UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Immigration and Inequality: Evidence from Sweden

Daniel Altsten

June 2021

Abstract

I empirically study the relationship between the share of foreign-born individuals and municipality income inequality measured by the Gini index. I use Swedish municipality level data from 2000 to 2019, showing a positive and significant relationship between the municipality level inequality and the share of foreign-born individuals. As predicted by existing theory, my results seem to be driven mainly by the share of immigrants more likely to be born in countries with lower human capital. These results motivate policy interventions to increase human capital levels for such foreign-born at their point of arrival. My results are robust to various specifications, including fixed effects models and various ways of defining municipality level inequality.

Supervisor

Andreea Mitrut

Keywords

Immigration, municipality, income inequality, heterogeneity.

Master's Thesis in Economics
Graduate School

Acknowledgment

I want to express plenty of gratitude to the wonderful people at the University of Gothenburg that have made my experience at graduate school truly pleasant. A special mention to my friend Chi Nguyen, whom I have shared countless exciting conversations with, and my advisor Andreea Mitrut, who provided invaluable help during the thesis project.

Contents

List of Figures

List of Tables

1 Introduction

Income inequality has been studied and linked to numerous social effects, such as; lower trust in the nation's institutions (Magnusson & Johansson, 2008), lower *interpersonal* trust (D'Hombres et al., (2013)), criminal activity (Atems, 2020; Becker, 1968), or declination of health (Pickett et al., 2015). Even the former United States president, Barrack Obama, denoted their rising inequality as "the defining challenge of our time." Sweden has a long history of being a notably equal country, but lately, several sources have discussed rising income inequality (OECD, 2014; LO, 2018). Figure 1 depicts, at the municipality level, inequality (measured by the Gini-coefficient[1]) year 2019 minus same estimate year 2000.

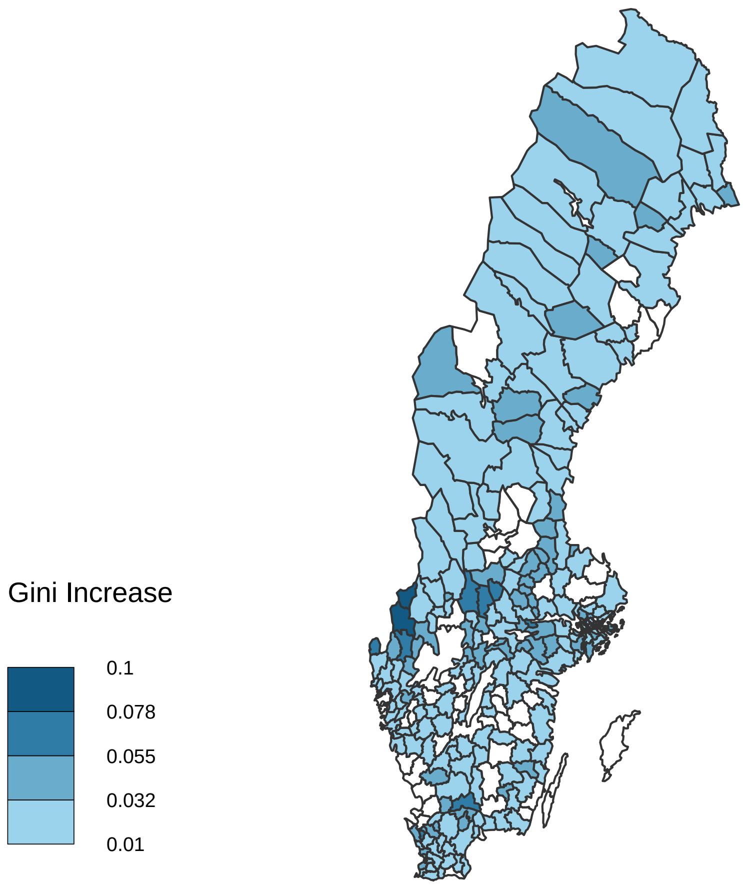

Figure 1: Inequality changes. | Gini-estimates are derived from the author's calculations using income data, originating at Statistics Sweden.

As observed, most municipalities have a higher inequality index year 2019, but there are clear regional variations. Nevertheless, designing effective policies against inequality rests, intuitively, on carefully identifying its determinants. One suggested inequality determinant is the level of immigration (Borjas, 2000; Hatton et al., 1998) - which is the topic of this thesis. More specifically, I will examine how the municipality-level proportion of foreign-born is associated with inequality during the period 2000-2019.

[1]Note how higher inequality likewise is represented by higher Gini-values. Details on the Gini-estimate are presented in section 3.2.

Figure 2 depicts the annual municipality-average of the foreign-born proportion (from now on denoted "PFB"), as well as respective continent proportions.

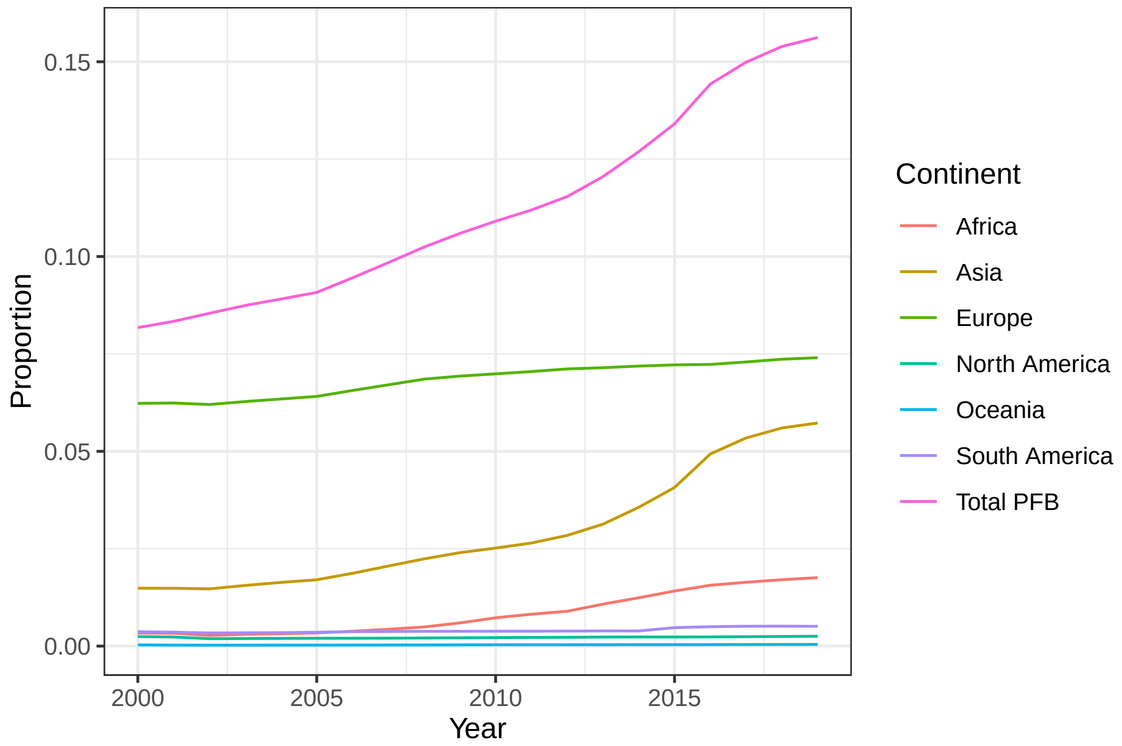

Figure 2: Foreign-born proportions. Note: Author's own calculations. Data originating from Statistics Sweden.

As observed, Sweden has transitioned rather quickly into a heterogeneous country. In 2000 municipalities hosted, on average, 8% foreign-born inhabitants - the same statistic was 15.4% in 2019. Figure 2 also shows that the PFB rise is primarily driven by Asian and African-born immigrants, while the remaining continents have had a relatively stable representation.

Both inequality and the PFB are trending in the same direction. Hence the correlation is naturally positive, but what mechanism does economic theory suggest between the two variables? The relation is somewhat complicated since immigration and inequality lack unison definitions. In this thesis, I choose to study the mechanism between *income inequality* and the proportion of foreign-born, to which one can apply and discuss human capital theory. There should be noted that human capital theory is designed explicitly for *wages*, not income. Income holds a few other factors, such as sick pay and unemployment benefits. However, one's actual wage is the dominant factor within one's income for most people. Further, human capital may decide whether one has a wage or not; hence theory relating to human capital is very relevant.

Becker (1968) discusses in his book: *Human capital*, cross-sectional variation in productivity between countries. Becker argues that such variation can be explained by countries' decisions to invest in their worker's education. Higher education increases productivity, i.e., higher educated workers have higher human capital. Following such a fundamental concept, Mincer (1974) proposed that human capital can also be used to determine one's individual wage. He argues that one's human capital can be estimated by the length of one's schooling (s) and workplace experience (exp). Individuals with higher human capital can expect a higher wage, culminating in the famous Mincer equation: $ln(W) = \beta_0 + \beta_1 * s + \beta_2 * exp + \beta_3 * exp^2 + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. Until this day, such simplistic empirical models have been shown to explain wages (Lemieux, 2006) efficiently. The implication is that *wage variation* arises if the inhabitants' human capital becomes increasingly heterogeneous. This essentially is the mechanism behind inequality since inequality is, in fact, different forms of wage data variation. Heterogeneous human capital levels within a population predict inequality.

It needs to be stated that human capital remains to be a rather abstract and hard-to-define concept. Researchers have since followed Mincer and proposed other variables that can be considered to increase one's human capital; such as health, i.e., a "health augmented Mincer wage equation" (Jorgensson, 1989), or*quality* of education (Kaarsen, 2014; Schoeman, 2012). For instance, Schoeman (2012) shows how education quality significantly explains a considerable amount of cross-sectional variation between immigrants' wages in their new native country. He finds that immigrants from the more developed region of the world (defined as OECD countries) have a significantly higher return to their education than immigrants from lesser developed areas (i.e., non-OECD).

Therefore, the mechanism proposed from economic literature is relatively intuitive; when migrants' human capital background is comparable to that of natives, there is no reason to expect wage inequality. Vice versa, when immigrants stem from backgrounds with notably better - as discussed by Advani et al. (2020) -, or lesser beneficial conditions, the total population becomes more heterogeneous, causing variation in wages - inequality.

This theory sets up for an interesting interplay of background characteristics, where it comes down to the relative group conditions that determine the outcome. Immigration naturally results in discussions of movements between countries. Therefore, theoretical predictions of how immigration affects inequality depend on which country-pair one studies. I will devote section 2 for a discussion regarding the theoretical effects of Sweden's immigration over the past 20 years.

In Sweden, there has been a substantial amount of research that indirectly can link immigration and inequality. For instance, Saco (2017) discusses that there is a significant wage gap between ethnically different[2], but similarly educated people. To which explanations such as discrimination, language barriers, or inability to adapt to the natives' culture (Carlsson et al., 2007; Basilio, 2017) have been proposed when determining immigrants' labor market success.

I will extend the literature by directly studying how the PFB is associated with inequality - a study that has not been done yet, to my knowledge. I will primarily assess the level of inequality by the Gini measurement. Given the human capital background of Swedish immigration over the past 20 years, I hypothesize the PFB to be positively associated with the Gini estimate. Further, I hypothesize that it is the income distribution's left tail that should correlate with the PFB. The Gini estimate inherently indicates disparity in a data vector and does not tell where the disparity arises in a distribution. Therefore I also use a range of left-tail indexes to conduct a robustness check. Such a robustness check allows for a more reliable examination regarding wherein the distribution correlation arises. Section 2 also reveals that human capital levels differ rather notably based on what continent the foreign-born stem from. Thus, I conduct a heterogeneity analysis, and group the foreign-born by their continent-origin and examine whether associations with inequality differ. In conjunction with the thesis's primary purpose, I have an extensive discussion about the technical complexity of estimating inequality using so-called

[2]Defined as being born in different countries.

"grouped" income data - more on this issue in section 3.

My primary conclusion is that the PFB is a robust determinant of income inequality, even controlling for fixed-, and various time-varying effects. In the heterogeneity analysis, I, instead, find that the European-born proportion significantly predicts equality. African-born proportion yields significant associations with inequality, while the Asian-born proportion is insignificantly associated with inequality. I conduct an exploratory analysis to examine quadratic effects. I find that for municipalities with more than 10% PFB, a negative squared PFB term is significant, somewhat indicating a non-linear effect.

The thesis structure is such that section 2 discusses the characteristics of Swedish immigration(2000-2019); again, this is necessary to give theoretical predictions. Section 3 discusses data collection, technical challenges, and general data descriptives. Empirical specifications can be found in section 4. Section 5 present a descriptive analysis of inequality development for municipalities with the most significant PFB changes. Section 5 also discusses why a quadratic relationship between the PFB and inequality could be appropriate.

2 Heterogeneity within the Swedish immigrants

Effects of immigration on inequality can manifest in many forms. Therefore, this section is devoted to describing the human capital background of Swedish immigration under the period 2000-2019. An essential procedure to state hypotheses about the relationship between immigration and inequality. As touched upon, human capital can be argued to incorporate many things, meaning different "human capital indexes" show different results. This thesis will use two indexes to build a perception about immigrants' human capital levels: the human capital index and a self-proposed education index.

As of 2018, the World Bank, based on the above-discussed research, created the human capital index (further on denoted HCI). I leave out the exact technical specification due to its complex nature. Nevertheless, the HCI-index incorporates the tripartite effect of schooling, quality, and health in a trial to assess the human capital for workers originating from different countries. The HCI somewhat indicates the expected human capital for some countries but ignores regional variation. It is not certain that a "representative" person immigrates to Sweden. Therefore, a meaningful discussion is whether there might be a selection bias; simply put, is the immigrants representative for their countries' estimated human capital levels? Speculating; people who have the financial resources to migrate may be better equipped than the average person. I do not have data on the education quality or health status of Swedish immigrants. Hence I can not replicate the HCI for Swedish immigrants. Nevertheless, as an attempt to control for such selection bias, I propose a weighted education index to control whether the countries *ranking* is similar for the two indexes. If the ranking is similar, we should observe a high correlation when plotting the indexes versus each other - indicating low selection bias. On the contrary, if the ranking is inconsistent, the correlation will be sparse - indicating a selection bias.

My education index is essentially weighting the actual education observed for Swedish immigrants towards their proportions. I use data from SCB. Five different education levels are available: "less than high school," "high school education," "less than three years of university education," "more or equal to three years of university education," and "Ph.D. level education." I code these education levels to 1, ..., 5 - one represents

”less than high school,” etc. For the 40 biggest immigration countries[3], I calculate the proportion of the total number of people who have immigrated over the period 2000-2019 with education levels 1,...,5. Finally, weighting the proportions to education-levels, and summing yields the index. Figure 3 showcases the HCI[4]-, and education index plotted against each other.

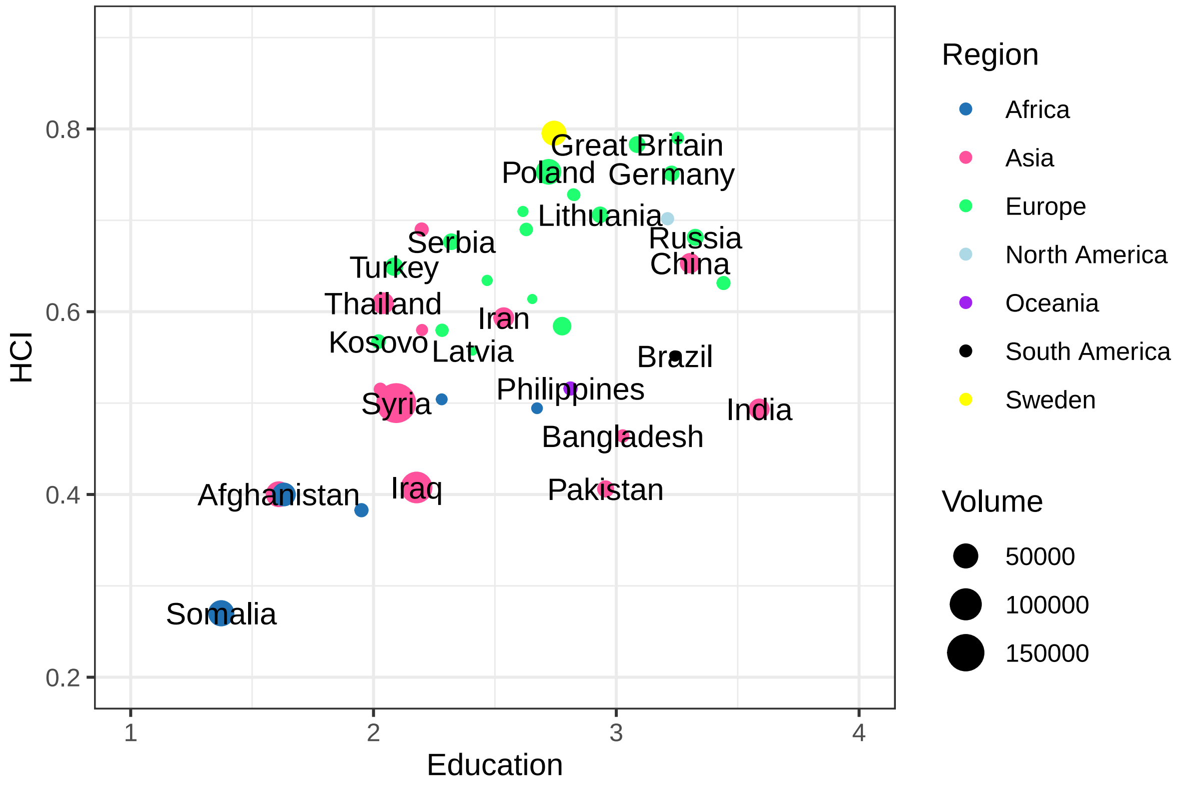

Figure 3: Estimates of immigrating nation’s human capital levels. Note: *Sweden is included for comparison, and the volume of Sweden is not according to scale.* Further, only a selection of countries are labeled to avoid clutter.

The correlation is high, indicating low selection bias, but there are certain exciting outliers. India $(3.6, 0.5)$ - having the highest estimated education index in Sweden but a relatively low HCI score. Somewhat signaling that the Indian-born group in Sweden is better equipped than what one could expect when using a macro index such as the HCI. Further, we can see that the Asian region is characterized by considerable variation in the education index, having a few notably high and low score countries. The European region HCI-indexes are more similar to each other but are also subject to considerable education-index variation. The African group has overall relatively low HCI-, and education indexes. Finally, we observe that Sweden has the highest estimated HCI-score of all countries but has a lower education index than a handful of countries. There are undoubtedly many other essential but unobservable factors that could be argued to impact one’s human capital, such as language skills, social networks, or psychological health. Nevertheless, I still argue that these estimates can give a meaningful description of the immigrants’ human capital backgrounds.

[3]I choose to present the top 40 countries since they represent a notable large proportion of this period’s total immigration: 89.3%.

[4]Three countries lack HCI scores and are therefore linearly extrapolated. See details in appendix 11.2.

To summarize, human capital theory, in an immigration context, emphasizes that it is the relative conditions between the natives and immigrants that influence the inequality outcomes. Figure 3 shows notable heterogeneity between the continents. Lower human capital estimates characterize the African region compared to that of Sweden. Therefore, given the theory of human capital, it is likely that an increasing share of African-born "treats" municipalities with heterogeneity - causing inequality. The European-born immigrants do stem from countries with comparable HCI estimates to that of Sweden's, but their observed education varies; overall, I do not expect this group to affect inequality significantly. The Asian region has, in general, lower HCI-indexes compared to Sweden, but similarly to the European-born, this region has a significant variation in education. Nevertheless, one can observe larger volumes from countries with lower education-, and HCI indexes. Hence, it is likely that this group, as a whole, also treats municipalities with human capital heterogeneity, predicting inequality.

3 Data

Data originate from Statistics Sweden and the Swedish public employment service - all variables are calculated over the period 2000-2019 for all 290 Swedish municipalities. Section 3.1 discusses the data on foreign-born, 3.2 discusses the available income data, and how to assess inequality from a discrete structure, 3.3 discusses alternative inequality indexes used for robustness test, and section 3.4 discusses my selection of appropriate covariates.

3.1 Foreign-born data

SCB summarizes the number of people in every municipality annually. Such spreadsheets contain information regarding the sum of foreign-born and their continent-origin. Two conditions define a foreign-born person: being born abroad and being registered as a municipality member. Therefore, one important excluded foreign-born group is asylum-seekers, which instead get registered when they have their asylum admission accepted. I calculate the annual PFB on municipality-level as well as the continent proportions and merge the 20 individual data frames. I present a few year-pooled descriptive statistics:

Variable	Mean	Sd	Min	Max	Average Increase$_{2000 \rightarrow 2019}$
PFB	0.11	0.06	0.02	0.43	0.07
Europe	0.07	0.04	0.02	0.39	0.01
Asia	0.03	0.02	0.00	0.21	0.04
Africa	0.01	0.01	0.00	0.06	0.01
South America	0.00	0.00	0.00	0.05	0
North America	0.00	0.00	0.00	0.02	0
Oceania	0.00	0.00	0.00	0.00	0

Table 1: Descriptive statistics for foreign-born proportions. Note: All estimates are truncated at two digits.

We note how the European-born group is the largest subgroup on average; they also have the largest variation in representation. Asian-born proportions have had the highest increase throughout the period of interest, while both European-, and Africa-born have had slight increases. American and Oceania-born proportions seem to be somewhat negligible in the grand scheme of things.

3.2 Income data, Inequality, and Simulation

To derive inequality indexes, I will use non-negative income data originating from SCB. As touched upon earlier, such income data contains one's wage - which is the dominant factor for most people-and sick pay, pension, activity compensation, and parental support.

SCB presents annually, at municipality-level, 26 different income intervals, in which the interval average and the number of people in such interval can be obtained. Intuitively, due to SCB "grouping," i.e., registering the number of people in every interval, but not people's specific income, such discrete data structure is called "grouped income structure." There is no pre-created data that can be downloaded. Instead, I use the statistical software R to code income vectors. Essentially I create a function $f(arg1, arg2)$ that takes two arguments: (1) the mean in each of the 26 intervals and (2) the number of people in every interval. Nest-looping $f()$ over all municipalities and years return income vectors that I use to assess inequality. Since Sweden have 290 municipalities, and my period stretches over 20 years, I create 5800 income vectors. To showcase the data structure, I pool all datapoints and calculate a few statistics:

Range	Interval	Proportion	Mean	Min	Max
1	[0, 0]	0.043	0.00	0.00	0.00
2	(0, 190]	0.04	84.32	56.00	111.00
3	(190, 390]	0.021	292.52	267.00	333.00
4	(390, 590]	0.023	502.2	473.00	552.00
5	(590, 790]	0.024	708.88	676.00	754.00
6	(790, 990]	0.044	910.02	887.00	936.00
7	(990,1190]	0.05	1100.13	1083.00	1134.00
8	(1190, 1390]	0.052	1301.55	1288.00	1315.00
9	(1390, 1590]	0.058	1501.66	1490.00	1513.00
10	(1590, 1790]	0.062	1701.01	1688.00	1713.00
11	(1790, 1990]	0.062	1899.65	1887.00	1911.00
12	(1990, 2190]	0.06	2099.28	2085.00	2109.00
13	(2190, 2390]	0.057	2298.81	2281.00	2312.00
14	(2390, 2590]	0.053	2498.15	2478.00	2511.00
15	(2590, 2790]	0.048	2698.44	2679.00	2712.003
16	(2790, 2990]	0.044	2898.12	2877.00	2914.00
17	(2990, 3190]	0.04	3096.69	3075.00	3112.003
18	(3190, 3390]	0.035	3296.6	3265.0	3321.0
19	(3390, 3590]	0.029	3496.19	3466.00	3528.00
20	(3590, 3790]	0.025	3695.85	3651.00	3730.00
21	(3790, 3990]	0.02	3896.03	3844.00	3932.00
22	(3990, 4990]	0.061	4412.39	4231.00	4506.00
23	(4990, 5990]	0.024	5423.5	5156.0	5682.0
24	(5990, 7990]	0.016	6760.48	6266.006	7408.00
25	(7990, 10000]	0.005	8817.54	8194.00	9656.00
26	(10000, ∞)	0.004	14654.52	10496.00	47073.00

Table 2: Descriptions regarding grouped income. Note: the intervals describe yearly income in thousands.

Table 2 shows that most densities are located in the middle intervals, which is intuitive. It's also noteworthy that many people have a zero income. It is more frequently occurring that a person has zero income than a really low one. To showcase an example of how a created income vector's density can look like, I plot the density of municipality Gothenburg, at the year 2000:

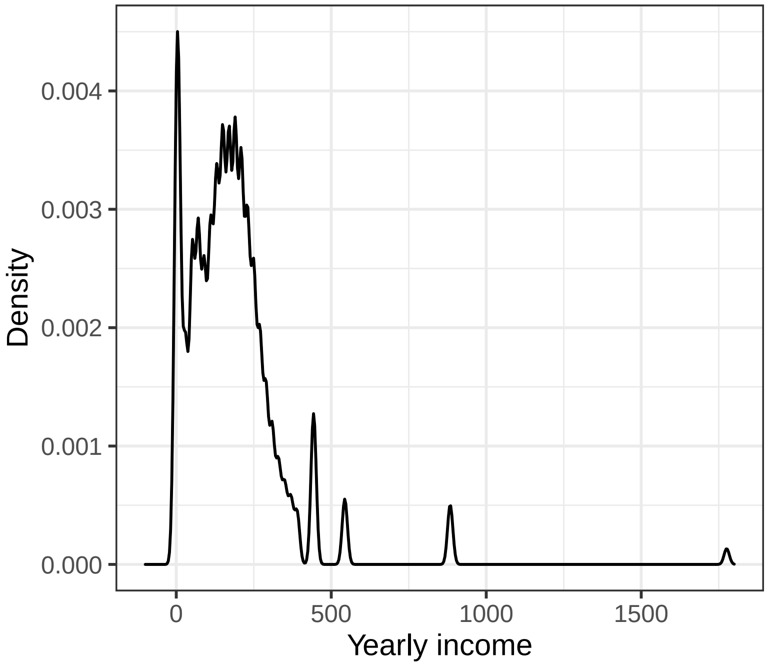

Figure 4: Density of Gothenburg income vector year 2000.

Likewise, one can observe that many people have a zero income, a large density is located in the left tail of the distribution, and the grouped data structure causes such ”spikes” in the density.

Now, when it comes to estimating inequality, what are the implications of using such discrete structure? I can accurately capture the total income in every interval. This is because the average income within an interval times the number of people in such interval yields equivalent value as summing every specific income, i.e., $N * \bar{X} = \sum_{i=1}^{N} X_i$, where x_i denotes the income of a person. Thus, assessing how the total interval wealth transfers over time can correctly be measured. Nevertheless, it is likely that estimating inequality will be underestimated since one ignores income variation *within* every interval - except in interval one. Interval one only consists of zero-incomes, therefore not containing any variation.

3.2.1 Simulation experiment

To estimate the magnitude of measurement error, I simulate continuous income data from a given distribution. I compare inequality indexes derived from simulated grouped and simulated non-grouped data. The results allow me to infer what the implications are of using a grouped income structure. This section will repetitively refer to four different distributions; hence, it can be beneficial to clarify them. The ” true” continuous distribution is the *actual* income distribution that exists in reality - this is unobserved. Instead, I observe from SCB a ”grouped” version of the true distribution; this is the income data I will use throughout the thesis. Then, I simulate a continuous distribution for experiment reasons, and finally, I transform the simulated continuous distribution into a grouped version to examine the effects of such grouping.

The eminent problem is to select a parametric distribution that allows for reliable simulations. Many different distributions have been argued to mimic income distributions well. Nevertheless, recently Jäntti et al. (2018) examined a total of 5000 grouped datasets and found that a generalized Beta distribution of the second kind (further on denoted GB2) has an astonishingly good fit. Hence, I will follow their findings and use this distribution. The probability density function; $p(x)$ of GB2 is given by:

$$p_{GB2}(x) = \frac{a(x/b)^{ap-1}}{bB(p,q)(1+(x/b)^a)^{p+q}}, x > 0 \tag{1}$$

With parameters: $a > 0$, $b > 0$, $p > 0$ and $q > 0$, also note how $B(p, q)$ represents the Beta function. Parameter estimation is conducted by maximum likelihood from R package *GB2*. Although, note how this distribution only applies to positive incomes.I propose that a uniform distribution with lower and upper bounds equivalent to zero is intuitive for zero incomes. Hence, the complete income distribution is well represented by such a mixture distribution:

$$p_{mix}(x) = w * p_{Uni}(x) + (1-w) * p_{GB2}(x), x \geq 0 \tag{2}$$

The weight, $w : w \in (0, 1)$ is set to the actual proportion of zero incomes observed in municipality i, at year t.

To summarize this experiment: the problem is that the "true" continuous income distribution is unobserved, instead I observe a grouped version. To infer what this implies for my results, I'll conduct the following steps:

1. I assume that the true continuous income distribution follows a GB2 distribution. Estimating the parameters of the GB2 distribution by inputting the observed grouped sample. Note that I do this for all municipalities i, at year t.

2. After the parameters have been estimated, for let's say municipality $i = 1$, at year $t = 2000$; I simulate a continuous distribution (simulation size:[5] $N_{sim} = 100000$). This distribution will mimic the grouped one. The point is that it's somewhat realistic. Simulation is conducted using R-functions $runif$ for the uniform distribution, and $rgb2$ for the GB2 distribution.

3. I then estimate inequality (Gini-index - details of Gini in next section) for the continuous simulated sample.

4. Then, I transform the simulated continuous sample into a group one - according to the boundaries presented in table 2, and derive inequality from the now grouped sample.

5. I compare the difference in estimated inequality between the continuous simulated sample and the transformed one.

6. Finally, going back to the original grouped distribution, I can now infer the effects of not observing the true continuous distribution.

[5]Since estimates from simulations are subject to a certain degree of uncertainty I discuss the simulation size in appendix 11.1.

As an example, I showcase the simulation for Gothenburg income data in the year 2000. Note how figure 4 showcased the density of the original grouped distribution. The blue line now represents the density of a simulated sample, using Gothenburg's estimated parameters.

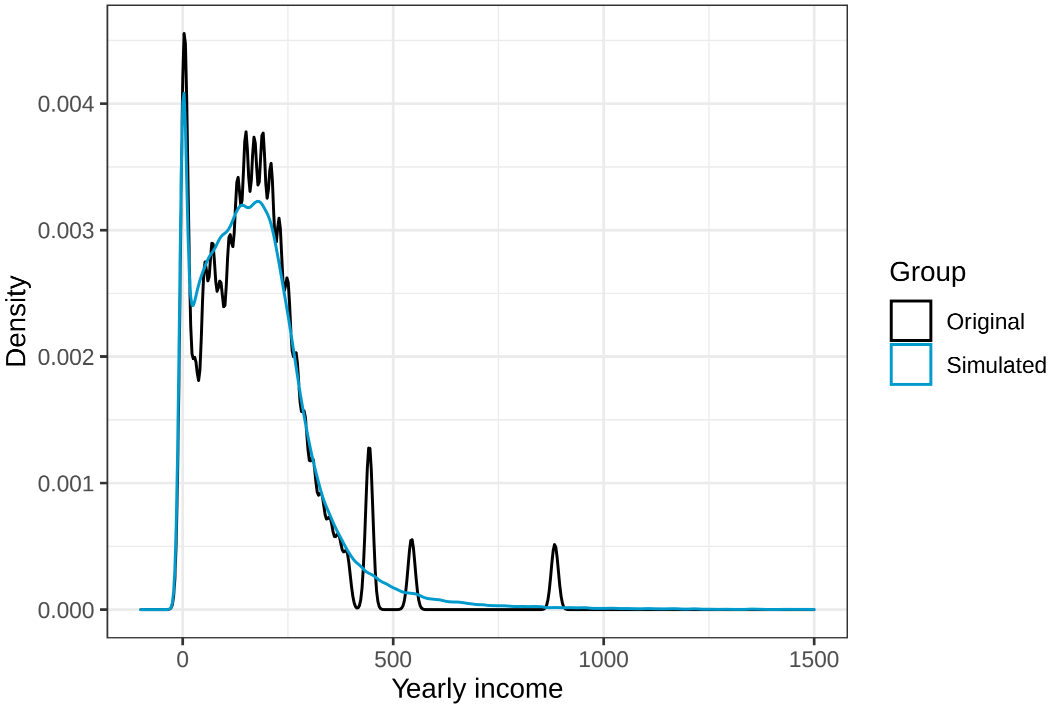

Figure 5: Density fit for Gothenburg income vector year 2000.

The results are rather appealing; such smoothed income distribution fits well, meaning it follows a realistic depiction of the income distribution.

The fundamental question is now, *if* I transformed my simulated Gothenburg sample according to the boundaries presented in table 1; how will measured inequality differ between the simulated continuous sample and the transformed sample? First, I present what inequality is by discussing my main inequality index, which is the Gini-measurement.

The definition of the Gini-coefficient is given by eq.3 and can be understood as half of the relative mean absolute difference. As touched upon in the introduction, the estimate $G : G \in [0, 1]$ indicates lower values for equal populations and vice versa. If all incomes, $X_i = \{X_1, ..., X_n\}$, are equivalent, the Gini estimate is zero - perfect equality. If all but one income is zero, society is characterized by perfect inequality.

$$G = \frac{\sum_{i=1}^{N} \sum_{j}^{N-1} |X_i - X_{j \neq i}|}{2N^2 \bar{X}}, X_i \geq 0 \tag{3}$$

We understand that the numerator drives up the inequality when the summation of the absolute income differences increases. The index is decreasing in the mean income; however, the relation is rather complicated. Suppose density within the right tail of the distribution increase in absolute numbers, meaning more people earn a high income (leading to a higher average). In that case, it may still increase the indexation since the numerator's "disparity effect" overtakes a larger denominator.

Returning to the grouping discussion, how will the estimate G be affected if one inserts the simulated continuous sample versus if one inserts a transformed grouped version? Examining the effect for the Gothenburg case; *before a transformation*, I estimate the Gini coefficient to 0.418, but after the transformation, it is reduced to 0.34. A natural result since, as argued, we should lose out on some of the variation within all but the first range. I repeat such an experiment for all municipalities $i = 1, ..., 290$, at year $2000, ..., 2019$, and calculate the reduction in the Gini-estimate when grouping. I showcase the tabulated results for the 5800 cases:

	Mean	Sd	Min	Max
Reduction in Gini	0.05	0.02	0.01	0.11

Table 3: Estimated effects of grouping.

Inequality is underestimated in each case (since $\min\{reduction\} > 0$). On average, Gini-estimates are reduced by 0.05 units or 14.5%. I also calculate the correlation between the sample of simulated continuous Gini-coefficients and the sample of transformed Gini-coefficients to examine whether the variation in the samples is similar. The correlation estimate shows 0.95, meaning the sample vectors tend to move together in the same direction.

Summarizing, these results indicate that Gini estimates derived from my original grouped data instead of the true continuous distribution will underestimate the actual inequality. My experiment indicates an underestimation of about 15%. The sample correlation was estimated high, indicating that inequality variation should be well captured by the grouped data. An important conclusion since I am using such inequality variation in regression frameworks, therefore standard errors and significance should be relatively trustworthy.

3.3 Alternative inequality indexes

In conjunction with the Gini estimate discussed above, I use a few alternative indexes to more precisely examine what tail of the income distribution is associated with a change in PFB. I showcase a descriptive table with basic statistics, where the statistics are calculated using pooled data. I discuss the index's characteristics and critical differences and finally present a figure showcasing the index's annual trend between 2000-2019.

Variable	Label	Mean	Sd	Min	Max	Skewness
Gini	Gini-coefficient	0.36	0.034	0.28	0.55	1.38
Theil	Theil-coefficient	0.2	0.04	0.13	0.5	2.37
L4	Share of incomes in four lowest groups.	0.13	0.02	0.06	0.26	0.91
L8	Share of incomes in eight lowest groups.	0.3	0.06	0.14	0.55	0.42
L13	Share of incomes in thirteen lowest groups.	0.6	0.13	0.31	0.9	0.24
Skewness	Skewness of the distribution.	1.94	0.87	0.1	16.61	3.39

Table 4: Descriptive statistics for inequality (independent) variables.

3.3.1 Theil's L index

One frequently used index that is aimed at capturing variation in the left tail of the distribution is the Theil's L index (Cowell, 2016):

$$T = \frac{1}{N}\sum_{i=1}^{N} ln\left(\frac{\overline{X}}{X_i}\right) \tag{4}$$

The benefit of such an index is that the index moves more with changes in the left tail of the distribution - due to the logarithmic nature - meaning this index is more efficient in capturing variation caused by left tail impacts. Hence, an excellent complement to the Gini that only assesses actual differences, independent of where in the distribution that such differences arise.

3.3.2 Skewness

$$\hat{\eta}_3 = \frac{\hat{\mu}_3}{\sigma^3} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2]^{3/2}} \tag{5}$$

More directly, to assess how the proportion of foreign-born affects the distribution of the income distribution, one can calculate the skewness of income data. Where positive skewness, in this context, means that a higher proportion has an income in the lower intervals.

3.3.3 L4, L8 & L13 - index

To further assess what tail of the distribution is affected by immigration, I simply calculate the proportion of income-takers in the 4,8, and 13 lowest income ranges; I denote these variables $L4, L8, L13$. These three variables decrease every year due to two reasons: first, inflation causes people's income to increase in nominal value, simultaneously SCB uses fixed interval boundaries. These two conditions cause less, and less people to be registered in the lower intervals. Nevertheless, it's interesting to examine whether municipalities with a more considerable increase in PFB also have a *lower decrease* of people in the lower-income groups.

Visualizing the annual municipality-average trend the entire set of inequality variables:

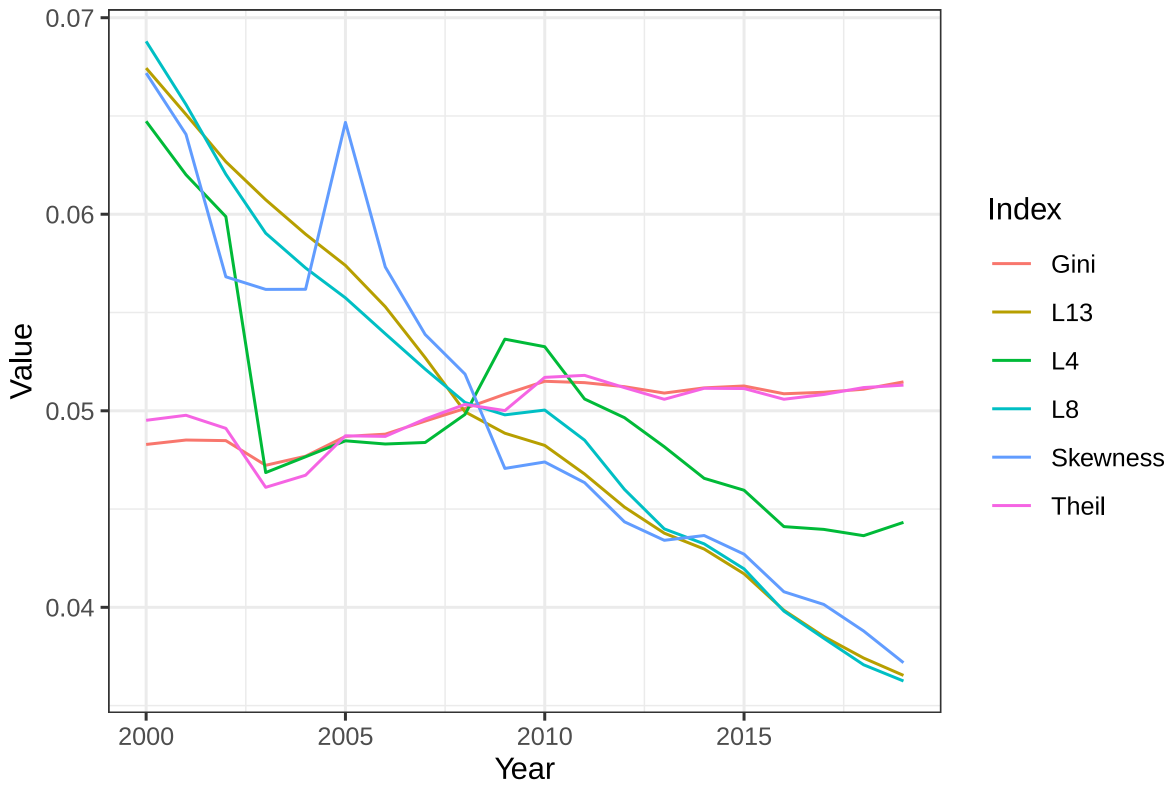

Figure 6: Inequality trend. Note: all indexes are normalized to fit an appropriate scale.

Overtime; $t = 2000, ..., 2019$, we note that the municipality average has increased for Gini and Theil - indicating increased income inequality. $L4, L8$, and $L13$, slowly go down due to inflation, whilst the skewness of income vectors become more right-tailed. It's notable how the $L4$-index is more volatile; $L4$ is greatly determined by the number of people having zero incomes which could be influenced by many factors such as unemployment effects or other macro-related shocks. $L8$, and $L13$ captures more intervals and are therefore less prone to variation within a single interval; consequently, they are more stable. Given predictions from section 2, the question is whether the L-indexes are estimated to be "less" negative and "less" right-skewed in municipalities with a higher stock of PFB. Repeating the descriptive procedure for my selection of independent covariates.

3.4 Independent variables

Variable	Label	Mean	Sd	Min	Max	Skewness
PFB	Proportion of foreign born	0.11	0.06	0.02	0.43	1.55
Meanage	Municipality mean age	42.67	2.76	35.7	50.5	-1.51
High_Educ	Proportion with uni education $\geq$ 3yrs	0.13	0.06	0.04	0.5	1.87
F_M	Municipality female to male ratio	0.99	0.3	0.87	1.1	-0.08
Meanwage	Municipality mean wage	230.64	49.68	138.09	590.65	1.11
Unemp	Total unemployed/population	0.08	0.03	0	0.23	0.1

Table 5: Descriptive statistics for independent variables.

A normal confounding effect that hinders convergence to an unbiased parameter is various time-varying variables. We are interested in how the proportion of foreign-born affects inequality, but several variables could invalidate the exogeneity assumption. I argue for the following covariates:

- I include *proportion of highly educated* since such share tends to stretch the right tail of the income distribution, causing considerable disparity between incomes. Being highly educated is defined by having a bachelor's degree at minimum.
- Controlling for *female-to-male ratio* since the gender distribution impacts the income distribution through social norms, systematic career selections with different income opportunities, and there might also be the eminent issue with wage discrimination.
- Also *mean age*, and *(mean age)*^{2} since the age distribution directly links to the income distribution - older persons tend to have higher incomes. However, the effect is likely to be diminishing; hence I also add a squared term. If a municipality is shocked by a large density of young people, the income distribution is naturally skewed positively.
- Then, I control for various *economic indicators*, inferring that municipalities with better conditions should have an easier time finding jobs for job searchers.
    1. *Meanwage*, due to more prosperous municipalities capturing people with higher incomes - stretching the distribution to the right, causing inequality.
    2. *Unemp*, as a proxy for beneficial economic conditions, unemployment is a good signal of the number of real opportunities that present themselves. It is intuitive that municipalities with high unemployment have equal incomes and vice versa.

I visualize the annual municipality-average trends in figure.7:

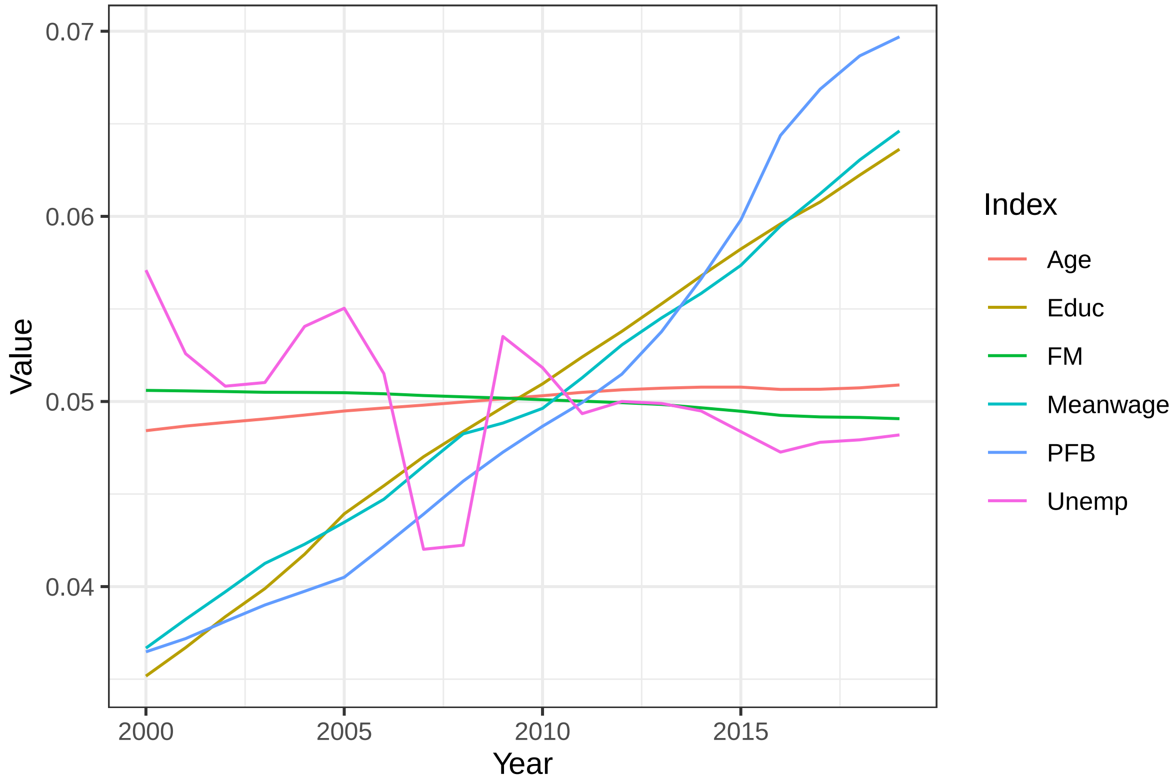

Figure 7: Covariate trends. Note: all indexes are normalized to fit an appropriate scale.

We observe that the PFB, mean wage and proportion of highly educated have a rising trend in Sweden. The population has transitioned into a slightly more male-heavy distribution, and we have an aging population. Unemployment is more volatile, which is natural - it should be noted that the last data point is registered before the COVID outbreak. Hence the unemployment shock is unnoticed for this data frame.

4 Empirical specifications

My baseline specification consists of a univariate model in which I model inequality (assessed by the Gini-coefficient) by the PFB. Dummies are included to control for time and municipality-fixed effects. I then update the model iteratively by adding covariate $\{2, ..., 6\}$ to observe changes of fitted parameter estimates.

	Dependent variable:					
	Gini					
	(1)	(2)	(3)	(4)	(5)	(6)
PFB	0.274***	0.282***	0.222***	0.224***	0.182***	0.189***
	(0.008)	(0.010)	(0.011)	(0.011)	(0.013)	(0.013)
Age		0.003	0.007***	0.006***	0.010***	0.011***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Age^2		−0.00004	−0.0001***	−0.0001***	−0.0001***	−0.0001***
		(0.00002)	(0.00002)	(0.00003)	(0.00003)	(0.00003)
High_Educ			−0.205***	−0.205***	−0.096***	−0.094***
			(0.016)	(0.016)	(0.018)	(0.018)
F_M				0.013	0.007	0.006
				(0.010)	(0.010)	(0.010)
Meanwage					−0.0002***	−0.0002***
					(0.00002)	(0.00003)
Unemp						−0.036***
						(0.010)
Observations	5,800	5,800	5,800	5,800	5,800	5,800
Adjusted R^2	0.999	0.997	0.997	0.997	0.997	0.997
R. Std. Error	0.006	0.006	0.006	0.006	0.006	0.006
F Statistic	70,857***	70,438***	74,268***	74,046***	76,020***	75,979***
Fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 6: Linear Panel Regression Models of Inequality. Note: All numbers are rounded.

All point estimates are derived with the OLS method with heteroscedasticity-robust standard errors. Analyzing the fitted parameters, table 6 shows that the PFB is significant on $\alpha = 0.05$, *independent of model specification*. It can be worth pointing out that fixed effect models adjust for the "initial stock" of PFB, meaning foreign-born before year of 2000 do not impact the estimates. Including the proportion of highly educated and municipality mean wages seem to be important when trying to control for the exogeneity condition since they impact the PFB's point estimate greatly. The model's explanatory power is very high ($adjR^2 = 0.997$) due to the high correlation between the vector PFB-, and Gini, as well as including many dummies. However, this should not be interpretable as a true cause-and-effect relationship. Instead, the mechanism behind the two variables needs to be further examined. I denote my main specification (model 6):

$$Gini_{it} = \beta_0 + \beta_1 PFB_{it} + \gamma \mathbf{X}_{it} + \delta_2 T_2 + ... + \delta_{20} T_{20} + \eta_2 M_2 + ... + \eta_{290} M_{290} + \varepsilon_{it} \quad (6)$$

Where $\mathbf{X}_{it}$ is the vector of covariates controlling for time-varying effects included in model 6. $T_2, ..., T_{20}$ are dummies included to capture yearly fixed effects. $M_2, ..., M_{290}$ are dummies included to capture municipality fixed effects.

5 Non-linear effects

One can speculate that heterogeneity in inhabitants' human capital goes down when PFB increases to some cut-off points. Meaning we see some segregation effect, as discussed by several sources, Aldén et al. (2015), Attenius(2020), or even touched upon by the Swedish civil minister himself (Baylan, 2017). At some "cut-off" point, a filtration effect occurs; primarily native Swedes and educated foreign-born decide to move out, leaving immigrants behind with lesser beneficial backgrounds. Such an effect should result in a quadratic relationship between inequality and PFB. I test this speculation by fitting eq. 6 again, including a squared PFB-term.

Contradictory, we observe that the squared term shows a highly significant but positive coefficient - indicating an exponential effect rather than a concave one. However, it may be that the variation is not structured properly when using the entire data set, meaning that a concave effect becomes more clear when subsetting on municipalities with a higher level of PFB. To examine this, I subset the dataframe on $PFB \geq \left\{ \frac{0}{20} \times \frac{4}{20}, \frac{1}{20} \times \frac{4}{20}, ..., \frac{20}{20} \times \frac{4}{20} \right\}$ to examine different parameter estimations for the squared term. I plot the point estimates:

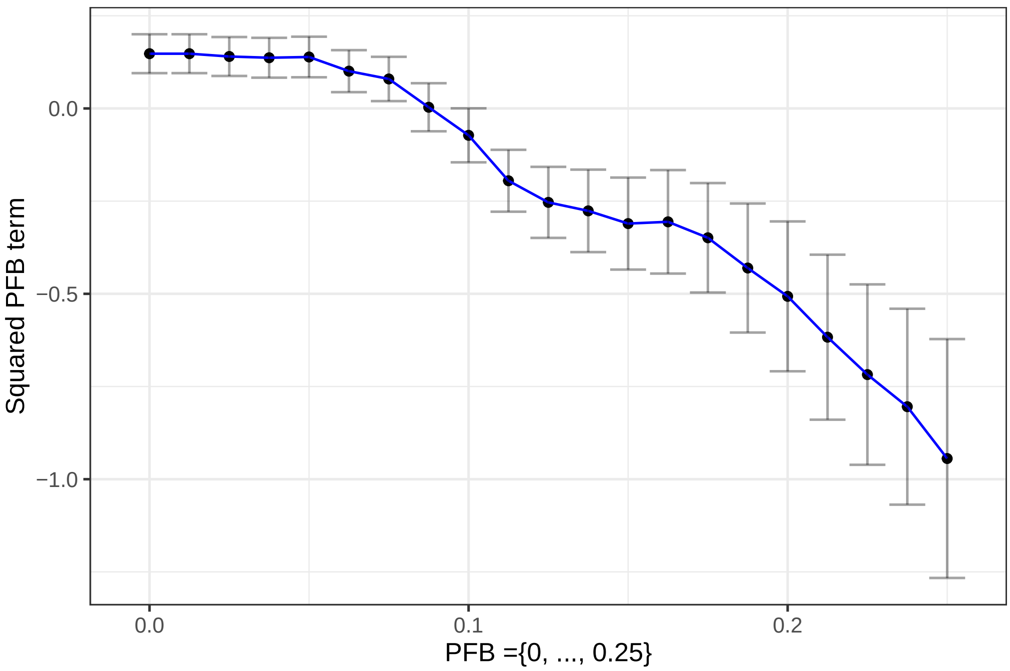

Figure 8: Subset point estimate for squared PFB-term. Note: Error bars contain the 95% confidence interval.

I observe a significant negative point estimate for a squared PFB-term when subsetting the data frame on municipalities with a proportion foreign-born higher than 1/10, somewhat indicating, but not proving, what Aldén et al. and others discuss. At some point, native Swedes and high human capital immigrants move away. Hence the inequality effect becomes quadratic when those who are staying become more homogeneous.

5.1 Further examination

I explore whether the initial stock of PFB in the year 2000 matters for the estimated effect of PFB on inequality. To do this, I calculate the average PFB year 2000 and classify all municipalities depending if they belong to the more PFB rich group or not. If we believe the quadratic effect, the estimated effect of PFB should be higher for municipalities with lower initial stock.

	PFB_{LOW}	PFB_{HIGH}
$\hat{\beta}_{PFB}$	0.25^{***}	0.11^{***}

Table 7: Heterogeneity assesment.

As aligned with the quadratic theory, the fitted effect of PFB to Gini is considerably higher for municipalities that had a lower than average PFB stock in the year 2000.

5.2 Outlier municipalities

To examine further evidence - I analyze actual effects for PFB outliers. More specifically, I examine the average inequality trend for the municipalities with the ten highest- and lowest PFB increases between 2000 and 2019. Again, to explore whether the initial stock matters, I conduct this analysis conditioned on whether municipalities had a higher or lower initial stock than the average year 2000. Therefore we get four subgroups with different conditions. Suppose parameters estimated in model specification six are reliable. In that case, we are likely to note a higher inequality increase for the municipalities with a higher PFB increase and vice versa. But (based on the previous section's result) also that the effect should be larger if the municipalities had a smaller initial stock.

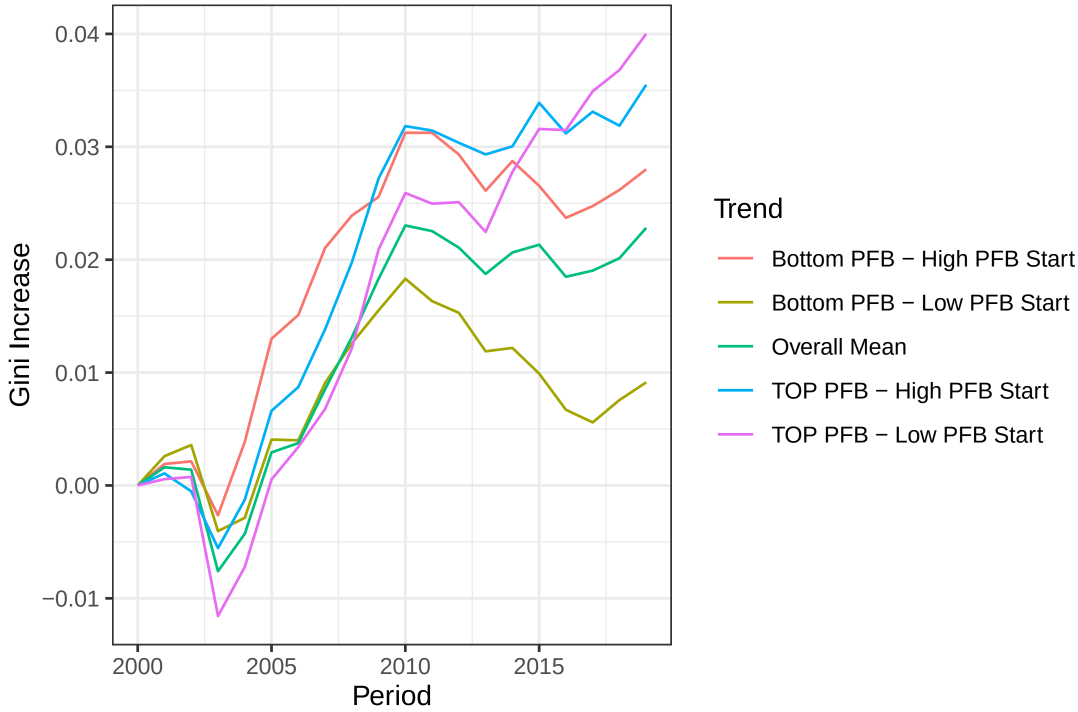

Figure 9: Inequality development for PFB outliers. *Note*: The y-axis represents Gini-deviations from the initial value year 2000.

Figure 9 does reveal exciting aspects. Municipalities with higher PFB increases have had a higher increase in the Gini estimate. The effect is higher for those municipalities characterized on a lower initial stock. Vice versa, the lowest inequality increase is observed for municipalities with low PFB increase and being conditioned on having a low initial stock. Whilst only being descriptive evidence, it is interesting that expected effects are, in fact, observed - adding another piece to the overall conclusion.

6 Robustness check

If an economic theory predicts correctly, a large density of immigrants with notably lower levels of human capital (as discussed in section 2) should cause inequality through a shock to the left tail of the income distribution. Hence, as a robustness check, I also examine whether left-tail indexes significantly correlate with PFB.

	Dependent variable:					
	Gini	*Theil*	*L4*	*L8*	*L13*	*Skewness*
PFB	0.189***	0.138***	0.200***	0.310***	0.397***	1.450
	(0.013)	(0.016)	(0.016)	(0.022)	(0.027)	(1.448)
Age	0.011***	−0.002	0.058***	0.140***	0.059***	−2.034***
	(0.002)	(0.003)	(0.004)	(0.005)	(0.006)	(0.244)
Age^2	−0.0001***	0.00002	−0.001***	−0.002***	−0.001***	0.024***
	(0.00003)	(0.00003)	(0.00005)	(0.0001)	(0.0001)	(0.003)
High_Educ	−0.094***	−0.149***	0.093***	0.378***	0.243***	−15.058***
	(0.018)	(0.024)	(0.035)	(0.048)	(0.064)	(1.461)
F_M	0.006	0.019	−0.021	0.049**	0.081***	−1.925
	(0.010)	(0.013)	(0.016)	(0.021)	(0.023)	(1.895)
Meanwage	−0.0002***	−0.0001***	−0.0004***	−0.0001	0.0001	0.011***
	(0.00003)	(0.00004)	(0.0001)	(0.0001)	(0.0001)	(0.002)
Unemp	−0.036***	−0.042***	0.107***	0.223***	0.392***	−3.117**
	(0.010)	(0.013)	(0.017)	(0.023)	(0.027)	(1.340)
Observations	5,800	5,800	5,800	5,800	5,800	5,800
Adjusted R^2	0.999	0.999	0.996	0.999	0.998	0.911
R. Std. Error	0.006	0.007	0.008	0.011	0.013	0.636
F Statistic	75,979***	16,639***	4,552***	14,5230***	40,901***	188***
Fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 8: Linear Panel Regression Models of Inequality.

All point estimates are derived with the OLS method with heteroscedasticity-robust standard errors. We observe that the PFB significantly impacts all inequality indexes. It does not predict the skewness of the distribution, though. Nevertheless, I conclude that the proportion of foreign-born is a robust determinant of inequality.

7 Heterogeneity analysis

There is a large degree of heterogeneity within the foreign-born. As discussed in section 2, there are, on the continent level, notable differences in estimated human capital. This section analyzes, therefore, how the proportions of different continents are associated with inequality.

	Dependent variable:					
	Gini	*Theil*	*L4*	*L8*	*L13*	*Skewness*
Europe	−0.029***	−0.026***	−0.056***	−0.099***	−0.085***	0.168
	(0.008)	(0.007)	(0.011)	(0.013)	(0.014)	(0.753)
Asia	0.012	0.015	0.019	0.072***	0.118***	−3.450***
	(0.009)	(0.011)	(0.013)	(0.016)	(0.020)	(1.078)
Africa	0.085***	0.107***	0.100***	0.077**	−0.057	13.904***
	(0.020)	(0.024)	(0.026)	(0.036)	(0.040)	(2.172)
Age	0.005**	−0.006**	0.051***	0.130***	0.046***	−2.078***
	(0.002)	(0.003)	(0.004)	(0.005)	(0.006)	(0.255)
Age^2	−0.0001***	0.0001	−0.001***	−0.002***	−0.001***	0.024***
	(0.00003)	(0.00003)	(0.00005)	(0.0001)	(0.0001)	(0.003)
High_Educ	−0.100***	−0.152***	0.087**	0.361***	0.201***	−13.823***
	(0.020)	(0.024)	(0.036)	(0.050)	(0.067)	(1.461)
F_M	−0.030***	−0.007	−0.058***	−0.010	0.003	−2.001
	(0.011)	(0.013)	(0.016)	(0.021)	(0.024)	(1.742)
Meanwage	−0.0003***	−0.0002***	−0.0005***	−0.0002***	−0.0001	0.011***
	(0.00003)	(0.00004)	(0.0001)	(0.0001)	(0.0001)	(0.002)
Unemp	0.003	−0.013	0.149***	0.284***	0.465***	−2.398*
	(0.010)	(0.013)	(0.017)	(0.023)	(0.027)	(1.268)
Observations	5,800	5,800	5,800	5,800	5,800	5,800
Adjusted R^2	0.999	0.999	0.996	0.998	0.997	0.910
R. Std. Error	0.006	0.007	0.009	0.012	0.013	0.639
F Statistic	70,898***	16,283***	4,1214***	11,192***	38,147***	185.854***
Fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 9: Heterogeneity assessment - Europe, Africa, Asia.

All point estimates are derived with the OLS method with heteroscedasticity-robust standard errors. These results tell a different story. The Europe-born proportion, that was deemed rather similiar in terms of human capital are estimated as a significant negative predictor of all inequality indexes. The Asian-born proportion are insignificant predictors of inequality (Gini and Theil), whilst positively predicting $L8$, and $L13$. African-born proportion positively predict the inequality indexes, as well as the lower proportion-indexes. The African-born proportion is also a strong determinant of a positively skewed income distribution, where again a positively skewed distribution essentially means more density in the lower-income intervals.

7.1 Outlier municipalities 2

Repeating a similar descriptive analysis from section 5.2, we can observe whether changes have occurred in line with what we would expect according to our fitted model parameters. I examine the inequality trend (measured by Gini) for the top ten municipalities with the highest increase in Africa-, Asia-, and European-born proportions. Again I include the overall trend for comparison (i.e., the yearly average for all municipalities).

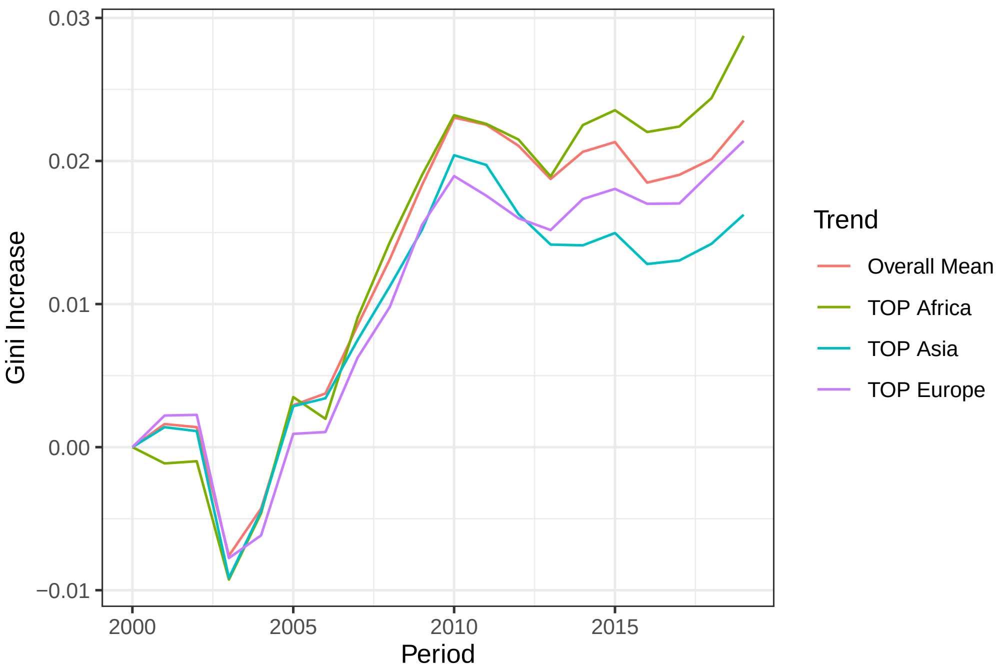

Figure 10: Inequality development for PFB outliers - Heterogeneity analysis.

Interestingly figure 10 show that municipalities with the highest increase in Africa-born proportion have had the highest increase in the Gini-estimate out of all groups. Since the European-born group is estimated to be a negative predictor, it makes sense that this sub-group has lower inequality than the overall trend. The more surprising result is that the Asian trend shows the least increase of all groups. The Asian-born proportions were insignificant in the regression, and their innate heterogeneity is likely the reason for such unclear results.

8 Conclusions

For the period 2000-2019, I conclude that there is a robust positive relationship between the proportion of foreign-born and inequality. These results are robust to various specifications. When conducting a heterogeneity analysis (grouping foreign-born by continents), I find that European-born immigrants are a negative predictor of the Gini-, and Theil index. The Asian-born proportion is insignificantly predicting Gini and the Theil-index. The African-born sub-group is the only group that significantly predicts all inequality indexes except for $L13$.

Further, descriptive evidence regarding top outliers shows multiple interesting things. Municipalities with the highest PFB increase have had notably high Gini-increase, especially if they were subject to low initial stock. The lowest Gini increase is observed for municipalities with low PFB increases with low initial stock. On the continent level, descriptive evidence shows that municipalities with the highest increase in African-born proportion have had the highest increase in inequality.

9 Discussion

In this thesis, I have shown that foreign-born proportion is a robust determinant of a range of inequality indexes. These indexes are specifically chosen for the type of immigration that Sweden has had over the past 20 years. The effect does seem to have a diminishing character - perhaps a sign of segregation; as Aldén et al. (2015) discussed, such speculation should be further studied. The design of the study can only examine correlations, and I do not claim causally interpretable parameters. Nevertheless, the proposed mechanism from economic theory has been shown to predict inequality outcomes accurately.

The immigration debate is essential. The latest NOVUS survey regarding Swedish political issues showed that 52% of the Swedes think immigration is "one of the most important issues," and 15% considered immigration "to be the most important issue." My results indicate that immigration groups with lower estimated human capital are an essential driver of inequality. Suppose Sweden chooses to take care of groups with similar characteristics in the future. For instance, if Europe sees a climate-related refugee wave. A policy-maker should then think about appropriate policies to handle such groups in the best way possible. Perhaps constructing a program aimed at increasing human capital at the point of arrival.

9.1 Caveats & Further studies

While meaningful conclusions can be made, it should be noted that this project suffers from shortcomings - estimating inequality with grouped income data come with some flaws. Estimating the magnitude of this problem, my experiment in section 3.2.1 signals a measurement error of about 15% on average for the Gini coefficient. However, that rests on the assumption of specific a parametric distribution; hence the actual error may deviate slightly. Thus, having access to continuous data would update the accuracy of the municipality inequality.

Further, human capital is an abstract concept, and there exists a multitude of different definitions. I use indicators that are a mixed analysis of macro data (HCI-index) combined with actual schooling observed of Swedish immigrants to build my perception of the general human capital background. Nevertheless, undoubtedly, there are many unobserved factors on individual level that can determine one's income. This dimension is not discussed since I do not have individual data, but could certainly be interesting to look into more. I group by continent to examine heterogeneous human capital levels among the foreign-born, but there are certainly variations *within* continents themsleves that could be interesting to study. For example, Asia is a remarkable heterogeneous continent, and more things could have been learned about this group with more detailed data.

10 References

1. Aart Kraay, The World Bank Human Capital Index: A Guide, The World Bank Research Observer, Volume 34, Issue 1, February 2019, Pages 1–33, https://doi-org.ezproxy.ub.gu.se/10.1093/wbro/lkz001.

2. Advani, Arun and Koenig, Felix and Pessina, Lorenzo and Summers, Andrew, Importing Inequality: Immigration and the Top 1 Percent (2020). CESifo Working Paper No. 8665, Available at SSRN: https://ssrn.com/abstract=3727739.

3. Aldén, Lina, Mats Hammarstedt, and Emma Neuman. "Ethnic Segregation, Tipping Behavior, and Native Residential Mobility." International Migration Review 49.1 (2015): 36-69. Web.

4. Almqivst, Alma. 2016. "Den ekonomiska ojämnlikheten i Sverige." Enheten för ekonomisk politik och arbetsmarknad. https://www.lo.se/home/lo/res.nsf/ vRes/lo_fakta _1366027478784_den _ekonomiska_ojamlikheten_i_sverige_pdf /File/Den_ekonomiska _ojamlikheten_i_Sverige.pdf.

5. Basilio, Leilanie, Thomas K. Bauer, and Anica Kramer. "Transferability of Human Capital and Immigrant Assimilation: An Analysis for Germany." LABOUR 31.3 (2017): 245-64. Web.

6. Bebonchu, Items. "Identifying the Dynamic Effects of Income Inequality on Crime." Oxford Bulletin of Economics and Statistics 82.4 (2020): 751-82. Web.

7. Becker, G.S. (1964) Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education. 3rd Edition, The University of Chicago Press, Chicago.

8. Borjas, George J. 1994. "The Economics of Immigration." Journal of Economic Literature 32:1667–1717.

9. Carlsson, Magnus, and Dan-Olof Rooth. "Evidence of Ethnic Discrimination in the Swedish Labor Market Using Experimental Data." Labour Economics 14.4 (2007): 716-29. Web.

10. Christl, Michael, Monika Köppl-Turyna, and Phillipp Gnan. "Wage Differences between Immigrants and Natives: The Role of Literacy Skills." Journal of Ethnic and Migration Studies 46.19 (2020): 4010-042. Web.

11. Cowell, Frank (2016) Inequality and poverty measures. In: Adler, Matthew D. and Fleurbaey, Marc, (eds.) The Oxford Handbook of Well-Being and Public Policy. Oxford University Press, Oxford, UK. ISBN 9780199325818 (In Press)

12. D'Hombres, Béatrice & Elia, Leandro & Weber, Anke. (2013). Multivariate analysis of the effect of income inequality on health, social capital, and happiness. 10.2788/68427.

13. Edström, Josefin., and Pokarzhevskaya Galina. "Lönegap mellan akademiker med svensk och utländsk bakgrund. " saco. 2017. https://www.saco.se/globalassets/saco/dokument/rapporter/2017-lonegap-mellan-akademiker-med-svensk-och-utlandsk-bakgrund.pdf

14. Eskelinen T. (2011) Relative Poverty. In: Chatterjee D.K. (eds) Encyclopedia of Global Justice. Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-9160-5_182.

15. Friedberg, Rachel M. "You Can't Take It with You? Immigrant Assimilation and the Portability of Human Capital." Journal of Labor Economics 18.2 (2000): 221-51. Web.

16. Gustavsson, Magnus, and Henrik Jordahl. "Inequality and Trust in Sweden: Some Inequalities Are More Harmful than Others." Journal of Public Economics 92.1-2 (2008): 348-65. Web.

17. Holbrow, Hilary J. "Detangling Capital from Context: A Critical Investigation of Human Capital Explanations for Immigrant Wage Inequality." Journal of Ethnic and Migration Studies 46.19 (2020): 4043-065. Web.

18. Iftikhar, Zainab, and Anna Zaharieva. "General Equilibrium Effects of Immigration in Germany: Search and Matching Approach." Review of Economic Dynamics 31 (2019): 245-276. Web.

19. Jorda, Vanesa, José Sarabia, and Markus Jäntti. "Estimation of Income Inequality from Grouped Data." ArXiv.org (2018): ArXiv.org, Aug 29, 2018. Web.

20. Jorgenson, D. W., and B. M. Fraumeni. 1989. "The Accumulation of Human and Nonhuman Capital, 1948–1984." In The Measurement of Savings, Investment and Wealth Studies in Income and Wealth, Volume 52, edited by R. Lipsey and H. Tice, 273-331. Chicago: University of Chicago Press.

21. Lundborg, Per. "Refugees' Employment Integration in Sweden: Cultural Distance and Labor Market Performance." Review of International Economics 21.2 (2013): 219-32. Web.

22. Kaarsen, N. 2014. "Cross-Country Differences in the Quality of Schooling." Journal of Development Economics 107: 215–24

23. Lemieux T. (2006) The "Mincer Equation" Thirty Years After Schooling, Experience, and Earnings. In: Grossbard S. (eds) Jacob Mincer A Pioneer of Modern Labor Economics. Springer, Boston, MA. https://doi.org/10.1007/0-387-29175-X_11

24. Mincer, Jacob. "Human Capital and Economic Growth." Economics of Education Review 3.3 (1984): 195-205. Web.

25. Novus. 2021. "Rapport: Viktigaste frågan" https://novus.se/novus-svensk-valjaropinion /valjarforstaelse/viktigaste-politiska-fragan/.

26. OECD. 2019. "OECD Economic Surveys: Sweden 2019". https://www.oecd-ilibrary.org/economics/OECD-economic-surveys-Sweden-2019_c510039b-en.

27. Pickett, Kate E, and Richard G Wilkinson. "Income Inequality and Health: A Causal Review." Social Science & Medicine (1982) 128 (2015): 316-26. Web.

28. Sanromá, Esteban, Raúl Ramos, and Hipólito Simón. "HOW RELEVANT IS THE ORIGIN OF HUMAN CAPITAL FOR IMMIGRANT WAGES? EVIDENCE FROM SPAIN." Journal of Applied Economics 18.1 (2015): 149-72. Web.

29. Schoellman, Todd. "Education Quality and Development Accounting." The Review of Economic Studies 79.1 (2012): 388-417. Web.

30. Warsono Waryoto, Usman Mustofa, Wamiliana, Wamiliana. (2018). ESTIMATION OF PARAMETERS OF GENERALIZED BETA OF THE SECOND KIND (GB2) DISTRIBUTION BY MAXIMUM LIKELIHOOD ESTIMATION (MLE) AND NEWTON-RAPHSON ITERATION. Far East Journal of Mathematical Sciences (FJMS). 105. 1-20. 10.17654/MS105010001.

31. World Bank. *Climate Change Could Force Over 140 Million to Migrate Within Countries by 2050: World Bank Report*. Collected 2020-11-30. https://www.worldbank.org/en/news/press-release/2018/03/19 climate-change-could-force-over-140-million-to-migrate-within-countries-by-2050-world-bank-report

11 Appendix

11.1 Simulation

11.1.1 Simulation size

In the simulation experiment, I essentially simulate random variables (Gini-estimates) from the pdf given by equation 2, using fitted parameters from a municipality i, at year t. One question worth talking about is the variation that stems from simulation. I use a simulation size of $N = 100000$ in the thesis, meaning I simulate a population with 100000 incomes from which I fit a Gini-estimate. The question is whether such sample size yields trustworthy Gini-estimates, preferable the estimates would be similar if one would have rerun the experiment. To test this, I input the fitted parameters for Gothenburg at year 2000, and run the simulation 100 times with different sample sizes. This gives a distribution of the fitted Gini-parameter for Gothenburg.

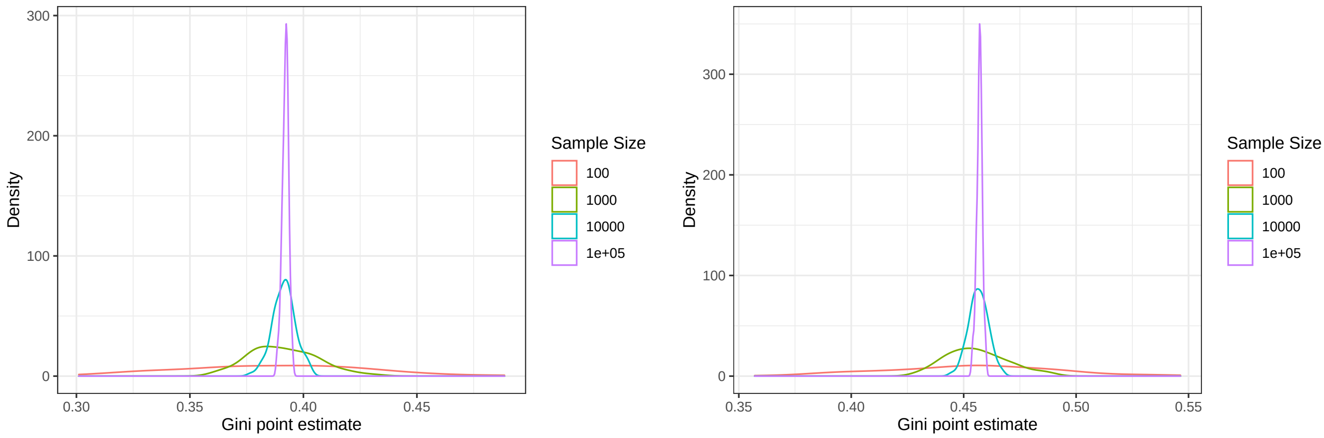

Figure 11: Left: Simulated Gini distribution & Right: Grouped simulated Gini distribution.

Figure 11 shows that the distribution narrows quickly, and at $N = 100000$, there's very little variation. This results indicater that Gini estimates derived under a sample size of $N = 100000$ are reliable. To examine whether different fitted parameters change the outcome, I try a few other examples but find similar results. These results are not included here to avoid clutter.

11.2 Missing data - HCI

11.2.1 Parameter training

Assessing heterogeneity from the HCI-index require datapoints for all countries - however 3 countries lack data: Somalia, Eritrea and Syria. I do find that the HCI index is highly correlated, $\rho = 0.9$, with life expectancy, as shown here:

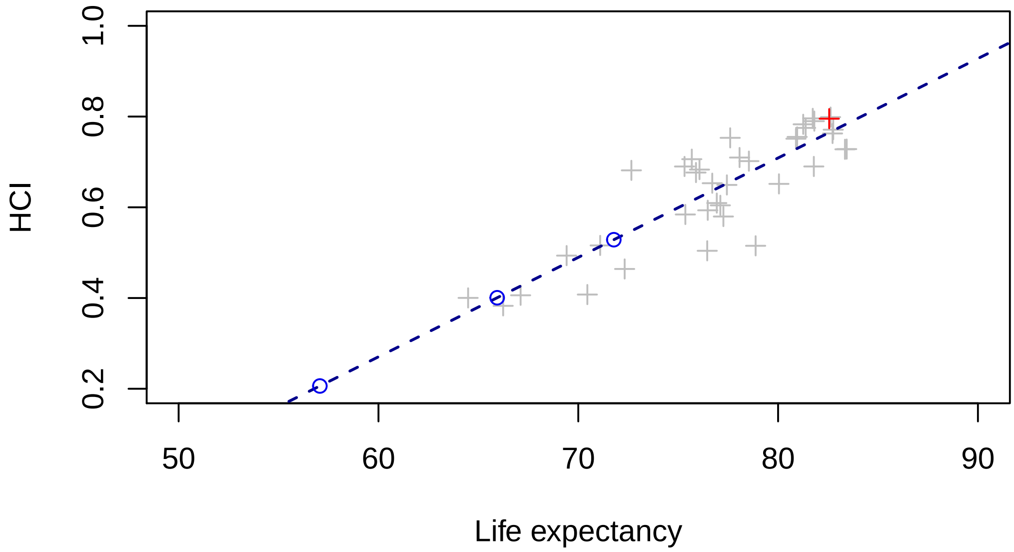

Figure 12: HCI vs Life expectancy | Predicted countries in blue.

Thus as an estimate, I fit a simple linear model: $HCI = \hat{\beta}_0 + \hat{\beta}_1 Lifeexp$, where the parameters are found on the subset of countries that do have data points ($N = 166$), and then linearly extrapolate the HCI-score for the countries with missing data - luckily there is data on their life expectancy. Data on life expectancy also originates from the World Bank, same as HCI-scores.