# Ethnic Enclaves, Self-Employment and the Economic Performance of Refugees

Henrik Andersson

UPPSALA
UNIVERSITET

# ETHNIC ENCLAVES, SELF-EMPLOYMENT AND THE ECONOMIC PERFORMANCE OF REFUGEES

HENRIK ANDERSSON

# Ethnic Enclaves, Self-Employment and the Economic Performance of Refugees[*]

Henrik Andersson[†]

March 18, 2018

### Abstract

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, effects 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.

**Keywords:** Immigration; Self-employment; Sweden; Foreign born; Ethnic Enclaves, Coethnics
**JEL classification:** C21; J15; M13; R23

# 1 Introduction

The segregation of natives and 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 employment and income (Beaman, 2012; Munshi, 2003; Edin et al., 2003; Damm, 2009; Bayer et al., 2008), 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.[1] Self-employment has particular importance; partly because self-employment rates tend to be higher for foreign born than for native born,[2] 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*.[3]

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

---

[1] Already at this point 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.

[2] See for example: "Immigrant's self-employment and entrepreneurship activities" (in "The missing Entrepreneurs 2017".

[3] 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 km$^2$).

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, information, 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, legal and institutional knowledge and contacts they provide, which is the key component of the ethnic enclave. Access to a large number of coethnics, regardless of qualities, does not seem to cause an increase in the probability of choosing self-employment.

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 neighbourhood variables, as well as individual characteristics. Anyone with a taxable business income[4] 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.

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 behaviour of coethnics, a simple linear regression will be biased. As a way of addressing this endogeneity concern I use a

---

[4]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.

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 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]

---

[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; Eliasson, 2014 and Åslund and Fredriksson, 2009).

[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 is 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 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.

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 processes. As noted, there are different possible mechanisms through which ethnic clusters might cause entrepreneurial 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 municipality 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 point. Given that only around 4.5 percent of the sample has 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, ethnics 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 conceptualising 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.[8] 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.

$$y_i = \mathbf{X}_{1,i}\mathbf{\Phi}_1 + \epsilon_{1,i} \tag{1}$$

$$w_i = \mathbf{X}_{2,i}\mathbf{\Phi}_2 + \epsilon_{2,i} \tag{2}$$

Income is a function of vectors $\mathbf{X}_{1,i}$ and $\mathbf{X}_{2,i}$, 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).

$$I^* = y_i - w_i = (\mathbf{X}_{1,i}\mathbf{\Phi}_1 + \epsilon_{1,i}) - (\mathbf{X}_{2,i}\mathbf{\Phi}_2 + \epsilon_{2,i}) \tag{3}$$

Based on Equation 3, a decision rule emerges: 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 $\mathbf{X}_{1,i}$, defined as $\mathbf{Z}_i \subseteq \mathbf{X}_{1,i}$, capturing individual and local requirements needed to be able to start a business. These can be for example institutional and legal

---

[8]For examples see Lucas (1978); Evans and Jovanovic (1989).

knowledge, a specific entrepreneurial skill set, access to capital or access to a consumer base. Define a minimum level of $\mathbf{Z_{min}}$ as necessary for any business income to be possible. Thereafter define equation 5:

$$
\begin{cases} y_i = 0, & \text{if } \mathbf{Z_i} < \mathbf{Z_{min}} \\ y_i \geq 0, & \text{otherwise} \end{cases} \tag{5}
$$

That is, positive business income is only possible with a certain level of individual and local qualities $\mathbf{Z_i} \geq \mathbf{Z_{min}}$. Now, access to coethnics can positively affect $\mathbf{Z_i}$, by transferring the necessary skills, legal knowledge, institutional know-how, or providing consumers, workers and capital. These qualities in turn drive the possibility for self-employment income. Define *EthnicEnclave* as the size of the enclave. One can thereafter write:

$$
\frac{\partial \mathbf{Z_i}}{\partial EthnicEnclave} > 0 \longrightarrow \tag{6}
$$
$$
\frac{\partial P[I^* > 0]}{\partial EthnicEnclave} > 0
$$

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 $\mathbf{Z_i}$ 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)

---

[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".

further notes 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]

$$\frac{\partial z_1}{\partial \#Coethnics} > 0 \; \longrightarrow$$
$$\frac{\partial P[I^* > 0]}{\partial \#Coethnics} > 0$$

(7)

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,

$$\frac{\partial z_2}{\partial \#Self - Employed\ Coethnics} > 0 \; \longrightarrow$$
$$\frac{\partial P[I^* > 0]}{\partial \#Self - Employed\ Coethnics} > 0$$

(8)

Equation 7 and Equation 8 provide the main hypothesis' of the paper, that access to a larger number of coethnics, or a larger number of self-employed coethnics, increases 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

---

[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. While the author does not claim this, it could be connected to the smaller possibilities within an enclave, or with niche market products.

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:

$$\frac{\partial^2 z_2}{\partial \#Self-Employed\ Coethnics^2} < 0 \qquad (9)$$

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 Eliasson (2014), who shows that having a coethnic local banker in the port of entry municipality increases the propensity of self-employment.[12] 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. This channel hence leads individuals away from self-employment, meaning that many coethnics in the same municipality of arrival might cause 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:

_____

[12]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).

$$\frac{\partial \mathbf{X_{2,i}}}{\partial \#Self - Employed\ Coethnics} > 0 \; ; \; \frac{\partial \mathbf{X_{2,i}}}{\partial \#Coethnics} > 0 \longrightarrow \qquad (10)$$

$$\frac{\partial w_i}{\partial \mathbf{X_{2,i}}} > 0$$

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. Similarly, an estimated negative, or zero effect, might reflect partial effects running in opposite directions. While this will make it harder to pin-point a certain mechanism, I will use the detailed register data and, to the extent that I can, rule out unlikely channels.

## 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 permanent resident in Sweden from 1990 to 2014. 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 3.2.1.

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 was born in, non-OECD coun-

tries. 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 most likely made the new arriving immigrant a 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 section A.

Table 1: Distribution of country of birth for final sample of immigrants who arrived in 1990-1991.

| Country of birth | Freq. | Percent | Cum, |
|---|---|---|---|
| Iran | 3,118 | 22.13 | 22.13 |
| Iraq | 2,052 | 14.56 | 36.69 |
| Lebanon | 1,897 | 13.46 | 50.15 |
| Ethiopia | 1,388 | 9.85 | 60.00 |
| Somalia | 1,343 | 9.53 | 69.53 |
| Syria | 1,201 | 8.52 | 78.06 |
| Yugoslavia | 969 | 6.88 | 84.93 |
| Vietnam | 919 | 6.52 | 91.46 |
| Romania | 692 | 4.91 | 96.37 |
| Bulgaria | 512 | 3.63 | 100 |
| Total | 14,091 | 100 | |

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

## 3.2 Empirical model

Given the sample selected, the target of the empirical estimation 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 0 otherwise.[13] The use of business income has several

---

[13]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

strengths. It shows primarily those actively involved in their business, since any establishments without any income stream are ruled out. Also, the alternative is usually labor surveys, performed at a given week or month of the year (in Sweden in November). This latter method likely creates a measurement error less prevalent when using yearly income streams. With that said, limitations follow with the use of business income. Obvious ones include that any informal business activity is ruled out. This also include any contributing family workers, which is included in the ILO definition of self-employment.[14] Given the nature of register data, there is, however, no (good) way to measure informal activity. Second, if a company is organized as a "sole trader" or a "trading partnership", owners are personally liable and any corporate income is also declared as the owners. However, larger companies are often set up as a limited company, in which case business income is not declared for the owner.[15] Owners of the latter legal form can hence not be detected. 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 becomes less important, and a lot of individuals have possibly moved. I chose five years as a midway case. I do however provide estimates for 3 to 7 year spans as

---

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.

[14]http://ilo.org/global/statistics-and-databases/ statistics-overview-and-topics/status-in-employment/current-guidelines/ lang--en/index.htm.

[15]Comparable terms for "sole trader" is independent contractor, "trading partnership" can be labelled general partnership and limited company can be described as a joint stock company. For more information on the Swedish types of business, see https://www. verksamt.se/web/international/starting/types-of-business.

well, giving similar estimates as the baseline case.

Sub-indexes include $c$ for country of origin, $m$ for municipality and $k$ for cohort (arriving 1990 or 1991). The main treatment variables are coethnics who are self-employed, living in the municipality of arrival, standardized by the municipality population and all other coethnics as share of the municipal population. I standardize with municipality population as the baseline case[16], however I also provide robustness check including different functional forms (see the Appendix, Section B).

In detail this implies that, for individual $i$, born in country $c$, arriving in municipality $m$, with cohort $k$, 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, 1990 (1991). 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 ($\mathbf{X_{it}}$), 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).

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

### 3.2.1 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). With municipality and cohort by country fixed effects, the relevant comparison is between country/cohort groups within municipality. Thinking of it in terms of a within transformation, the average level of municipality coethnics with self-employment is subtracted, hence if there is a strong ten-

---

[16]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 $P_m = \frac{\#Self-Employed\ Coethnics_m}{Population_m}$, 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.

dency for immigrant entrepreneurship, or entrepreneurship in general in a specific municipality, it should be accounted for using this model. 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 lenghts by various studies and reports (see Edin et al. (2003); Borevi and Myrberg (2010); Åslund and Fredriksson (2009); Read (1992); Invandrarverket (1997); Dahlberg et al. (2012)), below, I therefore provide only a short description and introduction.

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 Swedens 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. Im-

migration 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 were to go, previous research suggest 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 placement 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).

$$y_{icmk} = \alpha + \mathbf{X_{2,i}}\mathbf{\Delta} + \theta_{kc} + \epsilon_{imck} \tag{12}$$

The $\mathbf{X_{2,i}}$ 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. Age has shown to be positively correlated with self-employment, which also holds for marriage and sex. Men are more prone to start a business, and and the same holds for those with spouses. 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 $(\hat{y}_{icmk})$, becomes a measure for individual likelihood of self-employment.

Having done this regression, in step two, I regress the self-employed co-ethnics (as share of the population), living in the municipality of arrival on the predicted self-employment $(\hat{y}_{icmk})$, conditional on the full set of co-variates 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 is unobserved labor market characteristics on municipality level, especially suitable for a certain birth country, this 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.[18] 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.

Table 2: Comparing means for group of stayers and subsequent movers.

|  | Sample staying | Sample moving | Pr(T >t) |
| --- | --- | --- | --- |
| Share Self-Employed w. 5 yrs | 0.046 (0.0026) | 0.042 (0.0023) | 0.311 |

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

---

[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.

[18] Employment is measured using labor market surveys performed in November each year.

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 include summary statistics and t-statistics in Table 2, comparing the mean between those staying in the same municipality after five years, and those living in another municipality five years later. What the table show 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.

# 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.

At arrival, around half the sample is 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.[19] A surprisingly large number is that almost one in six has 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

---

[19]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.

## Table 3: Summary Statistics

| VARIABLE | 1990-1991 | | | 1995-1996 | | |
|---|---|---|---|---|---|---|
| | Arrival year Statistics | | | | | |
| *Individual characterstics* | N | Mean | Std.dev | N | Mean | Std.dev |
| Age | 14,091 | 30.58 | 8.43 | 13,992 | 35.58 | 8.43 |
| Married | 14,091 | 0.53 | 0.50 | 13,992 | 0.62 | 0.49 |
| Men | 14,091 | 0.63 | 0.48 | 13,992 | 0.63 | 0.48 |
| Children | 14,091 | 0.36 | 0.48 | 13,992 | 0.54 | 0.50 |
| #Children (| parent) | 5,065 | 2.22 | 1.28 | 7,606 | 2.28 | 1.31 |
| | | | | | | |
| University educated | 14,091 | 0.06 | 0.23 | 13,992 | 0.22 | 0.41 |
| Less than nine years of education | 14,091 | 0.82 | 0.38 | 13,992 | 0.47 | 0.50 |
| Big City | 14,091 | 0.16 | 0.37 | 13,992 | 0.35 | 0.48 |
| Share with Wage>0 | 14,091 | 0.16 | 0.36 | 13,992 | 0.41 | 0.49 |
| Wage (|Wage>0) | 2,205 | 317.27 | 343.49 | 5,695 | 791.99 | 713.81 |
| | | | | | | |
| *Self employment* | | | | | | |
| Share with Business Income | 14,091 | 0.001 | 0.03 | 13,992 | 0.03 | 0.17 |
| Business Income (| Business Income>0) | 10 | 360.60 | 356.75 | 405 | 433.02 | 457.19 |
| | | | | | | |
| *Municipality characterstics* | | | | | | |
| Pop | 14,091 | 66,870 | 103,742 | 13,992 | 129,646 | 141,323 |
| # coethnics | 14,091 | 392.45 | 920.74 | 13,992 | 1,163 | 1,738 |
| Share of population | 14,091 | 0.005 | 0.01 | 13,992 | 0.01 | 0.01 |
| # coethnics with wage>0 | 14,091 | 194.12 | 486.43 | 13,992 | 315.22 | 525.32 |
| # coethnics with business income | 14,091 | 15.43 | 41.69 | 13,992 | 45.99 | 81.83 |
| Share of population | 14,091 | 0.0002 | 0.0003 | 13,992 | 0.0003 | 0.0005 |
| Share of coethnics | 14,091 | 0.03 | 0.04 | 13,992 | 0.04 | 0.05 |
| At least 1 Coethnic w. Business Income | 14,091 | 0.53 | 0.50 | 13,992 | 0.83 | 0.38 |

*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 Income* shows the share with any positive declared (active) business income. Both *Business Income (| Business Income>0)* and *Wage (|Wage>0)* 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. *At least 1 Coethnic w. Business Income* is a dummy for the percentage in the sample who stays at a municipality with at least one self-employed coethnic. Exact sample restrictions is described in the text in section 3.1.

which 194 have a positive salary (*# coethnics with wage>0*), and 15 are self-employed (*# coethnics with business income*).[20] 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.[21] 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. Instead of about 1/6, more than 1/3 now lives in one of the three big municipalities. Extending "municipality" to "metropolitan areas", increases the share.

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, which represent 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 entrepreneurs are also slightly younger, but most importantly, the 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.

Regarding sector, unfortunately a sizeable part of the individuals owning a firm (189 individuals) does not have any information on sector. 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.[22]

---

[20]The number of coethnics is based on the working age population.

[21]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).

[22]Note that the sum of frequency of the sectors will not add up to the 611 self-employed

Table 4: Top sectors of establishment and characteristics of the self-employed.

| *Establishments* | Freq | Percent | |
|---|---|---|---|
| | | | |
| Restaurants | 119 | 24.3 | |
| Retail sale in non-specialized stores | 41 | 8.4 | |
| Retail sale in Tobacco store | 28 | 5.7 | |
| Hair Services | 18 | 3.7 | |
| Taxi Services | 14 | 2.9 | |
| Retail sale of fruits and Vegetables | 10 | 2.0 | |
| Other | 70 | 14.3 | |
| Unknown | 189 | 38.6 | |
| | | | |
| *Characteristics 90-91* | Obs | Mean | Std. Dev. |
| | | | |
| Age | 611 | 29.05 | 7.04 |
| Married | 611 | 0.42 | 0.49 |
| Sex | 611 | 0.88 | 0.32 |
| Children | 611 | 0.35 | 0.48 |
| #Children (| parent) | 219 | 2.0 | 1.1 |
| University educated | 611 | 0.05 | 0.23 |
| Less than 9 years of education | 611 | 0.82 | 0.38 |
| Big city | 611 | 0.15 | 0.36 |

Upper panel: Establishments for the self-employed. The lower panel shows individual characteristics for the 611 self-employed in the sample. For more information on the variables, see table 3.

## 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 of the sample. Romania and Bulgaria are the smallest, making up less than ten percent of the refugees. To show some of the important heterogeneity between the countries, Table 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, one can note that individuals from Syria and Lebanon are heavily overrepresented 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 within 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 boarders 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.

---

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.

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-balling the distribution shows that areas where the enclaves where larger, also seem to be places where new firms were started.[23]

---

[23]Spatial illustrations of the enclave size and self-employment situation for each of ten source countries is available upon request.

Table 5: Enclaves and self-employment statistics per country of birth.

| . | Iran | Iraq | Lebanon | Ethiopia | Somalia | Syria | Yugoslavia | Vietnam | Romania | Bulgaria |
|---|---|---|---|---|---|---|---|---|---|---|
| Total refugees 90-91 | 3,118 | 2,052 | 1,897 | 1,388 | 1,343 | 1,201 | 969 | 919 | 692 | 512 |
| As share of sample | 22% | 15% | 13% | 10% | 10% | 9% | 7% | 7% | 5% | 4% |
| Self-employed within 5 yrs. | 159 | 61 | 156 | 5 | 3 | 136 | 38 | 3 | 16 | 34 |
| As share of sample | 26.0% | 10.0% | 25.5% | 0.8% | 0.5% | 22.3% | 6.2% | 0.5% | 2.7% | 5.6% |
| As share of own country | 5.1% | 3.0% | 8.2% | 0.4% | 0.2% | 11.3% | 4.0% | 0.3% | 2.3% | 6.6% |
| *Municipality Characteristics* | | | | | | | | | | |
| # coethnics 90-91 | 848 | 221 | 167 | 226 | 56 | 238 | 1245 | 118 | 148 | 46 |
|  | (1438) | (380) | (231) | (398) | (88) | (369) | (1686) | (141) | (202) | (65) |
| As share of population | 0.83% | 0.28% | 0.35% | 0.25% | 0.13% | 0.37% | 1.20% | 0.60% | 0.47% | 0.08% |
|  | (0,0051) | (0,0018) | (0,0027) | (0,0018) | (0,0011) | (0,0061) | (0,01043) | (0,0072) | (0,0055) | (0,00066) |
| # coethnics with business income | 28.6 | 7.1 | 8.2 | 1.7 | 0.03 | 17.1 | 72.2 | 1.7 | 3.6 | 2.1 |
|  | (50.63) | (15.83) | (14.27) | (3.95) | (0.20) | (29.06) | (99.88) | (3.19) | (6.89) | (4.33) |
| As share of coethnics | 2.5% | 1.9% | 4.2% | 0.5% | 0.0% | 5.0% | 6.2% | 1.0% | 1.6% | 2.0% |
|  | (0.0277) | (0.0337) | (0.0619) | (0.0211) | (0.0041) | (0.0630) | (0.0447) | (0.0186) | (0.0223) | (0.0356) |
| As share of population | 0.0229% | 0.0054% | 0.0152% | 0.0009% | 0.0000% | 0.0238% | 0.0643% | 0.0053% | 0.0059% | 0.0016% |
|  | (0.00027) | (0.00007) | (0.00024) | (0.00002) | (0.000002) | (0.000469) | (0.00055) | (0.00007) | (0.00011) | (0.00003) |

*Notes:* Country of birth, self-employment and municipality characteristics. *#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.

23

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 (1) within five years
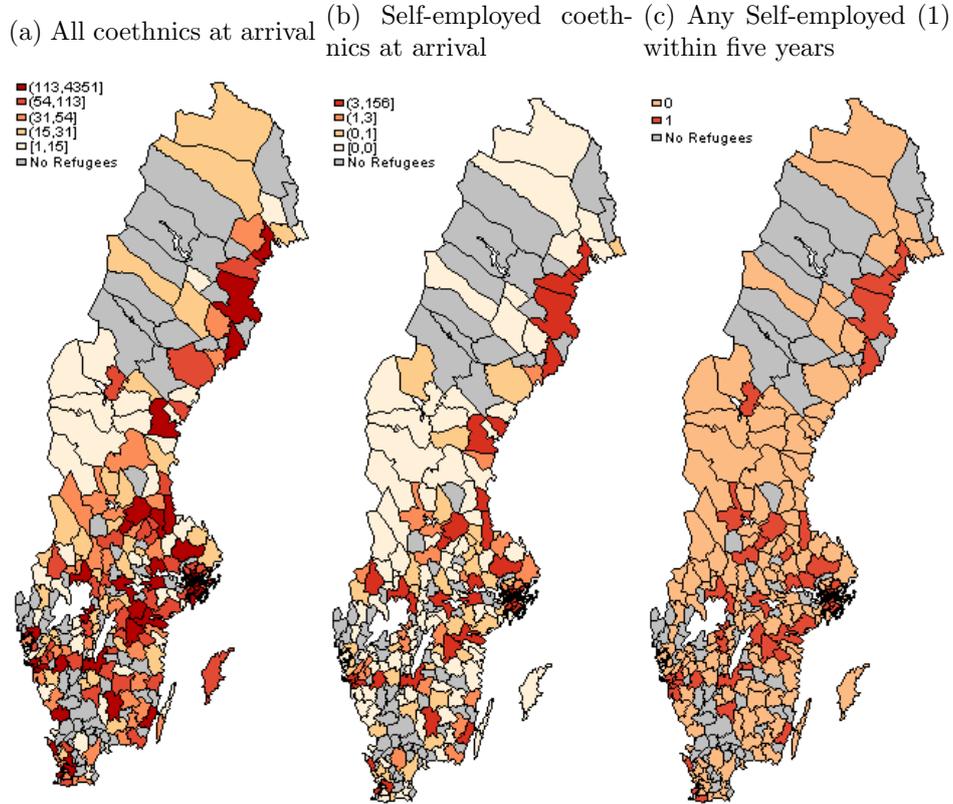


*Notes:* Map of Sweden, with administrative boundaries of municipalities. In Sub-figures 1a and 1b, the municipalities are colored based on the number of coethnics or number of self-employed coethnics, living in the municipality. Only municipalities to which at least 1 individual born in Iran arrived 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 exclude alternative stories as well as dig

somewhat 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[24], hence the coefficient represent 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.[25]

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 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.

While the estimates using the number of self-employed coethnics is both

---

[24][$X \sim (0,1)$].

[25]In Table 6, I only include specifications using no covariates or all covariates and fixed effects. In the Appendix, in Table A3, 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. Having business income or not within five years of arrival, regressed on the standardized municipality share of self-employed coethnics and share of all other coethnics, living in the municipality of arrival 1990-1991.

| VARIABLES | (1) Business Income or not | (2) Business Income or not | (3) Business Income or not |
|---|---|---|---|
| | Placement Policy Strategy | | "OLS" |
| # Self-employed Coethnics | | | |
| (As share of municipality population) | 0.0251*** | 0.0214*** | 0.0243*** |
| | (0.00466) | (0.00477) | (0.00467) |
| # Non-Self-employed Coethnics | | | |
| (As share of municipality population) | -0.0161*** | -0.0121*** | -0.0203*** |
| | (0.00393) | (0.00433) | (0.00388) |
| Observations | 13,992 | 13,992 | 13,992 |
| Mean Dep. Variable | 0.044 | 0.044 | 0.044 |
| Covariates and Fixed Effects | YES | YES | YES |

*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, age$^2$, 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.

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, niche or ethnic markets do 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 this 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 A3). The overall conclusion from the sensitivity results 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 seem to be a larger problem for the non-self-employed coethnics.[26]

---

[26]It's important to note that the comparison between column (2) and (3) is not an exact one. Since the paper has a reduced form research design, any comparison to demonstrate selection will be imperfect. 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.

## 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 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 of self-employed coethnics, 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 the sensitivity analysis in Table A3). 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 results.[27]

---

[27]Furthermore, in Table A5, 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: Regressing having business income sometime within the first five years after arrival on the standardized number of coethnics in the municipality of arrival in 1990 and 1991 (Different definitions on enclave).

| VARIABLES | (1) Business Inc. or not | (2) Business Inc. or not | (3) Business Inc. or not | (4) Business Inc. or not | (5) Business Inc. or not | (6) Business Inc. or not |
|---|---|---|---|---|---|---|
| #Unemployed Coethnics (As share of coethnic mun. pop.) | -0.000234 (0.00402) | | | | | |
| # Coethnics w. Capital Income (As share of mun. pop.) | | -0.000862 (0.00369) | | | | |
| # Coethnics w. High Capital Income (As share of mun. pop.) | | | 0.000478 (0.00399) | | | |
| # High Income Coethnics (As share of mun. pop.) | | | | 0.00300 (0.00281) | | |
| # High Educated Coethnics (As share of mun. pop.) | | | | | 0.00827*** (0.00312) | -0.000375 (0.00457) |
| # Self-employed Coethnics (As share of mun. pop.) | | | | | | 0.0129*** (0.00446) |
| Observations | 12,590 | 7,090 | 7,090 | 13,992 | 13,992 | 13,992 |
| Mean Dep. Variable | 0.044 | 0.044 | 0.044 | 0.044 | 0.044 | 0.044 |
| Covariates and Fixed Effects | YES | YES | YES | YES | YES | YES |

*Notes:* Regressing probability of self-employment on different characteristics of the ethnics enclave. Definitions of enclaves are: Column (1), unemployed coethnics as share of coethnics in municipality. Column (2), coethnics with capital income, and column (2) 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. Standard errors are clustered on municipality and country group level. For more information on covariates and fixed effects used see Table 6. *** p<0.01, ** p<0.05, * p<0.1.

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 is in fact 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 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, the disposable income of a certain birth country produces no significant effects. That is, 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 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, a 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.

A second reason why we might not see much of a niche market effect, is that it attracts a much broader market than one based on country of birth. A niche market might for example have an Arabic base rather than a specific Iraqi base. If my definition of the markets are too narrow, this might be the reason for not capturing any quantitative effect. 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 ten sample countries 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 A6, in the Appendix. Looking at the signs and sizes of the estimate in Table 8, clearly, there is

31

Table 8: Having business income or not within five years of arrival, regressed on the number of self-employed coethnics and all other coethnics, living in the municipality of arrival 1990-1991. Explanatory variables interacted.

| VARIABLES | (1) Business Inc. or not | (2) Business Inc. or not | (3) Business Inc. or not |
|---|---|---|---|
| # Non-Self-employed coethnics | 9.25e-06 | | |
| | (9.25e-06) | | |
| # Self-employed coethnics | 0.000471** | | |
| | (0.000189) | | |
| Interaction term | -1.24e-07* | | |
| | (6.57e-08) | | |
| Share non-Self-employed Coethnics, based on language | | | |
| (as share of mun. pop.) | | 0.00508 | -0.00774 |
| | | (0.00348) | (0.00514) |
| Share self-employed Coethnics, based on language | | | |
| (as share of mun. pop.) | | | 0.0180*** |
| | | | (0.00614) |
| | | | |
| Observations | 13,992 | 13,036 | 13,036 |
| Mean Dep. Variable | 0.044 | 0.044 | 0.044 |
| Covariates and Fixed Effects | YES | YES | YES |

*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 enclave is based on languages instead of birth country. The partition of the language groups is described in Table A6. 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. *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered on municipality and birth country level. See further Table 6 for information on covariates and fixed effects used.

no detectable positive effect from the number of language based coethnics at the municipality of arrival.

Pure income or financial resources seem not to be the driving factor of the effect of the enclave, neither is the effect driven by an especially difficult situation on the labor market at the municipality of arrival. 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 were meeting skilled coethnics is important because they provide and transfer knowledge 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?[28]

I study the effect on disposable income over the following twenty year period for the sample of refugees who arrived in 1990 and 1991.[29] The aggregated disposable income for the next twenty year period is assumed to

---

[28]The mediation approach based on the potential outcome framework is discussed in Imai et al. (2010).

[29]I drop the top 1 percent of income earners. The choice of twenty years is so to utilize the long term dimension of the data.

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:

$$y_{icmk} = \alpha_1 + \beta_1 \frac{\#SE\ Coethnics_{cmk}}{Population_{mk}} + \beta_2 \frac{\#nonSE\ Coethnics_{cmk}}{Population_{mk}} \quad (13)$$
$$+ \mathbf{X_i \Gamma_1} + \sigma_m^{mediator} + \theta_{kc}^{mediator} + \epsilon_{icmk}^{mediator}$$

$$Income_{icmk} = \alpha_2 + \rho y_{icmk} + \beta_3 \frac{\#SE\ Coethnics_{cmk}}{Population_{mk}} \quad (14)$$
$$+ \beta_4 \frac{\#nonSE\ Coethnics_{cmk}}{Population_{mk}} + \mathbf{X_i \Gamma_2} + \sigma_m^{outcome} + \theta_{kc}^{outcome} + \epsilon_{imck}^{outcome}$$

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 in the country ($Income_{icmk}$). $\beta_3$ 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).[30] Important to note is that for the mediation effect to have a causal interpretation, two assumptions are needed. First, the already discussed exogeneity assumption, which states that given pre-treatment characteristics and fixed effects, the size of the enclave is independent on 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

---

[30]This algorithm fits the two equations 13 and 14, and then simulate parameter values to arrive at the estimated effects.

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. I provide sensitivity checks below to further shed light on this.

Table 9: Causal mediation effect of self-employment on aggregate disposable income over twenty years (95 percent confidence intervals in parenthesis').

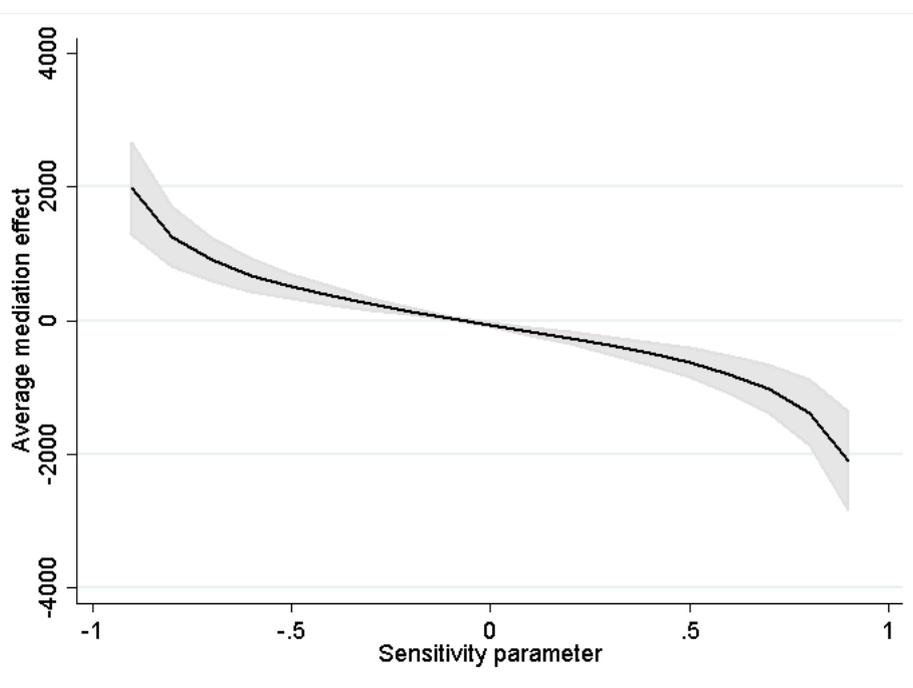| AVERAGE EFFECTS | (1)<br>Disposable Income |
| --- | --- |
| Causal Mediation Effect | -57.46 |
| | [-84.07, -33.42] |
| Direct Effect | -630.36 |
| | [-943.29, -296.48] |
| Total Effect | -687.82 |
| | [-997.45, -350.39] |
| % of Tot Effect Mediated through Self-Employment | 0.082 |
| | [.058, 0.16] |

*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.

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 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. Instead, it seems some part of the negative effect is mediated through the choice of

self-employment, as opposed to other labor market options.

Imai et al. (2010) has developed a sensitivity check to further analyse the importance of unobserved variation, affecting both income and the choice of self-employment. The method uses the correlation between residuals in Equations 13 and 14 ($\epsilon_{icmk}^{mediator}$ and $\epsilon_{icmk}^{outcome}$). 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 $\epsilon_{icmk}^{mediator}$ and $\epsilon_{icmk}^{outcome}$.

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 $p = corr(\epsilon_{icmk}^{mediator}, \epsilon_{icmk}^{outcome})$, which implies that negative values simulates

a situation where people with a potential higher income has lower probability of self-employment. Negative selection of this manner would for example be the case if those who become self-employed in general do so because they face difficulties and/or discrimination on the formal labor market. Note that this would have to be the case beyond the controlled for characteristics, most noticeably country of birth, municipality and education. Naturally, negative selection turns the average mediation estimates positive. 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 on income are negative, but given negative selection into self-employment, the story could in fact be the reverse.

The negative effect is, however, 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 un-employed.

(a) Mean Disposable Income

(b) Median Disposable Income



(c) Disposable Income, 25th percentile

(d) Disposable Income, 75th percentile
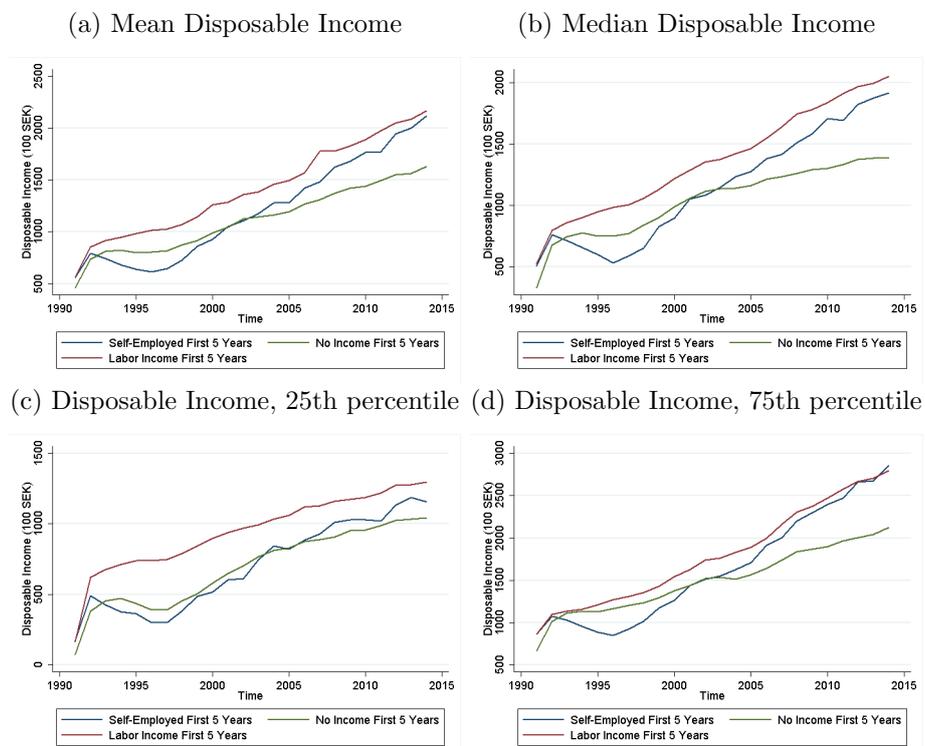


*Notes:* Disposable Income for the sample of refugees who arrived 1990-1991. The three lines represent three groups: one for those that had business income within the first five years, one for individuals who had other taxable labor income within the first five years, and one for those who had neither.

*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. Crudely, if we judge the outcome of working or becoming self-employed from the point of view of aggregated income of this almost 20 year long period, workers are generally more successful than those starting a firm. 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.

# 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 refugees who were placed in municipalities with a larger share of self-employed coethnics became self-employed to a larger degree. 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 behaviour of movers. I further rule out that the results are driven by coethnics with high

income, assets or that it is driven by 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 a story that information and knowledge from those already familiar with the market or on Swedish institutions is the most crucial mechanism, rather than the 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 affect 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

AGUILERA, M. B. (2009): "Ethnic enclaves and the earnings of self-employed Latinos," *Small Business Economics*, 413–425.

AKEE, R., D. JAEGER, AND K. TATSIRAMOS (2013): "The persistence of self-employment across borders: new evidence on legal immigrants to the United States," *Economics Bulletin*, 33, 127–137.

ALDRICH, H. AND R. WALDINGER (1990): "Ethnicity and entrepreneurship," *Annual review of Sociology*, 111–135.

ANDERSSON, L. AND M. HAMMARSTEDT (2015): "Ethnic enclaves, networks and self-employment among middle eastern immigrants in Sweden," *International Migration*, 53, 27–40.

ANDERSSON, M., J. LARSSON, AND Z. ÖNER (2017): "Ethnic enclaves, networks and immigrant self-employment: A neighborhood analysis of enclave size and quality," *IFN Working Paper*, 1195.

BAYER, P., S. ROSS, AND G. TOPA (2008): "Place of work and place of residence: Informal hiring networks and lobor market outcomes," *Journal of Political Economy*, 116, 1150–1196.

BEAMAN, L. A. (2012): "Social network and the dynamics of labor market outcomes: evidence from refugees resettled in the U.S." *Review of Economic Studies*, 79, 128–161.

BERTRAND, M., E. F. LUTTMER, AND S. MULLAINATHAN (2000): "Network effects and welfare cultures," *The Quarterly Journal of Economics*, 115, 1019–1055.

BOREVI, K. AND G. MYRBERG (2010): "Välfärdsstaten och de nyanlända: en flyktingplaceringspolitisk probleminventering," *MIM working papers series*, 79.

BORJAS, G. (1986): "The self-employment experience of immigrants," *The Journal of Human Resources*, 21, 485–506.

BURBIDGE, J. B., L. MAGEE, AND L. A. ROBB (1988): "Alternative transformations to handle extreme values of the dependent variable," *Journal of the American Statistical Association*, 83, 123–127.

CLARK, K. AND S. DRINKWATER (2002): "Enclaves, Neighborhood Effects and Employment Outcomes: Ethnic Minorities in England and Wales," *Journal of Population Economics*, 15, 5–29.

——— (2010): "Patterns of ethnic self-employment in time and space: evidence from British Census microdata," *Small Business Economics*, 34, 323–338.

CONSTANT, A. AND K. ZIMMERMAN (2006): "The making of entrepreneurs in Germany: are native men and immigrants alike?" *Small Business Economics*, 26, 279–300.

DAHLBERG, M., K. EDMARK, AND H. LUNDQVIST (2012): "Ethnic diversity and preferences for redistribution," *Journal of Political Economy*, 120, 41–76.

DAMM, A. (2009): "Ethnic Enclaves and Immigrant Labor Market Outcomes: Quasi-Experimental Evidence," *Journal of Labor Economics*, 27, 281–314.

EDIN, P.-A., P. FREDRIKSSON, AND O. ÅSLUND (2003): "Ethnic enclaves and the economic success of immigrants: evidence from a natural experiment," *The Quarterly Journal of Economics*, 118, 329–357.

ELIASSON, T. (2014): "Immigrant entrepreneurship and the origin of bankers," *IFAU Working Paper*, 19.

ERIKSSON, S. AND J. LAGERSTROM (2012): "Detecting discrimination in the hiring process: evidence from an Internet-based search channel," *Empirical Economics*, 43, 537–563.

EVANS, D. AND B. JOVANOVIC (1989): "An estimated model of entrepreneurial choice under liquidity constraints," *Journal of Political Economy*, 97, 808–827.

FAIRLIE, R. AND M. LOFSTROM (2015): "Immigration and entrepreneurship," in *Handbook of the economics of international migration*, ed. by B. Chiswick and P. Miller, Elsevier, 877–911.

FAIRLIE, R. AND C. WOODRUFF (2005): "Mexican entrepreneurship: a comparison of self-employment in Mexico and the United States," *NBER Working paper series*.

HICKS, R. AND D. TINGLEY (2011): "Causal mediation analysis," *The Stata Journal*, 11, 1–15.

HOCHGUERTEL, S. AND H. OHLSSON (2009): "Compensatory inter vivos gifts," *Journal of Applied Econometrics*, 24, 993–1023.

IMAI, K., L. KEELE, AND D. TINGLEY (2010): "A general approach to causal mediation analysis," *Psychological Methods*, 15, 309–334.

INVANDRARVERKET (1997): "Individuell Mångfald, Invandrarverkets utvärdering och analys av det samordnade flyktingmottagandet 1991-1996," .

KALNINS, A. AND W. CHUNG (2006): "Social capital, geography, and survival: Gujarati immigrant entrepreneurs in the US lodging industry," *Management Science*, 52, 233–247.

KERR, W. AND M. MANDORFF (2016): "Social networks, ethnicity, and entrepreneurship," *Harvard Business School Working Paper*, 16.

LE, A. T. (2000): "The determinants of immigrant self-employment in Australia," *International Migration Review*, 34, 183–214.

LIGHT, I. (1972): *Ethnic enterprise in America - Business and Welfare among Chinese Japanese and Blacks*, University of California Press Berkely.

LOFSTROM, M. (2002): "Labor market assimilation and the self-employment decision of immigrant entrepreneurs," *Journal of Population Economics*, 15, 83–114.

LOFSTROM, M. AND C. WANG (2009): "Mexican-American self-employment: a dynamic analysis of business ownership," in *Ethnicity and Labor Market Outcomes*, ed. by A. Constant, K. Zimmerman, and K. Tatsiramos, Emerald Books, 197–227.

LUCAS, R. (1978): "On the size distribution of business firms," *The Bell Journal of Economics*, 9, 508–523.

MUNSHI, K. (2003): "Networks in the modern economy: Mexican migrants in the U.S. labor market," *The Quarterly Journal of Economics*, 118, 549–599.

PORTES, A. AND R. BACH (1985): *Latin Journey - Cuban and Mexican Immigrants in the United States*, University of California Press.

READ, A. (1992): *Invandringspolitikens inriktning och resultat*, Riksdagens Revisorer, Stockholm.

ROY, A. (1951): "Some thoughts on the distribution of earnings," *Oxford Economic Papers*, 3, 135–146.

SIMOES, N., N. CRESPO, AND S. B. MOREIRA (2016): "Individual determinants of self-employment entry: what do we really know?" *Journal of Economic Surveys*, 30, 783–806.

ÅSLUND, O. AND P. FREDRIKSSON (2009): "Peer effects in welfare dependence - quasi experimental evidence," *Journal of Human Resources*, 44, 798–825.

ÅSLUND, O., L. HENSVIK, AND S. N. OSKAR (2014): "Seeking similarity: how immigrants and natives manage in the labor market." *Journal of Labor Economics*, 32, 405–441.

ÅSLUND, O. AND D.-O. ROOTH (2007): "Do when and where matter? initial labour market conditions and immigrant earnings," *The Economic Journal*, 117, 422–448.

WILSON, K. AND A. PORTES (1980): "Immigrant enclaves: an analysis of the labor market experience of Cubans in Miami," *American Journal of Sociology*, 86, 295–319.

YUENGERT, A. M. (1995): "Testing the hypotheses of immigrant self-employment," *The Journal of Human Resources*, 30, 194–204.

# Appendices

## A   Construction of the sample

As noted, I make use of GeoSweden, an individual level, full population database spanning from 1990-2014. The number of variables is vast, and span from demographic and socioeconomic information to housing characteristics. Essential to this study is 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[31], 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*. The

---

[31]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.

Swedish register data includes no information on *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.[32] 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 [https://www.migrationsverket.se/English/About-the-Migration-Agency/Facts-and-statistics-/Statistics/Overview-and-time-series.html](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

---

[32]The classification is based on the membership as of today (2016).

migration board per country (Note that labor market permits issued by the migration board in 1990-1991 are 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 4 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 be smaller, given that I focus on the number of self-employed 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 4: Probability of self-employment regressed on share of self-employed coethnics and share of all other coethnics. Coefficients from baseline regressions, 3-7 years between treatment and outcome.
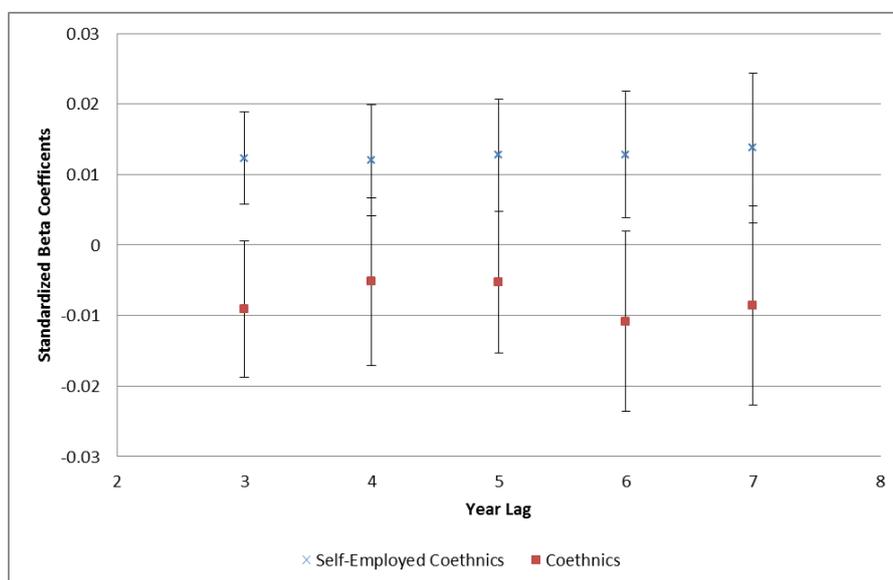


Vertical axis showing coefficient estimates doing the same regression as in Table 6, column 2. The dependent variable now defined at the year in question. Note the difference to the baseline estimates, where estimates are done capturing all individuals who at any point *up until* the fifth year lag had 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 Defintion 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. However, if agents actively seek coethnics within the municipality, standardizing with population might not be the most accurate method. In Table A1, column 1, I therefore look at the effect from regressing the probability of self-employment on the pure number of self-employed coethnics. The specification is the same as in Table 6, column 2. To take into consideration any extreme values or decreasing returns to scale[33], 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 ≈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.[34] The effect of self-employed coethnics is again significant and positive, while no significant effect can be detected for the other coethnics.

In the second column of Table A1, 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. The distribution of the explanatory variable is heavily skewed. The mean number of 15 self-employed coethnics should be compared to the median of only 1, while around 25 percent get placed with no self-employed coethnic, and a small sample in the top have over 100. In column 3 I therefore take away all individuals with more than 50 self-employed coethnics in the municipality of placement, which repre-

---

[33]Consider that the first self-employed coethnic could be more important than the tenth.

[34]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 several papers before this has included the transformation, including for example Hochguertel and Ohlsson (2009).

sents somewhat less than 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.

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 A2.

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 A3, 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.

Table A1: Probability of having business income regressed on self-employed coethnics and all other coethnics, as well as the inverse hyperbolic sine transformation of the explanatory variable.

| VARIABLES | (1) Business Income or not | (2) Business Income or not | (3) Business Income or not (Outliers Dropped) |
|---|---|---|---|
| # Self-employed Coethnics (Inverse Hyperbolic sine functional form) | 0.00943*** (0.00267) | | |
| # Non-Self-employed Coethnics (Inverse Hyperbolic sine functional form) | -0.00145 (0.00273) | | |
| # Self-employed Coethnics | | 0.000188 (0.000121) | 0.00156** (0.000612) |
| # Non-Self-employed Coethnics | | -1.63e-06 (6.59e-06) | -1.80e-05 (1.95e-05) |
| Observations | 13,992 | 13,992 | 12,611 |
| Covariates and Fixed Effects | YES | YES | YES |

*Notes:* Baseline regression using different functional forms. *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered on municipality - country group level. See Table 6 for information on covariates. Note that column (3) exclude all individuals who have more than 50 coethnics in their municipality of arrival (top 10 percent of sample).

Table A2: 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).

| VARIABLES | (1) Business Income or not Odds Ratios | (2) Business Income or not Odds Ratios | (3) Business Income or not Odds Ratios | (4) Business Income or not Odds Ratios |
|---|---|---|---|---|
| # Self-employed Coethnics (As share of municipality population) | 0.239*** (0.0433) | 0.232*** (0.0499) | 0.500*** (0.0911) | 0.480*** (0.102) |
| # Non-Self-employed Coethnics (As share of municipality population) | -0.183*** (0.0539) | -0.157*** (0.0589) | -0.400*** (0.118) | -0.321** (0.128) |
| Observations | 13,992 | 12,769 | 13,992 | 11,818 |
| Covariates and Fixed Effects | NO | YES | NO | YES |

*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.

Table A3: Baseline estimations. Having business income or not regressed on share of self-employed coethnics and general share of coethnics. Including different sets of covariates and fixed effects

| VARIABLES | (1) Business Income or not | (2) Business Income or not | (3) Business Income or not | (4) Business Income or not | (5) Business Income or not | (6) Business Income or not |
|---|---|---|---|---|---|---|
| # Non-Self-employed Coethnics (As share of municipality population) | 0.0251*** (0.00466) | 0.0265*** (0.00464) | 0.0175*** (0.00453) | 0.0176*** (0.00453) | 0.0214*** (0.00477) | 0.0215*** (0.00473) |
| # Self-employed Coethnics (As share of municipality population) | -0.0161*** (0.00393) | -0.0149*** (0.00403) | -0.00973** (0.00378) | -0.00998*** (0.00377) | -0.0121*** (0.00433) | -0.0121*** (0.00429) |
| Observations | 13,992 | 13,992 | 13,992 | 13,992 | 13,992 | 13,909 |
| Ind Controls | NO | YES | YES | YES | YES | YES |
| Country FE | NO | NO | YES | NO | NO | NO |
| Cohort by Country FE | NO | NO | NO | YES | YES | YES |
| Municipality FE | NO | NO | NO | NO | YES | YES |
| Country group Controls | NO | NO | NO | NO | NO | YES |

*Notes:* Linear Regression. *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered on municipality and country group level. Regarding covariates: on municipality level I include population and overall share of business owners and on country level share of employed coethnics as share of total amount of coethnics. For other covariates see Table 6.

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 survey definition provided by a yearly labor market survey. I do this for the dependent variable in Table A4. 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 A4: Estimates changing the definition of the dependent variable, using definition from labor market survey rather than having business income or not.

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

*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

Table A5: 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.

| VARIABLES | (1) Business Inc. or not | (2) Business Inc. or not | (3) Business Inc. or not | (4) Business Inc. or not | (5) Business Inc. or not | (6) Business Inc. or not | (7) Business Inc. or not | (8) Business Inc. or not |
|---|---|---|---|---|---|---|---|---|
| SUBGROUP | University educated (t+5) | Less than 9 years of educ. | Men | Women | Married | Non-married | Parents | Non-parents |
| # Self-employed Coethnics (As share of municipality pop.) | 0.0246** | 0.0231*** | 0.0381*** | 0.00155 | 0.0134*** | 0.0276*** | 0.0172** | 0.0219*** |
| | (0.0115) | (0.00605) | (0.00785) | (0.00332) | (0.00468) | (0.00769) | (0.00693) | (0.00563) |
| # Non-Self-employed Coethnics (As share of municipality pop.) | -0.0220* | -0.00817 | -0.0196*** | -9.43e-05 | -0.00452 | -0.0190*** | -0.00803 | -0.0136** |
| | (0.0116) | (0.00570) | (0.00648) | (0.00324) | (0.00434) | (0.00692) | (0.00627) | (0.00554) |
| Observations | 3,018 | 6,597 | 8,755 | 5,237 | 7,461 | 6,531 | 5,040 | 8,952 |
| Covariates and Fixed Effects | YES | YES | YES | YES | YES | YES | YES | YES |

*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.

54

# D  Categorization of Languages and Variables

Table A6: Languages and country of birth groups

| Language groups (in the sample) | Birth Country |
|---|---|
| Serbo-Croatian | Former Yugoslavia, Bosnia, Serbia, Croatia, Macedonia |
| Arabic | Somalia, Lebanon, Syria, Iraq, tunisia, Morocco, Algeria, Egypt |
| Persian | Iran, Afghanistan |
| Amharic | Ethiopia |
| Romanian | Romania |
| Bulgarian | Bulgaria |
| Vietnamese | Vietnam |