SCHOOL OF
ECONOMICS AND
MANAGEMENT

LUND UNIVERSITY

A Tale of Two Sets of Opportunities

An exploration of Inequality of Opportunity among natives and second-generation immigrants in Sweden

By

Irem Ucan

Master's Programme in Economic Development and Growth

Department of Economic History
Course code: EKHS42
Master's thesis (15 ECTS)
May 2023
Supervisor: Andreas Bergh
Examiner: Martin Dribe

Abstract

This paper examines the extent of income inequality that is attributed inherited circumstances such as social and family background, denoted as Inequality of Opportunity or IOp, for natives and second-generation immigrants in Sweden. The measurement of Inequality of Opportunity captures the between-type inequality, when individuals are grouped into types based on these predetermined circumstances, which involves parametrically estimating the average outcome of each type and comparing these averages between types. Using data from the Generations and Gender survey for Sweden, IOp is measured separately for natives and second-generation immigrants, as well for the whole sample. Furthermore, the measure of IOp is decomposed using Shapley decomposition to deduce the relative contributions of the inherited circumstances to the IOp measure. The results suggest that 16-17% of income inequality is due to inherited circumstances for natives, 14- 25% for second-generation immigrants and 15-16% for the whole sample. Gender and parental background such as education and occupation appear to be largest contributors to IOp both when the analysis is performed separately for natives and second-generation immigrants and jointly for the whole sample.

Keywords: Inequality of Opportunity, normative economics, Shapley decomposition, second-generation immigrants, income inequality

Acknowledgements

I would like to express my gratitude to my supervisor Andreas Bergh for the insightful discussions, his guidance and knowledge on the chosen topic of my thesis. Not only did he encourage me to realize my passion into a research paper, but also made the process truly rewarding.

I also would like to thank Andrés Palacio, Kerstin Enflo and Jutta Bolt for taking the time to discuss my thesis and highlighting important aspects to my chosen topic. Lastly, I want to express my appreciation to the Generations and Gender Programme for granting me access to the Generations and Gender survey for Sweden.

Table of Contents

List of Tables

List of Figures

1. Introduction

1.1 Background

With the increased migration to OECD countries over the past decades, ensuring successful migrant integration has become a key policy objective for many countries (OECD, 2016). This is especially the case for Sweden, which has a long history of providing a home for migrants (ibid.). However, policies to promote economic and social integration not only affects migrants but their children as well since one objective that migrants might have is to provide better opportunities for their children (Tasiran & Tezic, 2007). While the disadvantages faced by migrants in the labor market are well-documented, previous research is indicating a persistence of these intergenerational disadvantages (e.g. see Vilhelmsson, 2000; Rooth & Ekberg, 2003; Hammarstedt & Palme, 2012). For example, according to Vilhemsson (2000), second-generation immigrants have a lower chance of employment and a higher probability to be unemployed compared to their native counterparts.

However, should all forms of inequality be considered equally objectionable? On this question, Almås et al. (2010) find experimental evidence that individuals' view on fairness depends on the source of such inequality where individuals depict inequality aversion if these inequalities arise due to bad luck but not towards inequality that reflects individual choice. This would suggest that individuals distinguish between inequality due to inherited circumstances beyond individual control and inequality that reflect choice or effort. In this setting, does the observed income inequality between natives and second-generation immigrants reflect individual choice or simply bad luck? This question can be explored through the Inequality of Opportunity framework, which is concerned with inequalities that the individual has no control over. In broad strokes, outcomes such as income are partly influenced by individual effort and partly by inherited circumstances, such as social and family background, that are beyond the individual's control.

The advantage of this framework is that it can be used to estimate the share of outcome inequality that is due to inherited circumstances to explore whether individuals face equal opportunities in a country. Sweden is generally considered an egalitarian country that has low levels of inequality of opportunity (Ferreira & Gignoux, 2011; Björklund et al., 2012), however, it is important to examine whether these opportunities are enjoyed across the whole population including second-generation immigrants. This is especially relevant since, as of 2021, 14% of the Swedish population would be considered second-generation immigrants (Statistics Sweden, 2023)[1].

Lastly, while Björklund et al. (2012) claim that equal opportunities is "an ethical goal with almost universal appeal" (Björklund et al., 2012, p. 676) it also affects economic growth. Empirical evidence suggests that inequality of opportunity is negatively correlated with economic growth and inequality due to different levels of effort is positively correlated with economic growth (Marrero & Rodriguez, 2013). Thus, the exploration of the native- second-generation immigrant divide could shed light on a potential source of inefficiency that could have implications in terms of economic growth.

1.2 Aim and scope of the study

The aim of this study is to analyze inequality of opportunity in long-run income for natives and second-generation immigrants in Sweden. By analyzing natives and second-generation immigrants separately, I examine whether there are differences in inequality of opportunity between the two groups and which circumstances are relevant contributors to inequality of opportunity. Furthermore, by pooling the sample, I also obtain the latest estimate of the share of income inequality in Sweden that can be attributed to inherited circumstances.

The scope is limited to the exploration using the Inequality of Opportunity framework across two generations only and other approaches to intergenerational economic mobility are not regarded. Furthermore, this study is limited to individuals born between 1962-1981 in Sweden. Natives are considered those individuals that have two parents born in Sweden and second-generation immigrants constitute having one or two foreign-born parents. Lastly, the analysis of the observed outcome is limited to income for the time period 1994-2019.

1.3 Research questions

- To what extent does income inequality due to inherited circumstances differ for natives and second-generation immigrants in Sweden?
- To what extent do the effects of circumstances differ in income inequality due to inherited circumstances for natives and second-generation immigrants in Sweden?

[1] Second-generation immigrants are considered native-born individuals with one or two foreign-born parents.

1.4 Contribution of the study

While most papers on the native-second-generation immigrant divide study economic outcome inequality, that is, differences in outcomes such as income and education, this paper is the first to examine the extent inherited circumstances such as social and family background contribute to income inequality by performing separate analyses for natives and second-generation immigrants in Sweden. To my knowledge, no empirical study has estimated Inequality of Opportunity separately for natives and second-generation immigrants, in general or for Sweden, to explore whether these two groups face equal opportunities in the same country. The contribution of this study would not only be the improvement of the analysis on second-generation immigrants' economic integration but also to the literature on Inequality of Opportunity in the following aspects: data, methodology and analysis.

First, this study makes use of a longitudinal dataset that includes variables capturing both the early childhood and later adult experiences of individuals. Most empirical studies that attempt to measure Inequality of Opportunity are limited to a few circumstances due to data unavailability and thus this study would improve the understanding and relevance of early childhood experiences such as the effect of family structure on economic outcomes at adult age. Furthermore, the availability of panel data on income over 25 years (1994-2019) improves the accuracy of the analysis of the effect of inherited circumstances on the long-run income of individuals.

Second, the empirical specification employed in this study contributes to the literature in several ways. Most empirical studies that measure Inequality of Opportunity, both for a specific country and for cross-country comparisons, limit their analysis to father-son pairs i.e. only consider the characteristics of the male parent and observe the outcomes of male offsprings. This study includes both parents and both genders of the offsprings to not only increase the representativeness of the estimates for the whole population but also examine the robustness of the results obtained by other studies that only use father-son pairs, more specifically for Sweden. Furthermore, by observing the outcomes of individuals born between 1962-1981 in Sweden during the time period 1994-2019, the robustness of the results obtained by previous studies can be examined or detect a potential relative change in the national estimates.

Another important contribution to the literature in terms of empirical specification is the consideration of potential heterogeneity in the Inequality of Opportunity estimates, that is, whether individuals, belonging to two different groups in the population, with equal inherited circumstances face equal opportunities to achieve a certain outcome. The number of empirical studies that perform such type of analysis is scarce in the literature, yet it could be relevant to create a stronger bridge between the theory and measurement of Inequality of Opportunity. Furthermore, by examining potential differences in the effects of inherited circumstances on Inequality of Opportunity between natives and second-generation immigrants, the potential heterogeneity of intergenerational persistence can be identified which would increase our understanding of group-specific influences.

Lastly, this study performs comprehensive robustness checks using a recently proposed machine learning approach to the measurement of Inequality of Opportunity to examine the sensitivity of the results to alternative specifications which sheds important light on methodological specifications of the empirical studies that measure Inequality of Opportunity. Most articles in the literature don't justify their selection of models or variables in relation to their chosen empirical context and this study highlights the importance of such sensitivity analysis which could improve the cohesiveness of the empirical analysis in the literature.

1.6 Historical immigration in Sweden 1910-1960

The immigration patterns in Sweden during the 1910s and 1920s were mainly characterized by restrictive immigration policy along with emigration, especially to the U.S., surpassing immigration. It wasn't until the 1930s when Sweden had net immigration once the emigration to the U.S. had leveled off. Immigration increased during the Second World War, which consisted primarily of refugees from the Nordic countries and the Baltic States. This shifted post-war when labor force immigration increased in relation to refugee immigrants. While there was a surge of refugee immigration from Poland and the Baltic States in the late 1940s, from the 1950s until the early 1970s, the majority of immigration to Sweden was labor force (Hammarstedt & Palme, 2012).

The labor force migrants during the 1950s originated mainly from Finland, Western European countries such as Belgium and the Netherlands, and Southern European countries such as Italy and Greece. The migrants from Western Europe tended to have higher education levels compared to migrants from Finland and Southern Europe. Later, in the 1960s, an increasing number of migrants from former Yugoslavia began to migrate to Sweden and by mid-1960s, the largest sending countries of labor force migrants were Finland, former Yugoslavia, and Greece. Refugee migration remained low at the time and so was the migration from non-European countries (Hammarstedt & Palme, 2012).

1.7 Structure of the paper

In the following section, the theory and measurement of Inequality of Opportunity will be presented followed by the empirical specification employed in this study. After the methodology, the data source, the variable selection and the sample specification process will be described. The main empirical results and the robustness checks are reported in section 5, and analyzed in the following section. Lastly, the paper will be concluded with a general discussion.

2. Literature review

A detailed review of the theory and measurement of Inequality of Opportunity is presented in a systematic manner along with some empirical results from the literature. A central theme to the literature of Inequality of Opportunity is that there are many, often conflicting, conceptual and methodological approaches and formulations to what Inequality of Opportunity entails. Thus, the implication is the difficulty to formulate a unified framework.

2.1 The theoretical framework of Inequality of Opportunity

2.1.1 The origins of the theoretical framework

In broad strokes, the theory of Inequality of Opportunity[2], henceforth denoted as IOp, concerns with morally acceptable and unacceptable inequality in any given society (Roemer & Trannoy, 2016). The discussion of what constitutes as morally acceptable inequality originated from the political philosophical debate on the extent that the measure of inequality based on outcome was to be considered as morally acceptable since this measure did not hold individuals responsible for their choices per se, as proposed by welfare egalitarianism (ibid.). Hence, by incorporating the role of personal responsibility into egalitarian theory, beginning with the work of John Rawls (1958, 1971), the philosophical discussion shifted from equality of outcomes to equality of opportunities to be noted as 'luck egalitarianism' (ibid.). Since Rawls's first attempt to bring forth this important development in egalitarian theory, other contributions have been

[2] Or sometimes referred as Equality of Opportunity, EOp

added to the philosophical literature, most notably by Dworkin (1981a, 1981b), Arneson (1989) and Cohen (1989)[3]. From a welfarist perspective, luck egalitarians argue that a society should equalize the opportunities to achieve a certain outcome rather than equalize the outcomes.

On the theoretical conceptualization of IOp, the welfarist luck egalitarians attempt to provide a framework for a) what should be equalized and b) what individuals should be held responsible for (Roemer, 1993). There are two distinct views: responsibility for preferences (or the preference approach) as endorsed by Rawls (1971) and Dworkin (1981a, 1981b) and responsibility for control (or the control approach) as endorsed by Arneson (1989) and Cohen (1989).

Responsibility for preferences

Rawls's framework is based on his idea of the 'veil of ignorance', in A Theory of Justice, a thought experiment where individuals have no knowledge of their physical, social, and biological endowments. Rawls proceeds to consider these personal endowments to be a matter of luck and thus the distribution of these endowments to be 'morally arbitrary' (Rawls, 1971). Instead, justice is brought forth by maximizing the primary goods for those who are worse off in a society i.e. a maximin of a bundle of primary goods (ibid.). With Rawls's Difference Principle, inequality is permitted as long as transfers to those who are worse-off occur[4]. In an IOp context, primary goods are the inputs to obtain a certain outcome, or a life plan, and a society ought to equalize these primary goods. In such environment, the personal responsibility is then the extent the individual reaches this desired outcome or life plan (Roemer & Trannoy, 2016).

Dworkin (1981a, 1981b), on the other hand, emphasizes equality of resources rather than primary goods or welfare because Rawls's framework does not hold the individual responsible for their preferences. According to Dworkin (1981a), society should not be held responsible for the distribution of additional resources if the individual has expensive preferences or taste, especially if the individual does not 'identify' with this preference[5], in contrast to Rawls that

[3] Another important contributor is Sen (1980) with his idea of capability to functioning

[4] It should be noted that this is not the Pigou-Dalton Principle even if it resembles. Dworkin (1981b) highlights the issue with Rawls's Difference Principle because the better-off individuals might be worse off in certain circumstances at the expense of making the worse-off individuals better off based on the maximin condition.

[5] The example Dworkin (1981b) provides in which the individual might have a strong preference for sex which could inhibit the fruition of some other life plan. He denotes these preferences as handicaps.

don't make such distinction in individual choice. Furthermore, Dworkin (1981a) also considers an individual's mental capacity to formulate and realize such individual choice which could have implications on what personal responsibility constitutes. To conceptualize his framework, Dworkin (1981b) uses a hypothetical insurance market example, whereby individuals know their preferences but have no knowledge of their resources. In the framework, resources include both material and biological assets one holds e.g. from birth, which is beyond the individual's control. In such environment, the individual is then responsible for its preferences in terms of its willingness to take risks.

While Dworkin (1981b) maintains Rawls's initial argument of moral arbitrariness of the distribution of initial personal endowments, Dworkin (1981b) argues that the distribution of resources should be ambition-based and equalized over the aspects for which the individual has no control over. Lastly, on the matter of luck, instead of considering only the initial endowments as luck, as done by Rawls (1971), Dworkin (1981b) distinguishes between brute luck and option luck. Brute luck refers to situations that the individual could not predict or anticipate for, such as an economic crisis and should be considered beyond the individual's responsibility, and option luck, that is the when the individual takes a risk in the process of choice or preference i.e. risk assessment, however, it would not affect the equalization of resources per se since option luck prima facie belongs to his personal responsibility.

Dworkin's (1981a, 1981b) conceptualization of IOp sparked controversy in the political philosophy community and was criticized most notably by Arneson (1989). Arneson (1989) remarks two issues with Dworkin's (1981b) equality of resources, one to do with distinction of resources in relation to the individual and the second on personal responsibility. On the former, Arneson (1989) argues that Dworkin (1981b) made the cut between resources and responsibility at the wrong place. To clarify his argument on the former issue, Arneson (1989) provides the example, denoted as the 'slavery of the talented' problem, of how Dworkin's (1981b) framework would create a less than ideal scenario of equality in the following way: two individuals with identical preference for personal liberty (ownership over their body over their lifetime) are born with different levels of talent: low and high. In a society with a high demand for the high-talent individual, liberty for that individual is very expensive compared to the low-talented individual, hence the highly talented individual is less likely to achieve its life plan i.e. to pursue liberty. Thus, talent would be a resource that influences the extent an individual is able to fulfill its preference or life plan and it is unclear how the distribution of resources would be in order for both individuals to have equal opportunities to fulfill their preferences (ibid.).

On the latter issue of personal responsibility, Arneson (1989) argues that Dworkin's (1981b) claim that individuals are responsible for their preferences is ambiguous. Arneson (1989) begins with the premise that social and biological factors i.e. resources could influence individual preferences, and if the individual is only responsible for what is within their control, then this would imply that individuals are only partially responsible for their preferences because the formation of such preference was made partially through channels beyond what is in its control. Even if the preference formation occurred through channels beyond the individual's control, what the individuals should be held responsible for is the actions taken to realize such preferences (ibid.). Furthermore, if resources such as social and biological factors influence preferences, equality of resources would satisfy individual preferences to a lesser extent (ibid.).

Responsibility for control

Instead, Arneson (1989) proposes equality of opportunity for welfare. First, opportunity is defined as " a chance of getting a good if one seeks it" (Arneson, 1989, p. 85). Second, welfare is defined as preference satisfaction[6], whereby the greater the importance of the preference, the greater the welfare (ibid.). Hence in this framework, equality of opportunity for welfare is equal chance to satisfaction of a preference which that preference might offer compared to other options or preferences (ibid.). Arneson's (1989) framework is described in the following way: for each individual, a decision tree is constructed which contains all possible life paths the individual could take. Each life path is associated with a given preference satisfaction expectation, which also takes into account the possible options that an individual might encounter for each decision point. If every individual is faced with equivalent decision trees, then the expected value for each individual's ranking of options will be the same i.e. best, second-best set of options etc. In this scenario, opportunities are then ranked based on the welfare they can afford (ibid.).

However, equality of opportunity for welfare is only reached when individuals are faced with *effectively* equivalent decision trees[7] since individuals might differ in their awareness of their

[6] Arneson (1989) assumes that individuals are selfish i.e. have self-interested preferences

[7] The following conditions constitute effectively equivalent decision trees: a) options are equivalent and individuals have similar ability to negotiate for these options, b) options are not equivalent but this counteracts

options, their ability to choose rationally and to carry out their chosen option (Arneson, 1989). On personal responsibility, individuals are responsible for their choices to maintain such effectively equivalent sets of options. To clarify, two individuals might have the same opportunities for welfare at a given moment, and if one of the individuals chooses to deviate their behavior in a certain way[8] such that the other person now has a greater opportunity for welfare, then any inequality in their opportunities at a later time is deemed under the personal responsibility of the individual who deviated (Arneson, 1989). Hence, if individuals have equal opportunity for welfare i.e. effectively equivalent sets of options, any inequality of welfare is due to what is within the control of the individual (ibid.).

While Arneson's (1989) conceptualization eliminates what Cohen (1989) describes as "involuntary welfare deficiencies" (Cohen, 1989, p. 916), this is only one type of disadvantage. The objective of egalitarianism, according to Cohen (1989) is to eliminate involuntary disadvantage, that is "disadvantage for which the sufferer cannot be held responsible, since it does not appropriately reflect choices that he has made or is making or would make" (Cohen, 1989, p. 916). Thus, Arneson's (1989) formulation does not eliminate all forms of disadvantage since disadvantage encompasses more than welfare deficiencies (Cohen, 1989). Cohen (1989) provides an analog of physical disability to prove his point. Under Arneson (1989)'s formulation, the individual would be compensated with say a wheelchair to equalize opportunity to welfare to construct such effectively equivalent decision trees without knowing that individual's disposition to welfare. Thus, if an individual has naturally a disposition to welfare e.g. to be happy, physical disability does not necessarily have to be a hindrance to the opportunity for happiness, and such Arneson's (1989) framework falls short of what is actually equalized (ibid.).

Instead, Cohen (1989) proposes equal opportunity for advantage or equal access to advantage since equal opportunity for welfare does not necessarily mean equal access to welfare. While Cohen (1989) does not specify what advantage entails, 'access' in this context means both the opportunity and the capacity to obtain welfare. The opportunity-aspect follows the same reasoning as Arneson (1989) however the capacity-aspect is an important distinction compared

---

differences in negotiating abilities and c) options are equivalent and any difference in negotiating ability is due to factors within the individual's control.

[8] Which Arneson (1989) describes as "voluntarily chooses or negligently behaves" (p. 86) to refer to any potential deviation in behavior

to Arneson's (1989) framework because if the individual deviates from the effectively equivalent decision tress due to factors beyond their control i.e. involuntary disadvantages, then the individual should not be held responsible for the consequent inequality of welfare (ibid.).

To summarize the early formulations of the IOp theory, a central theme has been to distinguish what personal responsibility implies, whether these are preferences, choices or ambitions, and what is beyond the individual's control, whether this is primary goods, resources including both not limited to biological and social factors, or involuntary disadvantages. As seen, there are different interpretations of what personal responsibility means, which makes the identification of fair and unfair sources of inequality difficult and subject to constant revision (Brunori & Peragine, 2011). Furthermore, to complicate the matter even further, as noted by Dworkin (1981b), the distinction between preferences and resources is not clear-cut such that 'genuine' choice can be identified fully, which elicits the free will problem. While the discussion of free will is important in a distributive justice context, in an economic research setting, it is sufficient to view the degree of personal responsibility to be set by a given society, and thus avoid the discussion of free will, according to Roemer and Trannoy (2016). Nevertheless, the early formulations of IOp theory set the stage for economic research with Roemer (1993, 1998) as the frontrunner.

2.1.2 Roemer's definition and economic model of Inequality of Opportunity

The economic research on IOp began most notably with Roemer's (1993, 1998) definition and development of an economic model of IOp. On the definition of IOp, Roemer (1993) attempts to incorporate elements of Dworkin (1981a, 1981b), Arneson (1989) and Cohen's (1989) formulations to describe IOp in the following way: "equality of opportunity for X holds when the values of X for all those who exercised a comparable degree of responsibility are equal, regardless of their circumstances" (Roemer, 1993, p. 149). In Roemer's (1993) definition, circumstances are considered socioeconomic and biological factors, and two individuals are considered to have the circumstances if they share the same set of these factors. Which factors influence individual choice and what should be considered beyond their control is set by the society (ibid.). On what "comparable degree of responsibility" means, Roemer (1993) makes the distinction that comparable does not necessarily mean the same degree but rather that if two individuals with different sets of circumstances exercise different degrees of responsibility when their circumstances are taken into account, it might be considered as comparable.

Roemer (1998) proceeds to formulate an economic model that would capture his definition of IOp by introducing another element to the conceptualization of the framework: policy. Insofar, as the first wave of the IOp theory was concerned with distinguishing what is meant by personal responsibility and less so about the formulation of a public policy to apply the IOp framework, which is the point where Roemer (1998) makes an important contribution: to translate the political philosophical discussion into a public policy intervention (Pignataro, 2012). Roemer (1998) proposes an economic algorithm to calculate an equal-opportunity policy for a given society, which could also in a greater context be used as a method to assess and rank social policies based on their efficiency to equalize opportunities. While the main purpose of Roemer (1998)'s model was to translate the normative framework into a public policy intervention to 'level the playing field' (Roemer, 1998, p. 5), the model also became a cornerstone in the general model for the measurement of IOp (Pignataro, 2012).

Assumptions

Roemer (1998) considers a society whose population is partitioned into a set of types such that $\mathbf{T} = \{1,\ 2, \ldots ,\ T\}$, and each type is composed of a group of individuals that share the same set of circumstances[9]. The society chooses the set of circumstances which defines the types. Let the frequency of type *t* in the population to be p^t. The level of achieved outcome[10] by individual in type *t* is formalized as $u^t\ (x,\ e)$, where x denotes the amount of resources the individual consumes and *e* the effort expended to obtain the (desired) outcome[11]. It is further assumed that the outcome function, u^t, is monotone increasing in level of effort, *e.* [12]

Opportunity-equalizing policy

First, a society chooses a policy, φ , to allocate resources to its population, $\varphi = (\varphi^1, \ldots, \varphi^T)$ according to some allocation rule such that $\varphi^t(e)$ is the amount of resources the individual from type *t* receives if it expends effort 'e'. If each individual in type *t* faces the same allocation rule, there will be a distribution of exuded efforts within each type, $F^t_{\varphi^t}$ [13]. From this effort

[9] Circumstances are treated as environmental factors such as socioeconomic and biological which is beyond the individual's control (Roemer, 1993; 1998).
[10] The most common outcomes measures in the literature are income, health status and educational attainment.
[11] Effort is defined as the autonomous action within the control of the individual (Roemer, 1998)
[12] Unlike conventional utility functions, the implication of assuming that u^t to be monotone increasing is that the theory of IOp is insensitive to whether the individual obtains utility from expending the effort to reach desired outcome or not
[13] Non-negative real numbers

distribution for each type, a benchmark for ranking can be formulated for each type *t* facing the same allocation rule, such that the individual that expends the median degree of effort is regarded, $e^t(\pi, \varphi^t)$ where π denotes the individual at the π^{th} quantile of the effort distribution of type *t*. The benchmark is used to detangle effort, *e*, from circumstances beyond individual's control, thus identifying the degree of effort rather than the level of effort for each type *t*. Based on the ranking for each type *t*, the indirect utility function is defined as:

$$v^t(\pi;\ \varphi^t) = u^t(\varphi^t(e^t(\pi, \varphi^t)),\ e^t(\pi, \varphi^t) \qquad (1)$$

Second, assume a society has a set of social policies to choose between such that $\varphi \in \phi$, and the objective is to select the policy that equalizes opportunities across all types given the effort distribution for each type. Thus, the opportunity-equalizing policy would be one that maximizes the minimum level of opportunities i.e. the least advantaged individuals, across all types, for individuals that expend the degree of effort π within their type, or:

$$\varphi^{EOp} = \max_{\varphi \in \phi} \int_0^1 \min_{t \in \boldsymbol{T}} v^t(\pi\ ;\ \varphi^t\ )\ d\pi \qquad (2)$$

where EOp denotes Equality of Opportunity.

Through this formulation, inequality due to circumstances is ethically unacceptable, however, differences in outcome due to effort is acceptable (Roemer, 1998). In theory, such opportunity-equalizing policy would eliminate the influence of circumstances on outcomes, and thus outcomes would simply be a function of effort (Roemer & Trannoy, 2016).

However, Roemer (1998)'s approach assumes that the society accounts for all relevant circumstances in the definition of a type, and thus any cause of inequality that is not due to the type that the individual belongs to is due to effort (Roemer & Trannoy, 2016). The implications are: a) Roemer (1998)'s approach requires extensive information on the environmental factors that might be influencing the outcome and b) the types are defined sufficiently enough to account for all the potential circumstances in the set (Roemer & Trannoy, 2016; Pignataro, 2012). Hence, it is likely that the opportunity-equalizing policy will only equalize some observed inequalities but not all (ibid.).

Consider the following example: we define types based on one circumstance: parental education, and we are interested in income as our outcome variable. Based on Roemer (1998)'s approach, we might conclude opportunities for income are equalized or not equalized in a given society simply because the income distribution functions between the types are equal or not equal. In reality, there are circumstances other than parental education that might be influencing the chances for obtaining a certain income level which wouldn't be accounted for in the definition of the types (Roemer & Trannoy, 2016).

2.1.3 Compensation principle versus reward principle

Based on the theoretical framework, two distinct ethical principles have risen with regard to policy formulation: the compensation principle and the reward principle (Brunori & Peragine, 2011)[14]. The compensation principle regards that inequality of outcome due to circumstances is ethically unacceptable and should be eliminated or compensated by society[15]. On the other hand, according to the reward principle, inequality of outcome due to effort should not be compensated by society (ibid.). The reward principle can be divided into liberal and utilitarian reward where the former advocates equal transfers to individuals with equal circumstances and the latter makes transfers from individuals with effort levels that yield a low marginal utility to individuals with effort levels that yield high marginal utility in order to maximize the sum total of individual outcomes in a given society (Fleurbaey, 2008; Ramos & Van de gaer, 2016; Roemer & Trannoy, 2016). In order words, according to the utilitarian reward principle, resources should go where they can be utilized the best.

Not only do the principles influence policy orderings for IOp, but they are also incompatible (Brunori & Peragine, 2011). The implication of this incompatibility is that any measure of IOp and hence policy ranking will not fully satisfy both principles, but rather either fully one and partially the other or vice versa (ibid.). It should be noted that one principle is not superior to the other, but rather they differ in which aspect of IOp to focus on: what is beyond personal responsibility (circumstances) or within personal responsibility (effort). However, what is important be aware of are the **conceptual** differences of the IOp measures (Ramos & Van de gaer, 2021). Furthermore, as it will explained below, there are different measurement

[14] While these principles were first introduced in the fair division literature, they have been reinterpreted to the theory of IOp after Roemer (1998)'s opportunity-equalizing policy
[15] Roemer's policy formulation is compatible with the compensation principle

approaches to IOp and these approaches should reflect the mentioned ethical considerations (Jenkins, 1991).

2.2 Measurement of Inequality of Opportunity

2.2.1 The general model

The measurement of Inequality of Opportunity might have different purposes and the most common operations in the literature are quantifying, ranking and decomposing IOp (Roemer & Trannoy, 2016). Essentially, what matters to the measurement IOp is the differences in the impact of circumstances rather than the differences in circumstances (ibid.). In its general form, the measurement of Inequality of Opportunity can be seen as a two-step procedure: in the first stage, the distribution of the outcome of interest is transformed into a counterfactual distribution that reflects either only inequality that is ethically unacceptable i.e. beyond individual's control or in which there is no inequality of opportunity, which is an econometric-estimation process, and in the second step, an inequality index such as the Gini coefficient, general entropy indices some other index is applied to the counterfactual distribution to obtain the estimate of IOp (Roemer & Trannoy, 2016; Palmisano & Peragine, 2022).

The canonical model can be presented in the following way (e.g. see Ooghe et al. 2007; Ferreira & Peragine, 2015; Ramos & Van de gaer, 2016). First, consider an outcome distribution, say income, y, for a given population. Further assume that the determinants of y can be explained by a set of circumstances from vector $C \epsilon \Omega$, and effort, $e \epsilon \Theta$[16] such that the reduced-form model becomes:

$$y = g(C, e) \ , \tag{3}$$

where g denotes some function of unknown form and $g: \Omega \times \Theta \rightarrow R$ and C is a vector. Thus, in this model, individuals with equal circumstances and effort would obtain the same income.

Next, assume that the population can be partitioned into types, T_i based on Roemer (1998)'s formulation and tranches, T^k, based on the same degree of effort such that the outcome, denoted as y_{ij}, is generated by circumstances, C_i where $i = 1 \ldots n$ so that there are n types, and efforts,

[16] The notations for $C \epsilon \Omega$ and $e \epsilon \Theta$ are used to indicate that they belong to different sets.

e_j where $j = 1 \dots m$ so that there are m tranches. In this setting, the population can be represented in an $n \times m$ – dimensional matrix $[Y_{i\,j}]$, illustrated below:

$$Y = [Y_{i\,j}] = \begin{bmatrix} y_{1\,1} & y_{1\,2} & \cdots & y_{1m} \\ y_{2\,1} & y_{2\,2} & \cdots & y_{2\,m} \\ \cdots & \cdots & \cdots & \cdots \\ y_{n\,1} & y_{n\,2} & \cdots & y_{n\,m} \end{bmatrix}$$

where each row represents income vector for each type, based on the set of circumstances C_i, and each column (tranche) represents income vector for effort level j (Ferreira & Peragine, 2015; Ramos & Van de gaer, 2016; Palmisano & Peragine, 2022). In relation to the measurement approach to IOp, $[Y_{i\,j}]$ is the outcome distribution of interest, which is then transformed to a counterfactual distribution, $[\tilde{Y}_{i\,j}]$. How this counterfactual is constructed differs in the literature and the process can be classified based on three criteria: whether the measure is based on ex-ante or ex-post, parametric or non-parametric, and direct or indirect measurement method.

2.2.2 The construction of a counterfactual distribution

Ex-ante versus ex-post

The first distinction is whether the counterfactual should reflect IOp between the types (rows) or between the tranches (columns) in the outcome distribution matrix. Based on the theoretical ethical principles of compensation and reward, two versions have been developed in the literature: ex-ante and ex-post (Ramos & Van de gaer, 2016; Palmisano & Peragine, 2022). Ex-ante measures inequality between individual's opportunity sets and assume these to be due to circumstances whereas the ex-post measures assume that inequality in outcome is due to differences in effort levels (Fleurbaey & Peragine, 2013). Hence, in ex-ante IOp, the counterfactual distribution should reflect inequalities between the rows and ex-post between the columns[17].

The measurement of IOp should reflect the ethical principles in some regard (Jenkins, 1991). With respect to the compensation and reward principles and their relationship to ex-ante and ex-post, the compensation principle can either be ex-ante and ex-post, while the reward

[17] Ex-ante implying before effort has been expended and ex-post after.

principle is only an ex-ante approach. With regard to the compensation principle, ex-ante compensation attempts to equalize between opportunity sets (rows) based on their type as much as possible so that individuals can have the same opportunity to earn a certain income, without regard to differences in their effort levels i.e. if type a has more favorable circumstances than type b, then type a should be compensated so that both types have the same opportunity to reach a certain outcome. Meanwhile, ex-post compensation attempts to equalize such that individuals that expend the same level of effort should have the same opportunity for a certain outcome.

While ex-ante compensation and the reward principles are compatible, ex-post and ex-ante compensations are incompatible, and ex-post and reward principle is incompatible. This implies that in IOp policy formulation and ordering, and hence the measure of IOp needs to maintain either ex-post compensation only or ex-ante compensation/ reward.

Parametric versus non-parametric

The second criterion relates to the form of the function g, whereby in the general model, the form of the function is unknown. In other words, this has to do with how the effect of circumstances on the outcome is captured (Roemer & Trannoy, 2016). A parametric approximation imposes a functional form e.g. linear to capture the effect of circumstances on outcome while the non-parametric approximation does not (ibid.). Since the estimation of the counterfactual is based on the function $g(C, e)$, how the circumstances are assumed to influence the outcome matters for the counterfactual outcome distribution.

The most common parametric counterfactual in the literature is the one that estimates all inequality that is due to circumstances, including any potential correlation between circumstances and efforts (Ramos & Van de gaer, 2020). Meanwhile, non-parametric approach can take a more flexible form and instead groups individuals based on circumstances. The non-parametric approach is equivalent to parametric models with all possible interaction effects between the circumstances (Ramos & Van de gaer, 2021). The most common method for a non-parametric counterfactual involves averaging based on types.

Method suitability, whether parametric or non-parametric, depends to a larger extent on the dataset (Roemer & Trannoy, 2016). Since the parametric methods attempt to estimate the conditional expectation $E(y \mid C, e)$ rather than the conditional distribution $F(y \mid C, e)$, it is less data intensive compared to the non-parametric approach (Roemer & Trannoy, 2016; Pervaiz & Akram, 2018). However, since the parametric approximation assumes a certain functional form, it is also subject to the underlying assumptions of the selected functional form (Pervaiz & Akram, 2018).

Direct versus indirect

The last criterion is regarding the reference point of the counterfactual. Direct measures estimate IOp based on a counterfactual outcome distribution that only reflects inequality that is ethically unacceptable i.e. beyond individual's control, whereas the indirect method measures IOp in relation to a counterfactual outcome distribution where there is no inequality of opportunity (Ramos & Van de gaer, 2020). In former scenario, any inequality due to differences in effort has been removed and what remains is inequality due to circumstances or IOp.

Based on these three distinctions or criteria, authors in the literature have used various combinations of these such as direct- ex-ante, ex-post – parametric, non-parametric etc.[18] However, the measurement specification could influence the IOp estimate obtained. Using Spearman rank correlation estimation, Ramos and Van de gaer (2021) calculate the average Spearmen distance between the various combinations and find evidence that the difference between the ex-ante and ex-post measures of IOp is the largest among the three criteria. The extent of the direct/indirect divide seems to be conditional on the previous criteria and lastly, the parametric vs. non-parametric approximations yield similar results (smallest difference in the estimates).

2.2.3 Selection of inequality index

The second step in the measurement of IOp is to select an inequality index to apply the counterfactual distribution to. The two main groups of inequality indices used in the literature are the Gini coefficient and one or more members of the generalized entropy family[19], most notably the Mean Log Deviation (MLD or GE(0)) (Palmisano & Peragine, 2022).

[18] For a detailed list see Ramos and Van de gaer (2021).

[19] Along with the Gini coefficient and MLD, Björklund et al. (2012) and Hederos et al. (2017) also include the other members of the general entropy family i.e. Theil index (GE(1)) and CV^2 (GE(2)).

The Mean Log Deviation is defined as:

$$MLD(X) = \frac{1}{N}\sum_{i=1}^{N} ln\ \frac{\mu_X}{x_i} \qquad (4)$$

Where X denotes a distribution of say income, and μ_X the mean.

The main advantage of MLD is that it is perfectly decomposable into between-group and within-group components, which in an IOp-context can be used to distinguish between inequality due to circumstances (between) and inequality due to effort (within). Groups are defined according to the types proposed by Roemer (1998) (Palmisano & Peragine, 2022). However, MLD is also more sensitive to extreme outliers, especially to the bottom of the distribution, compared to the Gini index, and if the counterfactual is based on a smoothed distribution such as direct – ex-ante – non-parametric or indirect – ex-post – non-parametric, then MLD could underestimate IOp such that lower-bound estimates are obtained (Björklund & Jäntti, 2020; Ramos & Van de gaer, 2021; Palmisano & Peragine, 2022).

The Gini coefficient, on the other hand, is not as sensitive to extreme outliers as MLD and hence would avoid the issue of lower-bound estimates in the above-mentioned circumstances. However, the Gini coefficient is not strictly decomposable in the same way as MLD. Instead, when Gini is decomposed into within and between- components and distributions of the groups or types overlap, a residual term is obtained.

$$Gini\ =\ Gini_w +\ Gini_B +\ K \qquad (5)$$

Where $Gini_w$ denotes the within-group-component of the inequality index, $Gini_B$ the between-group, and K the residual term. In an IOp-context, the residual is the part of inequality that is jointly determined by circumstances and effort such that the effects of the two can't be separated. If all relevant circumstances are not included in the estimation, then it is likely that the Gini coefficient yields lower-bound estimates (Palmisano & Peragine, 2022).

2.2.4 Selection of circumstances and efforts

In an IOp setting, there are three major elements: outcome, circumstances, and effort. However, there are several caveats to how the normative framework is adapted to the measurement of IOp. First, empirical attempts to estimate IOp require considerations on which variables constitute as circumstances and effort, which has normative implications in terms of what personal responsibility entails and what should be compensated or rewarded (Ramos & Van de gaer, 2016). To illustrate the difference between the control approach and the preference approach, consider the variables age and gender. According to the control approach, age and gender are circumstances i.e. beyond individual's control and any inequality due to these circumstances should be considered ethically unacceptable. On the other hand, according to the preference approach, age and gender are determinants of preferences and thus should be regarded as factors of effort rather than a source of IOp (Roemer & Trannoy, 2016). Thus, the selection of variables to represent circumstances and effort is not an easy one and depends on which normative approach one adheres to.

The second caveat is that, in principle, all factors that influence the outcome variable and are beyond personal responsibility should be included in the measurement to truly reflect IOp. For the control approach, this means that the measurement of IOp should capture both the direct effect of circumstances on the outcome variable and indirect effect of circumstances through their influence on effort. Whereas, in the preference approach, the measurement of IOp should capture the effect of circumstances i.e. variables that don't influence preferences but has an effect on the outcome variable (Ramos & Van de gear, 2016). The extent that the indirect effect of circumstances is included in the preference approach depends on how the circumstances affect effort. If the circumstance only influences preference, then this should not be included.

However, generally, in an empirical setting, not all circumstances are observed which could lead to underestimation of IOp. Furthermore, raw effort is often unobserved and observed effort is often correlated with circumstances (Ramos & Van de gear, 2016). In the control view, observed effort should be cleaned out of these correlations with circumstances so that what remains is raw effort, hence the inclusion of indirect effect of circumstances, and in the preference view, these correlations are less of an issue as long as individuals identify with their preferences.

2.3 Empirical evidence

The applications in the empirical literature are extensive with different measurement approaches and operations, outcomes and selection of circumstances and effort variables. However, the empirical articles considered here are based on relevance for the empirical context of this study in terms of population and type of analysis. In general, Sweden scores low levels of IOp in cross-country rankings, indicating that, in relative terms, individuals have fairly equal opportunities (Ferreira & Gignoux, 2011; Brunori et al., 2019; Ramos & Van de gaer, 2020). The most frequently cited paper for Sweden in the literature is the empirical study by Björklund et al. (2012) who estimate IOp for males born between 1955-1967 in Sweden. Björklund et al. (2012) employ a direct-ex-ante-parametric approach to estimate IOp using a rich dataset including socioeconomic circumstances such as parental income and education, family structure and number of siblings, but also IQ test scores at the age of 18 to capture the indirect effect of circumstances on income. The authors apply both the Gini index and all of the members of general entropy indices to obtain the IOp estimates for Sweden.

A recent wave in the empirical literature attempt to estimate IOp separately for different groups. For example, Hederos et al. (2017) examine whether males and females face equal opportunities in Sweden by following a similar methodology to Björklund et al. (2012) and obtain lower IOp estimates for females compared to men. However, when gender is treated as a circumstance, the authors deduce that it is the largest contributor to IOp when the outcome is income. Davillas and Jones (2020) perform a similar analysis to Hederos et al. (2017) for health outcomes instead of income. The authors perform separate analyses based on gender and age cohorts and deduce that the relative effects of the circumstances to IOp vary across gender and age cohorts (ibid.).

Since separate IOp analyses for natives and second-generation immigrants have not been performed before in the literature before, neither for Sweden nor in other countries, it limits the empirical evidence for cross-reference. The closest proxies for the given empirical setting of this study are the studies by Behtoui (2006) and Tasiran and Tezic (2007). Tasiran and Tezic (2007) study the early labor market experiences of second-generation immigrants by using dynamic transition rate models (multinomial logit models) and deduce that parental background affects second-generation immigrants' education and overall labor market success. The authors include circumstance such as gender, parental education, income and ethnic origin, and whether the individual had one or two foreign-born parents. On the other hand, Behtoui (2006) use logistic regressions to estimate the probability to earn an income of ≥ 100 000 SEK, after controlling for individual characteristics such as education, gender, living in a large city, and marital status, for natives and second-generation immigrants and deduce that second-generation immigrants have, on average, ca 20% disadvantage compared to natives.

3. Methodology

3.1 The estimation of Inequality of Opportunity

While empirical studies have employed different approaches to constructing the counterfactual distribution and in general to estimate IOp, the most common method is direct – ex-ante – parametric (Brunori et al. 2013)[20]. Following the common approach in the literature, this study employs a regression-based method using a log-linear specification as proposed by Ferreira and Gignoux (2011)[21].

Assumptions and model

Individuals are partitioned into types, *T*, such that there are *t* types. Further assume that there a *J* circumstances such that the reduced form regression equation becomes:

$$\ln y_i = \beta_0 + \sum_{j=1}^{J} \beta_j C_i^j + \varepsilon_i \tag{6}$$

where y_i is the income of individual *i*, C_i^j denote each circumstance of individual *i,* and ε_i the error term of the regression.

From this, a counterfactual distribution is obtained by replacing individual income y_i with the type-specific mean that the smoothed distribution becomes:

$$\mu^t(y) = \exp\left[\sum_{j=1}^{J} \hat{\beta}_j C_i^j\right] \tag{7}$$

where $\hat{\beta}_j$ are the OLS coefficients from Eq. 6. In simpler terms, $\mu^t(y)$ is obtained through the predicted values from estimating Eq. 6. The error term of this estimation is assumed to reflect effort.

Once the counterfactual distribution is estimated, the next step is to use an inequality measure. In this study, Mean Log Deviation (MLD) is employed. Since the aim of this study is to estimate the share of inequality that is due to circumstances, the IOp estimates using MLD are expressed in relative terms, that is:

$$\theta_{IOp\ R} = \frac{MLD(\hat{Y})}{MLD(Y)} \tag{8}$$

$\theta_{IOp\ R}$ denotes the relative IOp estimate, and $\hat{Y}$ and Y the counterfactual and the actual income distributions respectively.

[20] It is also the method used by Björklund et al. (2012) and Hederos et al. (2017) to estimate IOp for Sweden, which would allow for comparison of the estimates obtained in this study.

[21] A log-linear specification to the parametric approach is the most common approach in the literature when the outcome variable is income (Ramos & Van de gaer, 2016)

Even though the methodology of this study is to a large extent influenced by the common practice in the literature, method suitability was also considered in relation to the empirical context of this study. First, while ideally, the IOp measurement should include both observed circumstance and effort variables to obtain the upper-bound estimates, it is very difficult to observe effort in a dataset and thus subject to biased estimates. As noted by Roemer and Trannoy (2016), raw effort can't be observed but is rather proxied, which imposes several issues. The most common proxies in the literature are hours of work and years of education. If hours of work correspond to desired amount of hours, then this is assumed to be a suitable proxy for effort. However, in the case that it is lower than desired for some involuntary reasons such as having to take part-time jobs because of unavailability of full-time jobs or unemployment due to bad luck, then it is likely to not reflect true effort (ibid). Instead, it is assumed that circumstances directly and indirectly affect the outcome, directly as specified in the regression model and indirectly through the effect of circumstances on effort (the residual of the regression estimation), hence by employing an ex-ante approach, the issues related to observed effort could be avoided (Brunori, 2016).

Parametric approach also coincides with the empirical context of this study in relation to sample size and considered population. One of the advantages of a parametric approach is that it is suitable when the sample size is small (Niehues & Peichl, 2014; Brunori et al., 2019). However, not only is a parametric approach employed, but a certain specification, namely linear. For the linear specification of the parametric approach, there are a few considerations to be made. One general disadvantage with linear models in an IOp-context is the assumption that the effect of circumstances on the outcome variable is fixed and additive, meaning that each circumstance is assumed to be independent of the other circumstances. The implication of this is the limitation of intersectional analysis such as being a female (as a circumstance) and coming from a certain socioeconomic background (Brunori et al., 2019). For certain countries, these interactions between the circumstances are significant, leading to downward bias of the IOp estimates. However, generally for countries characterized by low levels of IOp such as the Nordic countries, a linear specification is suitable (ibid.).

3.2 Decomposition of Inequality of Opportunity

One of the advantages of employing a direct-ex-ante-parametric approach with a linear specification as proposed by Ferreira and Gignoux (2011) is that the IOp estimate can be decomposed for each circumstance in the vector C to obtain the relative contribution of each circumstance to the IOp estimate, denoted as a Shapley decomposition. This involves first estimating the inequality measure for all permutations of the circumstances where each permutation represents a different distribution of opportunities for each individual and second calculating the average marginal effect of each circumstance to the total IOp (Davillas & Jones, 2020).

3.3 Limitations to methodology

The methodological specifications in this study do incur certain implications. First, unless all relevant circumstances are included in the regressions, including circumstances directly affecting the outcome and indirectly through the influence on effort, the IOp estimates will be lower bound. Furthermore, in parametric approaches, the residual term of the regression is meant to reflect raw effort however unobserved circumstances are automatically also included in the residual term of the regression. Hence, it is difficult to determine the extent that the residual is measuring effort. The implication of this is that when IOp is estimated using MLD, the within-group part of MLD might be inflated which would further underestimate the relative IOp estimate. Lastly, the decomposition of IOp using Shapley depends on the circumstances included in the regression, which means the relative contributions of the circumstances are sensitive to the inclusion and exclusion of circumstances, which could lead to inflated or deflated relative importance.

4. Data

4.1 Data source

The microdata used in this study was retrieved from the Swedish Generations and Gender Survey (GGS), which is part of the larger research infrastructure Generations & Gender Programme (GGP). The GGS combines both survey data, retrieved through either telephone interviews or postal/online questionnaires, and population register data to provide panel data for individuals aged 18-59. Apart from demographic information such as gender, age, economic activity and educational attainment, the respondents are asked questions regarding perceived quality and details about their family structures including their partnerships, children, and parents as well as opinions, attitudes and values on various topics such as fertility, family dynamics including gender roles and general trust. Information on family structure is obtained through a recall-based method i.e. in a retrospective manner.

Complementary information on income for the respondent and their partner for the time period 1990-2019, and the respondent's geographic location from birth and onward are reported from the population register, which was performed by Statistics Sweden. Using information from the population register implies a less risk of measurement bias compared to self-reported income. The GGS employed in this study is the second round of the survey which was conducted in 2021 with a target population of 30 000 individuals in Sweden. The total number of respondents in the second round was 8082 individuals born between 1962-2003 (response rate: 27%). It should be noted that the data is unbalanced both for the survey and population register data. Missing values are especially pronounced in information on parents, and to some extent income.

4. 2 Variables

The outcome variable in this study is long-run total market pretax income. This includes income from labor, business activity, realized capital gains, and social benefits such as sickness pay, unemployment insurance and parental leave pay. Given the nature of IOp, we are interested in lifetime or permanent income, however as noted by Böhlmark and Lindquist (2006) using current income as a proxy for lifetime income could lead to measurement issues such as attenuation bias and life cycle bias due to transitionary income variations and life-cycle variations in earnings. For this reason, similar to Björklund et al.'s (2012) specification, income was averaged over the years when the individuals were aged 32-38. Furthermore, to have comparable estimates, the incomes were adjusted to prices in 2005 using CPI estimates from Statistics Sweden to account for inflation over the years. Thus, the analysis is restricted to individuals born between 1962-1981 in Sweden, which reduces the original sample (N=8082) to 4004 individuals.

However, 78 of those 4004 individuals had missing values on income for one or more years when the individuals were 32-38 years old. While the majority was missing one to three years, there was no income data available for any year during the selected age range for a few individuals. The following procedure was conducted for the unbalance in the income data: for individuals that had one to three missing years the unweighted average was calculated for the observed years when they were 32-38 years old. If individuals had missing values throughout the selected age specification, it was averaged over the observed incomes when the individuals were 39-41 years old. These ages still avoid the potential measurement issues in proxying current income for lifetime income as noted by Böhlmark and Lindquist (2006). One individual had missing values for all years that income data was available and was excluded from the study.

One limitation of this adjusted averaging procedure due to missing values is the potential variation in the accuracy of the estimates for lifetime income. However, the concern is less about the effect of potential under- or overestimation of life-time income due to unobserved income since most of the missing values were present at the early ages i.e. 32-33 years old, and thus averaged over the years which would reduce the bias in lifetime income, but more on the effect on the income distribution of the sample. While missing values in income were present for all birth cohorts, nearly 40% of those 78 individuals were born between 1962-1966. To examine the extent of the effect of the averaging procedure due to missing values on the income distribution, the income distribution before and after the adjusted averaging procedure was computed using Kernel density estimation. The results indicate a fairly uniform distribution, with a slight reduction in the frequency around the mean post-adjustment[22]. However, the difference in average income before and after is less than 100 SEK. Overall, the effect of the adjusted averaging procedure appears to be minor.

From one source of life cycle bias to another. While most studies on intergenerational transmission processes such as IOp focus on father-son pairs (Björklund & Jäntti, 2020), this study includes both parents and both sons and daughters. Potential reasons as to why studies on IOp have excluded females could be due to lower labor market participation and more frequent labor market exits or absences for child reproduction compared to male counterparts (Heidrich, 2017). According to Heidrich (2017), childbearing could be viewed as a life cycle bias since the income trajectories over a female's lifetime are influenced by her decision to reproduce. Decision to reproduce, and the timing to do so, affect wages and are associated with lower labor market participation and increased participation in the public sector (ibid.). However, as noted by Heidrich (2017), while childbearing is strongly associated with females, childbearing is only of many potential sources of life cycle bias. For example, Nybom and Stuhler (2016) show that the lifetime incomes of men vary depending on educational attainment, indicating a life cycle bias due to heterogeneity in school decisions. In this respect, life cycle biases seem to affect both genders and are not a reason to exclude females (Heidrich, 2017).

[22] The results can be found in appendix A1

From gender to immigration, the purpose of this study is to examine whether natives and second-generation immigrants face equal opportunities in Sweden. GGS does not include information on the ethnic origin of the individuals born in Sweden or whether they are second-generation immigrants. However, the database does provide information on whether the parents of the offspring were born in Sweden or not. From this, an indicator was created for individuals with at least one foreign-born parent, and this is the the second-generation immigrant group. The comparison group is second-generation natives, that is individuals with two parents born in Sweden. There are 19 missing values for the variable for whether the father is born in Sweden or not and these are observations are excluded.

Unfortunately, the country of origin of the foreign-born parent(s) is unknown and thus imposes to some extent a limitation to the study due to the inability to account for potential ethnic heterogeneity in the second-generation immigrant group, especially for individuals with two foreign-born parents. However, according to results obtained by Hammarstedt and Palme (2012), the incomes of the different ethnic groups in the second-generation immigrant groups appear to converge, suggesting that the ethnic heterogeneity in the second-generation immigrant group in relation to income mobility is less of an issue compared to first-generation immigrants. Furthermore, since the ethnic background of the parents is unknown, individuals with two parents born in Sweden are considered natives even though some might be the grandchildren of immigrants. Hence it should be noted that the basis of the analysis is between second-generation immigrants and second-generation natives even though the term ‘natives’ is used throughout the study.

In this adjusted sample, 3490 individuals are considered natives and 491 individuals as second-generation immigrants. On the national representativeness of this ratio, Statistics Sweden estimates that in 2021, less than 1.5 million individuals had one or two foreign-born parents, and more than 6.8 million individuals had two parents born in Sweden (Statistics Sweden, 2023). The native-second-generation immigrant ratio in this adjusted sample is 88-12% and 84-16% for Sweden in 2021, indicating a fairly accurate representation of the actual population. 14% of the total population in Sweden was estimated to be second-generation immigrants in 2021.

Table 1. Summary list of variables

Variables	Explanation
Outcome:	
Income	Long-run total market pre-tax income. Adjusted to 2005 prices using CPI from Statistics Sweden. Averaged over when individual was aged 32-38.
Circumstances:	
Gender	Gender of offspring
Parental education	Highest level of education obtained by parents. The highest between father and mother was selected. Categorical: 1: compulsory school 2: more than compulsory school but no tertiary education 3: at least some tertiary education (≥2 years)
Parental occupation	Occupation of mother and father during offspring's childhood. Based on ISCO classification. Categorical 1. Craft and related trades workers, plant and machine operators and assemblers, and elementary occupations 2. Clerical support workers, service and sales workers, and skilled agricultural, forestry and fishery workers 3: Managers, professionals, and technicians and associate professionals
Number of siblings	Categorical 1: no siblings 2: 1-2 siblings 3: 3+ siblings
Family structure	Categorical 1: parents did not live in the same household during childhood 2: Both parents lived in the same household during childhood
Place of birth	Classification based on population density: 1: Small towns and rural municipalities 2: Large cities, commuting municipalities near large cities and medium-sized cities

The combination of the circumstances in this study yields T = 648. The selection of certain circumstances and their level of granularity was primarily based on Björklund et al.‘s (2012) and Hederos et al.’s (2017) empirical specifications, especially for family background, and adjusted to data availability. Hederos et al. (2017) include gender as a circumstance while Björklund et al. (2012) do not. Both studies include parental income as a circumstance which were not available in GGS, instead parental occupation was used[23]. According to Clark (2014) parental education and occupation along with income is an indication of social status. Furthermore, parental connections could influence access to certain jobs (Corak, 2013).

Other circumstances included in Björklund et al.’s (2012) and Hederos et al.‘s (2017) specifications are number of siblings and family structure. Family structure during childhood seems to affect long-run economic outcomes such as income and educational attainment at adult age, whereby growing up in a single-parent household is associated with lower incomes at adult age compared to living with both parents in the same household (Lopoo & DeLeire, 2014; Lerman et al. 2017). Similar results are found for number of siblings, where an increased number of siblings is associated with lower long-run income for the offspring at adult age (Mogstad & Wiswall, 2010.).

While Björklund et al. (2012) and Hederos et al. (2017) don’t include place of birth as a circumstance in their analysis, it is a frequently used circumstance in the literature according to Ferrerira and Gignoux (2011) and Brunori (2016). According to Aarberge et al. (2011), place of birth could be relevant in terms of locally available resources such as schools, childcare and interaction with peers, or in other words neighborhood effects. Furthermore, Heidrich (2017) finds empirical evidence that place of birth influences income mobility in Sweden[24].

On the level of granularity of the circumstances, the issue is to have a sufficient amount of observations for each circumstance and a sufficient amount of circumstances in relation to the sample size to avoid biased estimates, which is a constraint in empirical applications (Brunori et al., 2019). For example, consider the circumstances parental education and occupation. It is likely that individuals are uniformly distributed across types because certain circumstances such as parental education and occupation are typically strongly correlated, and so it is possible that

[23] Parental occupation is a typical circumstance in the literature (Björklund & Jäntti, 2020)
[24] Intergenerational mobility is another closely related approach which examines the relationship between the income of the parent(s) and the income of the offspring. In essence, intergenerational mobility is a one-dimensional analysis by only examining one circumstance and IOp a multi-dimensional version

there is not enough variability in the sample to have sufficient amount of observations for each type[25].

To deal with this issue in this study the following strategy was employed: 1) to limit the level of granularity of each circumstance to reflect sufficient difference in the categories but not too detailed since having too detailed circumstances might lead to upward bias (ibid.) and 2) a pair-wise correlation matrix was constructed for each category of each circumstance to examine if strong correlations between the circumstances (and the categories) exist[26]. Take the 'parental occupation' circumstances as an example, over 100 occupations based on the ISCO classification for both parents are reported in the study sample. Since ISCO has hierarchal structure, the categories of the circumstance are meant to reflect the occupational groups that share most similarities in terms of the nature of work, skill level, and educational requirements.

5. Empirical results

5.1 Descriptive statistics

Table 2. Long-run average incomes

	N	Mean	Std. dev	Min	Max
Panel A. Natives:					
1962-1966	1036	207 212	115 038	0	537 941
1967-1971	919	231 541	119 548	0	605 974
1972-1976	795	259 398	131 687	0	707 873
1977-1981	740	282 414	132 417	0	742 169
Pooled	3490	241 452	127 081	0	742 169
Panel B. Second-gen immigrants					
1962-1966	131	207 396	122 169	0	537 941
1967-1971	125	227 145	126 587	0	593 439
1972-1976	128	221 719	118 400	0	566 323
1977-1981	107	270 754	130 443	8020	664 203
Pooled	491	229 965	125 891	0	664 203
Panel C. All					
1962-1966	1167	207 233	115 806	0	537 941
1967-1971	1044	231 014	120 358	0	605 974
1972-1976	923	254 172	130 517	0	707 873
1977-1981	847	280 941	132 150	0	742 169
Pooled	3981	240 034	126 976	0	742 169

Expressed in SEK.

[25] In this study, this is not an issue because parental education and occupation are not strongly correlated (i.e. > ± 0.2)
[26] See appendix A2. for the correlation matrix. Most variables show weak correlation (< $\pm$ 0.3), with the exception of category 2 and 3 of the circumstance 'number of siblings'.

Overall, it appears that the average long-run income of natives (panel A) is higher than second-generation immigrants (panel B) both when the groups are pooled and separated based on birth cohorts. While the relative differences in average incomes are lower for the older birth cohorts, individuals born between 1962-1966 and 1967-1976, in all three groups, the relative differences increase for the younger cohorts. When the groups are pooled (panel C), the average long-run income for the whole sample is much closer to the one for natives since this group has substantially more observations. The average incomes appear to be the highest for the individuals born between 1977-1981 in all groups.

Table 3. Descriptive statistics on the observed circumstances

Circumstance	Natives (%)	Second-gen immigrants (%)	All (%)
Gender:			
Female	54.8	52.6	55.6
Male	45.2	47.4	45.4
Parental education:			
Compulsory school	14.3	17.1	14.6
More than compulsory, no tertiary	49.1	49.9	49.2
Some tertiary	36.6	33.0	36.2
Father’s occupation:			
Occupation G1	46.6	45.4	46.5
Occupation G2	19.8	16.9	19.5
Occupation G3	33.6	37.7	34.0
Mother’s occupation:			
Occupation G1	41.2	37.2	40.7
Occupation G2	49.0	44.5	48.5
Occupation G3	9.8	18.3	10.8
Nr. of siblings:			
No siblings	6.1	7.3	6.3
1-2 siblings	70.3	63.1	69.4
3+ siblings	23.6	29.6	24.3
Family structure:			
Both parents not present	17.4	25.1	18.4
Both parents present	82.6	74.9	81.6
Place of birth			
Small towns and rural	47	35	45
Large cities and mid-sized towns	54	65	55

The observed circumstances appear to be fairly similar both for natives and second-generation immigrants, as well as for the whole sample. The sample includes a slightly higher share of females compared to males. Most of the individuals have parents with more than compulsory education, and the occupation of the parents differ where the majority of the fathers were involved with craft and related trades workers, plant and machine operators and assemblers, and elementary occupations during the offspring childhood, and the mothers with clerical support, service and sales etc. Most individuals have 1-2 siblings and grew up in households with both parents present, however, this was slightly lower for second-generation immigrants compared to natives and the whole sample. Lastly, the majority of individuals in the sample were born in large cities and mid-sized towns, with the share of second-generation immigrants being higher compared to natives and the whole sample.

5.2 Main results

Before the main results are presented and interpreted, a few remarks should be made. First, by default, the IOp estimates are lower bound since it is likely that not all circumstances that could be relevant are included. Second, the results should be viewed with caution for the following reasons: the combination of unbalanced data and small sample size resulted in a low number of observations for the OLS estimation of the counterfactuals. The implication of this is that when MLD is applied to obtain the absolute and relative measures of IOp, including the whole sample for each group would result in biased estimates. For this reason, the IOp estimates that are presented only reflect the distribution for which the counterfactuals are estimated.

Table 4. OLS regression results for natives, second-generation immigrants and, the whole sample

Dependent variable: Log average income	Natives	Second-generation immigrants	All
Intercept	11.999***	12.711***	12.082***
	(0.156)	(0.473)	(0.132)
Male	0.477***	0.471***	0.469***
	(0.047)	(0.152)	(0.043)
Parental education G2	0.043	0.197	0.055
	(0.084)	(0.214)	(0.075)
Parental education G3	0.155	-0.064	0.146
	(0.099)	(0.270)	(0.090)
Father's occupation G2	-0.005	0.334*	0.038
	(0.058)	(0.191)	(0.062)
Father's occupation G3	0.002	0.161	0.027
	(0.057)	(0.229)	(0.057)
Mother's occupation G2	0.044	-0.595**	-0.020
	(0.082)	(0.243)	(0.069)
Mother's occupation G3	0.103	0.092	0.077
	(0.093)	(0.238)	(0.078)
Family structure	0.184**	-0.396*	0.120*
	(0.0819)	(0.221)	(0.007)
Nr. of siblings G2	-0.128	0.020	-0.114
	(0.088)	(0.239)	(0.076)
Nr. of siblings G3	-0.162*	-0.273	-0.1733**
	(0.096)	(0.204)	(0.083)
Large city and medium-sized town	-0.056	-0.330**	-0.082*
	(0.047)	(0.163)	(0.004)
Nr. of observations	636	78	714
Adjusted R^2	0.171	0.213	0.162

***significant at 1%, ** significant at 5%, * significant at 10%-level. Bootstrapped standard errors in parenthesis. 500 iterations. G refers to the category of the circumstance. See Table x. to identify the category.

The statistically significant intercepts could be indicating there are unobserved circumstances that are not included in the regressions, further confirming that these are lower-bound estimates. On the gender divide, being a man offers 0.469-0.477 log points advantage compared to being a woman in this sample. The results are statistically significant at 1%-level for all three groups.

On the relevance of parental characteristics, having parents that have completed at least secondary education appears to have a positive effect on the income of the offspring, except for second-generation immigrants in the 'at least some tertiary'-category in parental education, which yielded a negative effect. However, all of the results are statistically non-significant. The results for parental occupation are mixed and statistically non-significant. Father's occupation has a positive effect for second-generation immigrants and the pooled sample, and a negative effect for natives. Only the second category in father's occupation is statistically significant at 10%-level for second-generation immigrants. On the other hand, the second category in mother's occupation is statistically significant at 5% and negative for second-generation immigrants, positive and non-significant for natives, and negative and non-significant for the whole sample. The third category in mother's occupation was positive and non-significant for all three groups.

On the relevance of childhood and other background characteristics, living with both parents during childhood has a positive effect for natives and the pooled sample, and the results are statistically significant at 5% and 10% respectively. While it appears to have a negative effect for second-generation immigrants, the results are statistically significant at 10%-level. Furthermore, having 1-2 siblings offers a disadvantage for natives and the pooled sample, but an advantage for second-generation immigrants. However, the results are statistically non-significant. Having more than two siblings seem to offer a disadvantage for all three groups, but the results are only statistically significant at 5% for the whole sample and at 10%-level for natives. Lastly, having been born in an urban area is negatively correlated with income for all three groups but only statistically significant at 5% for second-generation immigrants and at 10%-level for the whole sample.

Table 5. The estimation of relative Inequality of Opportunity and the contribution of circumstances

	Natives	Second-gen immigrants	All
Inequality	0.171	0.251	0.160
Relative contributions (%)			
Gender	77.7	32.8	77.7
Parental education	6.2	5.3	6.1
Father's occupation	1.7	9.7	2.1
Mother's occupation	5.1	32.2	6.2
Family structure	7.1	5.8	4.3
Nr. of siblings	1.5	9.3	2.3
Place of birth	0.7	4.9	1.3

The relative IOp is measured using MLD.

Based on the counterfactual distribution estimated previously, the relative IOp estimates are obtained and decomposed for the relative contributions of the circumstances used in the counterfactual estimation. The relative IOp estimate for the whole sample was 0.160, indicating that 16% of the total income inequality in the sample is attributed to inequality due to circumstances. On the separate IOp estimates for natives and second-generation immigrants, both are higher than the estimate for the pooled sample, with 17% and 25% of the total income inequality could be attributed to inherited circumstances respectively.

On the decomposition of the IOp estimates, gender appears to be the largest relative contributor to IOp both when IOp is estimated separately for natives and second-generation immigrants, and for the whole sample. Parental education yielded similar results for all three groups, slightly lower for second-generation immigrants. While the father's occupation appears to be a minor source for IOp for natives and the whole sample, it is the third largest contributor to the IOp estimate for second-generation immigrants. In comparison to occupation of the father, the effect of mother's occupation on the IOp estimates is larger, especially for second-generation immigrants, which was the second largest relative contributor. Family structure during childhood seems to be a relevant source for IOp for natives, and less so for second-generation immigrants and the whole sample. On the other hand, number of siblings and place of birth appear to be more important for second-generation immigrants than for natives and the whole sample.

5.3 Robustness checks and extension

5.3.1 Robustness checks

Selection of methodology

The main results are robust to the inclusion and exclusion of various categories of each circumstance. Since a log-linear regression using OLS was employed for the construction of the counterfactual distribution, a dummy variable for each category of each circumstance was constructed, and to avoid multicollinearity, one category for each circumstance was omitted. To examine whether the results were sensitive to the inclusion and exclusion of a certain explanatory variable, the same estimations were repeated for various combinations of the explanatory variables and yielded similar IOp estimates. On most repetitions, the same IOp estimates were obtained, and any change was less than 0.001 points of the MLD index for all three groups (natives, second-generation immigrants, all).

However, as noted by Brunori et al. (2019), to avoid biased estimates, there should be sufficient amount of individuals in each type and sufficient amount of circumstances in relation to sample size. Given the small sample size, it is likely that there are types with no individuals or very few individuals and too many circumstances relative to the sample size, potentially increasing the sampling variance when the counterfactual distribution is estimated, which could lead to an upward bias. If the model is too complex in relation to the sample size, for example by including too many circumstances, the model would be able to perfectly explain the outcome variability in the sample but poorly predict out-of-sample, reducing the possibility to generalize the results for the population (ibid.). Brunori et al. (2019) propose *k*-fold Cross-Validation (CV) for model selection and evaluation to reduce the potential bias in the IOp estimates. The main advantage of this approach is that the model that minimizes the out-of-sample prediction error or the mean squared error (MSE) can be identified (ibid.). This is the model with the best predictive accuracy compared to other specified models.

In a *k*-fold CV, the sample is randomly divided into *k*-equal-sized samples: a test sample and *k-1* training samples, and the model is fitted to the training samples and then tested based on how accurately it predicts the outcome variable in the test sample by estimating the MSE (Brunori et al., 2019). The procedure is repeated by randomly selecting another subsample to be the test sample, e.g. a training sample in previous estimation, obtaining the MSE from changing the test sample, and then calculating the average MSE based on these repeated estimations. The test sample is meant to represent the out-of-sample observations and the folds indicate the number of groups that the test and training samples are divided into. Generally, a value between five and ten of *k* is selected and similar to Brunori et al. (2019) five folds are employed, *k*=5.

Following Brunori et al.'s (2019) reasoning and methodology, the main results are re-estimated by reducing the number of categories and circumstances. First, the number of categories for each circumstance in the baseline model was reduced to two by grouping each category in the main regressions for circumstances that had more than two categories, which yields 128 types. Then, each circumstance was removed one by one while keeping the remaining circumstances, and CV was performed for each model. The circumstance whose elimination yielded the lowest average root mean squared error, RMSE[27], for all groups was removed for the re-estimation of

[27] RMSE is simply the square root of MSE. RMSE is the preferred measure since it is more intuitive to understand compared to MSE. The lowest average RMSE is simply the unweighted average from the five estimations performed for each model.

the IOp estimates, which in this case was parental education[28]. In addition to the elimination of parental education as a circumstance based on the CV-approach, an alternative model is considered by excluding number of siblings for cross-reference. Thus, adjusting the groups for the circumstances and removing either parental education or number of siblings results in 64 types.

While the main purpose of using CV in this study is to select the model with the best predictive accuracy after the adjustments to the circumstances (and the elimination of a circumstance), to use *k*-fold CV could, to some extent, improve the underlying issue with the main regressions, which is that the data is only balanced for a subsample of the sample, and thus the relative IOp estimates are obtained based on those individuals. By choosing the model with the best predictive accuracy compared to the other models, the probability to obtain representative relative IOp estimates for the whole sample could increase.

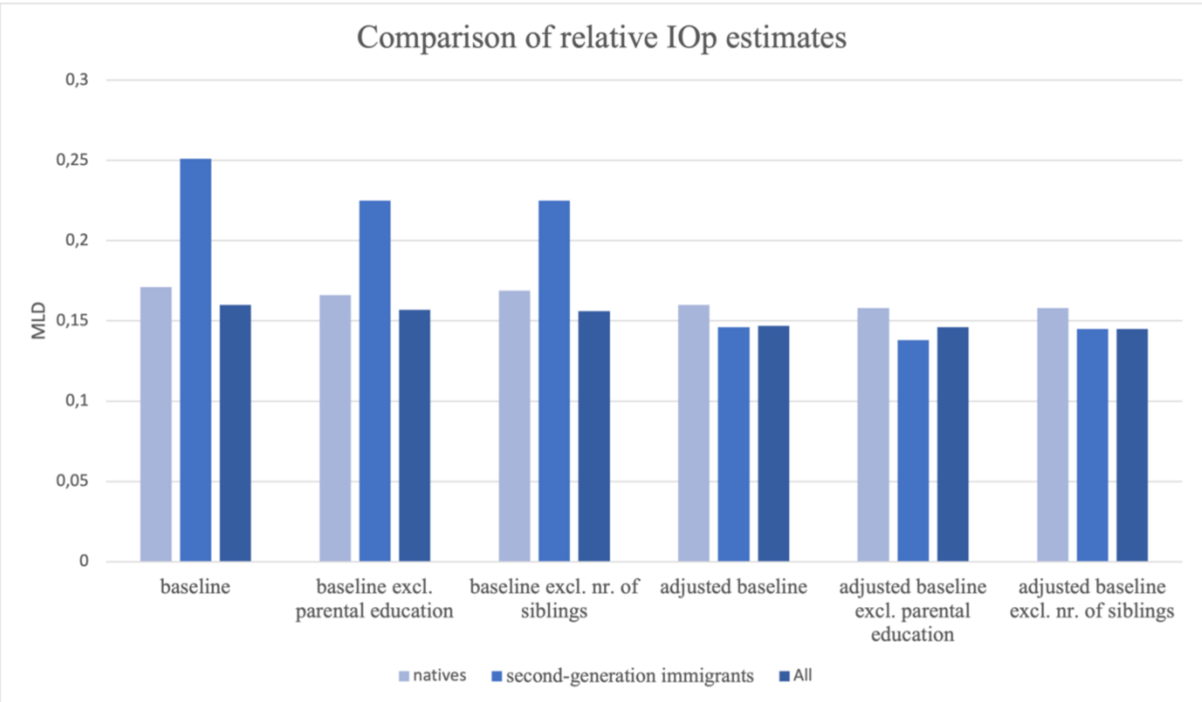


Figure 1. Comparison of baseline IOp estimates to alternative specifications

[28] See appendix A3 for the results from the CV

In addition to the IOp estimates in the main results (baseline), five additional specifications were considered for cross-comparison: one) removing parental education from the baseline estimates, two) removing number of siblings from the baseline estimates, 3) reducing the number of categories for each circumstance denoted as ‘adjusted’ baseline, four) removing parental education from the ‘adjusted’ baseline and five) removing number of siblings from the ‘adjusted’ baseline. As seen in figure 1, the IOp estimates for natives and the whole sample appear to remain fairly robust to the alternative specifications. The ‘adjusted’ specifications including the exclusion of parental education and number of siblings (the three specifications to the right in figure 1) resulted in a less than 10% reduction in relative IOp estimates. The reduction was even less when the two circumstances were removed from the baseline separately, <3% for both natives and the whole sample.

On the other hand, the IOp estimates for second-generation immigrants were reduced to more than half of the value in the main results which could indicate an overfitted model for the main results due to the number of circumstances in relation to the sample size. The potential implication is an upward bias of the relative IOp estimate for this particular group. The IOp estimates for second-generation immigrants and the whole sample remained level for the ‘adjusted’ baseline and when number of siblings was removed from the ‘adjusted’. While the exclusion of parental education and number of siblings from the ‘adjusted’ specification reduced the relative IOp estimates by 1% respectively for natives and the whole sample, the change was 6% for second-generation immigrants when parental education was removed and less than 1% when number of siblings was removed.

In addition to the previous cross-validations, it was also performed for the baseline, baseline excluding parental education and number of siblings[29]. The ‘adjusted’ baseline (specification 4), performed better than the baseline, baseline excluding parental education and number of siblings for natives and for the pooled group. Excluding parental education and number of siblings from the baseline model improved the average RMSE for second-generation immigrants. The model with the lowest average RMSE for natives and second-generation immigrants was the ‘adjusted’ baseline excluding parental education, and ‘adjusted’ baseline excluding number of siblings for the whole sample. ‘Adjusted’ baseline excluding parental education was close behind for the whole sample.

[29] See appendix A4 for the results from the CV

Selection of inequality measure

To examine for potential bias due to the selection of MLD as the inequality measure, the main regression results were also compared with the Gini coefficient. Following the same reasoning as for the main results, IOp is expressed in relative terms, where:

$$\theta_r = \frac{Gini(\hat{Y})}{Gini(Y_{within})+Gini(\hat{Y})+K} \qquad (9)$$

where θ_r is the relative IOp estimate, $\hat{Y}$ the counterfactual distribution, Y_{within} is the within-group inequality, or inequality due to effort, and K the residual term of the decomposition of the Gini index. Y_{within} and K are treated as the residual term of the counterfactual estimation, hence the Gini index is applied to this residual term and the counterfactual distribution to obtain the overall Gini index. Note that the dominator is the decomposable Gini (eq. 5) that is meant to reflect the within-group and between-group inequalities (Brunori et al. 2019).

The relative IOp estimates using the Gini coefficient were higher than the main IOp estimates, where it was 0.252 for natives, 0.360 for second-generation immigrants and 0.243 for the whole sample. The results from the Gini index are also compared to the ‘adjusted’ baseline excluding parental education model based on the previous analysis. The relative IOp estimates obtained from that adjusted regression are 0.242 for natives, 0.283 for second-generation immigrants and 0.235 for the whole sample. The relative IOp estimate for second-generation immigrants remained the highest when the Gini index is considered for both the main regressions and the ‘adjusted’ specification, while it changed from the highest to the lowest when MLD was employed to those models. One possibility is that it that MLD is underestimating the relative IOp for all groups, but especially for second-generation immigrants.

5.3.2 Extension: Foreign-born parents as a circumstance

While this study is unable to account for potential ethnic heterogeneity in the second-generation immigrant group, one consideration could be to examine the effect of having foreign-born parents. The purpose of this section is to explore the relevance of group-specific characteristics in inequality due to circumstances and the extent that the empirical analysis in this study and in general is sensitive to the selection of circumstances. Due to the nature of the main analysis, to adding having foreign-born parents as a circumstance was not possible since it would compromise the cohesiveness of the comparison between natives and second-generation immigrants if different circumstances were used. However, it might be a relevant circumstance for second-generation immigrants, for this reason, the regression was repeated by adding this particular circumstance. Given the potential upward bias in the IOp estimate for second-generation immigrants in the baseline results, the 'adjusted' specification excluding parental education will be used as a reference point since this model performed the best in comparison to the other two adjusted specifications.

The results from the OLS regression can be found in appendix A5. Having a foreign-born mother or father yielded a disadvantage both when only second-generation immigrants were considered and for the whole sample. However, only foreign-born mother was statistically significant at 10% for the second-generation immigrant group, and the remaining results statistically non-significant. Based on the counterfactual distributions the relative IOp estimates obtained for second-generation immigrants and the overall sample were 0.198 and 0.147 respectively, measured in MLD[30]. In comparison to the 'adjusted' specification excluding parental education, relative IOp increased for second-generation immigrants and remained the same for the whole sample both when MLD and the Gini coefficient were used. These results are robust to the exclusion of number of siblings from the 'adjusted' model and the combined exclusion of parental education and number of siblings from the 'adjusted' specification.

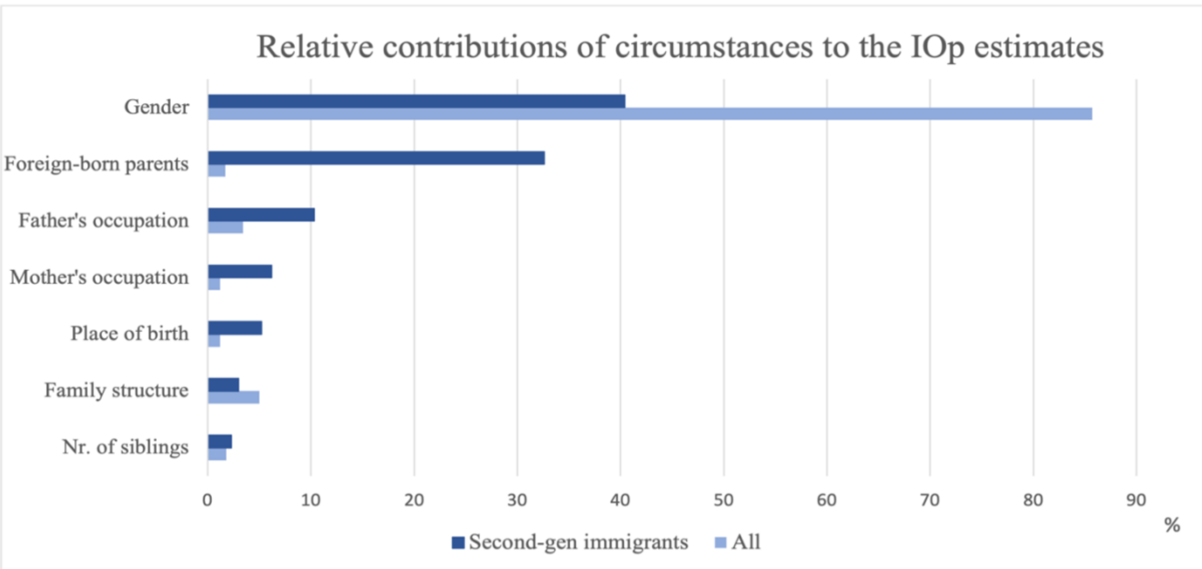


Figure 2. Relative contributions of circumstances to the IOp estimates when foreign-born parents is added as a circumstance. MLD employed.

[30] 0.292 and 0.234 respectively when measured with the Gini index

Not only did the relative IOp estimate increase for second-generation immigrants after foreign-born parents were added as a circumstance, but it became the second largest contributor to the IOp estimate, with over 30% of the total contribution to the IOp estimate. The effect of having foreign-born parents contributed less than 2% to the relative IOp estimate for the whole sample. Gender still had the largest relative effect on IOp for both groups, which has been consistent throughout this study. While the contribution of mother's occupation remained level for second-generation immigrants in comparison to the results obtained from the 'adjusted' specification excluding parental education (both models yielded 6% contribution respectively), father's occupation was reduced from 17% to 10%[31]. The relative effect of place of birth still remained higher for second-generation immigrants compared to the whole sample.

6. Analysis

In comparison of the IOp estimates for the overall sample to other empirical results, Björklund et al. (2012) obtain a relative IOp estimate of 0.158 in MLD and 0.263 using the Gini index. On the other hand, Hederos et al. (2017) obtain slightly higher estimates compared to Björklund et al. (2012), with 0.186 expressed in MLD and 0.296 in Gini respectively for individuals (both males and females are included) born between 1952-1964. Hederos et a. (2017) also perform a robustness check by including the age cohorts that Björklund et al (2012) use, i.e. 1955-1967, and obtain a relative IOp estimate of 0.163 measured in MLD. While the relative IOp estimate from the main regressions (0.160) is comparable to the result from Björklund et al. (2012), the result obtained from the 'adjusted' model excluding parental education is less than 7% lower from Björklund et al.'s (2012) estimate and ca 20% lower than Hederos et al.'s (2017) main estimate. The results using the Gini index were also lower in this study, 0.243 and 0.235 depending on which model specification is considered. However, similar to Björklund et al. (2012) and Hederos et al (2017), the relative IOp estimates using the Gini index were higher than the estimates obtained with MLD in this study.

Potential explanations for the difference between the relative IOp estimate obtained in this empirical study and those from Björklund et al. (2012) and Hederos et al. (2017) could be the difference in circumstances used in the regressions, the sample sizes and the birth cohorts. Both

[31] For IOp results from the 'adjusted' specification excluding parental education can be found in appendix A6

articles use IQ test scores at the age of 18 and parental income as circumstances, which are not accounted for in this study, and this study includes parental occupation and place of birth as circumstances that the other articles don’t account for. Furthermore, the sample sizes are substantially larger in the mentioned studies, with over 200 000 men in Björklund et al. (2012) and more than 350 000 individuals in Hederos et al. (2017). Lastly, this study also includes individuals born in later years compared to both articles, >1967. Overall, the estimates obtained in this study are not substantially different from estimates for Sweden.

Since there are no empirical articles that perform separate IOp estimates for natives and second-generation immigrants available for Sweden nor other Nordic countries the way employed in this study, makes the comparability of the estimates difficult. The closest proxy is the research by Behtoui (2006) that deduce a 20% disadvantage for second-generation immigrants in comparison to natives in Sweden. If the relative difference in the IOp estimates for natives and second-generation immigrants can be assumed as a measure of disadvantage, then based on the main regressions, second-generation immigrants have an increased 30% disadvantage both when MLD and the Gini index are considered. This estimated disadvantage disappears when the selected inequality measure is MLD for the ‘adjusted’ excluding parental education model, however if the Gini index is used for that particular model, then the difference is an increased disadvantage of 15% for second-generation immigrants. However, this type of analysis should be viewed with caution since the empirical strategy employed in this study differs from the one of Behtoui (2006).

On the decomposition of inequality, similar to Hederos et al. (2017), gender was the largest contributor to the relative IOp estimates both when performed separately for natives and second-generation immigrants, and for the whole sample. While parental income was an important source for IOp both in Björklund et al. (2012) and Hederos et al. (2017), in Hederos et al.’s (2017) analysis, parental education was also close behind. In this study, parental background was also important relative contributor for all three groups, but especially for second-generation immigrants. This would provide further evidence to Tasiran and Tezic (2007) that study the early labor market experiences of second-generation immigrants in Sweden. While the transition phase into the labor market could influence the later labor-market outcomes, this empirical study considers long-run incomes when individuals are 32-38 years old, indicating that the importance of parental education persists even after the transition phase.

Furthermore, place of birth was another important circumstance for second-generation immigrants, which Tasiran and Tezic (2007) also found evidence for.

However, there are potential limitations to the empirical results obtained in this study. First, due to unbalanced data, only a subsample for each group was included in the regressions, 18% of natives, 16% of second-generation immigrants and 18% of the whole sample were included. To estimate relative IOp for the full sample only using individuals with a complete set of circumstances for the counterfactual distribution would have led to biased estimates since the counterfactual reflects the average income of each type and we wouldn't be able to distinguish the types in comparison to the actual income distribution. Furthermore, it is unclear how comparable the IOp estimates are between the groups due to different sample sizes, especially since the size of the sample included in the regression for second-generation immigrants is 12% of the one included for natives.

The issue of unbalanced data was mainly driven by parental occupation. One potential issue mentioned by Neidhöfer et al. (2018) could be selection bias, that it is, those who were unable to report information on their parents could be from poorer economic backgrounds. While it is unclear if that is the case for this sample, there is some support for this potential explanation. Upon comparison of the income distributions of the full samples and the subsamples included in the OLS regressions, the curves shifted to the right, with a lower frequency of individuals with lower incomes and a higher frequency of high-earning individuals[32]. The implication is that it is possible that the average income for each type is higher compared to the full samples and it is unclear how representative the relative IOp estimates are for the full samples.

Another consideration relates to the statistical significance of the OLS estimations since most of the circumstances were statistically nonsignificant. The Variance Inflation Factor (VIF) tests for the OLS regressions don't indicate the presence of multicollinearity[33]. The question is whether the circumstances could still be contributing to IOp even if they are statistically nonsignificant. Unfortunately, the literature does not provide any guidance on this matter. Upon comparison with other empirical studies with similar methodologies, e.g. Ferreira and Gignoux (2011), Björklund et al. (2102) and Hederos et al. (2107), all three articles had more than one statistically nonsignificant result even for circumstances that the authors deduce to be important

[32] See appendices A7-9
[33] See appendix A10

sources of IOp such as parental education. Thus, it appears that the primary purpose of the OLS regressions is to obtain the counterfactual distributions for the estimation of IOp and the circumstances could still be relevant even if they are not statistically significant.

One potential explanation for the statistically nonsignificant results in this study could be insufficient sample sizes both when IOp is estimated separately and for the pooled sample. The lack of sufficient sample sizes could not only have been a source of influence for the statistical significance of the regression results but also the estimation of relative IOp. When the baseline estimates were compared to alternative specifications, especially when the number of categories for each circumstance was reduced, lower relative IOp estimates were obtained for all three groups, however, the change was most notable for second-generation immigrants. It is possible that the variability due to sampling variance was captured by the IOp estimate, indicating upward bias. Another potential evidence for such overestimation could be found in the change of the relative contribution of mother's occupation for second-generation immigrants, which was reduced from 32 to 6% when the baseline model was modified.

7. Concluding discussion

In this study, the extent income inequality attributed to inherited circumstances and the effects of these circumstances on income inequality due to inherited circumstances was explored for natives and second-generation immigrants in Sweden to determine whether they face equal opportunities. The empirical results obtained indicate that it depends on which model and inequality measure is considered. If MLD is used then, 16-17% of income inequality is due to inherited circumstances for natives, 14- 25% for second-generation immigrants and 15-16% for the whole sample. However, if the Gini index is considered then the results would be 24-25% for natives, 28-36% for second-generation immigrants and 24% for the whole sample. Regardless of which model or inequality measure is considered, as most empirical studies in the literature, these are lower-bound estimates of Inequality of Opportunity.

Furthermore, there are both similarities and differences in the relative effects of the circumstances for natives and second-generation immigrants. While gender and parental background were important contributors to IOp for both groups, the extent of the relative effects was different. The relative contribution of gender as a circumstance to IOp was higher for natives compared to second-generation immigrants and parental occupation remained higher for second-generation immigrants. However, as observed in the extension, the relative contribution of parental occupation to IOp was reduced when having foreign-born parents was added as a circumstance. This could be indicating the need for intersectional analysis, that is to add interaction term between foreign-born parents and parental occupation.

Albeit this study included circumstances that most empirical studies are unable to include in their analysis, there might be many factors unaccounted for in this study. Ethnic heterogeneity is one of them. All of these unobserved circumstances were captured by the residual term of the counterfactual distribution, along with raw effort and circumstances correlated with raw effort. However, the residual also captures another element: luck. The role and effect of unobserved luck is unexplored in the empirical literature even though it is an important part of the normative framework. This is an area that needs to be explored further in the literature.

This study also highlights some shortcomings with the empirical literature. Not only do the empirical studies need to improve the justification of selecting certain measurement approaches and circumstance and effort-variables to reflect the studied population but also consider type-specific effects or circumstances. If we ought to identify circumstances that are a source of inequality of opportunity as the theory suggests, then we might need to consider which circumstances are relevant to the outcome for the population, whether that it is within a population or the entire population. Furthermore, as seen in this study, the IOp estimates were to some extent sensitive to the selection of circumstances, and their level of granularity. The empirical literature is rich in various combinations of circumstances and measurement approaches, and it is unclear how comparable the results are, especially for cross-country comparisons. Another caveat is the extent of accuracy of these results in the empirical literature since most studies obtain lower-bound estimates however it is unclear how far off the estimates are from true Inequality of Opportunity.

Overall, the variation in the results in this study and variations in the measurement of IOp seem to be reflecting a variation that began with the origins of the framework. The challenge to create a unified framework still remains.

References

Aaberge, R., Mogstad, M. & Peragine, V. (2011). Measuring long-term inequality of opportunity. *Journal of Public Economics*. 95. pp. 193-204

Almås, I., Cappelen, A. W., Sorensen, Erik, O. Tungodden, B. (2010). Fairness and the development of inequality of acceptance. Science, 328 (5982), pp. 1176-1178

Arneson, R. J. (1989). Equality and equal opportunity for welfare. *Philosophical Studies*. 56. pp. 77-93

Bethoui, A. (2006). Unequal Opportunities. The impact of social capital and recruitment methods on immigrants and their children in the Swedish labor market. Dissertation. Linköping University, Department of Social and Welfare Studies. Linköping, Sweden

Björklund, A., Jäntti, M. & Roemer, J. E. (2012). Equality of opportunity and the distribution of long-run income in Sweden. *Social Choice and Welfare*. 39. pp. 675-696

Björklund, A. & Jäntti, M. (2020). Intergenerational mobility, intergenerational effects, sibling correlations and equality of opportunity: A comparison of four approaches. *Research in Social Stratification and Mobility*, 70, pp. 1-11

Brunori, P. (2016). How to measure inequality of opportunity: a hands-on guide. Life Course Centre Working paper series no 2016-04, pp. 1-53

Brunori, P., & Peragine, V. (2011). Compensation, reward, and the measurement of unfair inequalities. *Research on Economic Inequality*, 19, pp. 1–21.

Brunori, P., Peragine, V., & Serlenga, L. (2019). Upward and downward bias when measuring inequality of opportunity. *Social Choice and Welfare*, *52*(4), pp. 635–661.

Böhlmark, A. & Lindquist, M. J. (2006). Life-cycle variations in the association between current and lifetime income: replication and extension for Sweden. Journal of Labor Economics, 24(4), pp. 879-898

Cohen, G. A. (1989). On the currency of egalitarian justice. *Ethics*, 99, pp. 906-944

Corak, M. (2013). Income inequality, equality of opportunity, and intergenerational mobility. *The Journal of Economic Perspectives*, 27(3), pp. 79-102

Clark, M. (2014). The son also rises: surnames and the history of social mobility. Princeton, U.S.: Princeton University Press

Davillas, A., & Jones, A. M. (2020). Ex ante inequality of opportunity in health, decomposition and distributional analysis of biomarkers. *Journal of Health Economics*, 69, pp. 1-16

Dworkin (1981a). What is equality? Part 1: equality of welfare. *Philosophy & Public Affairs*, 10, pp.185-246

Dworkin (1981b). What is equality? Part 2: equality of resources. *Philosophy & Public Affairs*, 10, pp. 283-345

Ferreira, G. & Gignoux, J. (2011). The measurement of inequality of opportunity: theory and an application to Latin America. *Review of Income and Wealth*, 57(4), pp. 622-657

Ferreira, G. & Peragine, V. (2015). Equality of Opportunity: theory and evidence. IZA discussion papers no. 8994, pp. 1- 42

Fleurbaey, M., & Peragine, V. (2013). Ex Ante Versus Ex Post Equality of Opportunity. *Economica*, 80(317), pp. 118–130

Jenkins, S. P. (1991). The measurement of income inequality. In Osberg, I (ed.), *Economic Inequality and poverty: International Perspectives*, New York, U. S.: M.E Sharpe, pp. 3-38

Hammarstedt, M. & Palme, M. (2012). Human capital transmission and the earnings of second-generation immigrants in Sweden. *IZA Journal of Migration*, 1(4), pp. 1-23

Hederos, K., Jäntti, M. & Lindahl, L. (2017). Gender and Inequality of Opportunity is Sweden. *Social Choice and Welfare*, 4, pp. 605-635

Heidrich, S. (2017). Intergenerational mobility in Sweden: a regional perspective. *Journal of Population Economics*, 30(4), pp. 1241-1280

Lerman, R. I., Price, J. & Wilcox, W. B. (2017). Family structure and economic success across the life course. *Marriage & Family Review*, 53(8), pp. 844-758

Lopoo, L. M. & DeLeire, T. (2014). Family structure and the economic wellbeing of children in youth and adulthood. *Social Science Research*, 43, pp. 30-44

Marrero, G. A. & Rodriguez, J G. (2013). Inequality of opportunity and growth. *Journal of Development Economics*, 104, pp. 107-122

Mogstad, M., & Wiswall, M. (2010). Testing the Quantity-Quality Model of Fertility: linearity, marginal effects, and total effects. NYU working papers no. 5216, pp. 1-52

Neidhöfer, G., Serrano, J., & Gasparini, L. (2018). Educational inequality and intergenerational mobility in Latin America: A new database. *Journal of Development Economics*. 134. pp. 329-349

Niehues, J. & Peichl, A. (2014). Bounds of unfair inequality of opportunity: theory and evidence for Germany and the U.S. *Social Choice and Welfare*, 43, pp. 73-99

Nybom, M. & Stuhler, J. (2016). Heterogeneous income profiles and lifecycle bias in intergenerational mobility estimation. *Journal of Human Resources*, 51, pp. 165-181

OECD (2016). Working together: Skills and labor marker integration of immigrants and their children in Sweden. Paris: OECD Publishing

Ooghe, E., Schokaert, E. & Van de gaer, D. (2007). Equality of opportunity versus equality of opportunity sets. *Social Choice and Welfare,* 28, pp. 209-230

Palmisano, F., & Peragine, V. (2022). Inequality of Opportunity: Theoretical Considerations and Recent Empirical Evidence. In Chotikapanich, D. Rambaldi, A N. & Rohde, N. (eds.). *Advances in Economic Measurement: a volume in honour of D.S Prasada Rao.* Singapore: Springer Nature Singapore, pp. 349–386

Pervaiz, Z. & Akram, S. (2018). Estimating Inequality of Opportunities in Punjab (Pakistan): a non-parametric approach. *Pakistan Journal of Commerce and Social Sciences,* 12(1), pp. 136-152

Pignataro, G. (2012). Equality Of Opportunity: Policy and measurement paradigms. *Journal of Economic Surveys*, 26(5), pp. 800–834.

Ramos, X. & Van de gaer, D. (2016). Approaches to inequality of opportunity: principles, measures and evidence. *Journal of Economic Survey,* 30(5), pp. 855-883

Ramos, X. & Van de gaer, D. (2020). Measurement of inequality of opportunity based on counterfactuals. *Social Choice and Welfare*, *55*(3), pp. 595–627

Ramos, X., & Van de gaer, D. (2021). Is Inequality of Opportunity Robust to the Measurement Approach? *Review of Income and Wealth*, 67(1), pp. 18–36

Rawls, J. (1971) A theory of Justice. Oxford, United Kingdom: Oxford University Press

Roemer, J. E. (1993). A pragmatic theory of responsibility for the egalitarian planner. *Philosophy & Public Affairs,* 22(2), pp. 146-166

Roemer, J. E. (1998). Equality of Opportunity. Cambridge, Massachusetts: Harvard University Press

Roemer, J. E., & Trannoy, A. (2016). Equality of opportunity: Theory and measurement. *Journal of Economic Literature*, 54(4), pp. 1288–1332

Rooth, D-O. & Ekberg, J., (2003). Unemployment and earnings for second generation immigrants in Sweden. Ethnic background and parent composition, *Journal of Population Economics,* 16 (4), pp. 787–814.

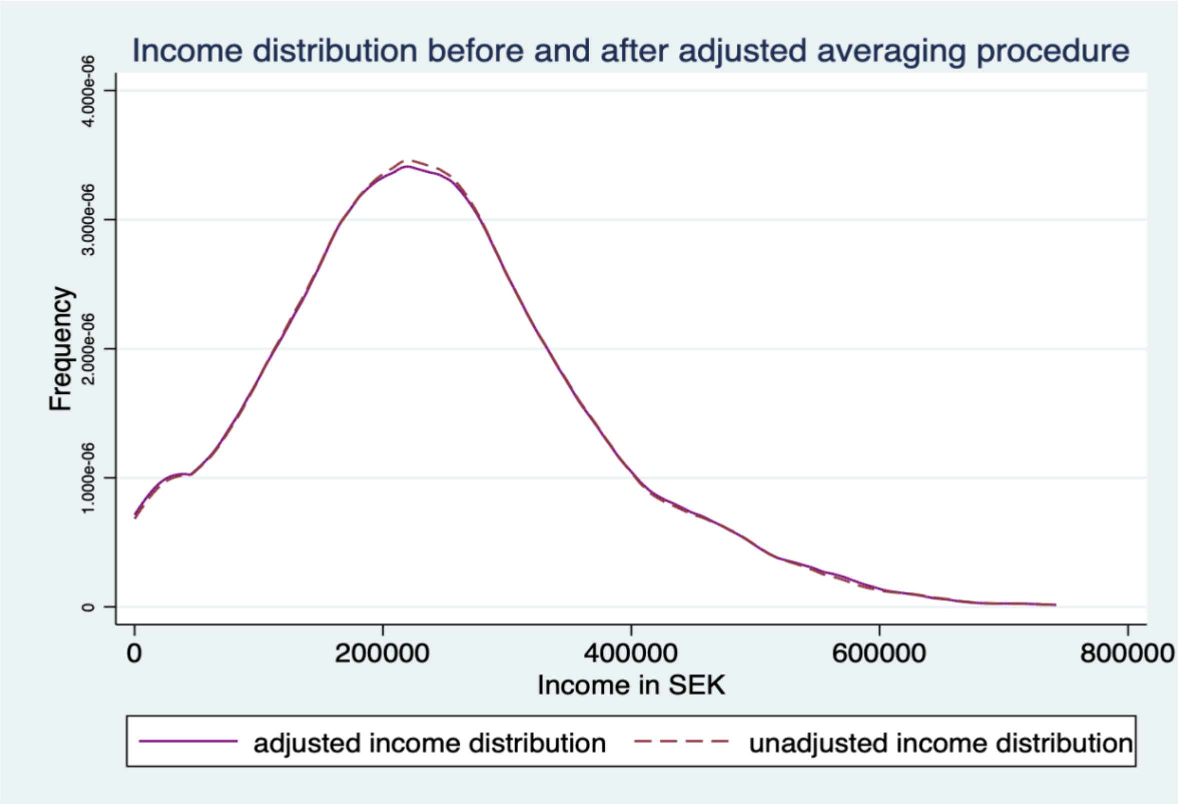
Statistics Sweden (2023). Befolkningsstatistik. Antal personer med utländsk eller svensk bakgrund (fin indelning) efter region, ålder och kön. År 2002 – 2022. URL: https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101Q/UtlSvBakgFin/ [Accessed: 19-04-2023]

Tasiran, A. & Tezic, K. (2007). Early labor-market experiences of second-generation immigrants in Sweden. *Applied Economics*, 39(7), pp. 809-824

Vilhelmsson, R. (2000). Ethnic differences in the Swedish youth labor market. Stockholm: The Swedish Institute for Social Research

Appendix

A1. Income distribution before and after adjusted averaging procedure



A2. Correlation matrices

A2.a Correlation matrix for the categories of the circumstances

	gender	p_ed1	p_ed2	p_ed3	focc1	focc2	focc3	mocc1	mocc2	mocc3	sib1	sib2	sib3	fam_str	pbirth1	pbirth2
gender	1.000															
p_ed1	0.002	1.000														
p_ed2	-0.055	-0.275	1.000													
p_ed3	0.054	-0.278	-0.847	1.000												
focc1	-0.073	0.143	0.308	-0.387	1.000											
focc2	-0.004	0.115	0.075	-0.138	-0.324	1.000										
focc3_	0.071	-0.222	-0.341	0.464	-0.658	-0.499	1.000									
mocc1	-0.060	0.256	0.100	-0.242	0.259	-0.010	-0.229	1.000								
mocc2	-0.029	0.090	0.392	-0.442	0.205	0.104	-0.270	-0.275	1.000							
mocc3	0.063	-0.236	-0.446	0.577	-0.351	-0.098	0.399	-0.299	-0.835	1.000						
sib1	0.003	-0.044	-0.039	0.063	-0.007	-0.031	0.031	-0.031	-0.068	0.069	1.000					
sib2	-0.035	-0.037	-0.005	0.025	0.007	0.005	-0.011	-0.044	-0.005	0.029	-0.255	1.000				
sib3	0.035	0.057	0.022	-0.053	-0.004	0.008	-0.003	0.046	0.034	-0.060	-0.180	-0.905	1.000			
fam_str	0.048	0.030	-0.056	0.039	-0.007	-0.001	0.007	-0.041	-0.034	0.060	0.024	0.172	-0.186	1.00		
pbirth1	0.015	-0.059	-0.071	0.104	-0.124	-0.061	0.162	-0.026	-0.094	0.108	0.146	-0.015	-0.049	-0.026	1.000	
pbirth2	-0.015	0.059	0.071	-0.104	0.124	0.061	-0.162	0.026	0.094	-0.108	-0.146	0.015	0.049	0.026	-1.000	1.00

A2.b Correlation matrix for log average income and the circumstances

	Log avr inc	P_ed	F_occ	M_occ	Fam_str	Sib	Place_birth
Log avr inc	1.000						
P_ed	-0.139	1.000					
F_occ	-0.010	0.455	1.000				
M_occ	-0.145	0.562	0.433	1.000			
Fam_str	0.102	-0.012	-0.006	-0.064	1.000		
Sib	0.068	-0.086	-0.013	-0.086	0.167	1.000	
Place_birth	0.012	-0.078	-0.132	-0.080	-0.035	0.094	1.000

A3. The CV results for the adjusted specifications

		RMSE					
Specification (only 2 categories)		est1	est2	est3	est4	est5	Average RMSE
adjusted baseline	natives	0,562	0,622	0,538	0,671	0,459	0,570
	immig	0,644	0,672	1,005	0,963	0,355	0,728
	all	0,585	0,435	0,618	0,571	0,65	0,572
Gender removed	natives	0,749	0,554	0,587	0,603	0,572	0,613
	immig	0,522	1,194	0,465	0,517	1,028	0,745
	all	0,647	0,572	0,530	0,777	0,576	0,620
Parental education removed	natives	0,724	0,498	0,628	0,549	0,414	0,563
	immig	0,404	0,243	0,785	0,997	0,520	0,590
	all	0,429	0,598	0,661	0,691	0,479	0,572
Father's occupation removed	natives	0,644	0,541	0,541	0,591	0,63	0,589
	immig	0,662	0,705	0,418	0,872	0,478	0,627
	all	0,633	0,607	0,518	0,662	0,569	0,598
Mother's occupation removed	natives	0,493	0,755	0,591	0,452	0,554	0,569
	immig	0,693	0,501	0,601	0,857	0,680	0,666
	all	0,587	0,571	0,684	0,518	0,577	0,587
Family structure removed	natives	0,494	0,609	0,609	0,491	0,635	0,568
	immig	0,417	0,482	0,431	0,634	1,485	0,690
	all	0,524	0,579	0,628	0,603	0,586	0,584
Nr. of siblings removed	natives	0,558	0,537	0,664	0,593	0,592	0,612
	immig	0,396	0,492	0,433	0,542	1,83	0,739
	all	0,555	0,667	0,597	0,604	0,428	0,570
Place of birth removed	natives	0,541	0,685	0,608	0,625	0,700	0,632
	immig	0,976	0,528	0,905	0,833	0,647	0,778
	all	0,634	0,625	0,689	0,653	0,656	0,651

A4. The CV results for the six models for comparison of relative IOp estimates

Alternative specifications		RMSE					
		est1.	est2.	est3	est4	est5.	Average RMSE
baseline	natives	0,652	0,707	0,438	0,573	0,511	0,576
	immig	0,427	0,465	0,965	0,466	1,018	0,668
	all	0,559	0,589	0,626	0,538	0,607	0,584
baseline excl. parental education	natives	0,591	0,61	0,504	0,561	0,601	0,573
	immig	0,428	0,782	0,321	0,736	0,951	0,644
	all	0,617	0,576	0,483	0,65	0,605	0,586
baseline excl. nr. of siblings	natives	0,585	0,435	0,618	0,571	0,65	0,572
	immig	1,297	0,627	0,523	0,364	0,421	0,646
	all	0,465	0,659	0,532	0,683	0,828	0,633
adjusted baseline	natives	0,562	0,622	0,538	0,671	0,459	0,570
	immig	0,644	0,672	1,005	0,963	0,355	0,728
	all	0,585	0,435	0,618	0,571	0,65	0,572
adjusted baseline excl. parental education	natives	0,724	0,498	0,628	0,549	0,414	0,563
	immig	0,404	0,243	0,785	0,997	0,520	0,590
	all	0,429	0,598	0,661	0,691	0,479	0,572
adjusted baseline excl. nr. of siblings	natives	0,558	0,537	0,664	0,593	0,592	0,612
	immig	0,396	0,492	0,433	0,542	1,83	0,739
	all	0,555	0,667	0,597	0,604	0,428	0,570

A5. The OLS regression results for the extension

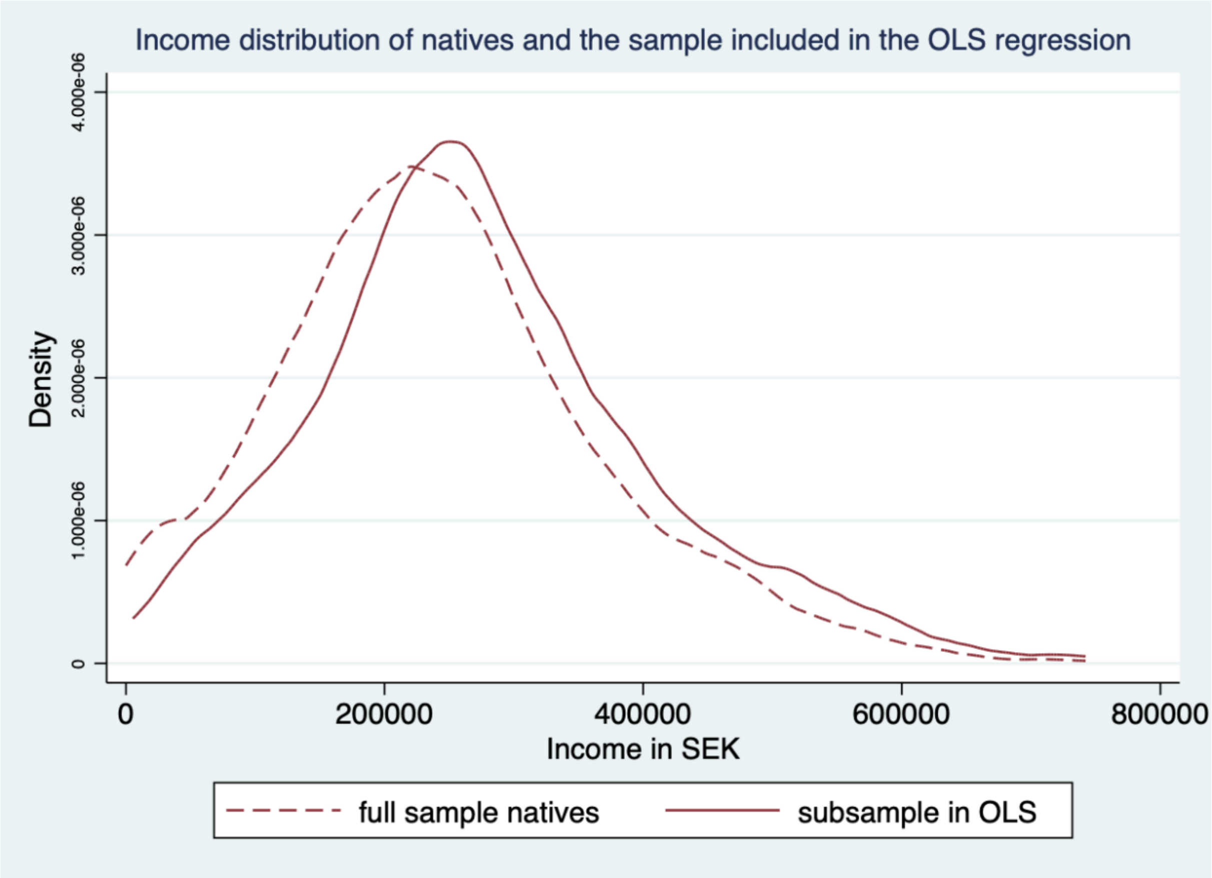
Dependent variable: Log average income	Second-gen immigrants	All
Intercept	13.117***	12.201***
	(0.516)	(0.137)
Gender	0.456**	0.470***
	(0.187)	(0.044)
Father's occupation	0.265	0.083*
	(0.175)	(0.049)
Mother's occupation	-0.346*	0.043
	(0.208)	(0.073)
Family structure	-0.191	0.140*
	(0.157)	(0.079)
Number of siblings	-0.150	-0.160**
	(0.174)	(0.080)
Foreign-born mother	-0.434*	-0.124
	(0.231)	(0.089)
Foreign-born father	-0.422	-0.031
	(0.268)	(0.089)
Large city and medium-sized towns	-0.079	-0.114*
	(0.296)	(0.068)
Nr. of observations	80	720
Adjusted R^2	0.135	0.151

A6. The corresponding table to figure 2

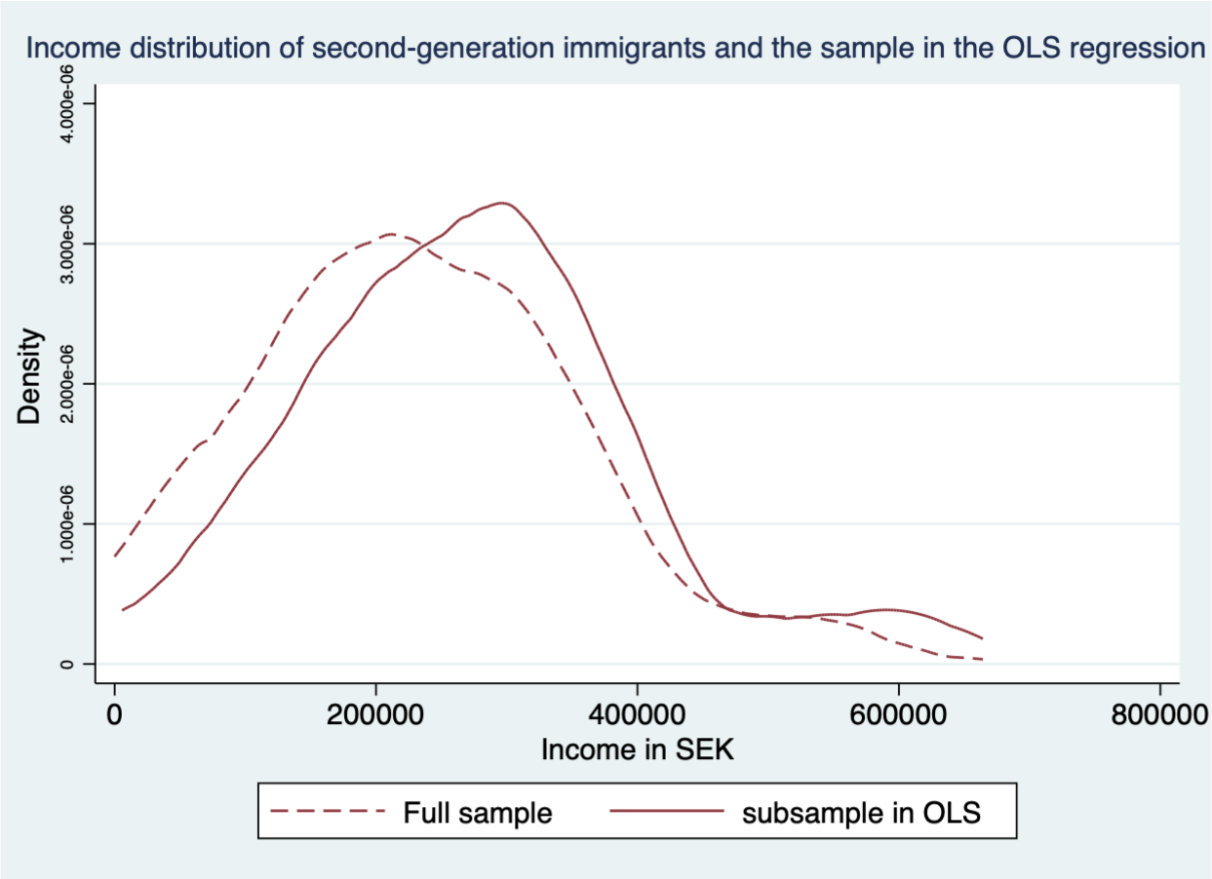
	Second-gen immigrants	All
Inequality	0.198	0.147
Relative contributions (%)		
Gender	40.4	85.7
Foreign-born parents	32.6	1.7
Father's occupation	10.3	3.4
Mother's occupation	6.2	1.2
Family structure	3.0	5.0
Nr. of siblings	2.3	1.8
Place of birth	5.2	1.2

MLD index employed for the IOp estimate and the circumstances were decomposed using Shapley. Certain circumstances were grouped.

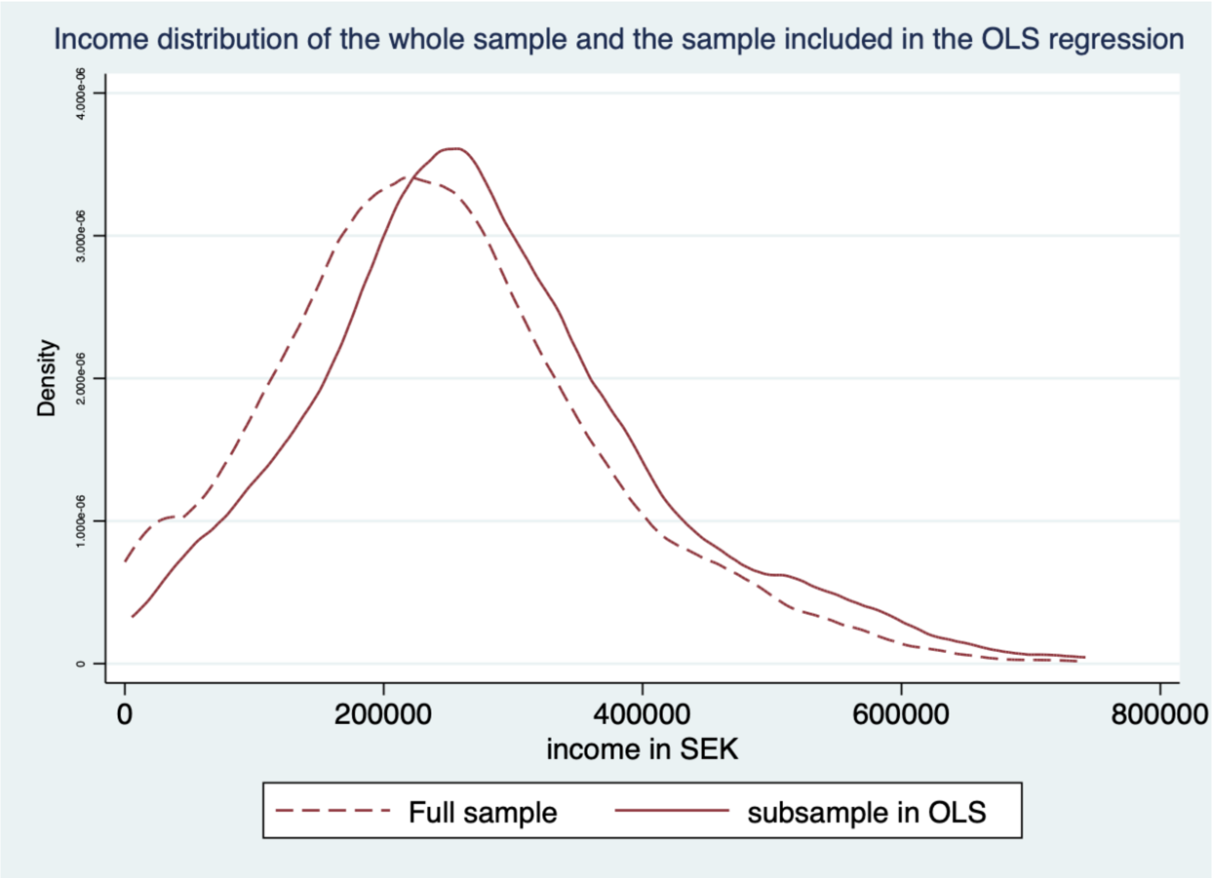
A7. Income distribution of natives (full sample) and the subsample included in the OLS regression



A8. Income distribution of second-generation immigrants (full sample) and the subsample included in the OLS regression



A9. Income distribution of the pooled sample(full sample) and the subsample included in the OLS regression



A10. Table. Variance inflator factor (VIF) values for the OLS regressions

Circumstances	Natives	Second-gen immigrants	All
Gender	1.02	1.15	1.01
Parental education G2	3.77	3.98	3.72
Parental education G3	4.81	3.91	4.72
Father's occupation G2	1.40	1.49	1.39
Father's occupation G3	1.86	1.89	1.83
Mother's occupation G2	4.01	2.24	3.51
Mother's occupation G3	4.87	3.67	4.40
Family structure	1.05	1.30	1.05
Nr. of siblings G2	6.67	3.79	5.90
Nr. of siblings G3	6.58	3.91	5.96
Place of birth G2	1.07	1.10	1.05
Mean VIF	3.36	2.76	3.14

VIF test is not available for bootstrapped regressions, instead it was performed on the underlying OLS regressions i.e. without the bootstraps.

VIF<5: low correlation
5<VIF<10: moderate correlation
VIF>10: strong correlation and a sign of multicollinearity