therefore support a qualitative story, in which meeting skilled coethnics matter greatly for self-employment entry, while niche, ethnic markets do not seem to matter for the outcome. Furthermore, there is a long term negative effect on income from being placed with a larger share of self-employed coethnics, an effect which is partly mediated through the choice of self-employment. While enclaves may foster self-employment, the overall effect on economic integration is not necessarily a one dimensional success story.

The next section discusses the mechanisms, section 3 introduces the sample and the empirical model, the data is described in detail in section 4 and the results are shown in section 5. Finally, section 6 concludes.

2 Mechanisms at work

As a simple way of conceptualizing the importance of ethnic enclaves in the self-employment decision process, I take my starting point in a Roy-model, which defines the choice of self-employment as a function of the expected outcome of different labor market options (Roy, 1951).
Such an argument has also been developed into more thorough models. Here, I restrict myself to a highly simplified version only to illustrate the link between enclaves and the choice of employment.

Assume first that the income from self-employment \( y_i \) is given by Equation 1, and other income \( w_i \) is given by Equation 2.

\[
y_i = X_1 \Phi'_1 + \epsilon_{1,i} \tag{1}
\]

\[
w_i = X_2 \Phi'_2 + \epsilon_{2,i} \tag{2}
\]

Income is a function of vectors \( X_1 \) and \( X_2 \), which are, broadly defined, capturing any individual or local characteristics affecting income. \( \epsilon_{1,i} \) and \( \epsilon_{2,i} \) are stochastic shocks. Define the function \( I^* \), as the difference between the outcomes (Equation 3).

\[
I^* = y_i - w_i = (X_1 \Phi'_1 + \epsilon_{1,i}) - (X_2 \Phi'_2 + \epsilon_{2,i}) \tag{3}
\]

Assuming for simplicity that one cares only about income, any individual opts for self-employment if the expected outcome from self-employment is larger than the alternative, or, in formal terms if:

\[
I^* > 0 \tag{4}
\]

How do enclaves enter this model? A simple way to think about it is that there are a number of barriers to starting a firm, which are necessary to surpass if a business is to be started. Assume that there is a subset of \( X_1 \), defined as \( Z \subseteq X_1 \), capturing individual and local requirements needed to be able to start a business. These can be for example institutional and legal knowledge, a specific entrepreneurial skill set, access to capital or access to a consumer base. Define a minimum level of \( Z_{\text{min}} \) as necessary for any business income to be possible. Thereafter define Equation 5:

\[
\begin{cases} 
  y_i = 0, & \text{if } Z < Z_{\text{min}} \\
  y_i \geq 0, & \text{otherwise}
\end{cases} \tag{5}
\]

That is, positive business income is only possible with a certain level of individual and local qualities \( Z \geq Z_{\text{min}} \). Now, access to coethnics can positively affect \( Z \); partly due to an increased consumer base, but plausibly also due to qualified coethnics transferring the necessary skills, legal knowledge, institutional know-how or capital for the business in

\[\text{For examples see Lucas (1978); Evans and Jovanovic (1989).}\]
mind, or by acting as consumers or workers. These qualities in turn drive the possibility for self-employment income. Define $EthnicEnclave$ as the size of the enclave. One can thereafter write:

$$\frac{\partial Z}{\partial EthnicEnclave} > 0 \rightarrow \frac{\partial P[I^* > 0]}{\partial EthnicEnclave} > 0$$

(6)

Taken together, equations 1-6, lead to the simple prediction that the size of the ethnic enclave increases the probability to enter self-employment.\(^9\)

Also, as discussed in the introduction, I attend to separate the effect of the quality of the enclave and the pure size of it. Assume that $Z$ includes two important qualities, $z_1$ representing an available consumer base, and $z_2$ representing different individual assets, such as legal knowledge and entrepreneurial-specific human capital. In both cases an individual needs to reach a certain level before being able to get any business income.

First, $z_1$ gives the demand for whatever product an individual wishes to produce. An ethnically clustered area can create a local demand for different sets of products (niche markets), for which coethnics likely hold large knowledge-based comparative advantages.\(^10\) Aldrich and Waldinger (1990) further note that consumers could have cultural preferences for dealing with coethnics. If self-employed individuals in Sweden open niche market businesses along ethnic lines, we expect the number of coethnics, living close by, to increase the consumer base.\(^11\)

---

\(^9\)The suggested link in Equation 6 is a partial effect, and only holds with certainty if the wage in Equation 2 is independent of, or decreasing in, the size of the ethnic enclave. This does not necessarily hold, in fact there are empirical papers suggesting the opposite. I return shortly to this complication in the next section (see Equation 10).

\(^10\)Light (1972) documents the importance of this phenomenon for several immigrant groups in the United States: “For instance, Chinese grocery stores feature exotic vegetables which most Americans cannot even identify. It is, therefore, no accident that only Chinese operate Chinatown grocery stores where exotic Chinese vegetables are sold”.

\(^11\)An important nuance is that while there are comparative advantages in selling products to a specific ethnic group, it can also put a cap on how much a firm can grow. An indication of this is Aguilera (2009), who finds that self-employed Mexican immigrants within enclaves have lower returns than non-enclave Mexican self-employed.
Furthermore, $z_2$ represent specific skills for starting a business in a certain country and place, including institutional and regulatory knowledge as well as specific skills on the process of self-employment. Network structures and information sharing within coethnics can here serve as an important tool to access better understanding on self-employment procedures. It is reasonable to assume that first and foremost self-employed coethnics, who have themselves gone through the same process, can inform and instruct newly arrived individuals on self-employment skills. Therefore, I define,

\[
\frac{\partial z_2}{\partial \#\text{Self}-\text{Employed Coethnics}} > 0 \quad \rightarrow \quad (8)
\]

\[
\frac{\partial P[I^* > 0]}{\partial \#\text{Self}-\text{Employed Coethnics}} > 0
\]

Equation 7 and Equation 8 provide the main hypotheses of the paper, that access to a larger number of coethnics, or a larger number of self-employed coethnics, increase the probability to become self-employed.

2.1 Some empirical considerations
The previous subsection provided a simple stylized picture of the relationship between an enclave and the probability of self-employment. Practically there are, however, a couple of complications to keep in mind. First, while more self-employed coethnics provide a larger source of information, they also mean a larger source of competition. This could decrease opportunities and have a negative effect on self-employment. Described in terms of the model:

\[
\frac{\partial^2 z_2}{\partial \#\text{Self}-\text{Employed Coethnics}^2} < 0 \quad (9)
\]

That is, the effect is positive but decreasing in the number of coethnics running a business. This points to the importance of testing different functional forms.
Second, individuals might suffer from liquidity constraints, which can be eased with access to coethnics with assets. An interesting historical example is rotating credit associations (Light, 1972; Aldrich and Waldinger, 1990). Historically in the US, many formal credit givers were not open to minorities, leading smaller groups of immigrants to swap and share credit within the group. A modern application is Elia-son (2014), who shows that having a coethnic local banker in the port of entry municipality increases the propensity of self-employment. To test for this in the current setting I will run regressions showing that being placed with more coethnics with larger levels of capital income do not cause a higher probability of self-employment.

Last, and most importantly, as has been noted, while a larger number of coethnics create access to an ethnic market, they can also increase formal labor market opportunities, indirectly causing a lower probability of self-employment. Similarly, a high number of self-employed coethnics could increase the options on the formal labor market for a newly arrived refugee, in being employed by the very self-employed he or she encounters. This mechanism is relevant since previous research has shown that coethnics tend to hire other coethnics (Åslund et al., 2014). Based on Equation 2 I get:

\[
\frac{\partial X_2}{\partial \# \text{Self} - \text{Employed Coethnics}} > 0; \quad \frac{\partial X_2}{\partial \# \text{Coethnics}} > 0 \rightarrow \frac{\partial w_i}{\partial X_2} > 0 \quad (10)
\]

That is \#Coethnics and \#Self - Employed Coethnics can increase potential formal labor market income. An implication for the empirical estimates is that a potential positive significant effect on self-employment from the number of self-employed coethnics, or coethnics in general, might be a lower bound of the effect.

\footnote{Naturally, this is also linked to discrimination, which in Sweden, as well as in other places, have been documented for labor market settings (Eriksson and Lagerstrom, 2012). Discriminated groups with larger obstacles to climb to the formal labor market could have more to gain from networking. Seen in this light, self-employment could be a strategy when wage labor is not available (Constant and Zimmerman, 2006).}
3 Empirical model and sample selection

3.1 Some brief notes on the sample

To estimate the effect of ethnic enclaves on the probability of self-employment, I make use of *GeoSweden*, a large and rich administrative database with yearly, individual information on every resident in Sweden from 1990 to 2014.\textsuperscript{13} The information is collected by Statistics Sweden, and is mainly based on population and tax registries.

The sample consists of working age (18-55 years old) foreign born adults, who arrived in Sweden 1990 or 1991. The choice of years is related to the identification strategy, which uses a refugee placement policy, that placed refugees in contracted municipalities. The policy was in place between 1985 and 1994, but reportedly became less encompassing after the unexpected increase in immigration from former Yugoslavia in 1992 (Åslund and Rooth, 2007). Given that the database does not stretch further back than 1990, the first two years of the 90’s will make up the sample of refugees. More on the refugee placement policy, and how it is used for identification, is found in section 3.2.

Only *refugees* were placed, and to make sure my sample is first and foremost made up of this group of immigrants, I add two restrictions. First, I limit myself to those arriving from, and who were born in, non-OECD countries. Second, I throw out anyone who, at arrival, already had a household member in the country. This household member had arrived in a prior year, which could mean that the new arriving immigrant was a tied family migrant, and therefore not subject to placement via the governmental program. Last, given the extensive number of countries with a very small number of arriving individuals and a small already present refugee stock, I make a last restriction to the top ten sending countries in 1990 and 1991. The total sample is made up of 14,091 individuals from the ten countries seen in Table 1. A more detailed discussion on the construction of the sample is found in the Appendix, in section A.

3.2 Empirical model

Given the sample selected, the target is to estimate the causal effect of different measures of the ethnic enclave on the probability of self-employment. The decision of self-employment for individual *i* is given by \( y_i \in \{0,1\} \). \( y_i = 1 \) if an agent declares positive business income, and

\textsuperscript{13} All residents with a registered address with the tax authorities.
Table 1. Distribution of country of birth for final sample of immigrants who arrived in 1990-1991

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Iran</td>
<td>3,118</td>
<td>22.13</td>
<td>22.13</td>
</tr>
<tr>
<td>Iraq</td>
<td>2,052</td>
<td>14.56</td>
<td>36.69</td>
</tr>
<tr>
<td>Lebanon</td>
<td>1,897</td>
<td>13.46</td>
<td>50.15</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>1,388</td>
<td>9.85</td>
<td>60.00</td>
</tr>
<tr>
<td>Somalia</td>
<td>1,343</td>
<td>9.53</td>
<td>69.53</td>
</tr>
<tr>
<td>Syria</td>
<td>1,201</td>
<td>8.52</td>
<td>78.06</td>
</tr>
<tr>
<td>Yugoslavia</td>
<td>969</td>
<td>6.88</td>
<td>84.93</td>
</tr>
<tr>
<td>Vietnam</td>
<td>919</td>
<td>6.52</td>
<td>91.46</td>
</tr>
<tr>
<td>Romania</td>
<td>692</td>
<td>4.91</td>
<td>96.37</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>512</td>
<td>3.63</td>
<td>100</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>14,091</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Data from GeoSweden. Sample restriction described in Section 3.1.

This definition is wide, but also has limitations. Most importantly, larger businesses are often set up as limited companies, in which case income is not declared for the owner, but for the corporation. The importance however turns out to be limited. In my sample, using survey variables on labor market status for 1995 and 1996 (not available for the years 1990-1991), only 13 people are registered to be involved in joint stock companies. Adding these as self-employed does not alter any conclusions.

In the baseline estimate, the dependent variable is measured within five years, and is cross-sectional in nature. While there is no scientific a priori reason for the use of exactly five years as lag, the choice is not without reason. On the one hand, I do not want a time horizon that is too short: within just a couple of years of arrival very few have likely had the time to establish a business. On the other hand, if I make the lag too long, the connection to the network in the assigned municipality likely

---

14 To be exact, only income that is active rather than passive is included. Passive income was added to the income statistics in 1991, so all income counted in 1990 is “active”. The concept is related to tax liability, and active income is in theory based on the agent having worked at least 600 hours during the relevant year or, performed the operation with own effort. In practice, when business owners declare income, they define their business as active or passive themselves. Looking at my sample, only 6 out of roughly 12 000 had passive income in 1995 or 1996. The use of only active income is therefore hardly a very large restriction.

15 A comparable term for a limited company is for example “joint stock company”. For more information on the Swedish types of business, see [https://www.verksamt.se/web/international/starting/types-of-business](https://www.verksamt.se/web/international/starting/types-of-business).
becomes less important, and a lot of individuals have possibly moved. I choose five years as a midway case. I do however provide estimates for 3 to 7 year spans as well, giving similar estimates as in the baseline case.

In detail this implies that, for individual $i$, born in country $c$, arriving in municipality $m$, with cohort $k$ (1990, 1991), I regress whether or not the individual got business income at some point within five years after arrival, on the municipal share of self-employed coethnics and share of other coethnics, in the municipality of arrival.\footnote{I standardize with municipality population in the baseline case. I do this based on a stylized idea of interactions: Consider a newly arrived refugee going out every day with a certain probability $P_m$ of meeting self-employed coethnics. Now, if I assume where an individual goes is independent of the number of self-employed coethnics, the probability of meeting any, would be $P_m = \frac{\#Self\text{-}Employed\ Coethnics}{Population_m}$, that is the number of self-employed coethnics as share of the municipality population. Further assuming all agents go out the same amount of days, I can use this definition as the treatment for the enclave. I however also provide robustness checks including different functional forms (see the Appendix, Section B).} Fixed effects are included for arrival municipality ($\sigma_m$) and the interaction of cohort and birth country ($\theta_{kc}$). I further include a vector of individual level covariates ($X_{it}$), including age, age$^2$, dummies for sex, university degree, if the individual moved during the arrival year, if he or she is married, if the individual has children, and how many. The full specification is seen in equation 11 (where SE = self-employed and nonSE = not self-employed).

$$y_{icmk} = \alpha + \beta_1 \frac{\#SE\ Coethnics_{cmk}}{Population_{mk}} + \beta_2 \frac{\#nonSE\ Coethnics_{cmk}}{Population_{mk}} + X_i \Gamma_1 + \sigma_m + \theta_{kc} + \epsilon_{icmk} \quad (11)$$

\textbf{Identification discussion}

The design in Equation 11 should take care of local labor market effects (e.g. more people owning a firm might just reflect a relatively better business climate). If there is a strong tendency for immigrant entrepreneurship, or entrepreneurship in general in a specific municipality, it should be accounted for with the fixed effects on municipality level. Adding the country by cohort fixed effects takes care of any general tendency within a certain country and cohort to become self-employed. The regression can hence be seen as a difference-in-differences, where the total treatment effect is given by comparing the difference in effects between country by cohort groups within a municipality, to the difference in effects between country by cohort groups within another municipality.

The fixed effects do, however, not address possible selection. To account for this I use a dispersal policy, which, conditional on a number
of observed individual characteristics, stripped away the possibility to choose your place of stay. The Swedish refugee placement program has already been described and discussed at lengths by various studies and reports (see Read (1992); Invandrarverket (1997); Edin et al. (2003); Åslund and Fredriksson (2009); Borevi and Myrberg (2010); Dahlberg et al. (2012)), below, I therefore provide only the key characteristics for identification.

The policy, which was in place between 1985 and 1994, aimed at geographical dispersion of refugees. An asylum seeker in one of these years, went through roughly the following process: After arrival and application, the migrant was placed in a refugee center run by the immigration board. In the center he or she took preparation courses, but was not allowed to work. After receiving a residence permit, the migrant was placed in one of the contracted municipalities, which during the time span of the study included almost all of Sweden’s 289 municipalities. According to Edin et al. (2003), there was no correlation between the location of the center and the port of entry. The municipality received state contributions to finance the reception of those arriving, however, migrants were allowed to move after placement, and any welfare contributions were not contingent on staying in the assigned location.

Besides the explicit target to limit the inflow to larger city regions, the immigration board was also supposed to match individuals in accordance with labor market characteristics. As has been documented prior, this ambition was undermined by the shortage of housing in many regions. Housing vacancies therefore became the most relevant (in some cases only) criteria, when assignment was decided (Borevi and Myrberg, 2010). Last, one should note that it was only refugees that were part of the distribution policy. Immigration due to for example family, or other reasons, was not part of the program.

As has been shown by Dahlberg et al. (2012) (see Figure 3B), the program succeeded in distributing refugees from larger to smaller cities. Given that it was aimed at strategically placing immigrants in a certain manner, it is evident that the policy cannot be seen as a randomized experiment. Despite this, it has been argued that the program can be seen as exogenous from the point of view of the arriving individual. There are a few reasons for this. First, even if immigrants were allowed to give preferences on where to go, previous research indicate that these suggestions were generally given little consideration (Borevi and Myrberg, 2010; Read, 1992). Second, as argued by Edin et al. (2003), since there were no contact between municipal officers and refugees, selection on unobservables is likely ruled out. Third, to the extent strategic place-
ment took place, it was based on information available in the Swedish data registers. The argument is therefore that placement was exogenous, conditional on observable characteristics.

Below I further provide an attempt to test if the design achieves exogenous variation in the explanatory variable. What one would like is for individuals who were treated with a larger enclave to be similar as compared to those who were placed in smaller enclaves, with regards to their ability or intent to become self-employed. I test for this in two simple steps. First, I use a linear regression model to predict the probability of self-employment as a function of individual characteristics (cf. equation 12).

$$y_{i\text{cmk}} = \alpha + X_{2i}\Delta' + \theta_{kc} + \epsilon_{i\text{mck}}$$ (12)

The $X_{2i}$ include age, age$^2$, sex, marital status, whether or not the individual have children, how many children, if he or she has a university degree, yearly disposable income in the arrival year, social assistance from the state, whether or not the individual is employed in the arrival year, whether or not the individual moves the initial year and a dummy for cohort and birth country. I further interact age and education status as well as education status and sex. Men are more prone to start a business, and self-employment has been shown to be positively correlated with age. Since a family could increase the propensity of business through family firms, I also include whether or not you have a child, and how many. The prediction ($\hat{y}_{i\text{cmk}}$), becomes a measure for individual likelihood of self-employment.

Having done this regression, in step two, I regress the self-employed coethnics (as share of the population), living in the municipality of arrival on the predicted self-employment ($\hat{y}_{i\text{cmk}}$), conditional on the full set of covariates and fixed effects used in Equation 12. Arguably there should be no effect on the size of the enclave if you have a higher probability of self-employment. The coefficient is negative, non-significant and as low as 0.001, which arguably is very low. Note also that, since the effect is negative, if there is any selection of those more prone to self-employment, they seem to choose municipalities with less coethnics. This is arguably less of a problem, since, if anything, it would imply an underestimation of the effect of enclaves on self-employment.$^{17}$

Furthermore, if there are unobserved labor market characteristics on municipality level, especially suitable for a certain birth country, this

---

$^{17}$I also correlate the prediction with the continuous number of coethnics as share of municipality population, which turns out to be almost the same (-0.003), and insignificant. Both estimates are available upon request.
could drive both the self-employment tendency for newly arrived as well as the number of coethnics, who came to the municipality in previous years. I therefore include robustness tests where I add controls on municipality by birth country level. The most important indicator is the municipality employment rate among coethnics. This is a quality indicator, which captures the municipality labor market integration of a specific country group. Indirectly, this further provides a test for whether individuals become self-employed due to poor labor market integration in a certain municipality. My robustness checks suggest that this mechanism is not the driving force behind the results observed. I return to this point in section 5.2.

Last, a somewhat different issue refers to the option of subsequent moving. Immigrants have no obligation to stay in the assigned municipality, and as will be demonstrated in the descriptive statistics, the option of moving is used. The main threat to identification here is that some entrepreneurs, who are placed with many self-employed coethnics, experience competition, or perhaps even a saturated market-place, and move to another municipality, where they instead can start a business. If this is the case, effects of coethnics at the arriving municipality will overestimate the effect on self-employment. To get some basic understanding of this, I compare the mean between those staying in the same municipality after five years, and those living in another municipality five years later. What the table shows is that there is no statistical difference between stayers and movers regarding the tendency to become self-employed within five years. Given the results shown, I deem subsequent movers not to be a threat to identification.

Table 2. Comparing means for group of stayers and subsequent movers five years after arrival

<table>
<thead>
<tr>
<th>Sample staying</th>
<th>Sample moving</th>
<th>Pr(T &gt; t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Share Self-Employed</td>
<td>0.046 (0.0026)</td>
<td>0.042 (0.0023)</td>
</tr>
</tbody>
</table>

Notes: Comparing the probability of self-employment for those moving to another municipality within the first five years, and those staying.

4 Describing the sample

Having introduced the research design, I now proceed by describing the characteristics of the sample. Table 3 includes a left panel with individual information for all the refugees at arrival (placement year) and the same follow up information five years later.
Table 3. Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>Std.dev</td>
</tr>
<tr>
<td>Individual characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>14,091</td>
<td>30.58</td>
<td>8.43</td>
</tr>
<tr>
<td>Married</td>
<td>14,091</td>
<td>0.53</td>
<td>0.50</td>
</tr>
<tr>
<td>Men</td>
<td>14,091</td>
<td>0.63</td>
<td>0.48</td>
</tr>
<tr>
<td>Children</td>
<td>14,091</td>
<td>0.36</td>
<td>0.48</td>
</tr>
<tr>
<td>#Children (</td>
<td>parent)</td>
<td>5,065</td>
<td>2.22</td>
</tr>
<tr>
<td>University educated</td>
<td>14,091</td>
<td>0.06</td>
<td>0.23</td>
</tr>
<tr>
<td>&lt; 9 years of education</td>
<td>14,091</td>
<td>0.82</td>
<td>0.38</td>
</tr>
<tr>
<td>Big City</td>
<td>14,091</td>
<td>0.16</td>
<td>0.37</td>
</tr>
<tr>
<td>Share with Wage&gt;0</td>
<td>14,091</td>
<td>0.16</td>
<td>0.36</td>
</tr>
<tr>
<td>Wage</td>
<td>2,205</td>
<td>317.27</td>
<td>343.49</td>
</tr>
<tr>
<td>Self employment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share with Business Inc.</td>
<td>14,091</td>
<td>0.001</td>
<td>0.03</td>
</tr>
<tr>
<td>Business Inc.</td>
<td>10</td>
<td>360.60</td>
<td>356.75</td>
</tr>
<tr>
<td>Municipality characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pop</td>
<td>14,091</td>
<td>66,870</td>
<td>103,742</td>
</tr>
<tr>
<td># coethnics</td>
<td>14,091</td>
<td>392.45</td>
<td>920.74</td>
</tr>
<tr>
<td>Share of population</td>
<td>14,091</td>
<td>0.005</td>
<td>0.01</td>
</tr>
<tr>
<td># coethnics with wage&gt;0</td>
<td>14,091</td>
<td>194.12</td>
<td>486.43</td>
</tr>
<tr>
<td># S-E coethnics</td>
<td>14,091</td>
<td>15.43</td>
<td>41.69</td>
</tr>
<tr>
<td>Share of population</td>
<td>14,091</td>
<td>0.0002</td>
<td>0.0003</td>
</tr>
<tr>
<td>Share of coethnics</td>
<td>14,091</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>≥1 S-E coethnic</td>
<td>14,091</td>
<td>0.53</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Notes: Big City implies staying in one of the three biggest cities, Stockholm, Malmö or Gothenburg. Share with Wage>0 counts those who declared any positive wage during the year. Similarly Share with Business Inc. shows the share with any positive declared (active) business income. Both Business Inc. and Wage are conditional on having some income; in the former case from business activity and in the latter from other labor market activities. Incomes are given in hundreds of Swedish SEK (in 1990 $1 ≈ 6 SEK). Municipality characteristics show information on municipality level. Hence # coethnics is the average number of coethnics in the municipality for a person in the sample. S-E coethnic shows the number of coethnics who are self-employed, measured as those who have any business income. ≥1 S-E coethnic is a dummy for the percentage in the sample who stays at a municipality with at least one self-employed coethnic. Exact sample restrictions are described in section 3.1 and section A.
At arrival, around half the sample are married, there are somewhat more men than women and the majority are so far not parents. The education variable tells us that only around six percent of the sample have a university degree, while a large majority have less than a high school education. A surprisingly large number is that almost one in six have some paid work during their first year. It should however be noted that the mean salary (for anyone with positive income) over the whole year is around 32,000 SEK (in 1990 around $6,000).

Looking at the key variables: \#coethnics means that an average immigrant in the sample comes to a municipality with 392 adult coethnics, of which 194 have a positive salary (\# coethnics with wage>0), and 15 are self-employed (\# S-E coethnics). Seen as share of the number of coethnics, on average about 3 percent of coethnics are self-employed, and seen as share of the full population, around 0.02 percent are self-employed coethnics. Last, not surprisingly, Share with Business Income says that only ten of the arriving migrants, were able to start a business within their first year in the country.

Five years later around 3 percent of the sample have some business income. It is here important to remember that this reflects the share of the entire sample, in which more than half are unemployed. Seen as a share of the employed, the rate of self-employment is around 7 percent. Also, the larger average population and share of people in big cities, suggests that an important part of the sample moves from their referred municipalities to larger metropolitan areas. All of this is expected and in line with previous research.

In Table 4, I continue by showing characteristics and type of establishment among those who became self-employed. 611 individuals get some business income within the five year interval, representing around 4.5 percent of the sample. The share of high and low educated seem to be the same as the sample at large, which also goes for the share of parents. The self-employed are also slightly younger, but most importantly, the

---

18 How to interpret this information is far from straightforward. Many of those with no formal education in 1990 may in fact be educated, but awaiting certification of their home country training. This is indicated by the fact that 22 percent of the sample have a university degree five years later, and that most of the sample now have more than 9 years of education. Some of this change is likely because of authorization of already existing human capital.

19 The number of coethnics is based on the working age population.

20 This can be compared to the national average at the time of 9 percent. This number increased somewhat during the following 2 decades, to 10.9 in 2010, which is low compared to most other countries. Note also that this is in line with cross country findings which show a distinct pattern where richer countries in general have a smaller share of self-employed among the working population (See World Development Indicators).
share of men is, overwhelming. Over 80 percent of the establishments are run by men. Also noticeable is that more people run businesses outside the big cities, as compared to where the general sample move.

Unfortunately a sizable part of the individuals owning a firm (189 individuals) have missing values on sector information. Of those left, most work in five sectors, which can be seen in the upper panel of Table 4. The biggest is restaurants, making up 24 percent of the businesses. Other important sectors include retail stores, hairdressers and cab-drivers.²¹

4.1 Country of origin

To supplement the basic individual information I include statistics on distribution of country of origin. As previously described, there are 10 countries represented in the sample. Of these, Iran, Iraq and Lebanon are the largest, making up more than half the sample. Romania and Bulgaria are the smallest, with less than ten percent of the refugees. To show some of the important heterogeneity between the countries, Table

²¹Note that the sum of frequency of the sectors will not add up to the 611 self-employed seen in the lower panel. This is because some have higher income from other labor, which means that I cannot tie the firm ID to the individual.
5 includes the frequency and relative frequency of the arriving refugees and the number who become self-employed at any point during the first five years. I also include the treatment, that is the size of the enclave.

A first thing to notice is the difference between number of refugees, and number of self-employed as share of the sample. Individuals from Iran make up 22 percent of the sample of refugees, but 26 percent of those who have business income within five years. In other terms, Iranians become more self-employed than what can be expected based on the relative frequency in the sample. Besides Iran, individuals from Syria and Lebanon are heavily over-represented as self-employed, whilst Somalis, Ethiopians, Vietnamese and Iraqis become self-employed less than expected from the relative frequency of refugees.

In general the above pattern is also reflected in the size of enclave. The average Syrian refugee for example arrives at a municipality with 230 coethnics, of which 5 percent are in self-employment. The average Somali on the other hand arrives at an enclave with 56 coethnics, of which less than 0.1 percent are self-employed. In other words, most Somalis arrive at a municipality were there are no self-employed coethnics. While not being causal evidence, the statistics for the different countries tell a story in line with the importance of enclaves: The countries with earlier large enclaves, also produce a higher share of self-employed among the refugees arriving in 1990 and 1991.

This pattern can also be shown using maps. In Figure 1, I show the distribution of the enclaves for the case of Iranians. In the map, the borders represent the administrative division of Swedish municipalities. Colored parts imply that at least one refugee born in Iran arrived to that very municipality in 1990 or 1991. No refugees from Iran arrived at the grey parts. The color is shaded where deeper colors of red represent larger enclaves. The Figure most to the left (Figure 1a), divides the sample in quintiles of number of coethnics, Figure 1b does the same but with number of self-employed coethnics, and Figure 1c shows a binary division: red for the municipalities where at least one of the arriving refugees in 1990 and 1991 started a business within the first five years. Note that this distribution is based on the municipality of arrival.

The map is interesting from two perspectives. First, there is a fairly strong geographical distribution of refugees. Iranians arrived to municipalities all over the country. Second, while it is far from a definitive proof, just eye-ball ing the distribution shows that areas where the enclaves were larger, also seem to be places where new firms were started.22

---

22Spatial illustrations of the enclave size and self-employment situation for each of ten source countries is available upon request.
<table>
<thead>
<tr>
<th>.</th>
<th>Iran</th>
<th>Iraq</th>
<th>Lebanon</th>
<th>Ethiopia</th>
<th>Somalia</th>
<th>Syria</th>
<th>Yugosl.</th>
<th>Vietnam</th>
<th>Romania</th>
<th>Bulgaria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total refugees 90-91</td>
<td>3,118</td>
<td>2,052</td>
<td>1,897</td>
<td>1,388</td>
<td>1,343</td>
<td>1,201</td>
<td>969</td>
<td>919</td>
<td>692</td>
<td>512</td>
</tr>
<tr>
<td>As share of sample</td>
<td>22%</td>
<td>15%</td>
<td>13%</td>
<td>10%</td>
<td>10%</td>
<td>9%</td>
<td>7%</td>
<td>7%</td>
<td>5%</td>
<td>4%</td>
</tr>
<tr>
<td>Self-employed within 5 yrs.</td>
<td>159</td>
<td>61</td>
<td>156</td>
<td>5</td>
<td>3</td>
<td>136</td>
<td>38</td>
<td>3</td>
<td>16</td>
<td>34</td>
</tr>
<tr>
<td>As share of sample</td>
<td>26.0%</td>
<td>10.0%</td>
<td>25.5%</td>
<td>0.8%</td>
<td>0.5%</td>
<td>22.3%</td>
<td>6.2%</td>
<td>0.5%</td>
<td>2.7%</td>
<td>5.6%</td>
</tr>
<tr>
<td>As share of own country</td>
<td>5.1%</td>
<td>3.0%</td>
<td>8.2%</td>
<td>0.4%</td>
<td>0.2%</td>
<td>11.3%</td>
<td>4.0%</td>
<td>0.3%</td>
<td>2.3%</td>
<td>6.6%</td>
</tr>
</tbody>
</table>

**Municipality characteristics**

<table>
<thead>
<tr>
<th>.</th>
<th>Iran</th>
<th>Iraq</th>
<th>Lebanon</th>
<th>Ethiopia</th>
<th>Somalia</th>
<th>Syria</th>
<th>Yugosl.</th>
<th>Vietnam</th>
<th>Romania</th>
<th>Bulgaria</th>
</tr>
</thead>
<tbody>
<tr>
<td># coethnics 90-91</td>
<td>848</td>
<td>221</td>
<td>167</td>
<td>226</td>
<td>56</td>
<td>238</td>
<td>1,245</td>
<td>118</td>
<td>148</td>
<td>46</td>
</tr>
<tr>
<td>(1,438)</td>
<td>(380)</td>
<td>(231)</td>
<td>(398)</td>
<td>(88)</td>
<td>(369)</td>
<td>(1,686)</td>
<td>(141)</td>
<td>(202)</td>
<td>(65)</td>
<td></td>
</tr>
<tr>
<td>As share of population</td>
<td>0.83%</td>
<td>0.28%</td>
<td>0.35%</td>
<td>0.25%</td>
<td>0.13%</td>
<td>0.37%</td>
<td>1.20%</td>
<td>0.60%</td>
<td>0.47%</td>
<td>0.08%</td>
</tr>
<tr>
<td>(0.0051)</td>
<td>(0.0018)</td>
<td>(0.0027)</td>
<td>(0.0018)</td>
<td>(0.0011)</td>
<td>(0.0061)</td>
<td>(0.01043)</td>
<td>(0.0072)</td>
<td>(0.0055)</td>
<td>(0.00066)</td>
<td></td>
</tr>
<tr>
<td># S-E coethnics</td>
<td>28.6</td>
<td>7.1</td>
<td>8.2</td>
<td>0.3</td>
<td>17.1</td>
<td>72.1</td>
<td>72.2</td>
<td>1.7</td>
<td>3.6</td>
<td>2.1</td>
</tr>
<tr>
<td>(50.63)</td>
<td>(15.83)</td>
<td>(14.27)</td>
<td>(3.95)</td>
<td>(0.20)</td>
<td>(29.06)</td>
<td>(99.88)</td>
<td>(3.19)</td>
<td>(6.89)</td>
<td>(4.33)</td>
<td></td>
</tr>
<tr>
<td>As share of coethnics</td>
<td>2.5%</td>
<td>1.9%</td>
<td>4.2%</td>
<td>0.5%</td>
<td>0.0%</td>
<td>5.0%</td>
<td>6.2%</td>
<td>1.0%</td>
<td>1.6%</td>
<td>2.0%</td>
</tr>
<tr>
<td>(0.0277)</td>
<td>(0.0337)</td>
<td>(0.0619)</td>
<td>(0.0211)</td>
<td>(0.0041)</td>
<td>(0.0630)</td>
<td>(0.0447)</td>
<td>(0.0186)</td>
<td>(0.0223)</td>
<td>(0.0356)</td>
<td></td>
</tr>
<tr>
<td>As share of population</td>
<td>0.0229%</td>
<td>0.0054%</td>
<td>0.0152%</td>
<td>0.0009%</td>
<td>0.0000%</td>
<td>0.0238%</td>
<td>0.0643%</td>
<td>0.0053%</td>
<td>0.0050%</td>
<td>0.0016%</td>
</tr>
<tr>
<td>(0.0002)</td>
<td>(0.00007)</td>
<td>(0.0002)</td>
<td>(0.00002)</td>
<td>(0.00002)</td>
<td>(0.0005)</td>
<td>(0.0006)</td>
<td>(0.00007)</td>
<td>(0.0001)</td>
<td>(0.00003)</td>
<td></td>
</tr>
</tbody>
</table>

*Notes: Country of birth, self-employment and municipality characteristics. Standard deviations in parenthesis’s. #coethnics represents the average number of coethnics in the arriving municipalities, in 1990 and 1991. Self-employed within 5 yrs is the arriving cohort that became self-employed.*
Figure 1. Size of ethnic enclave at arrival and municipalities with any self-employed from Iran within five years

(a) All coethnics at arrival
(b) Self-employed coethnics at arrival
(c) Any Self-employed within five years

Notes: Map of Sweden, with administrative boundaries of municipalities. In Sub-figures 1a and 1b, the municipalities are colored based on the number of coethnics or number of self-employed coethnics, living in the municipality. Only municipalities to which at least 1 individual born in Iran arrived to in 1990 and 1991, are colored. The coloring is based in quintiles or quartiles. Grey areas represent municipalities where no Iranians in my sample were placed 1990 or 1991. In the last Sub-figure (1c), a municipality is red if any individual born in Iran, who were placed in that municipality in 1990-1991, became self-employed over the next five years.

Source: GeoSweden (2017).

5 Results

I begin the section on results by presenting baseline estimates in section 5.1, showing a positive effect of self-employed coethnics on the probability to become self-employed. A large number of stability checks to make sure the results are stable can further be found in the Appendix (section B). In sections 5.2 and 5.3, I attempt to dig deeper into the story. Last, I provide a brief discussion on how the self-employed perform.
5.1 Baseline estimations

Table 6 shows the results when regressing a binary indicator of having positive business income at any point within five years after arrival on the municipality share of self-employed coethnics, living in the municipality of arrival and, the municipality share of all other coethnics. The main treatment variables are standardized\(^{23}\), hence the coefficient represents the effect of a standard deviation increase in the explanatory variable. Column (1) is a linear regression excluding all covariates and fixed effects, while column (2) adds individual controls as well as dummies for municipality of arrival and birth country by cohort. Standard errors are clustered on municipality and birth country level.\(^{24}\)

A first striking feature is that the estimations in column (1) and (2) are fairly stable with regards to the effect of self-employed coethnics. Adding covariates and fixed effects, changes the average effect very little. The effects are statistically significant, still at the 1 percent level when fixed effects are included. In the preferred specification, a standard deviation increase in the share of self-employed coethnics with business income gives a 2 percentage point increase in self-employment propensity. Given that only 4.4 percent of those who arrived in 1990-1991 had business income at some point within five years, the estimated effect is large (45 percent of the base-point). Here, it is important to keep in mind that around 25 percent of the refugees get placed with 0 self-employed coethnics, while only around 15 percent have a share of self-employed coethnics in their municipality of arrival which is higher than the standard deviation of (0.0003). A reasonable way to look at the treatment effect is therefore as an increase from a municipality with no or very little presence of self-employed coethnics, to a municipality with a large level of coethnics with a business. The coefficient hence reflect a large effect stemming from a fairly large treatment.

\(^{23}\)[X \sim (0, 1)].

\(^{24}\)In Table 6, I only include specifications using no covariates or all covariates and fixed effects. In the Appendix, in Table B.3, I show how the effect changes when adding different controls to the sample. As can be seen from this Table, results do not vary much by specification.
Table 6. Baseline estimations (c.f. Equation 11)

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S-E or not</td>
<td>S-E or not</td>
<td>S-E or not</td>
</tr>
<tr>
<td>Placement Policy Strategy</td>
<td>“OLS”</td>
<td></td>
<td></td>
</tr>
<tr>
<td># Self-employed Coethnics</td>
<td>0.0251***</td>
<td>0.0214***</td>
<td>0.0243***</td>
</tr>
<tr>
<td>(As share of municipality pop)</td>
<td>(0.00466)</td>
<td>(0.00477)</td>
<td>(0.00467)</td>
</tr>
<tr>
<td># Non-Self-employed Coethnics</td>
<td>-0.0161***</td>
<td>-0.0121***</td>
<td>-0.0203***</td>
</tr>
<tr>
<td>(As share of municipality pop)</td>
<td>(0.00393)</td>
<td>(0.00433)</td>
<td>(0.00388)</td>
</tr>
<tr>
<td>Observations</td>
<td>13,992</td>
<td>13,992</td>
<td>13,992</td>
</tr>
<tr>
<td>Mean Dep. Variable</td>
<td>0.044</td>
<td>0.044</td>
<td>0.044</td>
</tr>
<tr>
<td>Covariates and Fixed Effects</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

Notes: Baseline linear estimations regressing probability of self-employment within five years of arrival on the standardized share of coethnics with business income and municipality population share of all other coethnics in 1990 and 1991. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered on municipality and birth country level. Column (1) includes no covariates nor fixed effects and column (2) adds all covariates and fixed effects. Covariates on individual level include age, age², university education, sex, marital status, if the individual has children, number of children and if the individual moved within the first year of arrival or not. For exact specification of regression and covariates used, see equation 11 and section 3.2. In column (3) I regress the probability of self-employment on the standardized share of coethnics with business income and municipality population share of all other coethnics in 1995 (1996). The controls are the same, however municipality fixed effects are defined in 1995 (1996), rather than in the arrival year.

While the estimates using the share of self-employed coethnics are both sizeable and significant, the coefficient representing the quantity of all other coethnics is actually negative. This would imply that, given a certain share of self-employed coethnics, a larger share of other coethnics actually decreases the probability of self-employment. A standard deviation increase in the share of coethnics gives a significant drop in probability of self-employment with 1 percentage points. There are several possible interpretations of this coefficient, but most importantly, the overall number of coethnics does not seem to play a big role in understanding the connection between ethnic enclaves and self-employment in Sweden.

The estimates could be sensitive to many things. I attempt to account for these by re-estimating the baseline case using different techniques and samples. I use different time lags, different definitions of the treatment, including the absolute number of coethnics and an inverse hyperbolic sine transformation of coethnics, non-linear specifications (probit and logit), alternative definitions of the dependent variable and interaction
effects. I also test for the inclusion of additional control variables, including a quality indicator for the birth country at the municipality of arrival: the local employment rate within an ethnic group (see Table B.3). The overall conclusion from the sensitivity checks is that the positive effect from being placed with self-employed coethnics remains both positive and significant, while the effect of all, non-self-employed coethnics, stays significant and negative, or insignificant. The sensitivity checks are found in the Appendix, in section B.

Last, in addition to the preferred estimate in column (2), I add a third regression (column 3), which does not use the placement policy induced variation in 1990 and 1991. Instead, the enclave size is based on the municipality of residence in 1995 (cohort 1990) and 1996 (cohort 1991). This regression hence allows individuals to sort, and the size of the enclave will partly be a function of the individual selection on unobservables. I add this regression to get a better understanding of the importance of individual selection. What can be seen is that the effect of the share of self-employed coethnics is magnified a little, and become even more positive, whilst the effect of the share of other coethnics a lot more negative. Selection is hence more severe for the treatment using all non-self-employed coethnics. This is most consistent with coethnics selecting into areas based on labor market networks, and acting on information that could lead to non-self-employed labor, or, that individuals select into coethnic networks which hold alternative sources of support. I continue using the placement policy as my main strategy, but it can be noted that selection seems to be a larger problem for the treatment based on non-self-employed coethnics.\textsuperscript{25}

5.2 More on the channels

Moving on to the mechanisms, the baseline explanation is that meeting skilled coethnics, who have self-employment experience, matter for the tendency of newly arrived to become self-employed. This can be due to skill transfers or information on essential knowledge for running a business. While it is hard to exactly pin-point the importance of this

\textsuperscript{25}It’s important to note that the comparison between column (2) and (3) is not an exact one. In this case, the use of the treatment in 1990 and 1991 is based on port of entry treatment, while the design in column (3) reflect the contemporary effect several years after arriving. These effects are naturally not exactly comparable, and it is hence possible that part of the difference in coefficient size is a reflection of not only selection, but also the fact that the regressions are done under different contexts. Despite this weakness, the regression provides an interesting indication of selection.
story, I next provide a number of estimations to try to exclude alternative interpretations (see Table 7).

First, an important alternative story is that refugees could become self-employed because they lack skills required on the formal labor market, or because of discrimination. Note that this story requires discrimination that is specific to a country and a municipality. A hypothetical example would be if the situation is particularly difficult for Somalis in Gothenburg as compared to Somalis in Stockholm. A large share, or number, of self-employed coethnics could in this case reflect difficulties on the labor market.

Likely, if this mechanism is important, it should be reflected in a control variable measuring the share of coethnics who are employed in a specific municipality (I include such a specification in one of the sensitivity checks in Table B.3). I do however continue to test for this more extensively by looking at unemployed coethnics at arrival. I divide the number of unemployed coethnics with all coethnics, within a certain municipality. This will give a measure for how poorly a group of coethnics are doing, or how discriminated they are in their municipality of residence. Results for this can be found in column (1), Table 7. The coefficient implies that a standard deviation increase in the share of unemployed coethnics decreases the probability of self-employment with 0.002 percentage points. It therefore seems unlikely that the story of lack of formal requirements and discrimination is driving the baseline result.\footnote{Furthermore, in Table C.5, I redo the baseline regressions for different subgroups of the sample, most importantly including having a university degree or not. The estimates are based on education in 1995-1996, and show no statistical difference in effects between different levels of education. Hence, it does not seem that non-educated, who potentially lack relevant labor market skills, are reacting more to the effect of enclaves.}
Table 7. Redoing the baseline regressions (Table 6) using different definitions of the ethnic enclaves

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Unemployed Coethnics (As share of coethnic mun. pop.)</td>
<td>-0.000074 (0.000375)</td>
<td>-0.0000862 (0.000369)</td>
<td>-0.000478 (0.000399)</td>
<td>-0.000300 (0.00281)</td>
<td>-0.000375 (0.000457)</td>
<td>-0.000375 (0.000457)</td>
</tr>
<tr>
<td># Coethnics w. Capital Inc. (As share of mun. pop.)</td>
<td>-0.0000862 (0.000369)</td>
<td>-0.0000862 (0.000369)</td>
<td>0.000078 (0.000399)</td>
<td>0.000078 (0.000399)</td>
<td>0.000078 (0.000399)</td>
<td>0.000078 (0.000399)</td>
</tr>
<tr>
<td># Coethnics w. High Capital Inc. (As share of mun. pop.)</td>
<td>-0.0000862 (0.000369)</td>
<td>-0.0000862 (0.000369)</td>
<td>0.000078 (0.000399)</td>
<td>0.000078 (0.000399)</td>
<td>0.000078 (0.000399)</td>
<td>0.000078 (0.000399)</td>
</tr>
<tr>
<td># High Income Coethnics (As share of mun. pop.)</td>
<td>0.000300 (0.000399)</td>
<td>0.000300 (0.000399)</td>
<td>0.000300 (0.000399)</td>
<td>0.000300 (0.000399)</td>
<td>0.000300 (0.000399)</td>
<td>0.000300 (0.000399)</td>
</tr>
<tr>
<td># High Educated Coethnics (As share of mun. pop.)</td>
<td>0.0000375 (0.0000457)</td>
<td>0.0000375 (0.0000457)</td>
<td>0.0000375 (0.0000457)</td>
<td>0.0000375 (0.0000457)</td>
<td>0.0000375 (0.0000457)</td>
<td>0.0000375 (0.0000457)</td>
</tr>
<tr>
<td># Self-employed Coethnics (As share of mun. pop.)</td>
<td>0.0129*** (0.00446)</td>
<td>0.0129*** (0.00446)</td>
<td>0.0129*** (0.00446)</td>
<td>0.0129*** (0.00446)</td>
<td>0.0129*** (0.00446)</td>
<td>0.0129*** (0.00446)</td>
</tr>
<tr>
<td>Observations</td>
<td>12,590</td>
<td>7,090</td>
<td>7,090</td>
<td>13,992</td>
<td>13,992</td>
<td>13,992</td>
</tr>
<tr>
<td>Mean Dep. Variable</td>
<td>0.044</td>
<td>0.044</td>
<td>0.044</td>
<td>0.044</td>
<td>0.044</td>
<td>0.044</td>
</tr>
<tr>
<td>Covariates and Fixed Effects</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

Notes: Definitions of enclaves are: Column (1), unemployed coethnics as share of coethnic in municipality. Column (2), coethnics with capital income, and column (3) coethnics with high capital income, both divided by municipality population. “High” implies belonging to the top quartile of the source country income distribution in the country. Column (4) uses coethnics with high disposable income, and column (5) coethnics with a university education. Note that for column (3) and (4), only cohort 1991 is used, since capital income is not seen on individual level in the data in 1990. For more information on specification see Table 6.
Another explanation for the results is that it is not so much the knowledge related to the process of self-employment, but rather financial assets which is the main driver. Starting a business normally requires a certain amount of capital, which could be raised using the already self-employed coethnics. This explanation do, however, not fit with the results observed. In columns (2) and (3) of Table 7 I present results regressing having business income on, first, the share of the municipality population who are coethnics with any capital income and, second, the share of the municipality population who are coethnics, and belong to the 25 percent highest capital income earners among the coethnics (in the country). The first of these, using all with capital income, gives an insignificant, small, negative effect. For the top earners, the effect turns positive, but remains insignificant and very small in size. Generally, capital income among coethnics is therefore not a strong predictor of self-employment. Likely the reason is that many get financing from more conventional sources, most often banks. Such results have for example been provided in Eliasson (2014).

Last, other resources among coethnics could also matter. I therefore define two additional sets of high resource individuals using, first, disposable income and, second, university education. In the former case, I count the top quartile within the national distribution of a certain birth country to get the number of high income coethnics. In the latter, I simply count those with a university education. As in the baseline case, I further divide with municipality population. As can be seen from the estimates in column (4) of Table 7, it does not seem that arriving at a municipality with more of the richest coethnics causes a better chance for self-employment. On the other hand, there is a significant effect from living close to those with a university education. The effect is smaller than the baseline estimate of self-employed coethnics, but still economically significant. However, when adding a control for the standardized share of self-employed coethnics, the effect goes away (cf. column (6)).

5.3 Further evidence against the niche market channel
As hypothesized, if the probability of self-employment increases due to the presence of the sheer number of coethnics, this would be an indication of an ethnic (niche) market. Since I find no effect, a simple conclusion is that niche markets carry little importance in the case of Sweden. However, given partial effects running in different directions, I provide a couple of further indications speaking against the niche market story.
First, it could be that a niche market business is relevant to start only when none already exist. If a niche market becomes saturated quickly, just controlling for the number of self-employed coethnics might not be enough. In Table 8, column (1), I therefore include an interaction term, implying that I interact the number of self-employed coethnics at arrival with all other coethnics. Looking at the coefficients, there is no significant effect from the number of non-self-employed coethnics, when the number of self-employed coethnics are zero. While the sign of the coefficient has switched to positive, the effect is small and insignificant. Also, there is a significant effect of the number of self-employed coethnics, in cases when there are no other coethnics. This definitely points to a story were the niche market is of less importance. If businesses are started where there are few other coethnics, a niche market is of course highly unlikely.

Table 8. Effect of ethnic enclaves on self-employment using (i) interactions and (ii) language-based enclaves

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td># Non-Self-employed coethnics</td>
<td>9.25e-06</td>
<td>(9.25e-06)</td>
<td></td>
</tr>
<tr>
<td># Self-employed coethnics</td>
<td>0.000471**</td>
<td>0.000189</td>
<td></td>
</tr>
<tr>
<td>Interaction term</td>
<td>-1.24e-07*</td>
<td>(6.57e-08)</td>
<td></td>
</tr>
<tr>
<td>Non-Self-employed language coethnics (as share of mun. pop.)</td>
<td>0.00508</td>
<td>0.00774</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00348)</td>
<td>(0.00514)</td>
<td></td>
</tr>
<tr>
<td>Self-employed language coethnics (as share of mun. pop.)</td>
<td>0.0180***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00614)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>13,992</td>
<td>13,036</td>
<td>13,036</td>
</tr>
<tr>
<td>Mean Dep. Variable</td>
<td>0.044</td>
<td>0.044</td>
<td>0.044</td>
</tr>
<tr>
<td>Covariates and Fixed Effects</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

Notes: Column (1) interacts the absolute number of the two explanatory variables. Column (2) and (3) replicate the baseline regressions (Table 6), only the definition of the enclave is based on languages instead of birth country. The partition of the language groups is described in Table D.6. Note that column (2) includes only the use of all, non-self-employed coethnics as explanatory variable, and column (3) adds all self-employed coethnics. See Table 6 for more information on specification.

A second reason why we might not see much of a niche market effect, is that the definition of the ethnic enclave is too narrow. A niche market might for example have an Arabic base rather than a specific Iraqi base. In column (2) and (3), Table 8, I switch the definition of an ethnic group
and focus on language groups instead. The Swedish register data holds no information on spoken language, meaning that any information has to be inferred from the most spoken languages in the country of origin. I make a very strong assumption that all individuals from a certain country speak the largest native language. While this assumption is indeed highly restrictive, it is necessary to be able to do a comparison like this one at all. As an example, an individual arriving from Lebanon or Iraq will both be assumed to speak Arabic. I keep the sample countries as my observations, however, countries outside the sample enter the calculations through the size of the enclave. For example, while I have only Iranians in the sample, the size of the Persian enclave will consist of both Iranians and Afghans. The full definition of the language enclaves is found in Table D.6, in the Appendix. Looking at the signs and sizes of the estimates in Table 8, clearly, there is no detectable positive effect on self-employment from a larger enclave based on language.

Pure income or financial resources seem not to be the driving factor of the effect of the enclave, and the effect is not driven by an especially difficult situation on the local labor market. Furthermore, the number of coethnics, regardless of qualities, is not a predictor of the decision to become self-employed. Rather it seems to be specific knowledge on the self-employment process, which is the main driving mechanism. This is mostly consistent with a story which argues that meeting skilled coethnics is important because they provide and transfer knowledge and assistance important to start a business. Whether this is information about the legal framework, specific self-employment skills, language or other institutional knowledge is more difficult to pinpoint with the data at hand.

5.4 Extension: How do the self-employed perform?

Having shown a causal relation stemming from the share of self-employed coethnics to the probability of self-employment, a very important follow-up question relates to the performance of the self-employed. As noted in the first paragraph of this paper, physical segregation along ethnic lines is often discussed as a problem from the point of view of, for example, lack of access to host country skills. It is therefore important to investigate whether the channel of self-employment is a positive side of the residential concentration of coethnics, which then could counteract any negative effects of segregation.

I address this question using mediation analysis. In mediation analysis, the core idea is to move beyond the average causal effect, and
investigate possible mechanisms through so call mediators. In the case of enclaves, I ask; assuming that residential concentration of coethnics matters for economic outcomes, how much of that effect is mediated through the choice of self-employment?\textsuperscript{27}

I study the effect on disposable income over the following twenty year period for the sample of refugees who arrived in 1990 and 1991.\textsuperscript{28} The aggregated disposable income for the next twenty year period is assumed to be a function of the size of the ethnic enclave, and partly mediated through the choice of self-employment, which, as I have shown, is in turn a function of the enclave size. In essence, what is done to find the mediation effect is to fit the following two equations:

\[
y_{icmk} = \alpha_1 + \beta_1 \frac{SE \text{ Coethnics}_{cmk}}{Population_{mk}} + \beta_2 \frac{\text{non}SE \text{ Coethnics}_{cmk}}{Population_{mk}} + X_i \Gamma_1 + \sigma_{m}^{mediator} + \theta_{kc}^{mediator} + \epsilon_{mediator}^{icmk} \quad (13)
\]

\[
Income_{icmk} = \alpha_2 + \rho y_{icmk} + \beta_3 \frac{SE \text{ Coethnics}_{cmk}}{Population_{mk}} + \beta_4 \frac{\text{non}SE \text{ Coethnics}_{cmk}}{Population_{mk}} + X_i \Gamma_2 + \sigma_{m}^{outcome} + \theta_{kc}^{outcome} + \epsilon_{outcome}^{icmk} \quad (14)
\]

Equation 13 is the mediation equation, and furthermore exactly the same as the regression fitted in my baseline estimation. The choice of self-employment within five years of arrival ($y_{icmk}$) is the mediator, which is regressed on the treatment, the share of self-employed coethnics in the municipality of arrival, conditional on pre-treatment characteristics and fixed effects. The second equation, 14, gives the effect on the outcome of interest, in this case the aggregated disposable income over a twenty year period ($Income_{icmk}$). $\beta_3$ gives the direct effect of the enclave, conditional on the choice of self-employment, and the average mediation effect is given by the product of $\hat{\beta}_1$ and $\hat{\rho}$.

I run the model using the algorithm developed by Hicks and Tingley (2011).\textsuperscript{29} Important to note is that for the mediation effect to have a causal interpretation, two assumptions are needed. First, the

\textsuperscript{27}The mediation approach based on the potential outcome framework is discussed in Imai et al. (2010).
\textsuperscript{28}I drop the top 1 percent of income earners.
\textsuperscript{29}This algorithm fits the two equations 13 and 14, and then simulates parameter values to arrive at the estimated effects.
already discussed exogeneity assumption, which states that given pre-treatment characteristics and fixed effects, the size of the enclave is independent of the potential outcome of the mediator, that is the choice of self-employment. Second, given pre-treatment characteristics, fixed effects and the size of the enclave, the choice of self-employment is independent of the potential outcome; aggregated income over 20 years. The last assumption is naturally a very strong one. In fact, the underpinnings of the theoretical model laid out by Roy, is that individuals sort into self-employment based on expected income. Should there be any unobservable characteristics, driving both income and the choice of self-employment, estimates will be biased. The direction of the bias depends on what we choose to assume about the selection. Exactly what to expect here is not obvious. An analysis of Swedish male wage-earners showed that both individuals who earned below predicted income and wage earners earning above predicted income selected into self-employment (Andersson Joona and Waldensjö, 2013). I provide sensitivity checks below to further shed light on this.

Table 9 shows the average direct effect, the average mediation effect, and what percentage of the total effect which is mediated by self-employment. As can be seen, being placed in a municipality with a higher share of self-employed coethnics has a negative effect on the aggregate disposable income. Income is measured in 100 SEK, and the coefficient of -630 implies that a standard deviation increase in the share of self-employed coethnics decreases aggregate income with 63,000 SEK. This should be compared to an aggregate mean income over the full period of 2,1 million SEK, that is the effect constitute around 3 percent of the mean income. Of this, the choice of self-employment is associated with a lower income of around 5700 SEK, which in turn implies about 8.2 percent of the total effect of the enclave. There is hence an economic penalty connected to ending up in a municipality with more self-employed coethnics. If we believe the mediation estimates, this negative effect is not counteracted by the choice of self-employment.

30 Note that the analysis was conducted on Swedish-born individuals. It is possible the selection for foreign born functions differently.

31 A complication is that self-employment income is evidently easier to hide or mask than other labor income, since the latter is generally reported by a third party. To try and account for this I rerun the causal mediation analysis scaling the self-employment income. This is done by multiplying the part of the disposable income registered as self-employment income with 1.6 (1.6 is in accordance with methods used by Statistics Sweden). Rerunning Equations 13 and 14 with the new outcome gives a result with a similar conclusion as in Table 9. Both the direct effect of enclaves as well as the mediation of self-employment remain negative, although, as expected, the latter effect drops in size. Results are available upon request.
Table 9. *Causal mediation effect of self-employment on aggregate disposable income over twenty years (95 percent confidence intervals in parenthesis’)*

<table>
<thead>
<tr>
<th>AVERAGE EFFECTS</th>
<th>Disposable Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Causal Mediation Effect</td>
<td>-57.46</td>
</tr>
<tr>
<td></td>
<td>[-84.07, -33.42]</td>
</tr>
<tr>
<td>Direct Effect</td>
<td>-630.36</td>
</tr>
<tr>
<td></td>
<td>[-943.29, -296.48]</td>
</tr>
<tr>
<td>Total Effect</td>
<td>-687.82</td>
</tr>
<tr>
<td></td>
<td>[-997.45, -350.39]</td>
</tr>
<tr>
<td>% of Tot Effect Mediated through Self-Employment</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td>[.058, 0.16]</td>
</tr>
</tbody>
</table>

*Notes:* Causal mediation analysis. The two equations 13 and 14 are fitted. The algorithm thereafter follows the procedure set out in Hicks and Tingley (2011). The total effect represent the effect of the size of the ethnic enclave on aggregate disposable income, which is separated in a direct effect from the enclave and a mediated effect through the choice of self-employment. % of Tot Effect Mediated through Self-Employment is the share of the total effect given by the mediator. The top 1 percent of income earners are dropped from the sample.

Imai et al. (2010) has developed a sensitivity check to further analyse the importance of unobserved variation, in this case affecting both income and the choice of self-employment. The method uses the correlation between residuals in Equations 13 and 14 ($\epsilon_{\text{mediator}}^{icmnk}$ and $\epsilon_{\text{outcome}}^{icmnk}$). The more unobserved variation (and therefore selection into the mediator), the higher the correlation between the residuals. After fitting the mediation and the outcome equation, the method simulates parameter values of the mediation effect based on different values of the correlation between $\epsilon_{\text{mediator}}^{icmnk}$ and $\epsilon_{\text{outcome}}^{icmnk}$.
Figure 2. Average mediation effect for different values of selection

Notes: y-axis show average mediation effect, fitting the two equation in 13 and 14. The outcome is aggregated disposable income over a twenty year period in Sweden, the treatment is the share of self-employed coethnics in the municipality of arrival and the mediator is the choice of self-employment. The top 1 percent of income earners are excluded. x-axis show different values of correlation between the residuals of Equations 13 and 14. Average mediation effect simulated in accordance with method set out in Imai et al. (2010).

Source: GeoSweden (see Section 3 for further details).
In Figure 2, I plot the average mediator effect simulated under the different values of the correlation between the residuals. Income is again aggregated over a twenty year period. Correlation is simply defined as
\[ p = \text{corr}(\epsilon_{icmn}^{\text{mediator}}, \epsilon_{icmn}^{\text{outcome}}), \]
which implies that negative values simulates a situation where people with a higher income than predicted do not become self-employed. As can be seen from the figure, small negative correlations \((-0.06)\) flips the simulated average estimate to positive. In essence then, the estimate of a negative effect of self-employment depends on the importance of selection. This is an interesting follow-up question for future research. For now, the mediation estimates of self-employment at least show no indication of a positive counteracting effect to the negative effects of self-employment enclaves.

The negative effect is in line with more descriptive comparisons of self-employment the period in question. In Figure 3, I abstract from the placement policy and the size of the enclave, and divide the sample into three groups: 1) those who became self-employed at some point during their first five years, 2) those who had any non-self-employment taxable income within their first five years and 3) all the rest, which means the group that neither worked enough to get taxable income nor started a business. I show disposable income based on means, medians, the 25th percentile and the 75th percentile.
Figure 3. Disposable income 1991-2014 for self-employed, workers and unemployed.

(a) Mean Disposable Income

(b) Median Disposable Income

(c) Disposable Income, 25th percentile

(d) Disposable Income, 75th percentile

Notes: Disposable Income for the sample of refugees who arrived 1990-1991. The three lines represent three groups: one for those that had business income within the first five years, one for individuals who had other taxable labor income within the first five years, and one for those who had neither.

Source: GeoSweden (see Section 3 for further details).
The pattern of the figures show how the self-employed individuals dip in income relative to both those having worked anything over the first five years, but also relative to those having not worked at all, over a period of 1992 until slightly after 1995-1996. From this lower level the disposable income of the self-employed sample climbs in a higher pace than both the income of those who worked, as well as those who did not. Neither the mean nor the median development is, however, steep enough to reach the disposable income of those who worked. In fact, the average disposable income for the individuals with business income over the initial five year period is lower than that of the unemployed group still ten years after arrival.\footnote{A caveat is that low income among self-employed may be related to a higher living standard, as compared to similar income levels for the wage labor working individuals. I make no further adjustments to account for this, but it should be noted that income and living standards may differ in important ways (Johansson and Larsson, 2015).}

6 Conclusion

In this paper I estimate the causal effect of the size of the ethnic enclave in the port of entry municipality on the probability to become self-employed. I look at two cohorts of refugees, arriving in Sweden in 1990 and 1991, and use a spatial dispersal policy in place 1985-1994 to construct exogenous variation in sorting of the immigrants.

The results indicate that ethnic enclaves measured as the share of self-employed coethnics causes a higher chance of self-employment. A standard deviation increase in the municipality share of self-employed coethnics increases the probability of self-employment within five years with around 2 percentage points. Given that only 4.4 percent of the arriving cohorts chose self-employment within the time frame, the results are substantial. On the other end, being placed with a large number of coethnics, regardless of quality, have no, or even negative effects on the probability of self-employment.

The results are robust to the inclusion of a number of individual covariates as well as municipality and country by cohort fixed effects. The results are also robust to different specifications of the treatment, other definitions of the dependent variable, different time lengths, interaction effects, probit and logit specifications as well controlling for selection behavior of movers. I further rule out that the results are driven by coethnics with high income, capital assets, or low access to the local labor market, alternatively discrimination.
It hence seems ethnic enclaves do increase the probability of self-employment. This positive result aligns itself nicely with the so far mostly correlational papers studying the question. Also, results indicate that it is specifically the people with some detectable human capital, primarily owning a business but also having a university education, rather than coethnics in general, who are driving the results.

The estimates are hence first and foremost consistent with the story, which argues that information and knowledge from those already familiar with the market or on important institutions cause an increased entry into self-employment. This stands in contrast to the main alternative story; which promotes access to a large market of potential coethnic consumers or workers. In other words: quality seems to be more crucial than quantity.

While the share of self-employed coethnics at the municipality of arrival affects the probability of self-employment for refugees, on average, there is an income penalty from ending up in a larger enclave. Using causal mediation analysis, I show that there is a negative income effect from ending up with more self-employed coethnics, which is potentially further negatively effected by the choice of self-employment. The mediation analysis is however sensitive to selection into self-employment. Moreover, using descriptive statistics, I further find that those choosing self-employment during their first five years in the country, tend to stay economically behind for the full twenty year period, as compared to other employees. While it is therefore true that self-employment is fostered within enclaves, the outcome is not necessarily a success story.

An interesting follow-up task for research is to consider other outcomes and how they are mediated. Future projects should answer questions on when enclaves foster positive outcomes, and through which channels. Coming research also have an additional interesting task in recognizing what makes the firms successful, and particularly, if networking with the best firms also makes the newly arrived more successful.
References


Appendices

A Construction of the sample

As noted, I make use of GeoSweden, an individual level, full population database spanning from 1990-2014. The variables span from demographic and socioeconomic information to housing characteristics. Essential to this study are information on reported business income, labor market status, place of stay and immigration details, including country of birth and time of residence permit. Individuals are observed from the decision of residence permit, not before. It was also after the decision of residence permit that refugees were placed, making it the most relevant point in time. However, most individuals in the sample will likely have stayed in Sweden for several months before they are observed in the data. Each individual is further matched with a firm id from his or her largest source of income, making it possible to track not only if someone is registered as self-employed, but also characteristics of the firm. Place of stay is registered December 31 each year, but it is observable if an individual moved sometime during the year.

The sample of immigrants is constructed in the following manner: I start by including working age (18-55 years old) foreign born adults from identifiable countries\textsuperscript{33}, who arrived in Sweden 1990 or 1991. The database does not reach further back in time than 1990, which naturally makes this year a good starting point. The placement policy was in place until July 1994, however, I limit myself to 1990 and 1991. This limitation is made since previous research suggest the implementation was stricter during these years. According to Edin et al. (2003), between 1987 and 1991, 90 percent of refugees arriving were placed through the program.

It’s important to note that the exogenous source of variation, given by the refugee placement program, stems from the allocation of refugees.

\textsuperscript{33}Some origin countries are not visible in the data as single-country codes. As an example “Central-America” is its own coded entity, consisting of all the Caribbean and Central American states. Regional codes consisting of several countries of origin are not included in the sample. This is because cultural differences between the countries could imply that ethnic networks are less likely to evolve. It should, however, be noted that the number of individuals without single country coding is a very small share of all immigrants in a given year. In 1990, of all foreign born immigrants coming to Sweden, around 95 percent were born in an identifiable country.
The Swedish register data includes no information on \textit{allocated} place of stay, but rather the actual place of stay. Also, before 1997 the reason for immigration (type of residence permit) is unknown. Due to these limitations, several steps have to be taken to mimic the allocation of the program. First, to throw out any non-refugees, I restrict the sample to individuals who are neither immigrating from, nor are born in, countries which are members of OECD.\textsuperscript{34} Second, to get rid of any family migrants, I drop immigrants who at the time of arrival were married to someone who had already arrived in a prior year. This leaves me with individuals from 35 countries, of which many have very few observations, and several have no variation in the number of self-employed within five years. Since the estimation strategy will include fixed effects for birth country by cohort, and since I want to make sure that those included in fact were refugees, my last restriction limits the sample to the top ten sending countries 1990 and 1991. This accounts for more than 80 percent of the non-OECD sample. It further implies that no source country had fewer than 500 individuals arriving over the two years 1990 and 1991. The ten countries are seen in Table 1.

As a validation check, I can compare my countries to aggregate historical data available on the website of the Swedish Migration-board. This comparison will not be perfect: the GeoSweden data is based on country of birth, whilst the migration-board data is based on citizenship. In several cases these may not overlap. However, as a rough comparison, it is good to see that the numbers do not differ too much. In fact, on the top ten citizenships for asylum in Sweden 1990 and 1991, 9 of the countries are also in my sample. The only exception is that many Turkish citizens were granted asylum, a country that was excluded from my sample on the basis of being an OECD country, and that individuals with Romanian citizenship (included in my sample) were the 11th most likely group of refugees according to migration board data. One can further note that there were almost no guest students among the countries in the sample (See \url{https://www.migrationsverket.se/English/About-the-Migration-Agency/Facts-and-statistics-/Statistics/Overview-and-time-series.html}). Also, Migration-board statistics show that very few got labor market permits. While there were no permits for individuals from Somalia or Vietnam; individuals from Iran, Iraq, Lebanon, Ethiopia, Syria and Romania all received 3 or less work permits from the migration board per country (Note that labor market permits issued by the migration board in 1990-1991 are

\textsuperscript{34}The classification is based on the membership in 2016.
currently non-published. These numbers were sent to the author from the Migration Board, and are available upon request).

Assuming that the above made restrictions limit my sample to refugees, I am left with the problem of allocation. As noted above I observe the place of residence the last day of the year. Almost all refugees of course arrived earlier than the place of residence is observed in the data, and could therefore have moved in the course of the year. There is no clear cut solution to this problem, however, the data includes a variable counting the number of times an individual moves during a year. Should moving be structurally related to the self-employment decision and the size of the enclave, a simple solution is to add a dummy in the baseline regression, controlling for the choice of moving within the first year. I do this in the baseline estimations, and as it makes no difference to the estimates of the paper, I make no further adjustments.

B Stability of the estimates

B.1 Different time lags

The baseline section looked at the effect of self-employment at any point within five years. To make sure nothing special happened at this specific point in time, I further add a plot of coefficients based on 3-7 year lags. Unlike the baseline estimates these regressions use a dependent variable defined by being self-employed after 3-7 years, rather than within 3-7 years.

Figure B.1 shows the two main standardized coefficients including a 95-percent confidence interval. As can be seen, when using the share of self-employed coethnics the effect is positive for all year spans, with slightly increasing confidence intervals over time. The effect is significant on the 1 percent level for all years. The size of the effect changes very little, hence the effect is not isolated to a specific year. The size of the estimates are somewhat lower compared to the baseline estimates, but keep in mind that the mean of the dependent variable in this case will not be as large, given that I focus on the number of self-employed in a specific year. Looking at the coefficients for the share of non-self-employed coethnics, the estimates are all negative, similar to the baseline estimates.
Figure B.1. Baseline regressions, with time-spans of 3-7 years between treatment and outcome

Vertical axis showing coefficient estimates doing the same regression as in Table 6, column 2. The dependent variable now defined at the year in question. Note the difference to the baseline estimates, where estimates are done capturing all individuals who at any point up until the fifth year lag had business income. This figure shows the effect counting all that had business income after 3, 4, 5, 6 and 7 years. Estimates varies by years along x-axis: The label ‘3’ hence implies the coefficient when studying the outcome for the 1990 (1991) cohort in 1993 (1994). More information on specification of the regression is found in Table 6.
B.2 Definition of treatment and dependent variable, covariates and functional forms

Having studied the time dimension, I perform a battery of robustness checks to the baseline estimates below.

First, the baseline case used in Table 6 standardizes the explanatory variables with the municipality population. In Table B.1 I try out other functional forms. First, to take into consideration any extreme values or decreasing returns to scale\(^{35}\), I use an inverse hyperbolic sine function of the absolute number of self-employed coethnics and of all other coethnics. The hyperbolic sine function is given by: \(\ln(z + \sqrt{1 + z^2})\). The transformation has the nice feature of sustaining all zeros as zeros, while creating a log-like interval for the numbers larger than 0. More specifically, the transformation keeps values \(\approx 1\) close to the original value, while approaching a log approximation as the value increases. If large outliers with unusually many self-employed coethnics at arrival are driving the results, it should be picked up using this method.\(^{36}\) The effect is seen in column (1), and shows that the self-employed coethnics have a significant and positive effect, while no significant effect can be detected for the other coethnics.

In the second column of Table B.1, I use the absolute number of both my treatments, that is the pure number of self-employed coethnics and all other coethnics. As can be seen from the second column, using all coethnics produces a negative estimate and the number of self-employed coethnics produces a small non-significant estimate. However, just excluding outliers gives a different picture. In column (3) I therefore take away all individuals with more than 50 self-employed coethnics in the municipality of arrival, which represents around 10 percent of the full sample. There is now a significant effect, representing a coefficient somewhat smaller than the baseline. One additional self-employed coethnic increases the probability of self-employment with about 0.174 percentage points, or one standard deviation increase (6.1 individuals) gives a 1 percentage point rise in the probability of self-employment. Since the difference in effects is not very large, I proceed concluding that the baseline results holds even for other definitions of the treatment variable.

\(^{35}\)Consider that the first self-employed coethnic could be more important than the tenth.

\(^{36}\)The transformation of \(z\) gives approximately \(\ln(2) + \ln(z)\) for \(z > 2\). A more thorough discussion of properties can be found in Burbidge et al. (1988). Note also that other papers before this have included the transformation; see for example Hochguertel and Ohlsson (2009).
Second, as an additional test I further look into effects using a probit and a logit estimation, which both gives qualitatively similar results. These are found in Table B.2.

**Table B.1. Redoing baseline regression using different functional forms**

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S-E or not</td>
<td>S-E or not</td>
<td>S-E or not</td>
</tr>
<tr>
<td># Self-employed Coethnics</td>
<td>0.00943***</td>
<td>(0.00267)</td>
<td></td>
</tr>
<tr>
<td>(Inverse Hyperbolic sine)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># Non-Self-employed Coethnics</td>
<td>-0.00145</td>
<td>(0.00273)</td>
<td></td>
</tr>
<tr>
<td>(Inverse Hyperbolic sine)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># Self-employed Coethnics</td>
<td>0.000188</td>
<td>0.00156**</td>
<td></td>
</tr>
<tr>
<td>(Outliers Dropped)</td>
<td>(0.000121)</td>
<td>(0.000612)</td>
<td></td>
</tr>
<tr>
<td># Non-Self-employed Coethnics</td>
<td>-1.63e-06</td>
<td>-1.80e-05</td>
<td></td>
</tr>
<tr>
<td>(Outliers Dropped)</td>
<td>(6.59e-06)</td>
<td>(1.95e-05)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>13,992</td>
<td>13,992</td>
<td>12,611</td>
</tr>
<tr>
<td>Covariates and Fixed Effects</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

**Notes:** Baseline regression using different functional forms. Column (1) transforms the enclave definitions using an inverse hyperbolic sine transformation. Column (2) simply uses the absolute number of the enclaves (# Self-employed Coethnics and # Non-Self-employed Coethnics). Column (3) exclude all individuals who have more than 50 self-employed coethnics in their municipality of arrival (top 10 percent of sample). See Table 6 for information on specification.

**Table B.2. Regressing having business income or not on the standardized share of self-employed coethnics and standardized share of coethnics, using Probit (column 1 and 2) and Logit models (column 3 and 4)**

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds Ratios</td>
<td>Odds Ratios</td>
<td>Odds Ratios</td>
<td>Odds Ratios</td>
</tr>
<tr>
<td># S-E Coethnics</td>
<td>0.239***</td>
<td>0.232***</td>
<td>0.500***</td>
<td>0.480***</td>
</tr>
<tr>
<td>(As share of mun. pop.)</td>
<td>(0.0433)</td>
<td>(0.0499)</td>
<td>(0.0911)</td>
<td>(0.102)</td>
</tr>
<tr>
<td># Non-S-E Coethnics</td>
<td>-0.183***</td>
<td>-0.157***</td>
<td>-0.400***</td>
<td>-0.321**</td>
</tr>
<tr>
<td>(As share of mun. pop.)</td>
<td>(0.0539)</td>
<td>(0.0589)</td>
<td>(0.118)</td>
<td>(0.128)</td>
</tr>
<tr>
<td>Observations</td>
<td>13,992</td>
<td>12,769</td>
<td>13,992</td>
<td>11,818</td>
</tr>
<tr>
<td>Covariates and FE</td>
<td>NO</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
</tr>
</tbody>
</table>

**Notes:** Probit and logit estimations. Coefficients represent odds ratios. *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered on municipality and country group level. See Table 6 for information on covariates.
Third, in the baseline case, Table 6, I included two specifications, one without any covariates and fixed effects, and one with all of the preferred controls from the preferred specification. In Table B.3, I show several specifications, including different combinations of controls. In general what can be seen is that fixed effects for birth country is important for the size of the coefficient, but that the size of the main estimates varies little over specifications. In the last column I further add a quality control for the birth country at the municipality of arrival, that is the employment rate for coethnics at municipality level. Clearly this does not matter for the coefficient size.

A last robustness check is to consider a change in the dependent variable. So far I have used a definition relying on whether or not an individual has business income. Another way to capture this would be to use the definition provided by the Swedish registers, where an individual is counted a self-employed if being registered as one in a given point in time in November each year. I do this for the dependent variable in Table B.4. The definition is based on being registered as self-employed in 1995 or 1996. As can be seen from the coefficients, the estimates remain almost the same as in the baseline case.
### Table B.3. Baseline regression, including different sets of covariates and fixed effects

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-E or not</td>
<td>0.0251***</td>
<td>0.0265***</td>
<td>0.0175***</td>
<td>0.0176***</td>
<td>0.0214***</td>
<td>0.0215***</td>
</tr>
<tr>
<td>(0.00466)</td>
<td>(0.00464)</td>
<td>(0.00453)</td>
<td>(0.00453)</td>
<td>(0.00477)</td>
<td>(0.00473)</td>
<td></td>
</tr>
<tr>
<td># Self-employed Coethnics (As share of municipality pop)</td>
<td>0.0251***</td>
<td>0.0265***</td>
<td>0.0175***</td>
<td>0.0176***</td>
<td>0.0214***</td>
<td>0.0215***</td>
</tr>
<tr>
<td>(0.00466)</td>
<td>(0.00464)</td>
<td>(0.00453)</td>
<td>(0.00453)</td>
<td>(0.00477)</td>
<td>(0.00473)</td>
<td></td>
</tr>
<tr>
<td>S-E or not</td>
<td>-0.0161***</td>
<td>-0.0149***</td>
<td>-0.00973**</td>
<td>-0.00998***</td>
<td>-0.0121***</td>
<td>-0.0121***</td>
</tr>
<tr>
<td>(0.00393)</td>
<td>(0.00403)</td>
<td>(0.00378)</td>
<td>(0.00377)</td>
<td>(0.00433)</td>
<td>(0.00429)</td>
<td></td>
</tr>
<tr>
<td># Non-Self-employed Coethnics (As share of municipality pop)</td>
<td>-0.0161***</td>
<td>-0.0149***</td>
<td>-0.00973**</td>
<td>-0.00998***</td>
<td>-0.0121***</td>
<td>-0.0121***</td>
</tr>
<tr>
<td>(0.00393)</td>
<td>(0.00403)</td>
<td>(0.00378)</td>
<td>(0.00377)</td>
<td>(0.00433)</td>
<td>(0.00429)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>13,992</td>
<td>13,992</td>
<td>13,992</td>
<td>13,992</td>
<td>13,992</td>
<td>13,909</td>
</tr>
<tr>
<td>Ind Controls</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Country FE</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>Cohort by Country FE</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Municipality FE</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Country group Controls</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
</tr>
</tbody>
</table>

Notes: Baseline Regression estimated with different set of controls. Regarding covariates: on country level I include share of employed coethnics as share of total amount of coethnics. For notes on specification and variables see Table 6.
### Table B.4. Estimates changing the definition of the dependent variable, using definition from labor market survey rather than having business income or not

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registered as Self-employed</td>
<td></td>
<td></td>
</tr>
<tr>
<td># Self-employed Coethnics (As share of municipality population)</td>
<td>0.0108***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00390)</td>
<td></td>
</tr>
<tr>
<td># Non-Self-employed Coethnics (As share of municipality population)</td>
<td>-0.00662**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00308)</td>
<td></td>
</tr>
<tr>
<td>% Self-employed Coethnics (Inverse Hyperbolic sine functional form)</td>
<td>0.00457**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00191)</td>
<td></td>
</tr>
<tr>
<td>% Non-Self-employed Coethnics (Inverse Hyperbolic sine functional form)</td>
<td>0.000548</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00193)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>12,461</td>
<td>13,909</td>
</tr>
<tr>
<td>Mean Dep. Variable</td>
<td>0.027</td>
<td>0.027</td>
</tr>
<tr>
<td>Covariates and Fixed Effects</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

**Notes:** Estimations changing the definition of the dependent variable, using definition from labor market survey rather than having business income or not. *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered on municipality and country group level. See Table 6 for information on covariates.

### C Heterogeneity

In Table C.5 I redo the baseline analysis (Table 11), separated for 8 subgroups; university educated, those with less than nine years of education, men, women, married, non-married, parents and non-parents. A couple of interesting conclusions can be made. First, there is no distinguishable difference between educational groups, which implies at least no clear selection of low or high educated into self-employment. Second, the story holds for men, and men only. It should however be recalled that a big majority of the self-employed are men (more than 80 percent in the sample). Follow-up studies on networks among female and male self-employed are hence of great interest.
Table C.5. Heterogeneity estimations. Regressing probability of self-employment on the share of self-employed coethnics and all other coethnics, separately based on education, sex, marriage status and parenthood.

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>University</td>
<td>&lt; 9 years</td>
<td>Men</td>
<td>Women</td>
<td>Married</td>
<td>Non-married</td>
<td>Parents</td>
<td>Non-parents</td>
</tr>
<tr>
<td></td>
<td>educ. (t+5) of educ.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># Self-employed Coethnics</td>
<td>0.0246**</td>
<td>0.0231***</td>
<td>0.0381***</td>
<td>0.00155</td>
<td>0.0134***</td>
<td>0.0276***</td>
<td>0.0172**</td>
<td>0.0219***</td>
</tr>
<tr>
<td>(As share of municipality pop.)</td>
<td>(0.0115)</td>
<td>(0.00605)</td>
<td>(0.00785)</td>
<td>(0.00332)</td>
<td>(0.00468)</td>
<td>(0.00769)</td>
<td>(0.00693)</td>
<td>(0.00563)</td>
</tr>
<tr>
<td># Non-Self-employed Coethnics</td>
<td>-0.0220*</td>
<td>-0.00817</td>
<td>-0.0196***</td>
<td>-9.43e-05</td>
<td>-0.00452</td>
<td>-0.0190***</td>
<td>-0.00803</td>
<td>-0.0136**</td>
</tr>
<tr>
<td>(As share of municipality pop.)</td>
<td>(0.0116)</td>
<td>(0.00570)</td>
<td>(0.00648)</td>
<td>(0.00324)</td>
<td>(0.00434)</td>
<td>(0.00692)</td>
<td>(0.00627)</td>
<td>(0.00554)</td>
</tr>
<tr>
<td>Observations</td>
<td>3,018</td>
<td>6,597</td>
<td>8,755</td>
<td>5,237</td>
<td>7,461</td>
<td>6,531</td>
<td>5,040</td>
<td>8,952</td>
</tr>
<tr>
<td>Covariates and Fixed Effects</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

Notes: Heterogeneity results using subgroups of the cohorts. p<0.01, ** p<0.05, * p<0.1. Standard errors clustered on municipality - country group level. For more information see Table 6.
### D Categorization of languages

**Table D.6. Languages and country of birth groups**

<table>
<thead>
<tr>
<th>Language groups (in the sample)</th>
<th>Birth Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serbo-Croatian</td>
<td>Former Yugoslavia, Bosnia, Serbia, Croatia, Macedonia</td>
</tr>
<tr>
<td>Arabic</td>
<td>Somalia, Lebanon, Syria, Iraq, Tunisia, Morocco, Algeria, Egypt</td>
</tr>
<tr>
<td>Persian</td>
<td>Iran, Afghanistan</td>
</tr>
<tr>
<td>Amharic</td>
<td>Ethiopia</td>
</tr>
<tr>
<td>Romanian</td>
<td>Romania</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>Bulgaria</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>Vietnam</td>
</tr>
</tbody>
</table>
References


III. Migrating natives and foreign immigration: Is there a preference for ethnic residential homogeneity?

With Heléne Berg and Matz Dahlberg

*Acknowledgments:* The authors are grateful to Leah Platt Boustan, Florian Morath, Albert Saiz, Matti Sarvimäki, Susanne Urban and seminar participants at UC Irvine, UCLA, ETH/KOF in Zurich, Uppsala University, University of Verona, Statistics Norway, the 2015 IIPF conference in Dublin, Norface meeting at Rosersberg castle, UCFS meeting at Krusenberg, UCLS meeting in Uppsala, and the 2017 Urban Economic Association Meeting in Vancouver for helpful comments and discussions.
1 Introduction

Over the last decades, many European and other Western countries have witnessed increased immigration, with a drastic culmination in 2015; in this year alone, UNHCR estimated that around 1 million individuals reached the shores of Europe after having crossed the Mediterranean. In the wake of this experience, heated discussions have emerged on how and where to accommodate all refugees. In particular, a major political concern is the emergence of ethnically segregated neighborhoods. Aside from immigrants tending to one another, such a development is reinforced if the native population reacts by leaving or avoiding neighborhoods that become more ethnically diverse. The extent to which natives do so is the topic of this paper.

We study the migration behavior of the native population—here, native Swedes—when new immigrants arrive. We hypothesize that this may be manifested either in the form of native flight (immigration inducing natives to move out of a neighborhood), or in the form of native avoidance (immigration inducing natives to avoid moving into a neighborhood where more immigrants settle). Ultimately, the aim is to deduce from estimated migration responses whether natives prefer ethnically homogeneous neighborhoods. We approach this task by developing the so-called “shift-share” approach into, in several ways, a much improved identification strategy.

In order to create good policies to combat segregation, it is important to know both if natives change their behavior following immigration and, if so, why they do so. The maintained hypothesis in the literature on “white flight” is that migration responses are due to preferences for ethnically homogeneous neighborhoods (see, e.g., Saiz, 2007; Boustan, 2010; Saiz and Wachter, 2011; Sá, 2014). But newly arrived immigrants hold a number of different characteristics other than their ethnicity; the average refugee do, for example, typically have a lower education level and lower income than the native population. Which trait do the natives actually react on? Do they react on the ethnicity of the immigrants, as typically hypothesized in the earlier literature, or on the socio-economic part?¹

Thanks to comprehensive, detailed register data, we contribute in this paper by, aside from studying the if, examining the validity of the presumed ethnicity channel in a way that earlier literature has not been able to do. In particular, our data allows us to identify natives with

¹The data used in the paper allows us to observe country of birth and country of emigration. We do however not hold any data on self-proclaimed ethnicity, and therefore use source country to proxy for ethnicity.
different parental foreign background. Because many native-born individuals with non-Western parents are ethnically quite similar (in terms of country of origin) to current immigrants, yet in many cases socio-economically more similar to native-born individuals with Swedish-born parents, we use the parental information to explore the validity of the ethnicity channel. Consequently, by estimating the migration response of natives conditioning on their parents’ country of birth, we can examine whether there is support for the hypothesis that residential preferences are formed along an ethnic dimension.

Our paper also contains several methodological improvements. Our data holds information on each individual immigrant’s reason for residence permit—whether or not he or she arrived as a refugee, a tied mover or a labor migrant. We focus on refugees, which, besides being a highly topical and interesting group to study, arguably is more exogenous to the characteristics of the receiving city or neighborhood, as compared to labor, student or family migration. Note that much of the previous literature on white flight has focused on the US, in which often all immigration has been used as one common treatment. The generalization of all immigration as one concept could imply that we lose important nuances regarding mechanisms, but it might also come with issues from an empirical point of view. It can be argued that a larger share of immigrants to the US are pulled to specific places, whereas refugees, tend to have been pushed from their home country by wars and other catastrophes. Increases in the U.S type immigration could therefore, to a larger extent, be a function of regional chocks and pull factors, which affects both immigrants but potentially also the behavior of natives.

The information on reason for immigration is a unique feature of our data and constitutes a first distinct methodological improvement to related

---

2 Refugees in our paper includes all asylum related residence permits, most importantly “Geneva convention” refugees (in which case there is an individual reason for asylum) as well as those given protection due to conflicts and war.


4 According to the Migration Policy Institute, only 13 percent of all new U.S green card holders in 2016 were refugees, while almost half of all new permanent residents were refugees in Sweden (see https://www.migrationpolicy.org/article/frequently-requested-statistics-immigrants-and-immigration-united-states and https://www.migrationsverket.se/English/About-the-Migration-Agency/Facts-and-statistics-/Statistics/Overview-and-time-series.html).

5 Zimmermann (1996) provides a stylized economic definition of push and pull migration.

6 Consider for example a case where native US citizens increasingly appreciate Japanese food and culture. This could attract more Japanese into the States, while also making natives more inclined to live in Japanese-dense neighborhoods.
studies. Furthermore, this information is available on an annual basis, which allows us to exploit changes in refugee migration by incorporating neighborhood fixed effects.

We identify the causal effect of foreign immigration on the residential choice of natives by combining contemporary refugee migration into Sweden with previous immigrant settlement patterns resulting from a refugee placement policy that was in place in the earliest years of our study period. In short, the policy meant that refugees were not allowed to decide for themselves where to settle, but were assigned to a municipality by the Migration Board. We argue that this policy-generated settlement is yet another improvement to existing studies. The rationale for this is that settlement patterns of immigrants from the early 1990s, who subsequently attracted more recent push-driven refugee migrants, are more likely to be uncorrelated with neighborhood characteristics that matter for natives’ residential preferences than what would have been the case in the absence of the policy.

Ultimately, we construct an instrumental variable for changes in immigration based on the interaction of, on the one side, immigrant settlements during the placement policy era and, on the other, the timing of contemporary, refugee-driven immigrant shocks. Arguably, this results in an improvement to the typical shift-share instrument used earlier in the literature, where both initial immigrant settlement as well as contemporary immigration are left as is and therefore likely to be endogenous to the outcome (see, e.g., Altonji and Card, 1991; Card and DiNardo, 2000; Saiz, 2007; Carl and Siegenthaler, 2013; Chalfin and Levy, 2013; Sá, 2014).

A final contribution is that we acknowledge that, due to various constraints, far from all individuals are able to react on their residential preferences following an increase in immigration. Consequently, we focus the analysis on households characterized as having a high possibility to move. This is in accordance with assumptions made in theoretical models on the effects of immigration on native migration—but is a previously neglected aspect empirically—and it turns out to matter greatly for the results.

Many of the value-added features in this paper can only be implemented thanks to the detailed information in the data we have access to (GeoSweden). GeoSweden is a database that covers the full Swedish

---

7It can be noted that we have also applied the version of the shift-share instrument suggested by Jaeger et al. (2018); the two approaches provide very similar results.
8In our setting, we will define mobile actors as home owners rather than renters, as the rental market is characterized by long (sometimes extreme) queues, and renters in many cases compete for the same apartments as the newly arrived refugees.
population since 1990, and some of the valuable aspects of the data have already been covered; in particular, that we on a year-to-year basis can identify those immigrants that are granted a residence permit in Sweden based on refugee reasons, as well as those natives who indeed are mobile. Second, for each immigrant living in Sweden, there is information on the country of origin. Last, all variables come as an annual panel covering relatively small neighborhoods. While the panel structure allows for the fixed effects as mentioned above, the fine geographical resolution means that we can capture more nuanced residential preferences, as we are able to observe relatively short moves which are likely less costly than migration between larger units such as metropolitan areas.

Apart from the literature that directly estimates the extent to which the residential choices of natives are affected by immigration, our paper is related to an influential literature that has indirectly studied the response of natives to increased immigration by estimating effects on house prices (Saiz, 2007; Saiz and Wachter, 2011; Sá, 2014) and wages (Card, 1990; Altonji and Card, 1991).

The paper is also closely linked to the tipping-point literature that estimates at which potential share of immigrants in a neighborhood or a city the native population disproportionately starts to leave (Schelling, 1971; Card et al., 2008; Aldén et al., 2015). We instead focus on continuous native migration. However, to relate the results to previous work, we do also provide a set of tipping point-type estimates where we condition on initial share of immigrants. Finally, complementing the studies of the effects of residential segregation (Edin et al., 2003), our focus is on effects of immigration on residential segregation.

We reach four main conclusions. First, we do not find any evidence of neither native flight (i.e., that natives move out of neighborhoods following an increase in immigration to that neighborhood) nor native avoidance (i.e., that natives avoid moving into a neighborhood following an increase in immigration to that neighborhood) when studying the full population, regardless of tenure type.

Second, we do find that distinguishing between households with high/low possibility to move following an increased immigration is important; when studying the group of natives identified as indeed having

\[9\] In addition to the papers in the economics literature referred to above (e.g. Card, 1990; Altonji and Card, 1991; Saiz, 2007; Boustan, 2010; Saiz and Wachter, 2011; Sá, 2014), a substantial body in the sociology and geography literature studies this phenomenon; see Rathelot and Safi, 2014 and the references therein.
high possibilities to move, we estimate significant flight responses.\textsuperscript{10} We do, however, not find any effects of increased immigration on natives’ migration behavior for those identified as having a low possibility to move. This pronounced flight effect in the subsample of mobile natives potentially has implications for the interpretation in existing, related studies that—much due to data limitations—only have looked at aggregate (average) effects.

Third, we find that all natives, irrespective of their parents’ foreign background, react similarly to increased immigration. The likely interpretation of this is that a preference for ethnically homogeneous neighborhoods is not the dominant channel causing flight. Instead, further analyses indicate that natives have preferences for socio-economically homogeneous, or “better” neighborhoods.

Finally, conditioning on the initial immigrant share and thereby relating to tipping point estimates, we again find similar patterns irrespective of the natives’ parental foreign background. This is thus further evidence against the ethnicity channel, indicating that the tipping point literature might have focused on the wrong trait.

In the next section, we describe recent immigration patterns to Sweden. Section 3 then discusses the theoretical mechanisms through which we hypothesize that these patterns affect natives’ migration response, and, in particular describes our idea for examining whether there is any support for the ethnicity-based mechanism. While Section 4 lays out the strategy used to estimate these responses empirically, section 5 presents the data used to obtain the main results, which are provided in Section 6. Finally, we conclude.

2 Immigration to Sweden

The size and character of immigration to Sweden have changed over the last decades. In 1970, less than seven percent of the Swedish population were born in another country\textsuperscript{11}, and of those the large majority had arrived as labor immigrants from another Nordic or European country in the 1950s and 1960s. From the late 1970s/early 1980s, the immigration changed character; going from being mainly labor-induced, more refugees started to come. Consequently, there has since then been a

\textsuperscript{10}It is noteworthy that we find evidence of native flight, but not native avoidance. A possible interpretation is that natives mostly notice and consequently react on increased immigration into the neighborhood where they currently live.

\textsuperscript{11}Statistics Sweden, Yearbook of Sweden 2012, table 4.30 ”Population by country of birth”.
drastic change in both the number and the origin of the foreign-born population in Sweden. The changing pattern of the foreign born-population is clear from Figure 1. While the share with roots in the Nordic countries is decreasing over time, the share originating from non-European countries is increasing. In 1950, the approximately 200,000 foreign-born individuals living in Sweden constituted around 2.8 percent of the total population of around 7 million. By the end of 2017, the approximately 1,900,000 foreign-born individuals constituted more than 18 percent of the total population of around 10 million. More than half of these are born outside of Europe.

Figure 1. Number of foreign-born in Sweden by region of origin, 1950–2014.

Notes: Y-axis in units of thousands.
Source: Statistics Sweden.

Compared to most other European countries, Sweden has a relatively larger share of foreign-borns. According to statistics from Eurostat,12 in 2010, 47 million individuals in the EU 27-countries were not born in the country in which they resided. This amounted to almost ten percent of the total population. The majority of these, slightly more than 31 million, were born outside of the European Union. There is however a large variation in these numbers across the EU, ranging from Poland (with 1.2

percent foreign-born), Czech Republic, Hungary and Finland (all with around 4 percent foreign-born) to Austria (15.2 percent), Sweden (14.3 percent), Spain (14 percent) and Germany (12 percent).

Switching focus from stocks to flows, the annual immigration to Sweden during the period that we study, 1990–2010, is shown in Figure 2. Up until 2006, typically 50–60,000 individuals came each year. Then, from 2006 and onward, there has been a discrete increase in the number of immigrants, with a yearly average of around 100,000.

Figure 2. Total immigration to Sweden, 1990–2010

Source: GeoSweden (see Section 5 for further details).

3 Potential reactions of natives

The literature on residential segregation typically studies two types of reactions of the majority population to immigration of minorities: flight (immigration inducing the majority population to move out of a neighborhood), and avoidance (immigration inducing the majority population

---

13 The spike in the early 1990s is due to increased refugee immigration following the Balkan war, and the increase in 2006 is primarily related to an escalation of the Iraqi war.
to avoid moving into a neighborhood).\textsuperscript{14} For the analysis in this paper, it is necessary to distinguish between “native” and “white”. The concepts of native flight and avoidance are different from white flight and avoidance. The latter stems from a US tradition of research on the effects of racial diversity. Primarily due to a different data practice in how to classify individuals’ background, rather than focusing on racial diversity, we will study flight and avoidance due to increased diversity in terms of country of origin. Consequently, we refer to the potential reaction of the majority population as native flight and avoidance.

Our main definition of native is everyone born in Sweden. This means that our native group is quite heterogeneous in terms of their parental foreign background, a feature which we use in an attempt to disentangle the mechanisms behind the observed migration responses of natives. We continue with explaining this in more detail below.

3.1 Preference-based mechanisms
Why would increasing immigration affect natives’ location decisions? Scholars within sociology, economics and geography have lifted several potential mechanisms, where the dominating one is related to preferences for racial and/or ethnic homogeneity. Primarily sociologists have used attitude surveys to document racial and ethnic preferences. These might take the form of strict preferences for living with co-ethnics, or of aversion against perceived social unrest (Farley et al., 1978, 1994). Economists have incorporated this thought into their models by introducing a parameter capturing “distaste for immigrants” (or analogously, “preference for homogeneity”). An illustrative example is the set up in Sá (2014), where the preferences of the native population are modeled as:\textsuperscript{15}

$$U_{n,i} = V_{n,i} + f(h, x) - \delta I,$$

where $V_{n,i}$ measures the value individual $n$ attaches to the local amenities in neighborhood $i$, $f(h, x)$ is a function measuring utility from consumption of housing services ($h$) and of other goods ($x$), and $\delta$ captures natives’ preferences for immigrants $I$. The mobility response of na-
atives to immigration is derived by maximizing the utility function in (1) subject to the relevant budget constraint. This yields the intuitive prediction that native flight will increase if natives have a preference for homogeneity/a distaste for immigration (i.e., in terms of the model, if $\delta > 0$).

But what is the interpretation of the parameter $\delta$? Does it measure natives’ preferences for ethnicity, or their preferences for other traits that the newly arrived immigrants carry? In Sweden, newly arrived immigrants are to a large extent refugees. Particularly in the first years in the country, the average refugee has lower income and is less educated than the native population in the neighborhoods in which they locate. If natives have preferences for neighborhoods with homogeneous (high) levels of income and/or education, the change in the socio-economic composition in the neighborhood resulting from especially refugee immigration may drive native out-migration. In other words, if natives experience that the neighborhood status is dropping due to increased immigration, then observed native flight/avoidance might in fact be economic flight/avoidance.\footnote{We refer to this channel as preferences for homogeneity along the socio-economic dimension. Because refugees generally have lower socio-economic status, this is (empirically) equivalent to preferences against a lower composition of socio-economic traits. As an illustration of refugees generally having lower socio-economic status, we note from our GeoSweden data that the median refugee did not have any earned income in the first year after arrival.}

That immigrants’ socio-economic status might matter for natives’ locational decisions has of course been discussed earlier in the literature, see e.g. Boustan (2010); Saiz and Wachter (2011); Rathelot and Safi (2014); Sá (2014). Probably due to data restrictions, it has however never really been examined. Here, we contribute by disentangling this socio-economic channel from the commonly assumed ethnic channel, by using the detailed information in the Swedish register data about the foreign background of the parents of the native born.

In Sweden, the native-born individuals represent many different ethnic backgrounds on the parental side; some have Swedish-born parents, others have parents born in another Western country, and still others have parents born in non-Western countries who mostly arrived as refugees (or tied family members to refugees) before having children.\footnote{See https://www.migrationsverket.se/English/About-the-Migration-Agency/Facts-and-statistics-/Statistics/Overview-and-time-series.html for information on number and type of residence permits per country of origin from 1980 and onward.} Assume that ethnicity is the only characteristic among the new immigrants’ that matters for the migration decision of the natives—that is, natives have
a strong preference for ethnically homogeneous neighborhoods—so that \( \delta \) captures this dimension only (call it \( \delta^{\text{Ethnicity}} \)). Then we would expect the following hypotheses to hold true:

\[
\delta^{\text{Ethnicity}}_{\text{Swedish Parents}}, \delta^{\text{Ethnicity}}_{\text{Western Parents}} > \delta^{\text{Ethnicity}}_{\text{Non-Western Parents}}
\]  

That is, the mobility response within the group of natives who are (on average) ethnically more dissimilar to the newly arrived refugees (the native-born individuals with Swedish- and other Western-born parents) will be greater than the response within the group of natives who are (on average) ethnically more similar to the newly arrived refugees (the native-born individuals with parents born in a non-Western country). If there is a strong preference for ethnic homogeneity, we therefore expect \( \delta^{\text{Ethnicity}}_{\text{Non-Western Parents}} \) to be smaller than \( \delta^{\text{Ethnicity}}_{\text{Swedish Parents}} \). By relating our empirical results to the different \( \delta \)-coefficients in equation (2), we can examine the validity of the ethnicity-based channel vs. the socio-economic one.

3.2 Non-behavioral mechanisms

Aside from the two preference-based channels, there are non-behavioral mechanisms to consider. First, immigration may lead to changes in house prices that in turn may induce native flight and avoidance. Bousstan (2010) explains this clearly; in investigating historical white flight within the US, she sets up a model where house prices are a function of the number of inhabitants. Assuming an inelastic housing supply, immigration will initially cause prices to rise. Since locational decisions are likely to be affected by house prices, this will induce movement from the current population. Under such a scenario, part of the observed flight is therefore due to price increases rather than to behavioral effects induced by the preferences or the perceptions of the native majority. A similar reasoning can be found in for example Saiz (2007).

There is also the possibility of a reverse price effect, if the neighborhood status is (perceived to be) dropping with increased immigration. This could induce home owners who are worried about falling house prices to leave. However, the housing stock in high-immigration neighborhoods is typically characterized by a large share of rental apartments (see Section 5), and because the Swedish rental market is highly regulated, immigration cannot affect rental prices, neither up nor down. This is particularly true in the short-run perspective that our analysis take
(we consider native migration within one year of additional foreign immigration). Ultimately, we thus expect these non-behavioral mechanisms via house price changes to be rather small in the current setting. At the very least, they should not differ between the groups of natives with different parental background, meaning that the relative importance of preferences along the ethnic vs. socio-economic dimension can still be assessed as laid out above.

In addition to price effects, given that housing supply is not perfectly elastic, there is also a “mechanical effect” to consider. In the extreme case when housing supply is perfectly inelastic, irrespectively of residential preferences, a person can only move into a neighborhood if someone else has moved out. Thanks to the high frequency in our data, we are more or less able to rule out this mechanical effect for the case of flight; we know the place of residence on December 31\textsuperscript{st} of the year for each individual living in Sweden at that point in time. We thus observe immigrants as well as natives registered in a particular neighborhood on that very date, and can therefore with fairly good precision measure only native outflow that takes place after the arrival of new immigrants. This means that our measure of native flight is net of any such potential mechanical effect.

For the case of avoidance, however, no matter the data frequency, it is not possible to completely rule out that measured native avoidance is mechanically driven by a fixed housing supply. Specifically, when a person moves into a neighborhood where housing supply is fixed, there is one less apartment/house available for everybody else. Even if a native was contemplating moving there, the possibility might then not exist. This should, however, at most imply a (negative) 1:1 relation, meaning that we can rule out larger negative effects than that as being solely driven by such a mechanical effect.

3.3 Possibility to move
A prerequisite for deducing residential preferences from flight and avoidance estimates due to any mechanism is that people indeed are mobile. We recognize that this is far from true for everyone, meaning that some groups may not be able to react on their residential preferences. A contribution of our paper is that, to our knowledge, we are the first to take such mobility constraints into account.

The particular reason why some individuals cannot easily move depends on the institutional setting. In the current context, mobility constraints of individuals renting rather than owning their homes are likely
to be especially pronounced as a consequence of increased immigration. Several factors lead to this conclusion. First, renters are often resource constrained. Many renters are therefore referred to other rental apartments, should they wish to move. Second, municipalities are responsible for accommodating newly arrived refugees who are not able to find a place on their own. Usually this is done through municipality-owned rental apartments. These apartments make up a majority of the rental market and, in turn, a relatively large part of the total housing market. Access to these public (and often also to private) rentals requires queuing, in many municipalities for several years (or even decades, as in the case of Stockholm). This is true also for those already living in public rentals.

Given fixed short-run housing supply, competition for rental apartments is accentuated in high-immigration municipalities. In other words, a year with unusually large immigration leads to unusually few rental apartments available. Ultimately, this implies that, following increased immigration, moving to a new neighborhood will be particularly difficult for individuals living in rentals. As a last piece, this reasoning is conditional on that most individuals prefer to stay within the same municipality. Two reasons speak for this: First, long-distance moves are significantly more costly, not the least from a labor market point of view. Second, rental housing queues in Sweden are in almost all cases municipality based. A move to a new municipality may hence imply loosing a lot of queuing points.

To take mobility/budget constraints into account, based on the reasoning above, we focus much of the empirical analysis on the group owning their homes. Note that this is not to say that renters in general are less mobile. Rather, the combination between most immigrants occupying rental apartments, and the non-renter market being inaccessible for (budget constrained) renters, creates a municipality lock-in effect, where years of unusually high migration is followed by unusually few rental apartments available.

To sum up the discussion in section 3, if we observe substantial native flight among those that have a high possibility to move, this is most likely driven by preferences against living in an ethnically diverse neighborhood and/or in a socio-economic diverse neighborhood. The same is true for observed native avoidance larger than a (negative) 1:1 relation. Furthermore, if natives with varying parental foreign background react to a similar extent, this suggest that preferences are formed along

---

18 As documented in Andersson et al. (2010).
19 Although under certain circumstances, so-called “switching contracts” where two renters change apartments with one another can be approved.
socio-economic dimensions and, thus, that preferences for ethnically homogeneouse neighborhoods are (at most) of second order.

4 Econometric strategy

This section covers our econometric approach; we discuss the general set-up, the identification strategy, and our improvement compared to the earlier literature.

4.1 General set-up

Let us begin by defining native outflow, $outflow_{i,t}$, as the number of natives who leave neighborhood $i$ in year $t$. Analogously, we define native inflow, $inflow_{i,t}$, as the number of natives who move into $i$ in year $t$. In other words, $outflow_{i,t}$ is the number of natives who lived in $i$ in $t-1$ but lives in another neighborhood in $t$, whereas $inflow_{i,t}$ is the number of natives who did not live in $i$ in $t-1$ but does so in $t$. The two variables $outflow_{i,t}$ and $inflow_{i,t}$ are our main outcome variables, and our two parameters of interest are $\beta_{out}$ and $\beta_{in}$ in the following two equations:

\[
outflow_{i,t+s} = \alpha_{out} + \beta_{out} im_{i,t} + \epsilon_{out_{i,t+s}} \quad (3)
\]
\[
inflow_{i,t+s} = \alpha_{in} + \beta_{in} im_{i,t} + \epsilon_{in_{i,t+s}} \quad (4)
\]

where $im_{i,t}$ is the number of new immigrants in neighborhood $i$ in year $t$. Recalling the discussion from the previous section, we predict the following of $\beta_{out}$ and $\beta_{in}$:

**Empirical predictions.** If increased immigration cause...

... native flight, then $\beta_{out} > 0$.

... native avoidance, then $\beta_{in} < -1$

The geographic location of immigrants is not random, but might rather be correlated—either directly or via some unobserved neighborhood characteristic—with our outcome of interest, native migration. In other words, there is an endogeneity problem that must be solved. To

\footnote{Note that for the natives’ responses, we only consider migration within the country (i.e., not emigration responses).}
identify $\beta^{\text{out}}$ and $\beta^{\text{in}}$, we will use an instrumental variable that we consider substantially improves on the instruments typically used earlier in the literature (the so called shift-share instrument; see Altonji and Card, 1991, for the first use of this instrument). In short, the improvement is mainly attributed to two factors. First, we only consider refugee migration, arguably providing more exogenous variation in immigration than when conﬂated with other migration. Second, we make use of a Swedish refugee placement policy that was in effect in the early part of the period that we study, arguably generating a more exogenous historical allocation of immigrants than when they self-select the place of residency.

In the following, we discuss the general shift-share approach and our improvements to it.

4.2 Identification: Interaction between push-driven immigration and a historical placement policy

The instruments used in the earlier literature to solve the endogenous location choice of immigrants typically follow the shift-share strategy (see, e.g., Altonji and Card, 1991; Card and DiNardo, 2000; Saiz, 2007; Sá, 2014). The strategy builds on the observation that new immigrants tend to be drawn to places where former immigrants sharing their background have already settled. The idea is to instrument $im_{i,t}$ with the prediction $\tilde{im}_{i,t}$, defined as (exemplified by immigration to Sweden):

$$\tilde{im}_{i,t} = \sum_c \tilde{im}_{c,i,t} = \sum_c \left( \phi_{c,i,t^0} \times im_{c,SWE,t} \right),$$

where

$$\phi_{c,i,t^0} = \frac{im_{c,i,t^0}}{im_{c,SWE,t^0}}$$

is the fraction of immigrants from source country $c$ that arrived in Sweden and settled in neighborhood $i$ in some baseline period $t^0$. $im_{c,SWE,t}$ represents total immigration to Sweden from source country $c$ in year (or period) $t$. The instrument $\tilde{im}_{i,t}$ defined in equation (5) thus measures the contemporary immigration that would have been the result had the settlement of these immigrants and those who came in the baseline period been the same.
To implement the shift-share approach, source country $c$ and baseline period $t^0$ must be chosen, and it is in these two decisions that our methodological improvement lies. Next we discuss these two aspects in turn.

Definition of source country

In previous research, which mainly focus on US and UK data, typically all immigration has been used in the analyses. This can be problematic for a number of reasons.

First, given our aim to separate between ethnically and socio-economically motivated flight and avoidance, source country of the immigrants is a major concern. If native flight occur due to an increase of individuals from geographically and culturally distant nations, but not due to immigration from more similar countries, it would be more reasonable to believe the story is of ethnic nature.

Second, as we will use neighborhood fixed effects, identification in our shift share setting will come from variation within neighborhoods over time. The distribution of immigrants in the baseline years, $\phi_{c,i,t^0}$ only vary across neighborhoods, which means that identification of $\tilde{im}_{i,t}$ from $\tilde{im}_{i,t}$ requires inflow of immigrants from country $c$ to Sweden ($im_{c,SWE,t^0}$) to be time-varying. If the country-specific change in yearly inflow of immigrants is not large enough, we won’t be able to separate the predicted neighborhood level immigration in $t$ from that in $t + 1$. Now, country specific flows of refugees changes heavily from year to year. A hypothetical increase in a conflict or in the level of oppression should also increase the number of asylum seekers over time, sometimes with drastic changes from year to year. It is likely to be expected that other forms of migration, such as labor or student migration is somewhat more consistent over time.\(^{21}\)

Third, as noted, compared to other reasons for international migration, we also argue that the settlement of refugees is less driven by different pull factors of the neighborhood. This is not to say that destination countries of refugees are random, or that pull factors of certain countries do not enter the decision margin of refugees. However, we do claim that pull-factors affecting other forms of migration (e.g., work- or study-related migration) to a larger extent are related to city or neighborhood features in the destination country. For refugee migration, pull factors are more correlated with the liberal stand of asylum policies, which are national rather than local. In short, we thus argue that refugee migra-

\(^{21}\)This is at least the case in the Swedish setting, were large spikes or changes over time generally are related to changes in refugee migration (see for example Figure 2).
tion is more exogenous to neighborhood characteristics than other forms of migration.

Refugee migration constitutes a large part of the immigration to Sweden, implying that much of it is indeed push-driven. A unique feature of our data is that it includes all immigrants’ reason for immigration\(^{22}\) to Sweden; to work, as a refugee, as a tied family member or other reasons. By singling out refugees, we can restrict the analysis to push-type immigration driven by exogenous shocks.

The information on reason for immigration is available from 1997, and our period of analysis is 1997–2010. Individuals entering Sweden with refugee status during this period arrive from all source countries, but we drop those from OECD countries, since it is less likely that we observe flight from migration from for example Germany or Denmark. Also, many of these are likely Dublin cases with citizenship from other countries. We further drop Egypt and Eritrea. There are no/only 30 individuals arriving from Egypt/Eritrea in the baseline period,\(^{23}\) implying that \(\phi_{c,i,t,0}\) in equation (5) is not defined/will be highly imprecise. From the remaining source countries, at least 100 individuals or more arrived in the baseline period. The full list of these 34 countries and the frequency of refugees arriving in 1997–2010 are available in Table B.4 in the Appendix.

**Definition of baseline period**

As seen in Equation 5, the yearly national inflow of refugees from country \(c\) is scaled by the neighborhood share of immigrants in the baseline year, from the same country. Since the scaling is based on historical behavior, a remaining problem for identification is if the historical immigrant settlement patterns were guided by (unobserved) sticky or fixed factors that are correlated with natives’ migration decisions still today.\(^{24}\)

This is a problem that is left unsolved in the existing migration literature applying the shift-share approach, and one of our methodological improvements is to exploit a refugee placement policy that was in effect in Sweden from the beginning of 1985 to mid-1994. During this period, refugees could not decide themselves where to settle, but were assigned

\(^{22}\)Grund för bosättning in Swedish.

\(^{23}\)For definition of baseline period, see the next section.

\(^{24}\)Note that this is different from the problem of dynamic effects over time. The latter problem would happen in cases were immigration causes flight in the baseline period, which in turn sets a long term response in motion, that might still be in the process of evolving in the later year we are predicting. This problem has been discussed and addressed by Jaeger et al. (2018). In the Appendix we include a sensitivity check, which shows that the interpretation of our results remain when using the model suggested in Jaeger et al. (2018).
to a municipality through municipality-wise contracts, coordinated by the Immigration Board.\textsuperscript{25} The number of municipalities that had such a contract increased over time, and by 1991, 277 out of 286 were part of the program.

One of the main aims of the refugee placement program was to break the concentration of immigrants to larger towns (mainly Stockholm, Gothenburg and Malmö) and, instead, to achieve a more even distribution of refugees over the country. This aim was successfully fulfilled, as illustrated for example in Figure 3B in Dahlberg et al. (2012) and Table 1 in Edin et al. (2004).

Motivated by this, we choose for our baseline period \( t^0 \) the early years in our data in which the refugee placement program was in place, 1990–93 (our data starts in 1990). We think that this adds credibility to the instrument since, thanks to the placement program, the immigrant settlement pattern across neighborhoods back then is less likely to be driven by endogenous factors that also affect the migration pattern of natives following contemporary immigration increases (compared to a situation in which the policy had not existed). This is especially true conditional on neighborhood fixed effects and a set of neighborhood characteristics that we include in our estimation model. Put differently, it is not necessary that the program-generated placement of refugees across municipalities was random.\textsuperscript{26} What we argue is rather that, since the refugees that the municipalities received were effectively assigned to a specific apartment rather than choosing themselves where to live, conditional on a set of characteristics, the variation in immigration to a neighborhood \textit{within} a given municipality is likely to be exogenous to contemporaneous native flight and avoidance.\textsuperscript{27}

\textsuperscript{25}They were, however, allowed to move after the initial placement.

\textsuperscript{26}In fact, it was not entirely random, but rather determined by for example available housing (Dahlberg et al., 2012) and even party constellation in the municipal council (Folke, 2014). For a lengthier discussion of the exogeneity of the placement program with respect to municipal characteristics, we refer to Dahlberg et al. (2012).

\textsuperscript{27}A couple of caveats are to be noted here: First, for the years 1990–93, we have no information on reason for immigration. Instead, we use \textit{all immigrants} from the refugee-countries defined in the later time period \( t \). Unfortunately, this is a limitation since it was only refugees who were placed through the program. Also, there is evidence that the placement program became less strict and encompassing after 1992, mainly due to an unexpected and large increase in immigration from former Yugoslavia. We have still opted for using the years up to 1993 and to make no further restrictions to approximate who might have been a refugee. The core reason is due to the efficiency of the prediction. To be able to make a reasonable prediction we opt for maximizing the number of observations; when we estimate the first stage with only the years 1990 and 1991 we still get significant point estimates, but the instrument is not as powerful in terms of F-statistics. It can also be noted that when we apply the
We now proceed by specifying the details of our proposed estimation model, including the neighborhood characteristics upon which we condition the exogeneity assumption.

4.3 Estimation model
We analyze panel data, where the year of refugee immigration, \( t \) in equations (3) and (4), refers to years 1997–2009, while the year of migratory response by natives takes place in \( t + 1 \), implying that the effects are estimated for the years 1998–2010.\(^{28}\)

Besides instrumenting \( \text{im}_{i,t} \) with \( \tilde{\text{im}}_{i,t} \), our final estimation model differs from the basic equations in (3) and (4) in a few ways. First and most importantly, the panel structure of the data means that we can include neighborhood fixed effects,\(^{29}\) \( \mu_i \), and thereby exploit changes in immigration shocks within neighborhoods over time. Second, we include linear, quadratic and cubic controls for population size (\( \text{pop} \)) in \( t - 1 \). The purpose of these are to flexibly control for the fact that, in absolute terms, larger neighborhoods typically experience larger immigration inflows as well as larger population turnover in general. Third, since immigration of refugees could be correlated with immigration for other reasons, which in turn could lead to further migratory responses, we control for all non-refugee immigration from the refugees’ source countries in year \( t - 1 \).\(^{30}\) Fourth, we include time fixed effects to control for aggregate shocks that affect all neighborhoods in the same way in a given year. Finally, we control for a set of time-varying socio-economic characteristics of the neighborhood (measured in \( t - 1 \)); average disposable

---

\(^{28}\)We focus on the short-term perspective of one year because, at least in a quantitative sense, the estimated effects of immigration become less reliable the longer the native response is allowed to take. The reason is that immigration during and post year \( t \) is likely to be correlated, implying that native migration measured later may either be longer-run responses to immigration in year \( t \), or short-run responses to immigration after year \( t \).

\(^{29}\)A neighborhood is defined as a so-called SAMS; see the following section.

\(^{30}\)The main worry is that tied family migration arrives to the same neighborhoods as the refugees, causing an additional effect on native migration. Since we primarily worry about tied migration, we control for other types of migration only from the refugee countries we use to construct \( \text{im}_{i,t} \). We have however estimated a model with all other immigration as a covariate, with no important alterations to the baseline estimates. These results are available upon request.
income, the number of students, the per capita cost of social assistance and the number of public rental estates.\textsuperscript{31}

Letting the vector $\mathbf{X}$ include the variables for non-refugee immigration and the socio-economic characteristics, the first stage in our IV approach is:

$$im_{i,t} = \gamma \hat{im}_{i,t} + \sum_{p=1}^{3} \phi^p pop_{i,t-1}^p + \Gamma \mathbf{X} + \mu_i + \tau_t + \epsilon_{i,t}$$  (7)

The prediction $\hat{im}_{i,t}$ from this first stage is then used in the two equations capturing the migratory response of the native population:

$$outflow_{i,t+1} = \beta^{out} \hat{im}_{i,t} + \sum_{p=1}^{3} \delta^p pop_{i,t-1}^p + \Pi \mathbf{X} + \mu_i + \tau_t + \epsilon_{i,t+1}$$  (8)

and

$$inflow_{i,t+s} = \beta^{in} \hat{im}_{i,t} + \sum_{p=1}^{3} \delta^p pop_{i,t-1}^p + \Pi \mathbf{X} + \mu_i + \tau_t + \epsilon_{i,t+s}$$  (9)

Our approach thus identifies effects on native migration of immigration, within neighborhoods, over time, where the immigration stems from the interaction between contemporary year to year changes in the total inflow of refugees from specific countries and placement policy-induced immigrant settlement in a neighborhood several years before.

5 Data and descriptive statistics
In this section we present the data, which is obtained from the GeoSweden database, and how we define our key concept of “neighborhoods”. All data is collected and made anonymous by Statistics Sweden.

5.1 The GeoSweden database
The data used for the analysis comes from the database GeoSweden, which is administered by the Institute for Housing and Urban Research

\textsuperscript{31}The reason we date all variables in $t-1$ is to avoid a bad control problem—that is, that we control for things that are in fact responses to/implications of immigration.
at Uppsala University. The database, which is collected on a yearly basis, covers all individuals living in Sweden and is very comprehensive. It contains variables from several different registers such as the education, the income and the employment registers, and it contains information on individual characteristics such as year and country of birth, marital status, the number of children in the household, as well as the individuals’ level and type of education. It also contains pre-tax income from different sources, disposable income as well as various variables concerning the individual’s employment.

What is of extra importance for this paper is that the database includes detailed geographical information on where the individuals live, information on the date, from which country, and for what reason an individual immigrates to Sweden, along with annual information on migration patterns within Sweden.

We define a neighborhood to be a so-called SAMS (Small Areas for Market Statistics). A SAMS is a geographical unit that Statistics Sweden has defined to obtain a countrywide division of municipalities into homogeneous areas. Sweden consists of approximately 9,200 SAMS with an average population of around 1,000 individuals. In our sample, we have excluded SAMS that were not tractable throughout the study period, or that lack population at some point in time. This leaves us with 8,723 neighborhoods. The average number of SAMS per municipality is around 30 and the number of neighborhoods per municipality is highly correlated with the population of the municipality. We analyze the sensitivity of the first stage to the type of SAMS in Section 6.1.

5.2 Descriptives
Table 1 provides summary statistics of the variables used in the analysis, along with a clarifying description. As already noted, the average SAMS-neighborhood has around 1,000 inhabitants, but it spans between the very small places with only a couple of individuals to the large neighborhoods in inner Stockholm with around 20,000 inhabitants. Around 85 natives on average moves out of a neighborhood any given year, which represents about 8 percent of the population.

For the main endogenous immigration variable as well as its instrument (corresponding to \(im_{i,t}\) and \(\hat{im}_{i,t}\) in the above equations), the standard deviations are large relative to their means. This reflects the fact that roughly 85 percent of the observations contain zeros, which in turn is because many SAMS are very small. To get a sense of the variation in the data conditional on having positive immigration, Figure
3 shows the conditional distribution of these two immigration variables. As can be seen, the majority of neighborhoods have a fairly low level of immigration. Half the neighborhoods received 3 people or less, while 90 percent received 14 or less. The figures also suggest that the two distributions are highly correlated. This is indicative of a strong instrument, and we show below that this is indeed the case.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Key variables:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outflow</td>
<td>114,477</td>
<td>85.2</td>
<td>118</td>
<td>0</td>
<td>2,352</td>
</tr>
<tr>
<td>Inflow</td>
<td>114,477</td>
<td>85.2</td>
<td>121</td>
<td>0</td>
<td>2,716</td>
</tr>
<tr>
<td>Immigration (main)</td>
<td>114,478</td>
<td>0.82</td>
<td>4.7</td>
<td>0</td>
<td>313</td>
</tr>
<tr>
<td>Predicted immigration</td>
<td>113,503</td>
<td>0.81</td>
<td>3.6</td>
<td>0</td>
<td>251</td>
</tr>
<tr>
<td><strong>Control variables:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>114,478</td>
<td>1019</td>
<td>1236</td>
<td>1</td>
<td>20,285</td>
</tr>
<tr>
<td>Students</td>
<td>114,478</td>
<td>53.1</td>
<td>107.5</td>
<td>0</td>
<td>2,642</td>
</tr>
<tr>
<td>Disposable income</td>
<td>114,478</td>
<td>155,838</td>
<td>538</td>
<td>-107,050</td>
<td>5,688,067</td>
</tr>
<tr>
<td>Social assistance</td>
<td>114,478</td>
<td>8,700</td>
<td>22,500</td>
<td>0</td>
<td>108,200</td>
</tr>
<tr>
<td>Other non-OECD imm.</td>
<td>114,478</td>
<td>2.4</td>
<td>9.4</td>
<td>0</td>
<td>590</td>
</tr>
<tr>
<td>Public rentals</td>
<td>114,478</td>
<td>2.1</td>
<td>5.2</td>
<td>0</td>
<td>408</td>
</tr>
</tbody>
</table>

Outflow and Inflow measure the number of natives moving out of and into a given neighborhood in a given year. Immigration (main) is the main endogenous independent variable, measuring the annual number of refugees, and Predicted immigration is the instrument for this variable. Population denotes total SAMS population and students the number who receive some student contributions (majority of Swedish students). Disposable income and Social assistance are measured in SEK, other non-OECD immigration shows the number of non-refugee immigrants and Public rentals is the number of public rental estates. The unit of observation is SAMS-by-year, and the time span is 1997–2010.
Figure 3. Distribution of actual and predicted immigration

(a) Actual number of immigrants

(b) Predicted number of immigrants

Note: The figures show the cumulative distribution of immigration, actual (panel a) and as predicted by the instrument (panel b), conditional on positive immigration. The unit of observation is SAMS-by-year, and the time span is 1997–2010.

Source: GeoSweden.
6 Results

We now turn to the results. After establishing in Section 6.1 that our instrumental variable works well in the first stage regression, we provide the IV-estimates of the effects of foreign immigration on native migration in Section 6.2. By focusing on home owners, we study households that indeed have a fair possibility to move following increased immigration. Ideas about mechanisms are discussed and tested in Section 6.3, renters are studied in section 6.4 and we end with relating our results to the tipping point literature in section 6.5.

6.1 First stage

Table 2 shows the baseline estimation of the first stage as specified in equation (7); for the years 1997–2010, the inflow of refugees to neighborhood $i$ in year $t$ is regressed on the inflow as predicted by equation (5). An estimate of 1 implies perfect correlation; that is, a prediction based on the interaction of previous settlement patterns and current shocks of one more immigrant into neighborhood $i$ in year $t$ corresponds to an actual inflow of one more immigrant to that very neighborhood in that year. Because treatment is defined at the level of SAMS-by-year, our default is to cluster the standard errors at SAMS.$^{32}$

Column 1 presents raw correlations, while column 2 adds fixed effects and control variables according to the preferred model, based on the discussion in Section 4.3. We see in the latter that, conditional on last years’ population, socio-economic and demographic characteristics of the neighborhood, non-refugee immigration, as well as year and neighborhood fixed effects, one additional predicted immigrant is associated with 0.6 actual immigrants. The coefficient is highly significant, such that the instrument clearly fulfills the relevance condition. The model is also very stable; adding all the control variables, including the fixed effects, does not affect the estimate much (from 0.67 to 0.61; cf. columns 1 and 2).

$^{32}$For robustness, we have also re-estimated the model clustering at municipality.
Table 2. *First-stage estimates*

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No controls</td>
<td>Baseline specification</td>
</tr>
<tr>
<td>im</td>
<td>0.674***</td>
<td>0.608***</td>
</tr>
<tr>
<td></td>
<td>(0.0808)</td>
<td>(0.0859)</td>
</tr>
<tr>
<td>Observations</td>
<td>113,503</td>
<td>104,251</td>
</tr>
<tr>
<td>Number of SAMS</td>
<td>8,731</td>
<td>8,710</td>
</tr>
<tr>
<td>F-stat</td>
<td>69.55</td>
<td>49.98</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.170</td>
<td>0.170</td>
</tr>
</tbody>
</table>

*** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered on SAMS-level. Column 1 estimates the unconditional correlation, while column 2 estimates the first-stage according to the preferred specification as described in Section 4.3; it includes year and neighborhood fixed effects as well as linear, quadratic and cubic controls for population size, non-refugee immigration from the refugees’ source countries, average disposable income, the number of students, the per capita cost of social assistance and the number of public rental estates. All covariates are at the neighborhood level and measured in year \( t-1 \).

As a robustness check, in Table 3 we run the first stage for several different subsamples. First, we remove the 10 percent of neighborhoods with the smallest population (less than 322 individuals) and the largest population (more than 2043 individuals), respectively; see columns 1–2. Clearly, the estimations are more dependent on the larger neighborhoods. This is expected, as immigration is more consistent over time to larger neighborhoods. The coefficient is however highly statistically significant in both subsamples.

Gothenburg—the second largest city in Sweden—with its almost 800 neighborhoods is a clear outlier; very few municipalities have over 100, and Stockholm—the capital—has less than 200. We therefore exclude Gothenburg, with no big change in either power or significance; see column 3. Last, it is interesting to see how the first stage depends on the number of immigrants. Because the majority of neighborhoods in a typical year did not receive any refugees, dropping the top 10 percent of the distribution of immigrated refugees (as in columns 1–2 for population) would be too much of a restriction. Instead, we drop the top 10 percent of the sample, *given positive immigration*. In practice this implies any neighborhood receiving more than 14 immigrants. Just as when dropping neighborhoods with large populations, the first stage drops in power but, again, it is still highly significant; see column 4.

The first stage can be concluded as strong. The baseline estimate implies that an increase of 1 predicted refugee to a neighborhood is
associated with 0.6 more actual refugees to the very same neighborhood. It is highly stable for the inclusion of fixed effects as well as several control variables. It is also robust to the exclusion of segments of the sample, although the prime part of the variation is identified through larger neighborhoods.

Table 3. **Robustness of the first-stage estimate over different subsets of the sample**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Excl. least pop. n’hoods</td>
<td>Excl. most pop. n’hoods</td>
<td>Excl. Gbg</td>
<td>Excl. n’hoods with most im</td>
</tr>
<tr>
<td>im</td>
<td>0.622***</td>
<td>0.251***</td>
<td>0.608***</td>
<td>0.140***</td>
</tr>
<tr>
<td></td>
<td>(0.0882)</td>
<td>(0.0292)</td>
<td>(0.0896)</td>
<td>(0.0134)</td>
</tr>
<tr>
<td>Observations</td>
<td>78,997</td>
<td>93,635</td>
<td>95,327</td>
<td>103,168</td>
</tr>
<tr>
<td>Number of SAMS</td>
<td>6,770</td>
<td>7,927</td>
<td>7,960</td>
<td>8,709</td>
</tr>
<tr>
<td>F-Stat</td>
<td>49.84</td>
<td>74.13</td>
<td>45.96</td>
<td>108.16</td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td>0.186</td>
<td>0.0422</td>
<td>0.173</td>
<td>0.0392</td>
</tr>
</tbody>
</table>

*** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered on SAMS-level. Column 1 excludes neighborhoods in the bottom decile of the population size distribution, column 2 excludes neighborhoods in the top decile of the population size distribution, column 3 excludes Gothenburg (Gbg) and column 4 excludes neighborhoods in the top decile in the distribution of received immigrants (given positive immigration). See Table 2 for details of the estimated model.

6.2 Native flight and avoidance: Average effects

Moving to the estimated native flight and avoidance effects, Table 4 presents results from estimating the second-stage equations of outflow and inflow, as specified in (8) and (9), respectively. Any native residing in neighborhood \( i \) on the last day of \( t \), but living in another neighborhood \( -i \) on the last day of \( t+1 \) is counted as outflow from \( i \), while any native residing in neighborhood \( i \) on the last day of \( t+1 \) but in another neighborhood \( -i \) on the last day of \( t \), is counted as inflow into \( i \).
Table 4. Second-stage estimates of native flight and avoidance

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All natives</td>
<td>Home owners</td>
</tr>
<tr>
<td>OUTFLOW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>im</td>
<td>0.0645</td>
<td>0.347**</td>
</tr>
<tr>
<td></td>
<td>(0.158)</td>
<td>(0.156)</td>
</tr>
<tr>
<td>INFLOW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>im</td>
<td>-0.0847</td>
<td>0.170</td>
</tr>
<tr>
<td></td>
<td>(0.183)</td>
<td>(0.137)</td>
</tr>
<tr>
<td>Observations</td>
<td>104,250</td>
<td>104,250</td>
</tr>
<tr>
<td>Number of SAMS</td>
<td>8,710</td>
<td>8,710</td>
</tr>
<tr>
<td>Mean of Dep. Variable</td>
<td>85</td>
<td>39</td>
</tr>
</tbody>
</table>

*** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered on SAMS-level. Column 1 includes all natives and column 2 is restricted to native home owners. See Table 2 for details of the estimated model.

The left column of Table 4 includes all natives and shows neither signs of flight nor avoidance, as both coefficients are small and statistically insignificant. This null result is interesting, as it differs from previous literature despite the clear analogy of being based on the full population irrespective of their actual possibilities of moving. Capturing residential preferences through flight and avoidance is however only possible if people indeed are mobile. This is true in any institutional setting, although what determines mobility varies. In Sweden, as explained above, renting rather than owning your home constitutes a significant obstacle to moving—and especially so in a situation when the municipality has received and accommodated many immigrants.

As a way of getting closer to residential preferences and reactions to increased immigration, in the right column (and for much of the remaining analyses) we therefore restrict the sample to natives owning their home. It then becomes clear that the insignificant aggregate effects mask interesting heterogeneity. In particular, the estimated outflow effect among natives home owners is a statistically significant 0.35. The interpretation of this coefficient is that, when a neighborhood receives one more immigrant than on average, 0.35 additional natives move out.

In contrast to outflow, there is no statistically significant inflow effect of increased immigration among home owning natives. A possible explanation is that home owners mostly notice and consequently react on increased immigration into the neighborhood where they currently

---

33Results for natives renting their homes are provided in section 6.4.
live. Furthermore, a likely interpretation of the difference between the estimated flight and avoidance effects is that other immigrants and/or current renters are (at least partly) the buyers of the houses and apartments that the moving natives sell.

6.3 Is native flight determined by ethnically based preferences?
The pronounced flight effect in the subsample of home owners is interesting as such, in part because it potentially has implications for previous studies that mostly have looked at aggregate flight effects—which yet have been fairly in line with the effects in the group characterized as mobile above. We now make additional use of our data to consider the mechanism behind the estimated effects within this group.

Refugees come from a different ethnic background and typically also from lower socio-economic groups than the average native. In other words, if natives move due to increased immigration, they may do so either because they prefer ethnically homogenous neighborhoods, and/or if they have preferences for socio-economic homogeneity. As discussed in Section 3.1, we wish to examine if the commonly assumed ethnic channel is supported by the data, by grouping individuals according to their parental foreign background. While earlier work have speculated about which of ethnic vs. the socio-economic channel is the driving one, (see Saiz and Wachter, 2011; Sá, 2014; Rathelot and Safi, 2014), to our knowledge, we are the first to explicitly approach this question with relevant data.

As a group, native Swedes with non-western parents are on average ethnically more similar to the current immigrants, yet socio-economically more similar to natives with Swedish-born parents. This is the rationale for why we expect the relationship in equation (2) to apply, if natives indeed react on the immigrants’ ethnicity. Under such a scenario, estimated native flight would be higher among natives with native parents than among natives with non-Western parents. If, on the contrary, flight is observed to a similar extent among all natives irrespective of their parental background, then the main mechanism is more likely to be socio-economically driven.

We continue the focus on home owners and construct three groups of native home owners based on their foreign/ethnic background, and provide outflow and inflow effects for these respective groups in columns 2–4 of Table 5 (column 1 reproduces the average effect among home owners from above); column 2 contains those with native-born parents; column 3 contains those with at least one parent born in another Western
country\textsuperscript{34}; and column 4 contains those with at least one parent born in a non-Western country. As can be seen in Table 5, all flight estimates are positive and statistically significant. In comparing the magnitude across columns, it is clear that the relative magnitude is very similar across the groups of natives with different parental background; this is especially true when comparing the two groups that have native-born parents and non-Western born parents. The average number of movers are vastly different between the groups, and so, to simplify comparisons across the groups, the dependent variable is standardized with its own mean, prior to estimations.

Regarding inflow, the only (weakly) statistically significant effect is observed for the group in column 4—that is, for native home owners with at least one parent born in a non-Western country. But the point estimate is positive and thus, as above, there is no evidence of avoidance.

Treatment (increased immigration) is defined at the level of SAMS-by-year, and as noted above, we therefore cluster the standard errors by SAMS. Yet, the first phase of the placement program that defines our baseline period placed refugees to municipalities. We have therefore reestimated the model clustering at the municipality level. The number of clusters then decreases substantially, from around 8,700 to 290. Still, the first stage is hardly affected and remains statistically significant at conventional significance levels. For the second stage, the change in statistical significance varies; whereas the standard errors hardly change for the group of natives with parents born in other western and non-western countries, the estimate for natives with native-born parents is no longer significant with the municipality clusters.\textsuperscript{35} All in all, we trust the estimates obtained from clustering at the sams level, and note that despite some loss in precision, the comparison across groups of natives with different parental background mainly holds also when clustering at the municipality level.

\textsuperscript{34}Countries that are members of the OECD are defined as Western.

\textsuperscript{35}These results are available upon request. Natives with non-native parents are more likely to be concentrated to a few sams within a given municipality. Allowing for within-municipality correlation will then have little impact on the standard errors for these groups.
Table 5. Second-stage estimates of native flight and avoidance among home owners with different parental background

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All natives</td>
<td>Native</td>
<td>Western</td>
<td>Non-Western</td>
</tr>
<tr>
<td><strong>OUTFLOW</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(im)</td>
<td>0.00881**</td>
<td>0.00787**</td>
<td>0.0172***</td>
<td>0.00841**</td>
</tr>
<tr>
<td></td>
<td>(0.00396)</td>
<td>(0.00393)</td>
<td>(0.00636)</td>
<td>(0.00420)</td>
</tr>
<tr>
<td><strong>INFLOW</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(im)</td>
<td>0.00432</td>
<td>0.00375</td>
<td>0.00503</td>
<td>0.00799*</td>
</tr>
<tr>
<td></td>
<td>(0.00347)</td>
<td>(0.00353)</td>
<td>(0.00507)</td>
<td>(0.00447)</td>
</tr>
<tr>
<td>Observations</td>
<td>104,250</td>
<td>104,250</td>
<td>104,250</td>
<td>104,250</td>
</tr>
<tr>
<td>Number of SAMS</td>
<td>8,710</td>
<td>8,710</td>
<td>8,710</td>
<td>8,710</td>
</tr>
</tbody>
</table>

*** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered on SAMS-level.

Column 1 includes all native home owners, column 2 is restricted to native home owners with Swedish-born parents, column 3 is restricted to native home owners with at least one parent born in another Western country, and column 4 is restricted to native home owners with at least one parent born in a non-Western country. Dependent variables are standardized with its mean to facilitate comparisons across groups. See Table 2 for details of the estimated model.

That natives with different ethnic parental background display similar flight behavior thus quite strongly suggests that their residential preferences are not mainly shaped along the ethnic dimension. Rather, to the extent that immigrants on average are, or are perceived to be, less educated and poorer, increased immigration creates socio-economically more diverse neighborhoods, which may be the dimension along which natives’ residential preferences are shaped. In particular, the estimated effects for natives with native parents and natives with non-western parents are strikingly similar. The point estimate are in both cases close to 0.008, which indicates that one additional refugee in the neighborhood causes outflow, representing 0.8 percent of the mean number of movers within each group, per year.

To study this further, we use our model to study how immigration affects the income and education level in the neighborhood, respectively, among home owners in the same sub-groups as analyzed in Table 5. The first stage is the same as before, while the second stage is now given by:

\[
income_{i,t+1} = \beta^{inc} \hat{im}_{i,t} + \sum_{p=1}^{3} \delta^p pop^p_{i,t-1} + \Pi'X + \mu_i + \tau_t + \varepsilon^{inc}_{i,t+1} \quad (10)
\]
and

\[
\text{university}_{i,t+1} = \beta_{\text{univ}} \text{im}_{i,t} + \sum_{p=1}^{3} \delta_{p} \text{pop}_{i,t-1}^{p} + \Pi'X + \mu_{i} + \tau_{t} + \varepsilon_{i,t+1}^{\text{univ}},
\]

where the outcome variable is replaced with the average disposable income, \(\text{income}_{i,t+1}\), and the share of university-educated, \(\text{university}_{i,t+1}\), in the neighborhood. We define and estimate equations (10) and (11) for all home owners (that is, both native and non-native) as well as separately by the same sub-groups as analyzed in Table 5. Also, to once again facilitate comparisons across groups, the dependent variable is standardized with its own mean, prior to estimations.\(^{36}\)

The results, provided in Tables 6 and 7, show that the effect of increased immigration is that the income as well as the educational level decrease among all home owners irrespectively of their own or their parents' foreign background (although the income estimates for natives with foreign-born parents are statistically insignificant). Interestingly, the similar magnitudes of the effects for all home owners and all native home owners (cf. columns 1 and 2) imply that the socio-economic segregation is driven by natives from higher socio-economic groups moving out rather than by immigrants from lower socio-economic groups moving in. These results further strengthen the conjecture that preferences are formed along socio-economic rather than ethnic lines.

\(^{36}\)Note that when estimating the effect on the education level in the neighborhood (Equation 11, Table 7) we add a lagged control for the number of university educated in \(t - 1\) at neighborhood level.
Table 6. Second-stage estimates of the average disposable income among different types of home owners

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All natives</td>
<td>Native Western</td>
<td>Non-Western</td>
<td></td>
<td></td>
</tr>
<tr>
<td>im</td>
<td>-0.00186**</td>
<td>-0.00192**</td>
<td>-0.00130*</td>
<td>-0.00149</td>
<td>-0.000494</td>
</tr>
<tr>
<td></td>
<td>(0.000845)</td>
<td>(0.000874)</td>
<td>(0.000785)</td>
<td>(0.00113)</td>
<td>(0.000687)</td>
</tr>
<tr>
<td>Obs.</td>
<td>90,825</td>
<td>90,748</td>
<td>90,555</td>
<td>87,025</td>
<td>87,946</td>
</tr>
<tr>
<td>N'hoods</td>
<td>8,358</td>
<td>8,350</td>
<td>8,331</td>
<td>8,159</td>
<td>8,179</td>
</tr>
</tbody>
</table>

*** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered on SAMS-level.
The outcome variable is the neighborhood average disposable income in the respective groups. Column 1 includes all home owners, column 2 includes all native home owners, column 3 is restricted to native home owners with Swedish-born parents, column 4 is restricted to native home owners with at least one parent born in another Western country, and column 5 is restricted to native home owners with at least one parent born in a non-Western country. Dependent variables are standardized with its mean to facilitate comparisons across groups. See Table 2 for details of the estimated model.

Table 7. Second-stage estimates of the share university educated among different types of home owners

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All natives</td>
<td>Native Western</td>
<td>Non-Western</td>
<td></td>
<td></td>
</tr>
<tr>
<td>im</td>
<td>-0.00358***</td>
<td>-0.00344***</td>
<td>-0.00361***</td>
<td>-0.00359***</td>
<td>-0.00173***</td>
</tr>
<tr>
<td></td>
<td>(0.000888)</td>
<td>(0.000873)</td>
<td>(0.000923)</td>
<td>(0.00124)</td>
<td>(0.000584)</td>
</tr>
<tr>
<td>Obs.</td>
<td>95,551</td>
<td>95,551</td>
<td>95,551</td>
<td>95,551</td>
<td>95,551</td>
</tr>
<tr>
<td>N'hoods</td>
<td>8,709</td>
<td>8,709</td>
<td>8,709</td>
<td>8,709</td>
<td>8,709</td>
</tr>
</tbody>
</table>

*** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered on SAMS-level. The outcome variable is the neighborhood share of home owners in the respective groups that has at least some university education. Column 1 includes all home owners, column 2 includes all native home owners, column 3 is restricted to native home owners with Swedish-born parents, column 4 is restricted to native home owners with at least one parent born in another Western country, and column 5 is restricted to native home owners with at least one parent born in a non-Western country. Dependent variables are standardized with its mean to facilitate comparisons across groups. Covariates are the same as described in Table 2, with the addition of the lagged number of university educated.

6.4 Flight and avoidance among renting natives

The conclusion that changing socio-economic characteristics rather than ethnic heterogeneity seems to be the primary channel explaining natives’ migration behavior, pertains to the analysis above focusing on home owners, who in the current setting arguably are those who indeed can react on increased immigration. Table 8 instead presents flight and
avoidance estimates for natives in publicly provided rental apartments, again grouped according to parental foreign background. As the table shows, the effects among these groups are generally negative. That is, increased immigration leads to fewer renters moving out (consequently leaving less room for others to move in). This is in line with the argument above (see Section 3.3), that increased competition for public rentals in the wake of an inflow of immigrants causes lock-in effects among the initial renters.

An exception to the negative coefficients is the effect for natives with non-Western parents, who instead react by moving out of the neighborhood cf. column 4 of Table 8). One possible explanation to this result is that this group starts a housing career when new immigrants arrive. To test this, we look specifically at the outflow of native renters becoming home owners. Table 9 presents the results, showing that the only positive, significant effect is again among natives with non-Western parental background. Relative to the mean of the dependent variable, the magnitude is also similar to the overall effect for this group. Thus, the results are indeed consistent with this group making a housing career. Again, the estimated flight and avoidance behavior among renters are difficult to reconcile with an ethnically based mechanism.

### Table 8. Second-stage estimates of native flight and avoidance among renters with different parental background

<table>
<thead>
<tr>
<th></th>
<th>All Natives</th>
<th>Native Parental background:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Native</td>
<td>Western</td>
<td>Non-Western</td>
</tr>
<tr>
<td>OUTFLOW</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( im )</td>
<td>-0.00391</td>
<td>-0.0126***</td>
<td>-0.0127</td>
<td>0.0503***</td>
</tr>
<tr>
<td></td>
<td>(0.00496)</td>
<td>(0.00487)</td>
<td>(0.0107)</td>
<td>(0.0127)</td>
</tr>
<tr>
<td>INFLOW</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( im )</td>
<td>-0.00426**</td>
<td>-0.00618***</td>
<td>-0.00152**</td>
<td>0.0104</td>
</tr>
<tr>
<td></td>
<td>(0.00205)</td>
<td>(0.00207)</td>
<td>(0.000768)</td>
<td>(0.00992)</td>
</tr>
<tr>
<td>Observations</td>
<td>104,250</td>
<td>104,250</td>
<td>104,250</td>
<td>104,250</td>
</tr>
<tr>
<td>Number of SAMS</td>
<td>8,710</td>
<td>8,710</td>
<td>8,710</td>
<td>8,710</td>
</tr>
</tbody>
</table>

*** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered on SAMS-level. Column 1 includes all native renters, column 2 is restricted to native renters with Swedish-born parents, column 3 is restricted to native renters with at least one parent born in another Western country, and column 4 is restricted to native renters with at least one parent born in a non-Western country. See Table 2 for details of the estimated model.
Table 9. Second-stage estimates of native flight among renters with different parental background becoming home owners

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Parental background:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Natives</td>
<td>Native Western Non-Western</td>
<td></td>
<td></td>
</tr>
<tr>
<td>im</td>
<td>0.00412</td>
<td>-0.00785*</td>
<td>0.0184</td>
<td>0.0843***</td>
</tr>
<tr>
<td></td>
<td>(0.00571)</td>
<td>(0.00462)</td>
<td>(0.0159)</td>
<td>(0.0246)</td>
</tr>
<tr>
<td>Observations</td>
<td>104,250</td>
<td>104,250</td>
<td>104,250</td>
<td>104,250</td>
</tr>
<tr>
<td>Number of SAMS</td>
<td>8,710</td>
<td>8,710</td>
<td>8,710</td>
<td>8,710</td>
</tr>
</tbody>
</table>

*** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered on SAMS-level.

Column 1 includes all native renters-home owners, column 2 is restricted to native renters-home owners with Swedish-born parents, column 3 is restricted to native renters-home owners with at least one parent born in another Western country, and column 4 is restricted to native renters-home owners with at least one parent born in a non-Western country. See Table 2 for details of the estimated model.

6.5 Native flight and tipping points

Following the common approach in the native/white flight and avoidance literature, we have estimated a continuous model of migration behavior. This contrasts with the tipping point literature, which is based on a model in which natives’ migration behavior changes abruptly when the share of immigrants reaches a certain level (see, e.g., Schelling, 1971; Card et al., 2008). A question is then how our results relate to this latter literature hypothesizing that there is a neighborhood tipping point? To get a sense of this, we analyze how the flight coefficients vary with the initial share of immigrants. To this aim, the model is estimated separately depending on the neighborhood share of immigrants in the first year of our study period (1997). We focus on home owners and distinguish between natives of different parental foreign background.

Figure 4 shows the resulting estimates for each 1-percentage interval of the initial immigrant share from 4 to 25 percent (the minimum and maximum value, respectively). Although the estimates across the different initial levels are not significantly different from one another, there are signs of a discrete increase in the parameter estimate at an initial share of 18 percent or above. This is something to examine further in future research (as is the U-shaped pattern of the point estimates, which is not readily reconciled with one unique tipping point).\(^{37}\)

Furthermore, the three groups of natives with different parental foreign background display very similar patterns (cf. panels b-d). Interest-

\(^{37}\)Compare also with the tipping points found in Böhlmark and Willén (2017) and in Aldén et al. (2015) using Swedish data.
ingly, this is in line with the results found in previous sections, and thus with the conclusion that residential preferences do not seem to be formed along ethnic lines. The observed migration behavior of native-born individuals with varying foreign background is rather likely to be a reaction to changing socio-economic neighborhood characteristics. These results potentially have implications for the tipping point literature; significant tipping points for all native-born individuals irrespectively of their parental foreign background is suggestive of native tipping behavior based not, as commonly assumed, on ethnic/racial grounds, but rather on socio-economic grounds.

Figure 4. Second-stage estimates of native flight among home owners with different parental background, by initial immigrant share
(a) All natives
(b) Native parental background
(c) Western parental background
(d) Non-Western parental background

Note: The estimates are obtained from separate regressions depending on the share of foreign-born in the neighborhood in the initial year of the analysis (1997), for each 1-percentage interval between 4 and 25 percent. For comparison across groups, the dependent variable is normalized by dividing it with its own mean. Figure 4a includes all native home owners, Figure 4b is restricted to native home owners with Swedish-born parents, Figure 4c is restricted to native home owners with at least one parent born in another OECD country, and Figure 4d is restricted to native home owners with at least one parent born in a non-Western country.

Source: GeoSweden.
7 Concluding remarks

In this paper, we have applied detailed, comprehensive, register data to a refined shift-share methodological approach to answer whether native flight and avoidance are important phenomena in Sweden. In particular, using information in the data that allows us to identify native-born individuals that are to different degrees ethnically close to the newly arrived refugee immigrants (as defined via the heterogeneous parental backgrounds of the native-born), we have examined if there is any support for the hypothesis that natives prefer to live in ethnically homogeneous neighborhoods. Our study spans the period 1990–2010, which is an important and interesting period to study for at least two reasons; first, there was a large increase in refugee-based immigration to Sweden over this time period and, second, in the early part of the period, there was a refugee placement policy in Sweden which arguably can be used to improve the shift-share instrument.

Using push-driven refugee immigration to Sweden interacted with a settlement pattern of their countrymen in the early 1990s that was partly generated by this state-run placement policy, we reach four main conclusions.

First, we find no evidence of neither native flight nor native avoidance when using the full population in Sweden.

Second, when we look specifically at the group having a high possibility to move (home owners in the Swedish context), we do detect native flight. That is, home owning natives move out of neighborhoods experiencing an increase in immigration. We do however not find evidence of native avoidance, in the sense that natives do not move into these neighborhoods to a lesser extent. A possible interpretation of this discrepancy is that natives mostly notice and consequently react on increased immigration into the neighborhood where they currently live. We do not find any flight or avoidance effects among natives identified as having low possibilities to move (here renters). Hence, distinguishing between mobile/immobile households when examining the effects of immigration on native migration seems important, but is something the earlier literature has not been able to do.

Third, we find that the ethnic closeness between the native-born individuals and the newly arrived refugees does not matter for observed flight behavior; all Swedish-born individuals react in a very similar fashion to increased immigration. Preferences for ethnically homogeneous neighborhoods do therefore not seem to be the main channel causing flight. Rather, our analyses consistently indicate that natives have preferences for socio-economically homogeneous, or “better”, neighborhoods.
Fourth and finally, when conditioning on the initial share of foreign born, and thereby relating to the tipping point literature, we again find the same patterns regardless of the natives’ parental foreign background. This is thus further evidence against the ethnicity channel, which could indicate that the tipping point literature might have focused on the wrong trait.

If political decision-makers want to instigate policies to combat segregation following a change in natives’ migration behavior, it is important to know what the mechanisms are for the observed change; successful policies will likely differ depending on whether the main channel is ethnically based or socio-economically based. The results in this paper indicate that the native migration behavior is not mainly ethnically driven, but might rather be driven by socio-economic factors. More research is however needed before any firm policy conclusions can be drawn.

Several future extensions are of interest. First, while the one-year lag allows us to identify more precise quantitative causal effects, we acknowledge that this focus potentially misses flight behavior that takes place after a longer period of time. Longer time lags is therefore an interesting follow-up. Second, the focus on small neighborhoods could be extended to larger areas, such as municipalities: this is not the least interesting as a comparison to the the earlier literature, which mostly has studied larger geographical units. Third, our results indicate that the ethnically-based tipping point literature might have focused on the wrong trait. This is important to investigate further, for example by extending the analysis for native-born individuals with different ethnic backgrounds. Lastly, a possible alternative way to channel any preferences for homogeneity is via school choices. If parents perceive school quality to be dropping due to increased minority presence, an exodus from the neighborhood school could occur. Increasing school segregation is thus also an interesting topic for future research.
References


Appendices

A Using IV-estimator suggested by Jaeger et al. (2018)

In Jaeger et al. (2018), the shift share instrument is criticized for failing to account for dynamic effects. Their example is based on wages, but the example carry importance for any mechanism which might hold important dynamics over time.

Imagine an immigration shock in $t_0$ to city $i$. Due to a short term effect we get native flight. Now, the higher flight might put other forces in motion, such as a price decrease, continued flight, or possibly a mean reversion of prices. If we use a shift share instrument, we will make use of the serial correlation over time in immigrations residential location patterns. We will therefore, potentially, measure both the short term effect of immigration, as well as the slowly reversing dynamic process. The results hence become difficult to interpret.

The solution in Jaeger et al. (2018) is to add an extra lag to the model, and estimate the effect of both last years effect ($t-1$) and the current year effect ($t$). Since both are endogenous, we have to run two first stages. These are described in equations (12) and (13).

\begin{align}
    im_{i,t} &= \gamma_{1,1} \hat{im}_{i,t} + \gamma_{1,2} \hat{im}_{i,t-1} + \sum_{p=1}^{3} \phi_p pop^p_{i,t-1} + X\Gamma' + \mu_i + \tau_t + \epsilon_{i,t} \\
    \text{(12)}
\end{align}

\begin{align}
    im_{i,t-1} &= \gamma_{2,1} \hat{im}_{i,t} + \gamma_{2,2} \hat{im}_{i,t-1} + \sum_{p=1}^{3} \phi_p pop^p_{i,t-1} + X\Gamma' + \mu_i + \tau_t + \epsilon_{i,t} \\
    \text{(13)}
\end{align}

The two first stages in turn give the following second stage equations:

\begin{align}
    outflow_{i,t+1} &= \beta_{1}^{IV} \hat{im}_{i,t} + \beta_{2}^{IV} \hat{im}_{i,t-1} + \sum_{p=1}^{3} \delta_{p}^{IV} pop^p_{i,t-1} + X\Pi^{IV'} + \eta_i + \lambda_t + \epsilon_{i,t+1}^{IV} \\
    \text{(14)}
\end{align}
\[ \text{inflow}_{i,t+1} = \beta_1^{IV} \widehat{im}_{i,t} + \beta_2^{IV} \widehat{im}_{i,t-1} + \sum_{p=1}^{3} \delta_p^{IV} \text{pop}_p^p_{i,t-1} + \mathbf{X}^{IV'} + \eta_t + \lambda_t + \varepsilon_{i,t+1}^{IV} \quad (15) \]

Table A.1 shows the first stages (Equation 12 and 13). The F-statistic at the bottom indicate that neither of the equations suffer from weak instruments. The second stage can be seen for outflow in Table A.2 and for inflow in Table A.3. Comparing the results to the baseline estimates of Table 5, it is clear that the interpretation of the results stay the same. In other words, it does not seem dynamic effects are a threat to our short term estimates.

<table>
<thead>
<tr>
<th>Table A.1. First stage, multiple equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIABLES</td>
</tr>
<tr>
<td>( \widehat{im}_{it} )</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>( \widehat{im}_{it-1} )</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Observations</td>
</tr>
<tr>
<td>F-Stat</td>
</tr>
<tr>
<td>Number of SAMS</td>
</tr>
<tr>
<td>SAMS FE</td>
</tr>
<tr>
<td>Year FE</td>
</tr>
<tr>
<td>Covariates</td>
</tr>
</tbody>
</table>

*** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered on SAMS-level. Multiple first stages according to the model developed in Jaeger et al. (2018). Column 1 shows the estimates based on Equation 12 and column 2 the first stage according to Equation 13. For exact covariates see Table 2.
Table A.2. *Second stage according to Jaeger et al. (2018), Outflow*

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Outflow</th>
<th>Outflow</th>
<th>Outflow</th>
<th>Outflow</th>
</tr>
</thead>
<tbody>
<tr>
<td>(im_{it})</td>
<td>0.00827***</td>
<td>0.00717**</td>
<td>0.0177***</td>
<td>0.00811**</td>
</tr>
<tr>
<td></td>
<td>(0.00319)</td>
<td>(0.00312)</td>
<td>(0.00562)</td>
<td>(0.00378)</td>
</tr>
<tr>
<td>(im_{it-1})</td>
<td>0.00144</td>
<td>0.00187</td>
<td>-0.00141</td>
<td>0.000804</td>
</tr>
<tr>
<td></td>
<td>(0.00331)</td>
<td>(0.00325)</td>
<td>(0.00459)</td>
<td>(0.00440)</td>
</tr>
</tbody>
</table>

Observations 104,250 104,250 104,250 104,250
Number of SAMS 8,710 8,710 8,710 8,710
SAMS FE YES YES YES YES
Year FE YES YES YES YES
Covariates YES YES YES YES

*** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered on SAMS-level.

Estimating the effect of immigration on outflow among home owners using the econometric model developed in Jaeger et al. (2018). All specifications implements Equation 14, with changing definitions of the dependent variable over the columns. For exact covariates see Table 2.

Table A.3. *Second stage according to Jaeger et al. (2018), Inflow*

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Inflow</th>
<th>Inflow</th>
<th>Inflow</th>
<th>Inflow</th>
</tr>
</thead>
<tbody>
<tr>
<td>(im_{it})</td>
<td>0.00399</td>
<td>0.00349</td>
<td>0.00466</td>
<td>0.00709</td>
</tr>
<tr>
<td></td>
<td>(0.00280)</td>
<td>(0.00270)</td>
<td>(0.00472)</td>
<td>(0.00463)</td>
</tr>
<tr>
<td>(im_{it-1})</td>
<td>0.000901</td>
<td>0.000686</td>
<td>0.00101</td>
<td>0.00241</td>
</tr>
<tr>
<td></td>
<td>(0.00382)</td>
<td>(0.00378)</td>
<td>(0.00544)</td>
<td>(0.00444)</td>
</tr>
</tbody>
</table>

Number of SAMS 8,710 8,710 8,710 8,710
SAMS FE YES YES YES YES
Year FE YES YES YES YES
Covariates YES YES YES YES

*** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered on SAMS-level.

Estimating the effect of immigration on inflow among home owners using the econometric model developed in Jaeger et al. (2018). All specifications implements Equation 15, with changing definitions of the dependent variable over the columns. For exact covariates see Table 2.
B Source countries


<table>
<thead>
<tr>
<th>Country</th>
<th>Frequency</th>
<th>Arrival %</th>
<th>Residence %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iraq</td>
<td>40,537</td>
<td>43.31%</td>
<td>43.31%</td>
</tr>
<tr>
<td>Somalia</td>
<td>11,597</td>
<td>12.39%</td>
<td>55.70%</td>
</tr>
<tr>
<td>Yugoslavia</td>
<td>8,345</td>
<td>8.92%</td>
<td>64.61%</td>
</tr>
<tr>
<td>Bosnia</td>
<td>6,727</td>
<td>7.19%</td>
<td>71.80%</td>
</tr>
<tr>
<td>Iran</td>
<td>5,105</td>
<td>5.45%</td>
<td>77.26%</td>
</tr>
<tr>
<td>Afghanistan</td>
<td>4,347</td>
<td>4.64%</td>
<td>81.90%</td>
</tr>
<tr>
<td>Syria</td>
<td>3,954</td>
<td>4.22%</td>
<td>86.12%</td>
</tr>
<tr>
<td>Russia</td>
<td>2,676</td>
<td>2.86%</td>
<td>88.98%</td>
</tr>
<tr>
<td>Lebanon</td>
<td>2,563</td>
<td>2.74%</td>
<td>91.72%</td>
</tr>
<tr>
<td>Thailand</td>
<td>1,225</td>
<td>1.31%</td>
<td>93.03%</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>1,142</td>
<td>1.22%</td>
<td>94.25%</td>
</tr>
<tr>
<td>Croatia</td>
<td>887</td>
<td>0.95%</td>
<td>95.20%</td>
</tr>
<tr>
<td>Colombia</td>
<td>736</td>
<td>0.79%</td>
<td>95.98%</td>
</tr>
<tr>
<td>India</td>
<td>683</td>
<td>0.73%</td>
<td>96.71%</td>
</tr>
<tr>
<td>Peru</td>
<td>520</td>
<td>0.56%</td>
<td>97.27%</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>469</td>
<td>0.50%</td>
<td>97.77%</td>
</tr>
<tr>
<td>Pakistan</td>
<td>468</td>
<td>0.50%</td>
<td>98.27%</td>
</tr>
<tr>
<td>China</td>
<td>269</td>
<td>0.29%</td>
<td>98.56%</td>
</tr>
<tr>
<td>Uganda</td>
<td>187</td>
<td>0.20%</td>
<td>98.76%</td>
</tr>
<tr>
<td>Romania</td>
<td>165</td>
<td>0.18%</td>
<td>98.93%</td>
</tr>
<tr>
<td>Bolivia</td>
<td>164</td>
<td>0.18%</td>
<td>99.11%</td>
</tr>
<tr>
<td>Vietnam</td>
<td>160</td>
<td>0.17%</td>
<td>99.28%</td>
</tr>
<tr>
<td>Algeria</td>
<td>125</td>
<td>0.13%</td>
<td>99.41%</td>
</tr>
<tr>
<td>Poland</td>
<td>108</td>
<td>0.12%</td>
<td>99.53%</td>
</tr>
<tr>
<td>Morocco</td>
<td>86</td>
<td>0.09%</td>
<td>99.62%</td>
</tr>
<tr>
<td>Tunisia</td>
<td>78</td>
<td>0.08%</td>
<td>99.70%</td>
</tr>
<tr>
<td>Latvia and Lithuania</td>
<td>71</td>
<td>0.08%</td>
<td>99.78%</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>49</td>
<td>0.05%</td>
<td>99.83%</td>
</tr>
<tr>
<td>Estonia</td>
<td>38</td>
<td>0.04%</td>
<td>99.87%</td>
</tr>
<tr>
<td>Philippines</td>
<td>36</td>
<td>0.04%</td>
<td>99.91%</td>
</tr>
<tr>
<td>Gambia</td>
<td>31</td>
<td>0.03%</td>
<td>99.94%</td>
</tr>
<tr>
<td>Argentina</td>
<td>29</td>
<td>0.03%</td>
<td>99.98%</td>
</tr>
<tr>
<td>Slovenia</td>
<td>12</td>
<td>0.01%</td>
<td>99.99%</td>
</tr>
<tr>
<td>Brazil</td>
<td>11</td>
<td>0.01%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Number of refugees per emigration country who got a residence permit 1997-2010.
IV. Refugee immigration and the housing market

With Matz Dahlberg

Acknowledgments: The authors are grateful to Hélène Berg, Per-Anders Edin, Karin Edmark, Per Engström, Che-Yuan Liang and seminar participants at Uppsala University for helpful feedback and comments.
1 Introduction

Over the last decade, many European countries have experienced both an immigration-driven population growth and an increased ethnic heterogeneity in the existing population. Sweden constitutes a striking example of this development, with both a large influx of immigrants and an increased heterogeneity in terms of individuals’ countries of origin. First, from 2006 to 2016, the population in Sweden increased from about 9 to 10 million inhabitants, with 80% of the population growth being due to immigration.¹ Second, since the labor migration to Sweden drastically decreased during the 1970s, the refugee-driven, non-Western, immigration has been dominating. This trend is clearly visible also during the last decade. In Figure 1, which shows the share of all non-Swedish immigration to Sweden by origin region over the time period 2004-2014, it can be noted that almost 65% of the non-Swedish immigration to Sweden in 2014 had its origin in a non-OECD country.²

An important question is how this, largely refugee-driven, immigration affects the neighborhoods in which the immigrants settle. Earlier research has studied a vast number of effects of immigration, spanning from preferences to prices.³ This paper focuses on effects on the housing market, with the specific aim to examine how refugee immigration to Sweden affects house prices in small geographic areas.

House prices in general, and the relationship between immigration and house prices in smaller neighborhoods in particular, is interesting for several reasons. First, it informs policy makers on how immigration affects neighborhood dynamics and hence on how immigration is generally viewed at a geographic level where the actual interaction between immigrants and locals are most likely to take place. Second, policy makers learn about a potentially important margin in the determination of

¹The population growth can be calculated at https://www.scb.se/hitta-statistik/sverige-i-siffror/manniskorna-i-sverige/in-och-utvandring/ (visited on February 11, 2018). Number of residence permits is available at the website of the Migration Agency https://www.migrationsverket.se/English/Startpage.html.
²The reason we show the pattern for the period 2004-2014 is that this is the period we use in the empirical analysis. The statistics shown in the figure comes from own calculations based on data in GeoSweden (described later in the paper).
³To name a few, earlier research has for example looked at the effect of immigration and/or increased ethnic heterogeneity on labor market outcomes (Dustmann et al., 2017, 2016), preferences for redistribution (Luttmer, 2001; Dahlberg et al., 2012), trust (Alesina and La Ferrara, 2002), participation in social activities (Alesina and La Ferrara, 2000), charitable giving (Andreoni et al., 2016), collective action (Vigdor, 2004) and the size and mix of publicly provided goods and services (Alesina et al., 1999, 2000).
house prices. Third, individual owners of housing learn how immigration affects the value of one of their most important assets.

Figure 1. Share of all non-Swedish immigration to Sweden 2004-2014 by origin region.

Why should we expect immigration to affect house prices? There are at least three channels through which the effect can materialize. First, the most straightforward channel, emanating from a standard economic model of demand and supply, implies an upward pressure on housing prices. Given an inelastic housing supply, increased immigration yields more individuals and hence more potential bidders for each housing object on the market, which could increase prices. How large this upward pressure will be depends on how elastic housing supply is. To get an indication on the latter, we will also look into the effects of immigration on housing supply; if there is a significant and positive effect on housing supply, this will mitigate some of the upward pressure on prices coming from the increased demand.\(^4\) In addition, given that many refugees lack the resources to buy an apartment or a house, a second demand channel works via natives. If the arriving migrants occupy the presumably inelastic supply of rental apartments, a year with unusually large inflows

\(^4\)Using Spanish data, Fernández-Huertas Moraga et al. (2017) find that immigration has a positive effect on housing supply.
of immigration will imply unusually few rental apartments available for natives. This “crowding-out” effect can lead to more natives opting to buy rather than to rent, which will increase demand and hence prices.

Second, natives’ have preferences for neighborhood homogeneity (for example along ethnic or socio-economic lines), immigration might affect natives’ migration behavior. There is a literature examining whether increased immigration affects native population change (through native flight and/or native avoidance; see Saiz and Wachter, 2011, Sá, 2014, Fernández-Huertas Moraga et al., 2017, and Andersson et al., 2018, for examples of recent studies). The typical result in this literature is that immigration leads to a decrease in the native population (most likely through an increased out-migration of natives from the neighborhood, as indicated by earlier work), implying that there will be fewer bidders on each housing object on the market and hence a downward pressure on housing prices. Using Swedish data, Andersson et al. (2018) also find that the native flight has a clear socioeconomic component, with average disposable income among native homeowners dropping due to increased immigration. A decrease in the socio-economic level in the neighborhood could put a further downward pressure on house prices (i.e., both the number and the economic strength of the potential bidders decrease).

Third, if natives have expectations that immigration will affect local amenities negatively (e.g., through a decrease in the quality of local public goods provided, like schooling, or through increased crime in the neighborhood), this will also offset some of the direct demand effect following population growth through immigration. To get an idea on how important these types of expectations might be, we will examine how immigration affects expected house prices as measured by the initial (list) prices and by the time an object is on the market.

In our attempt to estimate the effect of immigration on house prices, we use yearly, encompassing, register data from Sweden covering the time period 2004-2014. To deal with the endogeneity of immigrants’ location choices, we adopt an instrumental variable that enhances the traditional shift-share instrument by using push-related immigration (defined through the refugee status of the immigrants as given in the registers) in combination with an initial settlement pattern that was determined through a refugee placement policy that was in effect in Sweden between 1985 and 1994. The panel structure of the data and the continuous large non-Western immigration to Sweden also allow us to use neighborhood fixed effects in the econometric specifications, which means that

---

5 A model of this type is provided in Accetturo et al. (2014).
6 The instrument was developed and used in Andersson et al. (2018), where also a lengthier discussion of the instrument is provided.
all identifying variation comes from a comparison within neighborhoods over time.

Currently, there are only a few papers in the earlier literature that specifically look at local (neighborhood) house price effects following immigration. The existing studies using data over small geographic areas typically find that immigration has a negative and significant effect on house prices (see Saiz and Wachter, 2011, for US evidence, Sá, 2014, for UK evidence, and Accetturo et al., 2014, for Italian evidence). The implied conclusion is that native migration responses and expectations about negative impacts on local amenities dominates the upward pressure on house prices coming from the increased demand for housing following immigration.\textsuperscript{7}

This paper complements the literature on immigration and house prices by adding a number of features. First, our detailed immigration data allows us to look specifically into a certain type of immigration, in this case refugee migration. This differs from the previous literature, which has treated immigration as one concept, regardless of source country and reason for migration. Refugees are a particularly interesting group given that they make up a large, and for long increasing, share of global migrants, but also due to the sometimes heated political discussions on the effects of refugees on recipient economies.\textsuperscript{8} Second, our detailed house price and individual-level data allow us to dig deeper into mechanisms, studying features such as list prices, time on market (counting days between the listing of an object and the signing of the final contract), and in which housing segment the newly arrived immigrants end up in. Third, we provide the first estimates for Sweden, a country that has experienced high (refugee) immigration for decades, but perhaps more importantly has a regulated rental housing market, including rents based on user values, and a large share of rental apartments owned by the municipalities.

Our baseline results show no effect of immigration on changes in housing prices. This result holds both for condominiums and owner-occupied houses, and is robust to alternative model specifications (such as adopt-

\textsuperscript{7}The papers that look at effects at more aggregate levels, like metropolitan areas, typically find positive effects of immigration on house prices; see, e.g., Saiz (2007) for evidence on US data, Tumen (2016) for evidence on Turkish data, Kurschner (2017) for evidence on German data, and Sanchis-Guarner (2017) for evidence on Spanish data. At more aggregate levels, the offsetting effects emanating from local interactions are probably much less important.

\textsuperscript{8}Refugees in this paper encompasses all types of asylum related residence permits, most importantly including conventional “geneva convention” refugees as well as refugees granted protection due to conflicts and war (“persons in need of protection”).
ing the IV-estimator suggested by Jaeger et al. (2018), adopting the specifications used by Saiz (2007) and (Sá, 2014), which controls for immigration and population size in different ways, and by using estimated hedonic prices as dependent variable). We also test the effect on the median amount of days an object stays on the market and changes in list prices, respectively, finding no significant effects. We do find a significant effect on housing supply, as measured by the number of objects on the market. This result is in line with a companion paper on native flight in Sweden, by Andersson et al. (2018), who find some flight among owners of housing, following immigration. Due to this flight mechanism, we would expect there to be more objects on sale following a year with larger than average levels of immigration.

It is finally worth noticing that our results are not in line with the negative results from immigration on house prices on neighborhood level found in the literature thus far. We can think of two possible reasons for this. The first reason is that Swedes do not, to the same extent as natives in other countries, practice native flight and avoidance and/or believe that the quality of local amenities drops because of immigration. There is some indicative data, which supports such a story. As few as 1.5 percent of Swedish respondents answered they wished not to have an immigrant or a foreign worker as a neighbor, when asked in the World Value Survey of 2005-2009.9 This number can be compared to 13.9% for Italy, 14.2% for the U.K. and 12.7% for the U.S. If Swedes are more tolerant to living in heterogeneous neighborhoods, this would explain the modest effects found in this study. The second reason is the Swedish institutional context. A large share of the rental market is dominated by public rentals, and new rental apartments have not been constructed in a pace to meet the population increase during the last decade. Municipalities are also responsible for housing refugees, should they not be able to find a place on their own, and most refugees end up staying in rentals during their initial years. The combination of a municipal responsibility for housing refugees and the restricted housing market might inflate a crowding-out mechanism, causing natives to buy instead of renting.

The rest of the paper is organized as follows: In the next section we describe the empirical strategy, in section 3 we describe the data sources and provide some descriptive statistics, section 4 presents the results and section 5 concludes.

9This is the fifth wave of the survey. Documents showing world value survey data is found at http://www.worldvaluessurvey.org/WVSContents.jsp.
2 Estimating the effect of immigration on house prices

The main goal of the paper is to estimate the causal effect of refugee immigration on housing prices in small neighborhoods. In this section we provide our strategy to achieve this.\textsuperscript{10}

We begin by defining a stylized OLS-equation, where the change in the log of the housing price between year $t$ and year $t+1$ for neighborhood $i$ ($\Delta \ln (P_{i,t+1})$) is regressed on the neighborhood-level inflow of non-western migrants in year $t$ ($im_{i,t}$).\textsuperscript{11} $\beta$ is the coefficient of interest, providing the effect on housing prices of one additional immigrant moving in.

$$\Delta \ln (P_{i,t+1}) = \beta^{OLS} im_{i,t} + \varepsilon^{OLS}_{i,t+1}$$ (1)

$\beta^{OLS} > 0$ is in line with immigration affecting housing prices positively, and $\beta^{OLS} < 0$ means immigration has a negative effect on prices.

2.1 Identification

The empirical problem to deal with is that the allocation of immigrants over space is non-random. Migrants might choose where to live based on different unobserved neighborhood characteristics, which may in turn be correlated with our outcome of interest, housing prices. If this is the case, $\beta^{OLS}$ will be biased, reflecting the effect of immigration as well as the unobserved characteristics. Estimations of immigration on house prices may also suffer from reversed causality. If prices are expected to drop, this may attract migrants with less capital or income. $\beta^{OLS} < 0$ could in this case reflect both the negative/positive effect of immigration on house prices, as well as the hypothesized negative effect of house prices on immigration.

As a first step in trying to isolate the causal effect of immigration on housing prices, we augment the naïve equation (1) with fixed effects and time-varying covariates, which we think might be correlated with both the location of immigrants and with prices:

\textsuperscript{10}When looking into mechanisms, for example when estimating the effect of immigration on housing supply and list prices, we use the same model as laid out in this section but with other dependent variables.

\textsuperscript{11}Effectively, since this specification is in first-differenced form, unobserved, neighborhood-specific fixed effects are controlled for.
\[
\Delta \ln(P_{i,t+1}) = \beta^{OLS} im_{i,t} + \sum_{p=1}^{3} \delta^{OLS}_p pop^p_{i,t-1} + X^{OLS'} + 
\eta_i + \lambda_t + \epsilon^{OLS}_{i,t+1}
\]  

(2)

Two of the added variables in Equation (2) are fixed effects meant to pick up unobserved covariates: \(\eta_i\) is a neighborhood fixed effect that, within the first-difference specification given in Equation (1), reflects a neighborhood-specific time trend that captures all (linear) trends within each neighborhood that might cause a spurious relationship between immigration and housing prices, and \(\lambda_t\) is a year fixed effect that captures aggregate shocks that affect the neighborhoods in the same way in a given year (in the application, we will use panel data that is measured annually). In addition, we add several time varying characteristics: \(pop^p_{i,t-1}\) flexibly captures, via a third-order polynomial, the importance of the population size\(^{12}\) and \(X\) is a vector containing the variables average disposable income, average social assistance paid out, the number of public rental estates and all other non-refugee immigration to the neighborhood.\(^{13}\) To avoid problems with bad controls, we use all time-varying characteristics lagged one year.

Equation (2) probably takes care of most of the characteristics affecting both immigration and housing prices. Having neighborhood fixed effects in the specification implies, e.g., that all identifying variation comes from comparisons within neighborhoods over time, which must be considered as a fairly tight specification. However, the OLS model is not able to deal with selection on time varying unobservables. We therefore use an instrumental variable strategy to achieve an exogenous sorting of immigrants.

The strategy builds on the model developed in Andersson et al. (2018), but is based on the commonly used shift share strategy (see for example Altonji and Card, 1991, Card and DiNardo, 2000, and Sá, 2014). This approach uses the fact that migrants, to some degree, tend to locate

\(^{12}\) Areas with larger populations are more likely to attract more individuals, and hence also more immigrants, and might also affect, via an increased demand, housing prices.

\(^{13}\) Since we focus on the effects of refugee-migration on housing prices, it is important to control for all other, non-refugee, immigration to the neighborhood since it might be correlated with both the location of refugees and housing prices. Since many of the (refugee) immigrants end up in public rental apartments, it might be important to control for the number of public rental properties in each neighborhood.
in the same neighborhoods as earlier waves of migrants from the same country or ethnicity. We can therefore predict immigration \((im_{it})\) to neighborhood \(i\) in year \(t\) using information from an earlier period \((t^0)\). Assuming the reasons for choosing to live in neighborhood \(i\) in the earlier years, is not related to characteristics of the neighborhood in later years \((t)\), the predicted immigration \((\tilde{im}_{i,t})\) can be used as a valid instrument for the actual level of immigration \((im_{i,t})\).

The construction of the instrument is shown in detail in Equations (3) and (4). To get the predicted level of immigration into neighborhood \(i\) in time \(t\), we first count the number of immigrants, emigrating from country \(c\) to neighborhood \(i\) in the baseline period, \(t^0\). This number is then divided by the total number of immigrants to Sweden, coming from the same country, in the same period (c.f. Equation (4)). This term yields the share of immigrants from country \(c\) in year \(t^0\) who chose to locate in a certain neighborhood \(i\). The share is then multiplied by the total number of immigrants from country \(c\), coming to Sweden, in the later year we wish to study \((im_{c,SWE,t};\ c.f.\ Equation\ (3))\). The operation leaves us with a predicted number of migrants from country \(c\) to neighborhood \(i\), in time \(t\). This prediction reflects the number of migrants that would have settled in neighborhood \(i\), had the settlement patterns over different neighborhoods for country \(c\) been exactly the same in the later time period \((t)\) as in the earlier years \((t^0)\). The last step of the operation is to sum over countries, leaving us with a prediction of the total number of migrants to a given neighborhood and time period \((\tilde{im}_{i,t})\).

\[
\tilde{im}_{i,t} = \sum_c \tilde{im}_{c,i,t} = \sum_c \left( \phi_{c,i,t^0} \times im_{c,SWE,t} \right), \quad (3)
\]

where

\[
\phi_{c,i,t^0} = \frac{im_{c,i,t^0}}{im_{c,SWE,t^0}} \quad (4)
\]

Now, note that, since we make use of neighborhood fixed effects, identification stems from variation within neighborhoods over time. \(\phi_{c,i,t^0}\) is constant over time, implying that identification of our first stage requires the total inflow of immigrants from country \(c\) to Sweden \((im_{c,SWE,t^0})\) to
be varying over time.\footnote{Also, as can be seen in Equation (3), \(im_{c,SWE,t_0}\) must be larger than zero for the term to be mathematically defined. Hence, a requirement for a country to be included is that there were in fact positive immigration to Sweden in the baseline years.} If the yearly change is small or non-existent, we will not be able to separate the predicted neighborhood level immigration in \(t\) from that in \(t + 1\). As noted, this paper focuses on the effects of refugee migration, which, besides being an interesting group to study, turns out to be beneficial also for the identification. The number of refugees, and from which countries, changes heavily over time. Escalations (relaxations) in intensity of conflicts or in oppression will increase (decrease) the number of asylum seekers over time, often with drastic changes. It is to be expected that other forms of migration, such as labor, student and family migration is somewhat more consistent over time. Compared to other reasons for international migration, we also argue that the settlement of refugees is less driven by different pull factors of the neighborhood.

In addition, as described in Equations (3) and (4), the total inflow from country \(c\) in \(t\) is weighted, or scaled, with the baseline year \((t_0)\) share of neighborhood immigration from the same country \(c\). This means that the selection mechanism into neighborhoods in the baseline years becomes an issue. A core critique against the shift share instrument has therefore been that individuals, even in the baseline years, can select into neighborhoods on characteristics that are sluggish over time and unobserved. We therefore use 1990-1993, which are beneficial for us, since during these years Sweden had a refugee placement program in place. The program, in short, meant that an arriving refugee was not allowed to choose where to live, but were instead placed in one of the contracted municipalities in Sweden. Note that most Swedish municipalities were part of the program.\footnote{Longer, more elaborate descriptions of the policy has been provided in earlier research (see for example Edin et al., 2003 and Andersson et al., 2018).} Using the years of the program, we believe, takes care of some of the possible baseline years selection on sluggish neighborhood characteristics.\footnote{Two things are to be noted here: First, for the baseline years, 1990-1993, we unfortunately cannot access information on the reason for immigration. Rather, we use all immigrants coming these years from the refugee-countries defined in the later time period. This is a limitation since only refugees were part of the placement program. Second, there is some evidence that the policy program became less encompassing after 1992 and 1993 due to the large scale Yugoslavian immigration. We have still opted for using the years 1990-1993, and there are two reasons for this: First, large parts of the Yugoslavian immigration to Sweden is not part of the sample since most of the refugees from Yugoslavia came during the 1990’s and not in our later time period. Second, for the efficiency of the prediction, we wish to include as many observations as possible; when we estimate the first stage with only the years 1990 and 1991 we}
2.2 Econometric specification

Combining the specification in Equation (2) with the instrument described in Equations (3) and (4) gives the econometric specification that we will adopt in the analysis:

\[ \text{im}_{i,t} = \gamma \widetilde{\text{im}}_{i,t} + \sum_{p=1}^{3} \phi_p \text{pop}_p^{i,t-1} + \mathbf{X}\Gamma' + \mu_i + \tau_t + \epsilon_{i,t} \quad (5) \]

\[ \Delta \ln(P_{i,t+1}) = \beta_{IV} \widetilde{\text{im}}_{i,t} + \sum_{p=1}^{3} \delta_{p}^{IV} \text{pop}_p^{i,t-1} + \mathbf{X}\Pi^{IV'} + \eta_i + \lambda_t + \epsilon_{i,t+1} \quad (6) \]

Comparing the specification in Equation (6) with the specifications used in Saiz (2007) and Sá (2014), it can be noted that while the dependent variable is the same in all three papers (change in log-prices) there are differences in how the immigration and population variables have been treated. While we have chosen to enter the annual inflow of refugee immigrants as a separate variable and then to control flexibly for the population size, Saiz (2007) normalize the annual inflow of immigrants with the population size (and have no additional controls for population size) and Sá (2014) normalize the annual change in the number of foreign born with the population size (and have no additional controls for population size). To examine if our results are sensitive to the chosen specification, we will re-estimate our baseline regressions using the specifications used in Saiz (2007) and Sá (2014) (these results are presented in the Appendix).

Estimates for the first stage (Equation (5)) are presented in section 4.1, while estimates for the second stage (Equation (6)) are provided in section 4.2. Sensitivity analyses, in which we check how sensitive the baseline results are to alterations in the baseline specifications, are presented in section A.\(^{17}\)

\(^{17}\)An important sensitivity check is the implementation of the improvement to the shift-share instrument suggested by Jaeger et al. (2018). Shortly, one issue with the
Before we proceed to the results, we will describe the data to be used in the empirical analyses.

3 Data

We combine information from two different data sources; one of them provides us with important and detailed information about immigration to Sweden and the other provides us with information about housing objects sold and their prices. In this section, we will first describe the two data sources and then present some descriptive statistics on immigration and house prices.

3.1 Data sources

First, to measure immigration and neighborhood characteristics, we make use of GeoSweden, a large individual level database, covering all residents in Sweden\(^{18}\), spanning from 1990-2014. The data is based on official records, such as tax records and civil registration, and is administered by the Institute for Housing and Urban Research at Uppsala University.\(^{19}\) Besides including socio-demographic and socio-economic information, the data holds a number of fairly detailed properties useful for our purpose. For each individual, we have information on the year of immigration, as well as information on the country of birth and which country the individual emigrated from. Each individual is also connected to a specific neighborhood, called small area for market statistics (SAMS), which is based on the address registered with the tax authorities on December 31 each year. Sweden as a country has little over 9,000 SAMS, spread over 290 municipalities. The number of SAMS (on average about 30) in each municipality is strongly connected to the population of the municipality, as the number of SAMS increases almost linearly with the population of the municipality.\(^{20}\) The average SAMS-area has a population of around 1000 individuals, although this number varies a lot between the biggest one in the Stockholm region with over

---

18 All residents who have a registered address with the tax authorities.
19 The data has been compiled and made anonymous by Statistics Sweden.
20 The only deviation from the almost linear relationship is Stockholm, which has a lot fewer SAMS than would be expected based on its population.
10,000 inhabitants, to some of the rural ones, in extreme cases with less than 10 inhabitants. Our definition of neighborhoods is the SAMS.

Second, for house price variables we use a database that covers the majority of all housing sales in Sweden over the period 2005-2014. The house price data comes from Svensk Mäklarstatistik AB, and is based on information reported by real estate agents after the close of a sale. The data includes the sales of dwellings in housing cooperatives (mainly apartments) and privately owned houses.\(^{21}\) More specifically, we have information on daily house sales with coordinates, housing characteristics (e.g., living area, number of rooms, plot area, year built, monthly fee), list prices, final prices, and list and contract dates.

3.2 Sample restrictions

In terms of neighborhoods, we make two key restrictions to arrive at the final sample. First, since some neighborhoods have unusually few inhabitants, changes in immigration could numerically impact these communities in strange ways. We therefore drop all neighborhoods with fewer than 100 inhabitants. Second, since we need objects to be sold on the housing market to be able to examine the question of interest, we restrict the sample to neighborhoods where at least 2 housing objects have been sold each year. This selection of neighborhoods leaves us with a balanced panel of 4,298 neighborhoods over 12 years (2004-2015).\(^{22}\)

For refugee migration to Sweden, we need to make three types of restrictions. First, in our database we cannot identify all sending countries. Some refugees will emigrate from smaller (in terms of number of immigrants) countries, which will be part of larger “region-codes”, encompassing several countries. These are excluded. Second, for the instrument to work we need to have some immigration in the baseline period (1990-1993); c.f. Equation (4). Two countries fail/almost fail this restriction; Egypt and Eritrea, and are therefore also excluded.\(^{23}\) Third, a

---

\(^{21}\) According to Mäklarstatistik, the data covers about 80 percent of all housing sales made in Sweden. Note that these are transaction prices, we have no information on housing rents, which is studied in some previous papers (Tumen, 2016; Sanchis-Guarner, 2017). Given the Swedish system of rents based on centrally negotiated user values, it is not clear that increased immigration would effect the level of rental prices.

\(^{22}\) Note that house prices exist for 2005-2015, and that we use data for 2004-2014 in the immigration database.

\(^{23}\) The former because immigration was zero, the latter because the state of Eritrea was not formally recognized until 1993 (and hence we have no information on immigration from Eritrea before mid-1993). 27 immigrants registered from Eritrea this year is simply too small a number to make reliable predictions.
smaller number of refugees will immigrate from OECD-countries. Some of these are likely Dublin-cases, and not initially from the country listed in the data. These are therefore also excluded.\textsuperscript{24} Comparing with aggregate figures from the Swedish Migration Agency, the 120,000 refugees we use to construct the immigration variable, \(im_{it}\), constitute about 2/3 of all refugee migration to Sweden during the period 2005-2014.

Regarding housing prices, we make a restriction to improve on the precision of the coordinates provided by the real estate agents. When reporting sold objects into the housing database, the agents write the exact address but use a drag-and-drop tool on a digital map to get the coordinates for the housing object. This implies that the coordinates provided by the agents may not be exactly correct. To improve on this, the coordinates provided by Google for each address given by the agents were collected and then compared with the coordinates given by the agents. We drop cases where coordinates differ by more than 100 meters, and use the coordinates given by Google for other observations.

3.3 Description of the variables in the empirical model

Table 1 describes the variables to be used in the empirical model. Starting with the main dependent variable, it can be noted that the median housing prices increases with approximately 6 percentage points any given year. This change hides considerable differences, with some, smaller neighborhoods experiencing drastic changes with between 200 and 300 percent over a year. These highly unusual numbers are not very informative for our purpose, and to avoid that outliers become the prime driver of the effects, we “winsorize” the dependent variable, censoring the top and bottom 1 percent of the distribution (\(p_{1}\) and \(p_{99}\)). In other words we put \(\Delta \ln(price) = p_{99}\) if \(\Delta \ln(price) > p_{99}\), and \(\Delta \ln(price) = p_{1}\) if \(\Delta \ln(price) < p_{1}\). The censored numbers are found in the same Table, and show a larger mean, with much less pronounced max and min values. In the results section, all price regressions use this censoring method.\textsuperscript{25}

Measured in actual prices, the median final \(m^2\) price is 15,622 SEK, which is slightly higher than the median list price. An object stays on the market for on average 35 days, and each neighborhood has on average a little more than 20 objects for sale per year. The latter variable shows a rather high standard deviation relative to the mean, which is

\begin{footnotesize}
\textsuperscript{24} According to the Dublin regulations, within the EU, the country a person is seeking asylum in first is also the country responsible for the application.
\textsuperscript{25} That said, the results are not dependent on the censoring. Not performing this transformation to the data leaves us with the same take-away as given by Table 5.
\end{footnotesize}
reasonable given that the variable is not standardized by the population of the neighborhood.

Turning to the main explanatory variable and the covariates, we find that, on average, 1.5 refugees immigrate to a neighborhood per year. This number does however hide a substantial amount of zeros, since a fairly large portion of the neighborhoods have no refugee immigration at all. The average population amounts to approximately 1,600 individuals, which makes our neighborhoods smaller than the neighborhoods used in studies on, for example, US or UK data. Lastly, it can be noted that the non-refugee migration is, on average, higher than the refugee-related migration. This is not surprising, since the non-refugee migration comprise both tied family members as well as work- and student-related migration.
Table 1. Descriptive Statistics

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>47,278</td>
<td>42,980</td>
<td>36,242</td>
<td>47,278</td>
<td>42,980</td>
</tr>
<tr>
<td>mean</td>
<td>15,622</td>
<td>0.06</td>
<td>0.05</td>
<td>21.82</td>
<td>-0.83</td>
</tr>
<tr>
<td>sd</td>
<td>10,485</td>
<td>0.24</td>
<td>0.29</td>
<td>35.17</td>
<td>42.04</td>
</tr>
<tr>
<td>min</td>
<td>229.4</td>
<td>-3.12</td>
<td>-3.13</td>
<td>2</td>
<td>-1,265</td>
</tr>
<tr>
<td>max</td>
<td>99,955</td>
<td>3.01</td>
<td>3.85</td>
<td>2</td>
<td>1,284</td>
</tr>
</tbody>
</table>

**DEPENDENT VARIABLES**

- Final price/m² (p50, SAMS) 47,278 15,622 10,485 229.4 99,955
- Δln(Final price/m²) (p50, SAMS) 42,980 0.06 0.24 -3.12 3.01
- Δln(Final price/m²) (Winsorized) 42,980 0.17 0.31 -0.63 0.73
- List price/m² (p50 in SAMS) 41,577 15,414 11,598 113.6 1.202e+06
- Δln(List price/m²) (p50, SAMS) 36,242 0.05 0.29 -3.13 3.85
- Objects on market 47,278 21.82 35.17 2 723
- ΔObjects on market 42,980 1.08 9.59 -151 226
- Days on market (p50, SAMS) 46,671 34.76 37.17 2 1,284
- ΔDays on market (p50, SAMS) 42,104 -0.83 42.04 -1,265 1,265

**IMMIGRATION VARIABLES**

- $im_{it}$ 47,234 1.528 5.837 0 298
- $\hat{im}_{it}$ 46,893 1.534 6.654 0 397.7

**COVARIATES**

- Population 47,234 1.639 1.571 114 21,173
- Disposable income 47,234 202,500 52,820 -73,000 1,036,700
- Social assistance 47,234 799 1,098 0 13,770
- Public rental estates 47,234 2.818 5.547 0 115
- Non-refugee immigration 47,234 10.54 22.87 0 720

**Notes:** Mean and standard deviations of final price and list price per m², as well as days on the market, are based on the median of all prices/days on the market in a SAMS a given year. Days on market shows number of days from the ad first appears on the sales website to the signing of the contract. $im_{it}$ is the endogenous refugee immigration and $\hat{im}_{it}$ the prediction based on the strategy laid out in section 2. Disposable income and social assistance is the average in Swedish SEK (1 $ \approx 7-9$ SEK). Social assistance is municipal aid to individuals with no income (not counting unemployment benefits). Public rental estates are all municipality owned rental buildings and non-refugee immigration shows all other immigration (with a residence permit), be it family, work or student related.
3.4 Immigration: Some further descriptions

To shed some further light on the type of immigration we are using, Table 2 lists the top ten sending countries in our sample between 2004 and 2014. More than 70 percent of the sample comes from three countries: Syria, Iraq and Somalia. The largest group by far are Syrians, with a clear majority arriving in 2012-2014, following the Syrian war. Immigration from Iraq and Somalia is somewhat more consistent over the studied time period, but also holds clear peaks; most of the Iraqi refugees arrived in 2006 and 2007, when the Iraqi conflict escalated, and similarly a disproportional share of Somalis arrived in 2009-2010. Besides the three big sending countries, there has also been large refugee immigration from Afghanistan and Iran. In total, the sample consists of 33 sending countries; these are listed in the note to Table 2.

<table>
<thead>
<tr>
<th>Emigration country</th>
<th>Freq.</th>
<th>Percent (of all refugees in sample)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syria</td>
<td>37,761</td>
<td>32%</td>
</tr>
<tr>
<td>Iraq</td>
<td>27,529</td>
<td>23%</td>
</tr>
<tr>
<td>Somalia</td>
<td>20,589</td>
<td>17%</td>
</tr>
<tr>
<td>Afghanistan</td>
<td>9,675</td>
<td>8%</td>
</tr>
<tr>
<td>Iran</td>
<td>9,275</td>
<td>8%</td>
</tr>
<tr>
<td>Russia</td>
<td>2,934</td>
<td>2%</td>
</tr>
<tr>
<td>Lebanon</td>
<td>2,658</td>
<td>2%</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>1,600</td>
<td>1%</td>
</tr>
<tr>
<td>Thailand</td>
<td>1,229</td>
<td>1%</td>
</tr>
<tr>
<td>Bosnia</td>
<td>1,153</td>
<td>1%</td>
</tr>
<tr>
<td>Other</td>
<td>4,646</td>
<td>4%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>119,049</td>
<td>100%</td>
</tr>
</tbody>
</table>

Note: “Other” includes (in order of frequency size): Uganda, Pakistan, Colombia, China, India, Bangladesh, Tunisia, Morocco, Romania, Bolivia, Vietnam, Algeria, Poland, Gambia, Croatia, Peru, Sri Lanka, Philippines, Bulgaria, Estonia, Brazil, Latvia, Argentina. Note also that the frequency in this Table reflects all refugees that came to Sweden from the above mentioned countries, during 2004-2014. The numbers will therefore not exactly match those used in the regressions, since a number of restrictions are made on the number of housing units sold in each SAMS per year, effectively dropping a number of SAMS. The total frequency of refugees will hence be somewhat smaller when running the regressions.

As noted, our geographical unit to capture neighborhoods is the SAMS, which in our sample has on average around 1,600 inhabitants. The small size of the units means that a sizable share of neighborhoods
will receive few or no refugees. Of the 4,298 neighborhoods identified in the sample, about 30 percent has zero refugee migration, and hence no variation over time. A smaller number of around 100 SAMS, have consistent, positive immigration for all eleven years. If we focus only on neighborhoods with any refugee immigration over time, the mean refugee immigration is still modest, with around 2 refugees per year.

3.5 Housing prices: Some further descriptions

For analysis, we make use of the yearly median level of neighborhood log prizes per m². For a better understanding of the price-spread, the distribution of nominal m²-prices is given in Figure 2; Figure 2a shows the relative distribution and Figure 2b shows the cumulative distribution. For most of Sweden the distribution is fairly compressed, with a clear majority of neighborhoods (around 75 percent) holding median m² prices under 20,000 SEK (around $2,500). The top segment of prices is largely concentrated to the Stockholm metropolitan area and to some extent Gothenburg. For example, more than 80 percent of the top five percent of the distribution (median price >35,000 SEK) is found in the Stockholm metropolitan area or Gothenburg. The top 1 percent (median prices >51,000 SEK) is almost exclusively sold in the Stockholm city center.
Figure 2. Housing prices per m$^2$ in Sweden, 2005-2015
(a) Distribution of housing prices 2005-2015

(b) Cumulative distribution

Note: The figures show the distribution of median SAMS-level housing-prices. The time span is 2005-2015.
Source: Own calculations based on data from the housing price database.
While the most expensive housing units are found in only a few cities, it is also of interest to understand to which geographic locations the immigrants arrive. Do they for example exclusively settle in relatively poor neighborhoods? To get a sense of this, we plot log housing prices per m², against refugee immigration (c.f. Figure 3a, where the horizontal red line indicates the mean sample value of log prices). While it is clear that more immigrants arrive to neighborhoods with lower than average housing prices, it can also be seen that refugee immigration takes place over the entire range of prices.

Examining the relationship between the change in log housing prices between \( t \) and \( t - 1 \) and the inflow of refugee immigration in \( t - 1 \) (and hence getting closer to our econometric specification), we get a slope coefficient, from a linearly fitted regression line, that is close to zero (0.0007); c.f. Figure 3b. This indicates only minor changes on housing prices due to refugee migration (and, if anything, a positive relationship). As we will see in the coming sections, this tentative conclusion is very similar to the results obtained in the 2SLS regressions.
Figure 3. Immigration and housing prices

(a) Log prices and immigration.

(b) Change in log prices ($\Delta t$) and immigration in $t-1$, including a fitted regression line.

Note: Figure (a) shows the correlation between refugee immigration and log housing prices in the neighborhood. The red solid line, representing the mean log price in the sample (9.44). Figure (b) shows the relation between the change in log prices between $t$ and $t-1$ and immigration in $t-1$. The solid blue line is a linearly estimated regression line, with a slope of 0.0006. Time period: 2005-2015.

Source: Own calculations based on data in GeoSweden and in the housing price database.
4 Results

In this section, we will present our results. In section 4.1 we show the first stage, followed by section 4.2, where we present the (baseline) 2SLS results on housing prices. In sections 4.3 and 4.4 we study effects on other outcomes, such as housing supply, list prices and days on market, and in section 4.5 we discuss our results.

4.1 First stage

Table 3 shows the first stage estimates, where the immigration to neighborhood \( i \) in year \( t \) (\( im_{i,t} \)) is regressed on the predicted level of migration to the same neighborhood and year (\( \hat{im}_{i,t} \)). While column (1) presents the results from a linear regression with no additional covariates and no fixed effects, column (2) presents the results from the preferred specification as given in Equation (5) (i.e., augmented with neighborhood and time fixed effects and with the time-varying covariates). In all estimations, standard errors are clustered on neighborhood level (SAMS).\(^{26}\)

The coefficient of the preferred specification in column (2) means that, on average, an increase of 1 predicted refugee to neighborhood \( i \) is associated with 0.218 more actual refugees in the very same neighborhood. Hence, there is a strong serial correlation over time, where refugees tend to locate in the same neighborhoods as individuals from the same country did, many years earlier. The coefficient is highly significant, and the F-statistic is high enough for us to conclude that we do not suffer from weak instruments. Comparing the estimates for the \( \hat{im}_{i,t} \)-variable in columns (1) and (2) shows that the preferred specification yields a point estimate that is fairly similar to the correlation in the parsimonious linear specification.

\(^{26}\)We have also redone the first stages and the baseline second stage (Table 5) with standard errors clustered on municipality level, giving no important changes to the results. Results are available upon request.
Table 3. First stage estimates (c.f. Equation (5))

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{im}_{i,t}$</td>
<td>0.325***</td>
<td>0.218***</td>
</tr>
<tr>
<td></td>
<td>(0.0430)</td>
<td>(0.0443)</td>
</tr>
<tr>
<td>Observations</td>
<td>46,893</td>
<td>42,630</td>
</tr>
<tr>
<td>Number of neighborhoods</td>
<td>4,263</td>
<td>4,263</td>
</tr>
<tr>
<td>F-Stat</td>
<td>57.3</td>
<td>24.1</td>
</tr>
<tr>
<td>Neighborhood FE</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>Year FE</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>Covariates</td>
<td>NO</td>
<td>YES</td>
</tr>
</tbody>
</table>

Notes: Significance levels: *p<0.1, **p<0.05, ***p<0.01. Standard errors clustered on SAMS-level. Covariates include population, population$^2$, population$^3$, average disposable income, average social assistance, non-refugee migration and number of public rental properties. All covariates defined in $t-1$.

While the baseline first stage estimate in column (2) in Table 3 provided a strong first stage, we continue with examining the robustness of this estimate by re-estimating the first stage equation on different subsamples of the full sample. The results from these robustness analyses are presented in Table 4. First, in column (1), we drop the two largest municipalities/cities in Sweden, Stockholm and Gothenburg. In these cities the number of neighborhoods in different ways depart significantly from those in other municipalities.\textsuperscript{27} Second, in the last two columns we drop the least populated neighborhoods in the country (column 2) as well as the most populated neighborhoods (column 3). Looking at these first stage estimates, we note that neither Stockholm, Gothenburg, nor the least populated neighborhoods are essential for the robustness of the first stage. The largest neighborhoods do however play a more important role; when excluding the neighborhoods with the 25% largest populations, the first stage point estimate drop to 0.09 (although still highly significant). There could be several explanations for this. First, immigration to more populated areas is likely more consistent over time. Second, as noted in the descriptive statistics, a fairly large number of SAMS have no immigration at all over the time period in question. These neighborhoods are naturally more common in the lower quartiles of the population distribution. We therefore continue with the estima-

\textsuperscript{27}Stockholm has very few neighborhoods (SAMS-areas), given the population being much larger than in any other city in Sweden. Gothenburg has around 800 SAMS-areas, which is many times more than any other municipality in Sweden.
tion of the second stage, noting that the first stage is strong, but that an important part of the variation stems from the more populated neighborhoods.

Table 4. First stage estimates: Robustness tests

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Drop Sthlm and Gbg</td>
<td>Drop n’hoods with small pop</td>
<td>Drop n’hoods with large pop</td>
</tr>
<tr>
<td></td>
<td>im(_{i,t})</td>
<td>(smallest 25%)</td>
<td>(largest 25%)</td>
</tr>
<tr>
<td>(\hat{im}_{i,t})</td>
<td>0.196*** (0.0559)</td>
<td>0.216*** (0.0452)</td>
<td>0.0855*** (0.0293)</td>
</tr>
<tr>
<td>Observations</td>
<td>37,640</td>
<td>35,252</td>
<td>30,021</td>
</tr>
<tr>
<td>Number of n’hoods</td>
<td>3,764</td>
<td>3,612</td>
<td>3,126</td>
</tr>
<tr>
<td>F-Stat</td>
<td>12.25</td>
<td>22.84</td>
<td>8.52</td>
</tr>
<tr>
<td>Neighborhood FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Year FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Covariates</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

Notes: First stage robustness regressions. Note that Column 1 excludes the two largest municipalities/cities in Sweden: Gothenburg (“Gbg”) and Stockholm (“Sthlm”). See Table 3 for information on covariates used. Significance levels: *p < 0.1, **p < 0.05, ***p < 0.01. Standard errors clustered on SAMS-level.

4.2 Baseline results: Effect of refugee migration on house prices

Next we turn to the 2SLS estimation of the baseline model to get estimates of \(\beta^{IV}\) (c.f. Equation (6)). These results are presented in column (2) in Table 5. We however start by showing the OLS results in column (1), where we regress the change in log final prices between \(t + 1\) and \(t\) on the endogenous refugee immigration in \(t\). The specification further includes the full set of neighborhood and time fixed effects as well as the time-varying covariates. The estimate is insignificant and close to zero. The literal interpretation of the coefficient is that, conditional on the covariates, a year with one more refugee migrant to a neighborhood than on average, is associated with a change in house prices of 0.03 percentage points.

Since the OLS-estimate cannot be considered to provide an estimate of the causal relationship between immigration and housing prices, we turn to the 2SLS estimates. From column (2) it is however clear that the coefficient remains small and insignificant. The coefficient seen in
column (2) is in fact even smaller than the OLS, representing a 0.008 percentage point rise in prices in response to an increase of one refugee.\footnote{The results in Table 5 are censored, as described in Section 3. Re-estimating the results without censoring does, however, not change the results. This also holds if we cluster standard errors on municipality level instead of the SAMS.}

Since it is possible that the mechanisms dictating apartments (condominiums) and owner-occupied houses are different, we also provide results separated for these groups.\footnote{For example, one possible channel is a rental crowding out effect. According to this, immigrants occupy rental apartments that would otherwise have been rented by natives, causing the latter group to buy a place instead of renting, and thereby to increase prices. Should this logic be present, we likely expect the effect to be more pronounced for condominiums, which is a more relevant substitute for a rental apartment.} As can be seen from columns (3) and (4), we do however not detect any big difference between the two types of housing. In both cases, we estimate an insignificant effect, with a point estimate very close to zero.

Table 5. Baseline results: Effects of immigration in $t$ on housing prices in $t + 1$ (c.f. Equation (6))

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1) OLS</th>
<th>(2) 2SLS</th>
<th>(3) 2SLS</th>
<th>(4) 2SLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIABLES</td>
<td>All</td>
<td>All</td>
<td>Condominiums</td>
<td>Owner-occupied houses</td>
</tr>
<tr>
<td>$im_{i,t}$</td>
<td>0.000355</td>
<td>8.47e-05</td>
<td>-0.000306</td>
<td>-0.000118</td>
</tr>
<tr>
<td>(0.000236)</td>
<td>(0.000733)</td>
<td>(0.000914)</td>
<td>(0.000871)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>42,940</td>
<td>42,630</td>
<td>42,630</td>
<td>34,093</td>
</tr>
<tr>
<td>Number of n’hoods</td>
<td>4,294</td>
<td>4,263</td>
<td>4,263</td>
<td>3,617</td>
</tr>
<tr>
<td>Neighborhood FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Year FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Covariates</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

Note: Effects of immigration in $t$ on $\Delta$ log prices in $t + 1$ (according to Equation (6)). Column (3) and column (4) runs the second stage separately for condiminiums and owner-occupied housing. See the notes to Table 3 for covariates and other details.

To examine if the baseline results we have obtained in this section are sensitive to different alterations of the econometric specification, we have put it to test by (i) estimating the model using the IV-estimator suggested by Jaeger et al. (2018) (which they argue is an improvement of the traditional shift-share instrument), (ii) estimating a model with immigration normalized by the population size (using econometric specifications similar to those used by Sá (2014) and Saiz (2007)), (iii) controlling for covariates and population size in $t$ and $t + 1$ and (iv) estimating a model in which we have calculated hedonic prices to control for more characteristics than just the size of the housing unit. It turns out
that the results obtained stay small and insignificant for most of these alterations.\footnote{When using the econometric specifications in Sá (2014) and Saiz (2007), we actually get (insignificant) positive estimates. If anything, a couple of sensitivity results would hence suggest a positive effect of immigration on neighborhood prices. These sensitivity results are presented in the Appendix.}

The results in this section hence indicate no effect of refugee immigration on housing prices, despite potential downward pressure on prices following native outmigration among those with above average education and earnings as found by Andersson et al. (2018). These results stand in contrast to the negative effects found in studies on US (Saiz and Wachter, 2011), UK (Sá, 2014) and Italian data (Accetturo et al., 2014). We return to why this might be in Section 4.5. Before that, we use some of the detailed information available in the database, and study the effect of immigration on other related housing characteristics.

4.3 The effect of immigration on housing supply

As discussed in the introduction, there can be an offsetting effect on prices from demand if immigration has an effect on housing supply. To examine if this is the case, we estimate the baseline model (as given in Equations (5) and (6)) but use change in the supply of housing as the dependent variable. To be exact, we use two different measures. The first is calculated from the house price database and measures the change in total number of objects on the market in each neighborhood and year. The second is constructed from information in our database GeoSweden and measures the change in the number of housing estates that turn up in the registers each year. This second measure covers all estates in each neighborhood (rental apartments as well privately owned apartments and houses), but only provides a measure of all estates, but not the exact number of apartments.\footnote{The id-number is on each estate, but we have no information on how many apartments there are in each estate with rental or private apartments.}

When running the baseline specifications with the two variables measuring new housing construction as dependent variables, we get positive coefficients when using the number of objects on the market (c.f. Table 6), indicating that immigration has a stimulating effect on supply. This is in line with parallel results found in Fernández-Huertas Moraga et al. (2017), but also with native flight found in Andersson et al. (2018). In the latter, it is found that immigration causes increased outmigration specifically by those owning a house or an apartment. Given an in-
creased flight of housing owners, it should naturally follow that more objects than usual are sold on the local housing market.

Table 6. Effects of immigration in $t$ on sales and new housing construction in $t + 1$

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2SLS</td>
<td>2SLS</td>
</tr>
<tr>
<td>$\Delta$Houses</td>
<td>0.143**</td>
<td>-0.0842</td>
</tr>
<tr>
<td></td>
<td>(0.0727)</td>
<td>(0.0577)</td>
</tr>
<tr>
<td>Observations</td>
<td>42,630</td>
<td>38,367</td>
</tr>
<tr>
<td>Number of neighborhoods</td>
<td>4,263</td>
<td>4,263</td>
</tr>
<tr>
<td>Mean of dep var</td>
<td>1.08</td>
<td>1.87</td>
</tr>
<tr>
<td>Neighborhood FE</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Year FE</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Covariates</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

Notes: Column (1) measures the effect on the change in the number of houses for sale on the market. Column (2) measures the effect on the change in the total number of estates/properties. For other notes, see notes to Table 3.

4.4 Expected effects of immigration on local amenities (effects on list prices and time on market)

Another potentially interesting mechanism might emanate from expectations on the side of the actors on the housing market. If the actors think that the new inflow of immigrants affects local amenities in a way that might capitalize into housing prices, they might react already on these expectations. To examine this, we look at the effects of immigration on initial list-prices. If we see a significant and negative effect already in this stage, this indicates that immigration has an effect on the expectations in the neighborhood, which might translate into lower final prices.

The first column in Table 7 shows the effect of immigration on list prices. It is clear that, if anything, immigration seems to have a positive effect on expectations in the sense that it has an upward pressure on list prices. The effect is however both small and insignificant.

List prices say something about the expectations from the sellers’ point of view. To bring the buyers and their potential expectations into picture, we examine the effects of immigration on the days on market. If immigration has a positive effects on the number of days an housing
object is on the market, this might indicate that the buyers are more hesitant to buy because their expectations about housing prices have been affected by the immigration. From the results, presented in the second column in Table 7, we do find a slightly positive effect. The effect is, however, not statistically significant, and we hence find no clear support for buyers being more hesitant as a response to increased immigration.

Table 7. Effects of immigration in \( t \) on list prices and time on market in \( t + 1 \)

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1) ( \Delta \ln(\text{list price}_{t+1}/\text{m}^2) )</th>
<th>(2) ( \Delta \text{DoM} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{im}_{it} )</td>
<td>0.000491 (0.00123)</td>
<td>0.0315 (0.134)</td>
</tr>
<tr>
<td>Observations</td>
<td>46,893</td>
<td>41,755</td>
</tr>
<tr>
<td>Number of n’hoods</td>
<td>4,263</td>
<td>4,263</td>
</tr>
<tr>
<td>Mean of dep var</td>
<td>0.30</td>
<td>-0.83</td>
</tr>
<tr>
<td>Neighborhood FE</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Year FE</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Covariates</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

Notes: Column (1) gives the effect on change in list prices. Column (2) estimates the effect on the change in the median number of days on the market (between ad posted and final sales). For other notes, see the notes to Table 3.

4.5 Discussion: How can we understand a null effect of immigration on housing?

Our analysis so far has revealed that refugee immigration into a small neighborhood causes no clear effect on neither final nor list housing prices on a one-year basis. The remaining question is then how we should understand such a result? This question becomes particularly pressing given the results seen in earlier research (Sá, 2014; Saiz and Wachter, 2011; Accetturo et al., 2014).

As already discussed, there are a number of factors potentially pulling prices downward. Natives can have preferences for living with other natives, or, framed in a different manner, preferences for not living with immigrants. Associated with this mechanism is also an expectation that local amenities, or public goods, drop in quality as a response to in-
creased immigration. Both mechanisms would lead to fewer bidders, causing prices to drop.

A first hypothesis would hence be that the factors pushing the prices down are less pronounced in the Swedish setting. This argument is not without merit. As few as 1.5 percent of Swedish respondents answered that they wished not to have an immigrant or a foreign worker as a neighbor, when asked in the World Value Survey of 2005-2009.32 This number can be compared to 13.9% for Italy, 14.2% for the UK and 12.7% for the US If Swedes are more tolerant to living in heterogeneous neighborhoods, this would explain the modest effects found in this study. Also, in the already mentioned companion paper by Andersson et al. (2018), flight was estimated for a sub-group of the population, home owners. For the population at large, the authors found no effects on native migration from increased immigration.

A second hypothesis would instead be that, assuming mechanisms putting downward pressure on prices are the same in Sweden as in other countries, the factors pulling prices upwards are relatively stronger. There are factors opposing this story as well as speaking in favor of it.

In this paper we have taken an explicit focus on refugees, as opposed to immigrants as a general group. In Sweden, refugees stay almost exclusively in the rental segment, where the rental price is often not determined through a market procedure.33 When we turn to our individual-level data in GeoSweden, where we can observe the type of housing each individual in Sweden lives in, we find that a small fraction, (approximately 9% when looking at the data for the years 2004-2013) of the immigrants live in either an owner occupied house or a condominium the year after arrival. The rest are registered to different forms of rental buildings. Note also that the number actually owning a place, is most likely only a fraction of the nine percent. The database tells us an individuals’ officially registered address, however, this include individuals who rent owner occupied houses or condominiums as well. All in all, only a small fraction ends up in the privately owned segment of the housing market, indicating that a positive push on prices from the immigrants’ side should be very small.34

---

32 This is the fifth wave of the survey. Documents showing world value survey data is found at [http://www.worldvaluessurvey.org/WVSContents.jsp](http://www.worldvaluessurvey.org/WVSContents.jsp).
33 To get an apartment in the dominating public rental sector, there is a queuing system (in which the first individual in the queue is offered an apartment as soon as one becomes available). However, since the municipalities are responsible for providing refugees with housing, refugees can sometimes bypass the queue for municipal rentals.
34 Also, when looking in our data, we note that very few individuals have any labor earnings the year after arrival; the median labor income is 0 and very few have
Immigrants are, however, not the only group putting upward pressure on prices. Since we know that refugees largely occupy rental apartments on their first year in the country, a side effect is that fewer rentals will be accessible for natives. Combining this feature with a municipal responsibility to house refugees, means that a year with unusually many refugees, is followed by a year with unusually few rental apartments to move into. This “crowding out” effect might cause some natives to buy instead of rent, pushing prices upwards. It is, however, not necessarily the case that this crowding out effect would be bigger or more pronounced in Sweden than in other countries. Since we do not know how many rental apartments that are constructed, it is hard to fully evaluate this claim.\textsuperscript{35}

5 Conclusions

In this paper we estimate causal effects of refugee immigration on neighborhood level housing prices. We use high quality, individual level register data from Sweden, studying the period 2004-2015. To construct exogenous variation in the geographical sorting of migrants, we use an instrument developed in Andersson et al. (2018). The instrument predicts contemporary refugee migration using neighborhood sorting in 1990-1993, a period during which Sweden used a refugee dispersal program, which placed individuals in contracted municipalities.

The main conclusion of the paper is that there is no effect of refugee immigration on local neighborhood prices. The results are robust to several alterations of the model, such as using hedonic price functions and adopting the instrumental variable approach suggested in Jaeger et al. (2018). Also, we find no indications of immigration affecting neither list prices nor how many days an object stays on the market.

We do find a positive effect on housing supply, measured as the change in number of objects on the market. We explain this with native flight among home owners, which has been documented in Andersson et al.

\textsuperscript{35}It should be noted that another possible mechanism to explain the results could work through a selection of more expensive apartments on the market. If immigration causes native flight of individuals with the highest socio-economic status, this implies that a year with unusually large immigration will be followed by unusually many above average priced housing units on the market. If such a mechanism was the main driver, we would, however, expect the estimate to change when applying hedonic price equations (taking the quality of the apartments and houses into consideration). However, as seen from the sensitivity analysis in the Appendix, the results remain stable even after applying hedonic price equations.
If more owners of apartments and houses move out of the neighborhood in the year following an immigration shock, we would of course also expect there to be more objects for sale on the market.

The null effect of refugee immigration on local (neighborhood level) housing prices stand in contrast to the negative effects found in the earlier literature, primarily using data from the US (Saiz and Wachter, 2011), the UK (Sá, 2014) and Italy (Accetturo et al., 2014). We hypothesize that this can be either because the channels that intermediates negative effects of immigration on housing prices are less pronounced in Sweden, or that the channels causing a positive effect are more pronounced. To support the former story, we for example show aggregate statistics from world value survey, suggesting Swedes are less averse to living close to immigrants, compared to most other populations of the world. For the latter story, we suggest a “crowding-out” mechanism, which entails that arriving refugees occupy the largely inelastic rental stock, leading natives to buy instead of rent, in turn causing prices to rise. The combination between a municipal responsibility to accommodate refugees and a large publicly owned rental stock could possibly enhance the crowding-out mechanism in the Swedish case.

Our results suggest further research is needed to understand the effect of immigration on housing prices. For example, given the non-negative results in this paper, it would be interesting to set up comparisons between different housing regimes. Cross-country, or state-wise comparison could be a fruitful way forward.
References


Appendices

In this appendix we present results from sensitivity analyses. To examine if the baseline results are sensitive to different alterations of the econometric specification, we have (i) used an alternative IV-estimator, (ii) adopted the specifications used in Saiz (2007) and Sá (2014), (iii) dated our covariates differently, and (iv) constructed housing prices based on estimated hedonic housing price functions.

A Using IV-estimator suggested by Jaeger et al. (2018)

In Jaeger et al. (2018), the shift share instrument is criticized for failing to account for dynamic developments over time. The example in their paper is based on the labor literature, and therefore focuses on the development of wages. Theoretically, however, one could worry about similar mechanisms in the case of housing prices.

Imagine an immigration shock in $t_0$ to neighborhood $i$. Due to a short term demand effect housing prices in $i$ increase. Now, the higher prices, and larger presence of immigrants in itself, cause some natives to opt out of the neighborhood, which sets prices off to a slow path of decrease again. If we use a shift share instrument, we will make use of the serial correlation over time in immigrants’ residential location patterns. Moving forward to $t$, the year in which we measure the effect of immigration on prices in the same neighborhood $i$, we will now measure the short term effect of immigration as well as the slowly reversing dynamic process that was set in motion in the baseline year.

The solution put forward by Jaeger et al. (2018) is to add an extra lag to the model, and estimate the effect of both last year’s effect ($t - 1$) as well as the effect in the current year ($t$). Since both are endogenous, we run two first stages:

$$im_{i,t} = \gamma_{1,1} \tilde{im}_{i,t} + \gamma_{1,2} \tilde{im}_{i,t-1} + \sum_{p=1}^{3} \phi_p \text{pop}_i \tilde{p}_t \tilde{t} - 1 + X\Gamma' + \mu_i + \tau_t + \epsilon_{i,t}$$

(7)
\[ i m_{i,t-1} = \gamma_{2,1} \tilde{i m}_{i,t} + \gamma_{2,2} \tilde{i m}_{i,t-1} + \sum_{p=1}^{3} \phi_p pop_{i,t-1}^{p} + X \Gamma' + \mu_i + \tau + \epsilon_{i,t} \]

The two first stages in equations (7) and (8) in turn gives the following second stage equation:

\[ \Delta \ln(P_{i,t+1}) = \beta_{1IV} \tilde{i m}_{i,t} + \beta_{2IV} \tilde{i m}_{i,t-1} + \sum_{p=1}^{3} \delta_p IV pop_{i,t-1}^{p} + X II IV' + \eta_i + \lambda_t + \epsilon_{i,t+1} \]

Table A.1 puts our baseline estimate (c.f. column 2 in Table 5) to the test of the method developed in Jaeger et al. (2018). The first column uses the refugee immigration that we use throughout this paper (and hence in column 2 in Table 5). In the middle (equation (7)) and bottom (equation (8)) panel we can see that the first stages are significant. The second stage, presented in the upper panel, shows the same results as in the baseline estimations; the point estimates are close to zero and statistically insignificant. Also, there is no effect of the lagged immigration, showing very little evidence for a (long-lasting) dynamic process. In column 2, we extend our analysis to all immigration; this means adding several countries which are not refugee sending countries, as well as including labor migrants, students and family migrants. Also when using this more extended definition of immigrants do we get a point estimate that is close to zero and statistically insignificant. We conclude that the results are stable to using the model suggested in Jaeger et al. (2018).
Table A.1. *Estimating effect of immigration in $t$ and $t-1$ on final log price in $t+1$. 2SLS according to Jaeger et al. (2018)*

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refugee Migration</td>
<td>0.00127</td>
<td>0.00105</td>
</tr>
<tr>
<td>All immigration</td>
<td>(0.000950)</td>
<td>(0.000997)</td>
</tr>
<tr>
<td>$im_{it}$</td>
<td>-0.00201*</td>
<td>-0.000477</td>
</tr>
<tr>
<td></td>
<td>(0.00105)</td>
<td>(0.000449)</td>
</tr>
<tr>
<td>$im_{it-1}$</td>
<td>-0.201*</td>
<td>-0.0477</td>
</tr>
<tr>
<td></td>
<td>(0.105)</td>
<td>(0.0449)</td>
</tr>
</tbody>
</table>

**SECOND STAGE (Effect on final $\Delta \text{log m}^2$ price)**

**FIRST STAGE, effect on $im_t$**

| $im_t$                      | 0.273***                   | 0.229***                   |
|                            | (0.0558)                   | (0.0703)                   |
| $im_{t-1}$                 | -0.183***                  | -0.169                     |
|                            | (0.0399)                   | (0.107)                    |
| F-Stat                     | 14.27                      | 5.28                       |

**FIRST STAGE, effect on $im_{t-1}$**

| $im_t$                      | -0.152*                    | -0.177***                  |
|                            | (0.0558)                   | (0.0817)                   |
| $im_{t-1}$                 | 0.444***                   | 0.565***                   |
|                            | (0.110)                    | (0.0612)                   |
| F-Stat                     | 10.53                      | 82.39                      |

Observations: 38,367, 38,367
Number of n’hoods: 4,263, 4,263
Neighborhood FE: YES, YES
Year FE: YES, YES
Covariates: YES, YES

Notes: Second stage and first stage, effect of immigration on $\Delta \text{log contract prices}$. 2sls-method in accordance with the method developed in Jaeger et al. (2018). Column (1) uses the same sample as in the baseline regression. Meaning that we focus exclusively on refugees. In column (2) we increase the sample to include other immigration as well, including non-western, family, labor market and student. Note that this also expands the number of countries. Besides the ones listed in Table 2, we also add immigration from the Nordic countries, Belgium, Greece, UK, Germany, Spain, France, Turkey, US, Chile and Australia. For specification of covariates see Table 5. Notice that the first column include a control for all non-refugee migration. This variable is, for obvious reasons, dropped in column 2. Significance levels: *$p<0.1$, **$p<0.05$, ***$p<0.01$. Standard errors clustered on neighborhood-level.
B Adopting specifications used by Saiz (2007) and Sá (2014)

As mentioned earlier, the specification we use in this paper differs somewhat from both the specification used in Saiz (2007) and the specification used in Sá (2014); the dependent variable is the same in all three papers (change in log-prices), but there are differences in how the immigration and population variables are dealt with. While we have chosen to enter the annual inflow of refugee immigrants as a separate variable and then to control flexibly for the population size, Saiz (2007) normalize the annual inflow of immigrants with the population size, giving rise to the following specification (c.f. his equation (I); he has no city-specific fixed effects in his specification):

\[
\Delta \ln(P_{i,t+1}) = \beta \frac{\hat{m}_{i,t-1}}{\text{pop}_{i,t-2}} + X' \Pi + \lambda_t + \varepsilon_{i,t+1}
\]  \hspace{1cm} (10)

and Sá (2014) normalize the annual change in the number of foreign born with the population size, giving rise to the following specification (c.f. her equation (11); she has neighborhood-specific fixed effects in her specification):

\[
\Delta \ln(P_{i,t+1}) = \beta \frac{\Delta FB_{i,t}}{\text{pop}_{i,t-1}} + X' \Pi + \eta_i + \lambda_t + \varepsilon_{i,t+1}
\]  \hspace{1cm} (11)

where \(\Delta FB_{i,t}\) is the annual change in the number of foreign born in each neighborhood. The X-variables in both papers include time-varying local covariates but not any population variables.

To examine if our results are sensitive to the chosen specification, we re-estimate our baseline regressions using the specifications used in Saiz (2007) and Sá (2014); we estimate their models both with and without neighborhood fixed effects (which, in these models, is equal to including neighborhood-specific time trends). From the results, presented in Table B.2, we draw three main conclusions: First, the specification used in Sá yields a larger point estimate than what we get with the specification used in this paper, but the standard errors are also fairly large (giving a marginally significant point estimate in the specification without neighborhood fixed effects). But what is more, the point estimates are
positive (Sá finds negative effects in her analysis on UK data). Second, the specification used in Saiz is sensitive to the inclusion of neighborhood fixed effects (going from a positive and marginally significant effect without neighborhood fixed effects to a statistically insignificant and much smaller, but negative, point estimate with the fixed effects included); the specification used in this paper and the specification used in Sá are less sensitive to the inclusion/exclusion of neighborhood fixed effects. Third, the results in Table B.2 are largely in line with our baseline results in the sense that there are no clear evidence of an effect of immigration on housing prices; if anything, it has a positive effect.

**Table B.2. Effects of immigration on housing prices when using specifications in Saiz (2007) and Sá (2014)**

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SECOND STAGE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Delta FB_{i,t} )</td>
<td>9.080*</td>
<td>8.233</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \frac{\Delta FB_{i,t}}{Pop_{i,t-1}} )</td>
<td>(5.374)</td>
<td>(6.529)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \frac{im_{i,t}}{Pop_{i,t-1}} )</td>
<td>5.482*</td>
<td>-0.298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \frac{\Delta FB_{i,t}}{Pop_{i,t-1}} )</td>
<td>(3.294)</td>
<td>(6.197)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| FIRST STAGE                 |         |         |        |         |
| \( \Delta FB_{i,t} \)      | 0.0700***| 0.0527***|        |         |
| \( \frac{\Delta FB_{i,t}}{Pop_{i,t-1}} \) | (0.0177) | (0.0151)|        |         |
| \( \frac{im_{i,t}}{Pop_{i,t-1}} \) | 0.0872***| 0.0678***|        |         |
| \( \frac{\Delta FB_{i,t}}{Pop_{i,t-1}} \) | (0.0187) | (0.0205)|        |         |

F-Stat                       | 16.08   | 12.25   | 21.72  | 10.96   |
Observations                 | 38,367  | 38,367  | 42,630 | 42,630  |
Number of N’hoods            | 4,263   | 4,263   | 4,263  | 4,263   |
N’hood FE                    | NO      | YES     | NO     | YES     |
Year FE                      | YES     | YES     | YES    | YES     |
Covariates                   | YES     | YES     | YES    | YES     |

**Notes:** Second and first stage, effect of immigration, on ∆ log final housing prices. Column (1) and (2) adopts the specification in Sá (2014), with column (1) excluding neighborhood fixed effects and column (2) including neighborhood fixed effects. For specification see Equation 11. Column (3) and (4) adopts the model in Saiz (2007), with column (3) excluding fixed effects and column (4) including fixed effects. Significance levels: *p<0.1, **p<0.05, ***p<0.01. Standard errors clustered on neighborhood-level.
C Using covariates dated in \( t \) and \( t + 1 \)

In the baseline specification, we estimate the effects of immigration in \( t \) on the change in (the log of) housing prices between \( t \) and \( t + 1 \) using the covariates dated in \( t - 1 \); the reason for doing this is to avoid problems with bad controls. However, it can be that a change in covariates in \( t \) and \( t + 1 \) (in particular changes in population size and other immigration) might have a direct effect on prices. To examine if this is important, we have re-estimated our models using the covariates dated in \( t \) and \( t + 1 \). From the results, presented in Table C.3, it is clear that this yields very similar results as in the baseline analysis.

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Covariates in ( t )</th>
<th>Covariates in ( t + 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( im_{it} )</td>
<td>2.17e-05 ( (0.000624) )</td>
<td>0.000933 ( (0.000775) )</td>
</tr>
<tr>
<td>Observations</td>
<td>46,893</td>
<td>42,630</td>
</tr>
<tr>
<td>Number of n’hoods</td>
<td>4,263</td>
<td>4,263</td>
</tr>
<tr>
<td>Neighborhood FE</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Year FE</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Covariates</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

Notes: Second stage, effect of immigration in \( t \) on \( \Delta \) log contract prices in \( t + 1 \). Column (1) shows second stage controlling for all covariates (including population) in \( t \) and and column (2) controlling for all covariates in \( t + 1 \). Significance levels: \(*p<0.1\), \(**p<0.05\), \(***p<0.01\). Standard errors clustered on SAMS-level.

D Hedonic price equations

In the baseline analysis, we only control for the size of the housing unit when estimating the effects on house prices (we do this by using the price per m\(^2\)). A worry is however that the objects that enter the market in different time periods differ in other dimensions than in terms of m\(^2\). In the final sensitivity analysis, we therefore base the housing prices on estimated hedonic price functions to better control for more characteristics than just the size.\(^{36}\)

\(^{36}\)The reason that we are not doing this in the baseline analysis is that we loose some observations because of missing information on some of the other housing characteristics.
We first estimate separate hedonic price functions for apartments and houses. The estimated hedonic price function for apartments is given by:

\[
\ln P_{j,t} = \beta_0 + \beta_1 \text{AREA}_{j,t} + \beta_2 \text{AREA}^2_{j,t} + \beta_3 \#\text{ROOMS}_{j,t} + \\
\beta_4 \#\text{ROOMS}^2_{j,t} + \beta_5 \text{FEE}_{j,t} + \tau_t + \epsilon_{j,t} \tag{12}
\]

where \text{AREA} is living area in square meters, \text{ROOMS} is number of rooms (both included also in squared terms), \text{FEE} is the monthly fee, and \tau_t is a time fixed effect. The results from this specification is given in the first column in Table D.4.

The estimated hedonic price function for houses is given by:

\[
\ln P_{j,t} = \gamma_0 + \gamma_1 \text{AREA}_{j,t} + \gamma_2 \text{AREA}^2_{j,t} + \gamma_3 \#\text{ROOMS}_{j,t} + \\
\gamma_4 \#\text{ROOMS}^2_{j,t} + \gamma_5 \text{PLOT}_{j,t} + \mu_t + \varepsilon_{j,t} \tag{13}
\]

where \text{PLOT} is the m$^2$ plot area. The results from this specification are given in the second column in Table D.4.

To get house prices net of these characteristics, we then calculate house prices as:

\[
\ln P_{j,t}^{\text{net}} = \hat{\beta}_0 + \hat{\epsilon}_{j,t} \tag{14}
\]

\[
\ln P_{j,t}^{\text{net}} = \hat{\gamma}_0 + \hat{\varepsilon}_{j,t} \tag{15}
\]

Based on these housing prices net of characteristics, we then calculate the median housing price in each neighborhood in each year and use these as the dependent variable in the baseline econometric specification. From the results, presented in Table D.5, we get an estimate that is similar to the one in the baseline estimations.
### Table D.4. Hedonic price functions

<table>
<thead>
<tr>
<th>Variables</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Apartments</td>
<td>Houses</td>
</tr>
<tr>
<td>Living area</td>
<td>0.0133***</td>
<td>0.0090***</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0001)</td>
</tr>
<tr>
<td>Living area(^2)</td>
<td>0.0000***</td>
<td>-0.0000***</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>Number of rooms</td>
<td>0.1836***</td>
<td>0.2616***</td>
</tr>
<tr>
<td></td>
<td>(0.0045)</td>
<td>(0.0035)</td>
</tr>
<tr>
<td>Number of rooms(^2)</td>
<td>-0.0156***</td>
<td>-0.0159***</td>
</tr>
<tr>
<td></td>
<td>(0.0007)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td>Monthly fee</td>
<td>-0.0002***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td></td>
</tr>
<tr>
<td>Plot area</td>
<td></td>
<td>-0.0000***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0000)</td>
</tr>
<tr>
<td>Constant</td>
<td>13.0343***</td>
<td>12.2430***</td>
</tr>
<tr>
<td></td>
<td>(0.0063)</td>
<td>(0.0093)</td>
</tr>
<tr>
<td>N</td>
<td>686,475</td>
<td>454,500</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td>Year fixed effects</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

*Notes:* Hedonic price functions estimated using all available observations. Significance levels: *p* < 0.1, **p** < 0.05, ***p** < 0.01.

### Table D.5. Effect of immigration in \( t \) on hedonic prices in \( t + 1 \)

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Hedonic Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>( im_{it} )</td>
<td>0.000767</td>
</tr>
<tr>
<td></td>
<td>(0.000844)</td>
</tr>
<tr>
<td>Observations</td>
<td>42,630</td>
</tr>
<tr>
<td>Number of Sams</td>
<td>4,263</td>
</tr>
<tr>
<td>Sams FE</td>
<td>YES</td>
</tr>
<tr>
<td>Year FE</td>
<td>YES</td>
</tr>
<tr>
<td>Covariates</td>
<td>YES</td>
</tr>
</tbody>
</table>

*Notes:* Second stage, effect of immigration on hedonic prices. See Equation 12 and 13 for information on the estimation of the different hedonic indexes. Significance levels: *p* < 0.1, **p** < 0.05, ***p** < 0.01. Standard errors clustered on SAMS-level.
References


Economic Studies

1987:1 Haraldson, Marty. To Care and To Cure. A linear programming approach to national health planning in developing countries. 98 pp.


1989:3 Choe, Byung-Tae. Some Notes on Utility Functions Demand and Aggregation. 39 pp.


<table>
<thead>
<tr>
<th>Number</th>
<th>Author</th>
<th>Title</th>
<th>Year</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>Apel, Mikael</td>
<td>Essays on Taxation and Economic Behavior</td>
<td>1994</td>
<td>144 pp</td>
</tr>
<tr>
<td>19</td>
<td>Dillén, Hans</td>
<td>Asset Prices in Open Monetary Economies: A Contingent Claims Approach</td>
<td>1994</td>
<td>100 pp</td>
</tr>
<tr>
<td>20</td>
<td>Jansson, Per</td>
<td>Essays on Empirical Macroeconomics</td>
<td>1994</td>
<td>146 pp</td>
</tr>
<tr>
<td>21</td>
<td>Banks, Ferdinand E.</td>
<td>A Modern Introduction to International Money, Banking, and Finance</td>
<td>1995</td>
<td>313 pp</td>
</tr>
<tr>
<td>22</td>
<td>Dufwenberg, Martin</td>
<td>On Rationality and Belief Formation in Games</td>
<td>1995</td>
<td>93 pp</td>
</tr>
<tr>
<td>23</td>
<td>Lindén, Johan</td>
<td>Job Search and Wage Bargaining</td>
<td>1995</td>
<td>127 pp</td>
</tr>
<tr>
<td>24</td>
<td>Shahnazarian, Hovick</td>
<td>Three Essays on Corporate Taxation</td>
<td>1996</td>
<td>112 pp</td>
</tr>
<tr>
<td>25</td>
<td>Svensson, Roger</td>
<td>Foreign Activities of Swedish Multinational Corporations</td>
<td>1996</td>
<td>166 pp</td>
</tr>
<tr>
<td>26</td>
<td>Sundberg, Gun</td>
<td>Essays on Health Economics</td>
<td>1996</td>
<td>174 pp</td>
</tr>
<tr>
<td>28</td>
<td>Fredriksson, Peter</td>
<td>Education, Migration and Active Labor Market Policy</td>
<td>1997</td>
<td>106 pp</td>
</tr>
<tr>
<td>29</td>
<td>Ekman, Erik</td>
<td>Household and Corporate Behaviour under Uncertainty</td>
<td>1997</td>
<td>160 pp</td>
</tr>
<tr>
<td>30</td>
<td>Stoltz, Bo.</td>
<td>Essays on Portfolio Behavior and Asset Pricing</td>
<td>1997</td>
<td>122 pp</td>
</tr>
<tr>
<td>31</td>
<td>Dahlberg, Matz</td>
<td>Essays on Estimation Methods and Local Public Economics</td>
<td>1997</td>
<td>179 pp</td>
</tr>
<tr>
<td>32</td>
<td>Kolm, Ann-Sofie</td>
<td>Taxation, Wage Formation, Unemployment and Welfare</td>
<td>1997</td>
<td>162 pp</td>
</tr>
<tr>
<td>33</td>
<td>Boije, Robert</td>
<td>Capitalisation, Efficiency and the Demand for Local Public Services</td>
<td>1997</td>
<td>148 pp</td>
</tr>
<tr>
<td>34</td>
<td>Hort, Katinka</td>
<td>On Price Formation and Quantity Adjustment in Swedish Housing Markets</td>
<td>1997</td>
<td>185 pp</td>
</tr>
<tr>
<td>35</td>
<td>Lindström, Thomas</td>
<td>Studies in Empirical Macroeconomics</td>
<td>1998</td>
<td>113 pp</td>
</tr>
<tr>
<td>36</td>
<td>Hemström, Maria</td>
<td>Salary Determination in Professional Labour Markets</td>
<td>1998</td>
<td>127 pp</td>
</tr>
<tr>
<td>37</td>
<td>Forsling, Gunnar</td>
<td>Utilization of Tax Allowances and Corporate Borrowing</td>
<td>1998</td>
<td>96 pp</td>
</tr>
<tr>
<td>38</td>
<td>Nydahl, Stefan</td>
<td>Essays on Stock Prices and Exchange Rates</td>
<td>1998</td>
<td>133 pp</td>
</tr>
</tbody>
</table>


Lindhe, Tobias. **Corporate Tax Integration and the Cost of Capital.** 2002. 102 pp.


<table>
<thead>
<tr>
<th>No.</th>
<th>Author</th>
<th>Title</th>
<th>Year</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>141</td>
<td>Oscar Erixson</td>
<td>Economic Decisions and Social Norms in Life and Death Situations</td>
<td>2013</td>
<td>183</td>
</tr>
<tr>
<td>142</td>
<td>Pia Fromlet</td>
<td>Essays on Inflation Targeting and Export Price Dynamics</td>
<td>2013</td>
<td>145</td>
</tr>
<tr>
<td>143</td>
<td>Daniel Avdic</td>
<td>Microeconometric Analyses of Individual Behavior in Public Welfare Systems, Applications in Health and Education Economics</td>
<td>2014</td>
<td>176</td>
</tr>
<tr>
<td>145</td>
<td>Karolina Stadin</td>
<td>Employment Dynamics</td>
<td>2014</td>
<td>134</td>
</tr>
<tr>
<td>146</td>
<td>Haishan Yu</td>
<td>Essays on Environmental and Energy Economics</td>
<td>2014</td>
<td>132</td>
</tr>
<tr>
<td>147</td>
<td>Martin Nilsson</td>
<td>Essays on Health Shocks and Social Insurance</td>
<td>2015</td>
<td>139</td>
</tr>
<tr>
<td>148</td>
<td>Tove Eliasson</td>
<td>Empirical Essays on Wage Setting and Immigrant Labor Market Opportunities</td>
<td>2014</td>
<td>144</td>
</tr>
<tr>
<td>149</td>
<td>Erik Spector</td>
<td>Financial Frictions and Firm Dynamics</td>
<td>2014</td>
<td>129</td>
</tr>
<tr>
<td>151</td>
<td>Selva Bahar Baziki</td>
<td>Firms, International Competition, and the Labor Market</td>
<td>2015</td>
<td>183</td>
</tr>
<tr>
<td>152</td>
<td>Fredrik Sävje</td>
<td>What would have happened? Four essays investigating causality</td>
<td>2015</td>
<td>229</td>
</tr>
<tr>
<td>153</td>
<td>Ina Blind</td>
<td>Essays on Urban Economics</td>
<td>2015</td>
<td>197</td>
</tr>
<tr>
<td>154</td>
<td>Jonas Poulsen</td>
<td>Essays on Development and Politics in Sub-Saharan Africa</td>
<td>2015</td>
<td>240</td>
</tr>
<tr>
<td>155</td>
<td>Lovisa Persson</td>
<td>Essays on Politics, Fiscal Institutions, and Public Finance</td>
<td>2015</td>
<td>137</td>
</tr>
<tr>
<td>156</td>
<td>Gabriella Chirico Willstedt</td>
<td>Demand, Competition and Redistribution in Swedish Dental Care</td>
<td>2015</td>
<td>119</td>
</tr>
<tr>
<td>158</td>
<td>Johannes Hagen</td>
<td>Essays on Pensions, Retirement and Tax Evasion</td>
<td>2016</td>
<td>195</td>
</tr>
<tr>
<td>159</td>
<td>Rachatar Nilavongse</td>
<td>Housing, Banking and the Macro Economy</td>
<td>2016</td>
<td>156</td>
</tr>
<tr>
<td>160</td>
<td>Linna Martén</td>
<td>Essays on Politics, Law, and Economics</td>
<td>2016</td>
<td>150</td>
</tr>
<tr>
<td>161</td>
<td>Olof Rosenqvist</td>
<td>Essays on Determinants of Individual Performance and Labor Market Outcomes</td>
<td>2016</td>
<td>151</td>
</tr>
<tr>
<td>162</td>
<td>Linuz Aggeborn</td>
<td>Essays on Politics and Health Economics</td>
<td>2016</td>
<td>203</td>
</tr>
<tr>
<td>No.</td>
<td>Author</td>
<td>Title</td>
<td>Year</td>
<td>Pages</td>
</tr>
<tr>
<td>-----</td>
<td>------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
<td>-------</td>
</tr>
<tr>
<td>163</td>
<td>Glenn Mickelsson</td>
<td>DSGE Model Estimation and Labor Market Dynamics</td>
<td>2016</td>
<td>166</td>
</tr>
<tr>
<td>164</td>
<td>Sebastian Axbard</td>
<td>Crime, Corruption and Development</td>
<td>2016</td>
<td>150</td>
</tr>
<tr>
<td>165</td>
<td>Mattias Öhman</td>
<td>Essays on Cognitive Development and Medical Care</td>
<td>2016</td>
<td>181</td>
</tr>
<tr>
<td>166</td>
<td>Jon Frank</td>
<td>Essays on Corporate Finance and Asset Pricing</td>
<td>2017</td>
<td>160</td>
</tr>
<tr>
<td>167</td>
<td>Ylva Moberg</td>
<td>Gender, Incentives, and the Division of Labor</td>
<td>2017</td>
<td>220</td>
</tr>
<tr>
<td>168</td>
<td>Sebastian Escobar</td>
<td>Essays on inheritance, small businesses and energy consumption</td>
<td>2017</td>
<td>194</td>
</tr>
<tr>
<td>169</td>
<td>Evelina Björkegren</td>
<td>Family, Neighborhoods, and Health</td>
<td>2017</td>
<td>226</td>
</tr>
<tr>
<td>170</td>
<td>Jenny Jans</td>
<td>Causes and Consequences of Early-life Conditions. Alcohol, Pollution and Parental Leave Policies</td>
<td>2017</td>
<td>209</td>
</tr>
<tr>
<td>171</td>
<td>Josefine Andersson</td>
<td>Insurances against job loss and disability. Private and public interventions and their effects on job search and labor supply</td>
<td>2017</td>
<td>175</td>
</tr>
<tr>
<td>172</td>
<td>Jacob Lundberg</td>
<td>Essays on Income Taxation and Wealth Inequality</td>
<td>2017</td>
<td>173</td>
</tr>
<tr>
<td>174</td>
<td>Irina Andone</td>
<td>Exchange Rates, Exports, Inflation, and International Monetary Cooperation</td>
<td>2018</td>
<td>174</td>
</tr>
<tr>
<td>175</td>
<td>Henrik Andersson</td>
<td>Immigration and the Neighborhood. Essays on the Causes and Consequences of International Migration</td>
<td>2018</td>
<td>181</td>
</tr>
</tbody>
</table>