The effects of circumstances on long-term income opportunities:

Re-examining evidence from the baby-boomers generation in Sweden

Preliminary Version, Latest Version here

Hélène Le Forner, Arnaud Lefranc, Alain Trannoy, Francesco Andreoli *

2021-02-28

Abstract

This paper provides a new approach to the empirical estimation of inequality of opportunity (IOP). Previous studies usually conclude that, in most countries, unfair inequality, arising from circumstances outside the realm of individual control, represent a small share of total income inequality. This result is largely driven by observational constraints that hamper the full observability of relevant circumstances and lead to lower-bound estimates. This also stands at odds with estimates of intergenerational and siblings correlation, which indicate a strong influence of differences in family background on individual success. This paper bridges the gap between these different approaches and relies on family and municipality fixed-effect models to account for shared unobservable circumstances alongside a rich set of individual characteristics. We apply this methodology to the estimation of IOP in Sweden. Our results point to a larger share of IOP than previously estimated and reaching up to 46% of observed income inequality.

JEL classification: D63, I24

Keywords: Inequality of Opportunity, Siblings Approach, Circumstances

*This project has received funding from the French National Research Agency and NORFACE consortium through the Dynamic of Inequality across the Life Course (DIAL) programme (Grants number 462-16-020 / ANR-17-DIAL-0002). Le Forner and Trannoy acknowledge support from the "Investissements d'Avenir" French Government program managed by the French National Research Agency (reference: ANR-17-EURE-0020) and from Excellence Initiative of Aix-Marseille University - A*MIDEX. Lefranc acknowledges support from CY Initiative of Excellence (Labex MME–DII - grant ANR–11–LBX–00023–01). We are very grateful to Anders Björklund and Markus Jäntti for giving us access to the data set. We thank participants at seminars at IMCHILD (DIAL) and GUODLCCI (DIAL) Workshop.

1 Introduction

Inequality of opportunity can be broadly defined as inequalities due to circumstances, *i.e.* factors that individuals cannot be deemed responsible for. Measuring inequality of opportunity has been addressed from different perspectives (see Björklund & Jäntti 2020, for a detailed review). First, inter-generational mobility studies (see Jäntti & Jenkins 2015, for a review) focus on the process of transmission of opportunities from parents to children, considering only one dimension of parents' social status, for instance, income or education hence neglecting other family circumstances.

Second, sibling correlations are used to capture the share of variance due to all circumstances, observed or not, shared by siblings (Solon 1999). One drawback is that it does not account for circumstances at the individual level (genetic endowment, birth order, or gender). Besides, it might be that factors common to siblings affect them differently due to an unequal allocation of resources according to birth order (Price 2008), to gender (Blau et al. 2020), or due to a differences in the sensitivity of siblings to those resources (*e.g.* birth spacing in the presence of sensitive periods (Cunha et al. 2010)). Hence it only provides a lower bound for inequality of opportunity.

Third, inequality of opportunity (henceforth IOP) studies (see Roemer & Trannoy 2016, for a survey) aim at disentangling the part of inequalities due to individual's responsibility from the part due to circumstances. Empirically, they predict the outcome from observed circumstances and compute the share of inequality due to those circumstances. This method has several strengths; it can raise the lower bound provided by siblings because it includes observed circumstances at the individual level such as gender or birth order which has been shown to be a large determinant of individual's outcomes (Black et al. 2011) ; besides, it includes a multi-dimensional economic status of parents (Clark 2015). The empirical literature on attempting to measure inequality of opportunity has blossomed in the past ten years without reaching a consensus about how to tackle the issue of unobserved circumstances. If family socio-economic circumstances are generally present in the data set, other information about neighbors, parental relationships, cousins, aunts and uncles, grandparents, and so on are usually missing. Major incidents or events during childhood that can have a primordial effect on individual trajectories are not reported, and generically, the list of available circumstances will always be incomplete. Biography, interviews, qualitative data may fill in these gaps but will be unfortunately only available for renowned people.

This paper aims to bridge the gap between these different approaches leading to improvements in the computation of inequality associated with circumstances. The measure of inequality of opportunity associated with the list of available circumstances clearly represents a lower bound measure. (Niehues & Peichl 2014) have proposed considering an upper bound approach by modeling individual fixed effects and counting these as circumstances . All luck factors that impact individuals in a specific way along their life (genetic luck, incidental luck - *e.g.* illness, meetings, results of lotteries, etc.) will be captured by the individual fixed effect. However, at the same time, personality traits that would be instead put on the effort side will be placed on the circumstance side. For example, grit, taste for effort, risk-taking, and the will to bounce back are likely to be included in the fixed effect, leading to an overestimation of the role of circumstances. We agree that the only way to tackle the unobservability of important circumstances is to propose a range of possible values for the share of circumstances in the observed outcome inequality. The usefulness of this approach will of course depend on how narrow this range will be. Working on increasing the lower bound and decreasing the upper bound would be critical to improve our knowledge about how circumstances shape inequality.

In this respect, the present paper submits two suggestions to reduce the gap between the upper and lower bound estimates of the share of inequality of opportunity. The two proposals exploit the fact that observations of siblings' outcomes are informative about the individual's circumstances. We import some methods (sibling correlation, for example) from the literature on siblings into inequality of opportunity measurement (see Björklund & Jäntti (2020) about comparing the aims, methods, and results of the two pieces of literature). The intergenerational mobility literature differs from the inequality of opportunity approach in several important ways. One is that the former literature strongly focuses on family characteristics in unequal opportunities while the latter does not limit its attention to the family and considers other channels of transmission of inequality of opportunity: *e.g.* school, city, networks, region, religion, race, gender. In a nutshell, one can say that the equality of opportunity literature has an encompassing perspective, whereas the intergenerational mobility approach provides an in-depth account of family influences.

The main novelty of the paper is to introduce siblings' outcomes as a circumstance in order to increase lower bound estimates of IOP. The rationale is simple: the success of your sibling may affect your chances of success. Another indirect reason is that siblings' outcome tells us something about the common factors, some in the data, and some not in the data, that you share with your siblings. The more delicate question is to know whether these common elements are always circumstances: As an example, the beliefs and values transmitted by parents brought. We propose to go a step further in saying that the common effort among all siblings is also a circumstance.

Here we resort to a normative statement that is a value judgment. We will not advocate this value judgment forcefully because our piece is not philosophical. As usual in normative economics, we make explicit this value judgment and explore where it will lead us about methodology and results. However, as a rationale behind this value judgment, it may be presented as a companion of the Roemerian value judgment that anything correlated to a circumstance is a circumstance. For Roemer (1998), the individual is held responsible only for the effect of her effort net of the impact of individual circumstances. This idea has been developed without hinging on the specific role of the family. Below we propose to refine this notion by taking into account that family is the crucial locus of making opportunities or missing opportunities. We add a new idea more or less implicit to the siblings' correlation literature but not formally explicit. Any shared characteristic among siblings is a circumstance because they cannot be held responsible for something jointly shared with others within the family environment. If it is jointly shared, it is inherited or determined by some other exogenous forces in one way or another. You might only be held responsible for something unique to you. It is a necessary requirement, not a sufficient condition. For example, your height is typical, but it is a circumstance. We apply this idea to what is called effort, which may gather other dimensions such as aspiration, ambition, beliefs, and values. We state that common characteristic and then joint effort at the family level among siblings is inherited through some channel, maybe nature through common genes –because of recombination, siblings share about 50 percent of the same DNA, on average–, but clearly a lot of nurture through education and transmission of values, mimicking or antagonizing parents.

Family is a micro-society. To a much lower extent, this normative statement can be further adapted to other communities of people who share time and resources and are engaged in active and daily relationships. One can think of schools, parishes, cities, or even small regions where people are tied through some specific culture. Here we bluntly explore the municipality link for people born in the same municipality can share some comment experience. We explore this second suggestion's implications to propose a tighter upper bound than Niehues & Peichl (2014)'s ones.

We implement our methodology on one of the champions, if not the world's equal opportunity champion, Sweden. In doing so, we don't make it easy because the gains about reducing the gap between the bounds are likely to be quite low. On the other side, we benefit from an impressive administrative data set drawn from Swedish registers, including 35% of the population born in Sweden in 1941-60 (see Björklund et al. (2009, 2012), for a description of the dataset).

Our methodology is in two steps. It is governed by the fact that we mainly focus on inequality of opportunity in permanent income. The first step aims to get a reliable measure of permanent income. The second step regresses the permanent income on a bunch of circumstances accounting for sibling and municipality at birth outcomes for reasons explained above.

First, we disentangle the share of the variance of current individual income due to their place of birth (at the municipality level), their family circumstances, measured as their siblings' mean income, and their time-invariant characteristics, and a transitory component. Hence, we can decompose the fixed effect introduced by (Niehues & Peichl 2014) into three components: geographic, family, and individual. We show that our family component relates to the sibling's correlation.

We find that half of the variance of current income is due to the time component. 11% of the permanent income variance is due to family circumstances, and this share goes up to 16% when we constrained the sample to brothers. Constraining the sample to native Swedes, we do not find that the municipal component explains variance's sizeable share.

In the second stage, we show that siblings' unobserved shared factors can be accounted for in the estimation to construct a reliable minimum bound for IOP measures of permanent income. Using OLS, we first predict the individual's permanent income from a municipality fixed effect and their siblings' income. Second, we estimate the share of income inequality due to this predicted income using the standard measures of inequality: the Gini, the Theil index, the MLD, and the squared coefficient of variation. We add to this specification observed circumstances at the family level, such as parents' permanent income and education, family size, and circumstances at the individual level, such as gender and birth order. We also include IQ and non-cognitive skills - available only for men - as measures of individual talent and can be qualified as socio-genetic variables.

In another estimation, we estimate that a polynomial of siblings' income explains 25% of the inequality in permanent income measured by the Gini index and the municipality at birth. In contrast, this share goes up to 30% when observed circumstances are also included. According to our findings, the previous estimates of inequality of opportunity in Sweden were rather underestimated. For instance, in the men sample, (Björklund et al. 2012) find that about one-third of the Gini is explained by circumstances, against 46% in our study, of which two-thirds for social circumstances alone and the remaining third for inequalities related to socio-emotional skills.

The outline of the paper is as follows. Section 2 describes the family's model we have in mind by making explicit our ethical assumptions. The data-set is presented in section 3. We go on by presenting our estimation strategy and how one can derive lower and upper bounds for the share of equality of opportunity in section 4. Results are further presented and discussed before in section 5, and section 6 gathers some concluding comments.

2 Model

Our objective is to offer a conceptual model to account for inequality of opportunity in the determination of individual income, in a context where individual outcomes are partly determined by shared family influences.

In accordance with the equality of opportunity literature, we assume that individual income y is determined by two main factors : effort, denoted e, and circumstances, denoted $circ$.

Individuals, indexed by i are nested in families indexed by j.[1] Let N_j denote the number of siblings in family j. Let N denote the individual population size and m the total number of families with $N = \Sigma_{j=1}^{m} N_j$, where $N_j = \{i = 1, .., n_j\}$

We extend the Solon (1999) decomposition to the partitioning of income determinants into effort and circumstances. Although each factor is individual-specific, they are partly determined by the family environment and shared across siblings. Thus we assume that both circumstances and effort can be written as the sum of a family component and an individual deviation from the family effect.

We note e_{ji} the effort of individual i in family j. We have in mind that at the theoretical (not only statistical) level the following relation is true:

$$e_{ji} = h_j + \epsilon_i \quad \text{for all } i \in N_j \tag{Eq . 1}$$

where h_j is the inherited effort common to all siblings in family j and ϵ_i is the idiosyncratic effort. $cov(h_j, \epsilon_i) = 0$ and ϵ_i is distributed across families with the same law with mean 0 within the family and finite variance σ_ϵ^2

Similarly, circumstances can be decomposed into a shared family effect and an individual idiosyncratic deviation :

$$circ_{ji} = circ_j + circ_i \quad \text{for all } i, j \tag{Eq . 2}$$

[1]In the rest of the analysis, we further consider that individuals are nested in family f and in municipality m. Here, to simplify, we note j the family f leaving in a municipality m.

where $circ_j$ is the inherited circumstances common to all siblings (e.g. parents' or neighborhood characteristics) in family j and $circ_i$ is the idiosyncratic circumstances (e.g. gender, birth order). $cov(circ_j, circ_i) = 0$ and $circ_i$ is iid across families according to the same law with 0 mean within the family and finite variance $\sigma^2_{circ_i}$. It means that the $circ_i$ should be expressed in deviation with the mean circumstance of the siblings of the family. For instance the deviation between birth's order of ego and the mean order of siblings in the family.

The above framework allows to derive new insights for the empirical assessment of inequality of opportunity, when circumstances and effort are imperfectly observed. Although the above decompositions are fairly general, Equality of Opportunity theory allow to place additional restrictions on the terms that appear in the above equations. In this paper, we make the following two claims.

Claim 1 *Normative statement: All characteristics shared in a family are circumstances. In particular, h_j is a circumstance and is included under $circ_j$. Shared luck will also be a circumstance.*

Claim 2 *Roemerian statement: the correlation between ϵ_i and $circ_i$ or $circ_j$ is a circumstance.*

Now consider a third determinant of individual outcome, luck, denoted η_i, and assumed to be iid across individuals, uncorrelated to all other variables (individual and mean effort, individual and mean circumstance), with mean 0 and finite variance. η_i is really a noise when $circ_i$ and ϵ_i are computed as deviations from their within-means.

Taking together the different components of income and their decomposition into family and indvidual effects, and using the above two claims, we can now write individual income y_{ji} as :

$$y_{ji} = \alpha + circ_j + circ_i + \epsilon_i + \eta_i \text{ for all } i, j \qquad \text{(Eq . 3)}$$

The issue is to find an econometric strategy which makes possible to retrieve the outcome's part related to circumstances as defined in the two claims.

3 Data and Estimation Sample

Following Björklund & Jäntti (2012); Björklund et al. (2009), we use a rich data set merging several Swedish data sources. We use multi-generational Swedish registers, including approximately 35% of population born in Sweden in 1941-60. This enables us to identify biological siblings. We merge this data set with statistics Sweden's income register for 1968 to 2007, providing the total income from all sources of income (work, self employment, capital, real estate ; as well as some transfers from 1974 onward). Finally, we use the Swedish Military Enlistment Battery, which provides a measure of intellectual capacity, along with socio-emotional skills at age 18-19, essentially for men. The IQ Test consists of four different parts (synonyms, inductions, metal folding and technical comprehension). Socio-emotional skills measures characteristics such as responsibility, independence, persistence, emotional stability and social skills ; they have the advantage to be measured from a psychologist through a 25-minute interview, and therefore may be less subject to measurement errors than measures assessed by the mother, commonly used in the literature. Both assessments are graded on a Stanine scale from 1 to 9, and standardized for a mean of five. These measures of abilities have been shown to have strong labor market returns (Lindqvist & Vestman 2011).

Our sample includes all cohorts born between 1941 and 1960, and all observations years from 1965 to 2007. We restrict our analysis to families with at least two siblings. To get an accurate measure of income, we restrict the sample to individuals from 30 to 45 years-old, for whom we observe a non negative income for at least 10 years. Our sample includes 10 million of observations, from 640 000 individuals (observed on average 15 times), they come from about 265 000 families, born in around 1000 municipalities. For family circumstances, we observe parents' income and their education, and their age at birth, along with family size. Summary statistics are reported in Table 1. The log of parents' income has been cleaned from age, year and gender effects, table 1 reports the residual of this regression.

Table 1: Summary Statistics of Main Variables

Panel A: At the family level

	All					Men					Women				
	mean	sd	min	max	count	mean	sd	min	max	count	mean	sd	min	max	count
Share of Girls in the Sibship	0.49	0.32	0	1	264886	0.29	0.24	0	1	8746	0.70	0.25	0	1	8413
Family Size	2.69	1.00	2	15	264886	2.69	1.00	2	11	8746	2.71	0.99	2	12	8413
Log of Father's Income	0.07	0.75	-8	4	264886	0.07	0.77	-7	3	8746	0.05	0.76	-7	3	8413
Log of Mother's Income	-0.04	0.82	-7	3	264886	-0.05	0.87	-6	2	8746	-0.05	0.82	-6	2	8413
Father's education	9.15	2.94	7	20	264886	9.14	2.92	7	20	8746	9.11	2.93	7	20	8413
Mother's education	8.72	2.44	7	20	264886	8.75	2.48	7	20	8746	8.69	2.41	7	20	8413
Observations	264886					8746					8413				

Panel B: At the individual level

	All					Men					Women				
	mean	sd	min	max	count	mean	sd	min	max	count	mean	sd	min	max	count
Age	37.53	4.59	30	45	10022390	37.51	4.59	30	45	327388	37.58	4.58	30	45	313553
Woman	0.49	0.50	0	1	640941	0.00	0.00	0	0	327388	1.00	0.00	1	1	313553
Year of Birth	1951.55	5.08	1941	1960	640941	1951.50	5.09	1941	1960	327388	1951.60	5.06	1941	1960	313553
Birth Order = 1	0.38	0.49	0	1	640941	0.38	0.49	0	1	327388	0.38	0.49	0	1	313553
Birth Order = 2	0.38	0.49	0	1	640941	0.38	0.49	0	1	327388	0.38	0.49	0	1	313553
Birth Order = 3	0.16	0.36	0	1	640941	0.16	0.36	0	1	327388	0.16	0.36	0	1	313553
Birth Order = 4	0.05	0.23	0	1	640941	0.05	0.22	0	1	327388	0.05	0.23	0	1	313553
Birth order > 5	0.03	0.17	0	1	640941	0.03	0.17	0	1	327388	0.03	0.17	0	1	313553
General ability	5.11	2.06	0	9	163791	5.11	2.06	0	9	163779	0.00	0.00	0	0	12
Non-cognitive skills	4.98	1.98	0	9	163791	4.98	1.98	0	9	163779	0.00	0.00	0	0	12
Brother's General Ability	4.98	2.02	0	9	256213	4.97	2.03	0	9	130966	4.99	2.02	0	9	125247
Brother's Non Cognitive Skills	4.88	1.92	0	9	256212	4.88	1.92	0	9	130963	4.88	1.91	0	9	125249
Mother's age at birth	27.22	5.16	14	48	640941	27.21	5.15	15	48	327388	27.23	5.17	14	48	313553
Father's age at birth	30.21	5.34	14	49	640941	30.19	5.33	14	49	327388	30.22	5.36	14	49	313553
Observations	10022390					327388					313553				

Panel C: Income Variables, at the individual level

	All					Men					Women				
	mean	sd	min	max	count	mean	sd	min	max	count	mean	sd	min	max	count
Taxable Income	215515.77	132996.02	100	34570973	10022390	258677.98	153856.33	102	34570973	5136778	170134.62	85896.98	100	12156990	4885612
Taxable Permanent Income	213919.04	105357.78	354	10330617	640941	257236.47	117334.76	354	10330617	327388	168690.30	65655.04	467	3588810	313553
Participation	15.64	1.01	10	16	640941	15.69	0.95	10	16	327388	15.58	1.07	10	16	313553
Observations	10022390					5136778					4885612				

Notes: This table reports the summary statistics for variables of interest in our sample. The log of parents' income has been cleaned from age, year and gender effects, this table reports the residual of this regression.

Source: All individuals for whom we observe at least 10 times their income after the age of 30, we keep only the observations from the age of 30 and for which income is strictly positive. To compute the family effect, we restrict the sample to families for which at least two individuals respect the sample criteria.

4 Estimation

4.1 Framework

Rather than observing a perfect measure of long-run income, we observe annual income. We take the logarithm of income $\tilde{y}_{mfit}$ and clean it from age, years, and gender effects, along with interacted effect of those variables (U_{mfit}).

$$\tilde{y}_{mfit} = U_{mfit}\beta + y_{mfit} \qquad \text{(Eq . 4)}$$

We propose to breakdown the individual's income into four components: a geographical one, associated to the individual's place of birth at the municipality level a_m, a family component b_{mf}, an individual component c_{mfi} and a transitory component ε_{mfit}.

$$y_{mfit} = a_m + b_{mf} + c_{mfi} + \varepsilon_{mfit} \qquad \text{(Eq . 5)}$$

a_m is a permanent component, common to all individuals born in municipality m, b_{mf} is a permanent component, common to all individuals born in family f, which captures the deviation from the municipality, c_{mfi} is a permanent component unique to individual i, which captures deviations from the family average and the municipality average. ε_{mfit} is a transitory component capturing deviations from the permanent income, for a given age and a given year.

4.2 Decomposing the Variance

Our four components are independent from each other, hence the variance of income can be decomposed into the variance of the four components:

$$\sigma_y^2 = \sigma_a^2 + \sigma_b^2 + \sigma_c^2 + \sigma_\varepsilon^2 \qquad \text{(Eq . 6)}$$

The share of income variance that can be attributed to municipality at birth is $\frac{\sigma_a^2}{\sigma_y^2}$, and can be thought as a measure of the importance of community, neighborhood, amenities such as school or transports effects. The share of income variance that can be attributed to family, net from municipality effects is $\frac{\sigma_b^2}{\sigma_y^2}$, and can be thought as a measure of the importance of all variables, observed or not, that are shared by siblings such as parents' characteristics, it also catches siblings spillovers. Let's note that the common measure of siblings' correlation ρ can be computed from the sum of these two first components when siblings are born in the same municipality

$$\rho = \frac{\sigma_a^2 + \sigma_b^2}{\sigma_y^2}$$

The share of income variance attributed to the third component c_{mfi}, $\frac{\sigma_c^2}{\sigma_y^2}$, catches all circumstances that are not shared by siblings, e.g. genetic endowment, birth order[2], it also accounts the family resources that may not be allocated equally across siblings, depending on their birth order (Price 2008), or their gender (Blau et al. 2020). It also accounts for heterogeneity in the effects of circumstances across siblings, circumstances shared by siblings but that affect them differently, because of differences in sensitivity of siblings to family's characteristics or shocks. Literature has shown that family shocks may have different effects due to the existence of sensitive period (Cunha et al. 2010), or due to differences in gender (Autor et al. 2019; Briole et al. 2020; Chetty et al. 2016). On the other hand, it also catches income variance that is due to the individual's effort which does not vary over time. Hence, including this component into a measure of IOP would give an upper bound (Niehues & Peichl 2014), and excluding it from the circumstances would give a lower bound (Björklund & Jäntti 2012; Björklund et al. 2009).[3] Finally, the income variance due to transitory shocks over time are given by $\frac{\sigma_\varepsilon^2}{\sigma_y^2}$[4], and catches for example the effect of luck (see Lefranc et al. 2009).

[2]It could also account for gender, but here, we have cleaned the effect of gender on income.

[3]We are aware that we may catch in this individual component persistence of shocks affecting individuals that persist over time, and are not due to individual's characteristics per se. To deal with that we also estimate the decomposition of the variance, accounting for this persistence using an AR(1) process. Results are available in Appendix A1. Not accounting for the persistence of the transitory shock means that we over-estimate the variation due to the individual component (σ_c) and we over-estimate the true variance of permanent income ($\sigma_a + \sigma_b + \sigma_c$), hence we under-estimate the variation of the permanent income due to the municipality and the family background (Björklund et al. 2009).

[4]In the case of the AR(1), this component would also catch the persistence of transitory shocks, see section A1.

We estimate equations Eq . 5 to Eq . 18 using the Restricted Maximum Likelihood (REML) method.

4.3 Estimating IOP measures

In a second stage, we estimate IOP measures. Until now, the literature has only used observed circumstances to compute these measures (Björklund & Jäntti 2020; Björklund et al. 2012; Hederos et al. 2017). We show that unobserved circumstances that are shared by individuals born in the same municipality a_m and by siblings b_{mf} can be accounted for in an estimation, so that a reliable minimum bound for IOP measures can be constructed. Since we are interested in family background, that is assumed to have a fixed effect on individual's income and by circumstances that do not vary over time, we focus on the permanent income.

Using OLS, we first predict the individual's income from a municipality fixed effect, and their siblings' average income. We also include observed circumstances.

$$y_{mfi} = f(a_m) + g(b_{mf}) + \gamma_1 W_{mf} + \gamma_2 X_{mfi} + \nu_{mfi} \tag{Eq . 7}$$

where $a_m = \bar{y}_r$ is the mean income of the individuals born in the same municipality, and $b_{mf} = \bar{y}_f$ is the mean income of the individuals' siblings. For $f(.)$ and $g(.)$, we take polynomial of order four[5]. W_{mf} accounts for all observed circumstances at the family level such as parent's permanent income and education, family size, and X_{mfi} accounts for circumstances at the individual level such as gender, birth order and parents' age at birth. We also include IQ and socio-emotional skills for men. Until now, the literature has only included the observed circumstances W_{mf} and X_{mfi} (Björklund & Jäntti 2020; Björklund et al. 2012; Hederos et al. 2017). Including functions of circumstances that are shared by individuals born in a same municipality or in a same family enables us to construct a reliable minimum bound for IOP measures.

[5]Taking a set of dummies indicating centiles give similar results.

We can then predict $\hat{y}_{mfi}$, and compute the share of the inequality due to these circumstances, taking the exponential of $\hat{y}_{mfi}$, and computing $\frac{I(\hat{Y})}{I(Y)}$, where Y denotes the exponential of y, e.g. the level of income. We use indices commonly used in the literature (Björklund & Jäntti 2020; Björklund et al. 2012), namely the Gini, the Theil index (GE(1)), the mean log deviation (MLD = GE(0)) and the squared coefficient of variation ($CV^2 = 2 \times GE(2)$). For each regression, we also report the R^2.

To interpret the R^2, let's note W_{mf} the part of $a_m + b_{mf}$ we observe in our data and z_{mf} the part of b_{mf} we do not observe. We have:

$$a_m + b_{mf} = \delta W_{mf} + z_{mf} \qquad \text{(Eq . 8)}$$

Hence,

$$\frac{\sigma_a^2 + \sigma_b^2}{\sigma_y^2} = \delta^2 \frac{\sigma_W^2}{\sigma_y^2} + \frac{\sigma_z^2}{\sigma_y^2}$$

$$\Leftrightarrow \rho = R^2_{y/W} + \frac{\sigma_z^2}{\sigma_y^2} \qquad \text{(Eq . 9)}$$

where $R^2_{y/W}$ is the R^2 of the regression of the income y on the family circumstances we observe W_{mf}. Therefore, the R^2 of the regression of the income on the observed circumstances should not be compared to the R^2 of the regression of the income y on a_m and b_{mf} but to the siblings correlation, deduced from the decomposition of the variance (see Solon 1999).

4.4 Lower Bound and Upper Bound

As announced in the introduction, our paper is at the crossroad of different research streams and incorporate various tools coming from different horizons. This subsection clarifies the relationship between our approach and previous ones about what ingredients to include in IOp estimation. We then deepen the comparison with what Niehues & Peichl (2014) did, using their terminology of lower and upper bounds of IOp.

The discussion hinges on the following Table 2 that merges two streams of research. The EOp literature focuses on the distinction between effort and circumstances, which figure in columns. The sibling and, more generally, the intergenerational mobility literature spots on what is shared within the family. A more pedestrian distinction is about what is observable and not. This distinction is meaningful for both circumstances and effort. Still, practically, effort variables are seldom observed, a case in point since our dataset does not have any effort information. It is also useful to distinguish whether effort is time-invariant or not. This explains that we have only materialized six cells instead of eight in Roman numerals.

Table 2: Summary of the literature and what it picks up

	CIRCUMSTANCES		EFFORT
	Observed	Non Observed	Non Observed
Shared	I	II	V Any effort common to siblings is due to circumstances
Non Shared	III	IV	VI time-invariant effort; time-variant effort

The variance analysis on which the intergenerational literature relies is investing the first line of this table, and siblings and twin correlation represent an attempt to capture cells I and II. Plus a bit of cell V, shared effort among siblings, but maybe like Mr. Jourdain, without knowing it.

Conversely, the IOP literature is spotting on case III, which is observed and non-shared circumstances. We will quote the well-crafted study of IOp on observable variables (I + III) by Björklund et al. (2012) for Sweden by authors who have also contributed to the siblings and twin correlation (Björklund & Jäntti 2012).

In a nutshell, our proposition can be described as adding up the IOp measures for these two veins and considering I+III+II+V as a recommended benchmark for EOp. On the one hand, this measure can be viewed as an improved lower bound viz the admitted measure I+III. On the other hand, starting from the upper bound proposal by Niehues & Peichl (2014) with their individual fixed effect in panel analysis which captures all time-invariant circumstances, I+II+III+IV, plus all time-invariant effort variables, that is, V and VI, our approach can be also be qualified, maybe misleadingly, of relaxing upper bound. In the remaining subsection, we compare our methodology with that of Niehues & Peichl (2014) regarding lower and upper bound in our case when effort variables are not observed. We first present their method for current incomes, then go on by presenting our methodology for permanent incomes and then show how we can derive lower and upper bounds for current incomes.

4.4.1 Lower and upper bounds for current income (Niehues & Peichl 2014)

To simplify the notations, we denote X'_{mfi} the whole set of circumstances we can observe at the family and individual level (hence including both X_{mfi} and W_{mf}).
Let's start by computing the lower bound, starting from the estimation:

$$\log Y_{it} = y_{it} = \delta' X'_{mfi} + u_t + \epsilon_{it}$$

where Y denotes the level of income, and y denotes the logarithm of the income throughout the section. Then, they compute a parametric estimate of the smooth distribution.[6]

$$\widehat{y_{it}} = \hat{\delta}' X'_{mfi}$$

IOp lower bound is then obtained as the ratio of inequality corresponding to the smooth distribution to the inequality of current income distribution (in the direct approach to EOp)

$$\underline{IOp}(Y_{it}) = \frac{I(\widehat{Y_{it}})}{I(Y_{it})} \qquad \text{(Eq . 10)}$$

For the upper bound, Niehues & Peichl (2014) starts with the estimation of the individual fixed-effect:

$$y_{it} = u_i + u_t + \varepsilon_{it}$$

[6]To simplify the notations, we neglect the existence of heteroscedasticity, we should otherwise add a term $\frac{\sigma^2}{2}$, with σ^2 the variance of ϵ_{it}. This additional term lets relative indices of inequality invariant since all smooth incomes are multiplied by a scale factor $e^{\frac{\sigma^2}{2}}$.

which allows them to compute the parametric estimate of the smooth distribution

$$\widehat{u_i} = \bar{y}_i$$

where $\widehat{u_i}$ actually coincides with the estimated permanent income $\bar{y}_i$; the IOp upper bound writes :

$$\overline{IOp}(Y_{it}) = \frac{I(\bar{Y}_i)}{I(Y_{it})} \qquad \text{(Eq . 11)}$$

where $\bar{Y}_i$ is the exponential of $\bar{y}_i$.

4.4.2 Lower and upper bounds for permanent income

For permanent income, one estimates the permanent income. Computing the lower bound, we estimate:

$$\bar{y}_i = \delta X'_{mfi} + \eta_i$$

and then, we estimate a parametric estimate of the smooth distribution :

$$\hat{\bar{y}}_i = \hat{\delta} X'_{mfi}$$

and similarly, the lower bound estimate of IOp for permanent income reads :

$$\underline{IOp}(\bar{Y}_i) = \frac{I(\widehat{\bar{Y}_i})}{I(\bar{Y}_i)} \qquad \text{(Eq . 12)}$$

where $\bar{Y}_i$ is the exponential of $\bar{y}_i$, which is actually the estimated individual fixed-effect ($\hat{u}_i$).

Beginning with our decomposition relation and introducing explicitly time fixed effect u_t to make it easy the comparison with Niehues & Peichl (2014)'s procedure:[7]

[7] In the main analysis, we clean the income from the time fixed effects before estimating IOP.

$$\log y_{it} = a_m + b_{mf} + c_{mfi} + u_t + \varepsilon_{it} \tag{Eq . 13}$$

we proceed to the estimation of

$$\bar{y}_i = \gamma X'_{mfi} + f(a_m) + g(b_{mf}) + \varsigma_i$$

which enables to obtain a parametric estimate of the smooth distribution, denoted $\hat{\hat{\bar{y}}} > \hat{\bar{y}}$.

$$\hat{\hat{\bar{y}}}_i = \hat{\gamma} X'_{mfi} + f(\widehat{a_m}) + g(\widehat{b_{mf}})$$

Our lower bound of inequality of opportunity for permanent incomes is then computed as

$$\underline{IOp}(\bar{Y}_i) = \frac{I(\widehat{\widehat{\bar{Y}_i}})}{I(\bar{Y}_i)} \tag{Eq . 14}$$

4.4.3 Implication for the estimation of lower and upper bound of current income

The results are simpler to expose for the upper bound since the upper bound result for the current and permanent income are nested. For the lower bound they are not and a new estimation is needed.

4.4.3.1 New upper bound for current income

Using (Eq . 11) and (Eq . 14) and for relative indices of inequality we get:

$$\begin{aligned}\overline{\overline{IOp}}(Y_{it}) &= \frac{I(\hat{\hat{\hat{Y}}})}{I(Y_{it})} \\ &= \frac{I(\bar{Y}_i)}{I(Y_{it})}\frac{I(\hat{\hat{\hat{\bar{Y}}}})}{I(\bar{Y}_i)} \\ &= \overline{IOP}(Y_{it})\frac{I(\hat{\hat{\hat{Y}}})}{I(\bar{Y}_i)} \qquad \text{(Eq . 15)}\end{aligned}$$

where $\hat{\hat{\hat{Y}}}$ is the estimated permanent income from circumstances including the municipality (a_m) and the family effects (b_f), and $\bar{Y}_i$ is the permanent income (average income).
Since $\frac{I(\hat{\hat{\hat{Y}}})}{I(\bar{Y}_i)} < 1$, we can write:

$$\overline{\overline{IOp}}(Y_{it}) < \overline{IOP}(Y_{it})$$

Hence, $\overline{\overline{IOp}}(y_{it})$ is a lower upper bound than that computed by Niehues & Peichl (2014), but omits circumstances at the individual level that does not vary over time.

4.4.3.2 New lower bound for current income

We have to proceed to a further estimation of the following regression

$$y_{it} = \gamma' X'_{mfi} + f(a_m) + g(b_{mf}) + u_t + \iota_{it} \qquad \text{(Eq . 16)}$$

from which we derive a parametric estimate of the smooth distribution

$$\widehat{\widehat{y_{it}}} = \widehat{\gamma}' X'_{mfi} + f'\widehat{(a_m)} + g'\widehat{(b_{mf})}$$

which differs from $\widehat{y_{it}} = \hat{\gamma} X'_{mfi}$ by integrating the municipality and family effects. The new lower bound is computed as

$$\underline{\underline{IOp}}(Y_{it}) = \frac{I(\widehat{\widehat{Y_i}})}{I(Y_{it})} > \underline{IOp}(Y_{it}) \qquad \text{(Eq . 17)}$$

and then an upper lower bound than that proposed by Niehues & Peichl (2014).

5 Estimation Results

5.1 Decomposition of the variance

Table 3: Variance decomposition

	All Sample	Men	Women
var(Municipality)	0.00316	0.00633	0.00278
	(0.000189)	(0.000397)	(0.000193)
var(Family)	0.0333	0.0612	0.0361
	(0.000548)	(0.00120)	(0.000936)
var(Individual)	0.248	0.258	0.204
	(0.000677)	(0.00127)	(0.00103)
var(Residual)	0.276	0.248	0.305
	(0.000129)	(0.000162)	(0.000204)
N	9800751	5023373	4777378

Notes: This Table shows the decomposition of the variance, explained in section 4.2.
Source: All individuals for whom we observe at least 15 times their income after the age of 30, we keep only the observations from the age of 30 and for which income is strictly positive. To compute the family effect, we restrict the sample to families for which at least two individuals respect the sample criteria.

We first decompose the variance of individual's current income into four components: a geographical one, associated to the individual's place of birth at the municipality level (a_m), a family component (b_{mf}), an individual component (c_{mfi}) and a transitory component (ε_{mfit}) for the whole sample. The total variance of the income for the whole sample is around 0.541, and is similar across genders. Nearly half of the variance in current income can be attributed to the transitory component. The other half, which corresponds to the variance of the permanent income, is largely explained by the individual component. 5,5% of the variance of the current income is due to the family component, and the variance attributed to the geographical component is close to zero. This is line with previous research, Lindahl (2011) finds that neighborhood correlation is very small in Sweden. Looking at the permanent income, the share explained by the family component is around 11%.[8]

Because individuals may be affected by municipality characteristics or by family characteristics differently according to gender, we split the sample according to gender, and estimate the share of the variance due to municipality or to family for individuals of the same gender. For instance, the family component now accounts for variables that are shared by brothers and by sisters, respectively. The effect of municipality seems to be larger for men, this is line with previous research showing that boys are more sensitive to neighborhood quality (Autor et al. 2019). Brother's correlation is also larger that sister's correlation, it is about 21% against 16% for women, consistent with Björklund & Jäntti (2012)'s findings. This is consistent with the literature showing that men are more sensitive to family background (Autor et al. 2019; Chetty et al. 2016). The share of the variance of income due to the individual-specific component is more important for men than for women, while the variance due to the transitory component is more important for women, this indicates that women displays more variability in their life-cycle income profile. This is explained by parental leave and income shocks associated to child birth (Böhlmark & Lindquist 2006).

In the rest of the analysis we focus on permanent income.

[8]The decomposition of the variance between the individual component and the transitory component relies on the assumption of the persistence of shocks. When we use an AR(1), we show that the persistence of shock is large, and the variance attributed to the individual component is over-estimated when we do not account for the persistence of shocks across periods (see section A1). Variation due to the family background and to municipality are however not affected. Nevertheless, since assuming the absence of auto-correlation leads to over-estimate the individual-specific component, it under-estimate the relative importance of the variance of permanent income due to family background and to municipality, because we erroneously includes the persistence in the transitory variation in the variance of the permanent income (Björklund et al. 2007, 2009).

5.2 Inequality of Opportunity Indicators

As stated by Björklund & Jäntti (2020), one limit of the approach of the decomposition of the variance is that we are not able to explicitly know from which variables it comes from. Using IOP measures, we are able to know how much each observed circumstance accounts for inequality of opportunity. Besides, using IOP measures enables us to raise the lower bound of siblings correlation i.e. the share of the variance due to variables shared by siblings, by including variables that are not shared by siblings (birth order, age of parents at birth, IQ, socio-emotional skills). Finally, IOP measures are more flexible with respect to the measure of inequality (Jenkins 1991). Until now, researchers have only used observed circumstances, we show that unobserved circumstances that are shared by individuals born in the same municipality (a_m) and by siblings (b_{mf}) can be accounted for in estimation, so that a reliable minimum bound for IOP measures can be constructed.

We first present the results for the whole sample. Next, we use a sub-sample of men, for whom we observe their IQ and their socio-emotional skills, two important factors of labour market outcomes (Lindqvist & Vestman 2011).

Tables A.3 and 5 report the Gini index of the estimated income (in level) due to circumstances $\hat{Y}$, the second column shows the ratio between this Gini and the Gini index computed on the observed income Y, $\frac{I(\hat{Y})}{I(Y)}$ (see section 4.3). The six next columns show the same estimations for the GE(0), GE(1) and CV2 indices, respectively. The last column shows the R^2 of the regression. We first show the IOP measures on the observed circumstances, added one by one, and then the IOP measures based on our family and municipality components.

5.2.1 Main results

Table A.3 shows the results for the main sample. We see that 16% of the Gini index is attributed to the log of father's income, this share goes up to 21% when we take a quadratic term of the log of father's income. Adding the log of mother's income and its quadratic term raises the share of the Gini index attributed to family circusmtances to 23,2%. Adding the family size, this share increases by one point of percentage. Adding the parents' education, the share of Gini index explained by family circumstances is around 25%. Adding circumstances at the individual level, namely birth order and age of parents at birth do not increase much the share of the Gini explained by circumstances. This is similar to Hederos et al. (2017)'s findings who find that 22,9% of the Gini is attributed to family and individual circumstances when the sample includes both men and women.

Turning to other indices - GE(0), GE(1) and CV2, the share of inequality is poorly explained by circumstances, only 6% of those indices are explained by observed circumstances. This is lower than what Hederos et al. (2017) find when using types and a shapley decomposition, they find that less than 14% is due to circumstances where other indices than the Gini index are considered.

Looking at the R^2, the observed family circumstances account for less than 4% of the variance of permanent income, this is very low compared to the 11% of the variance income due to family circumstances, found in section 5.1. It indicates that a large share of inequalities due to variables common to siblings

Table 4: IOP indicators $y_{mfi.}$, permanent income

	Gini	Gini(share)	GE0	GE0(share)	GE1	GE1(share)	CV2	CV2(share)	R2
log income father	0.037	0.169	0.003	0.029	0.003	0.033	0.006	0.031	0.023
+log father income (sqrt)	0.047	0.213	0.004	0.039	0.004	0.047	0.008	0.046	0.029
+ log mother income	0.050	0.228	0.005	0.043	0.005	0.052	0.009	0.051	0.033
+log income mother (sqrt)	0.052	0.232	0.005	0.044	0.005	0.054	0.010	0.053	0.034
+ family size	0.053	0.241	0.005	0.047	0.005	0.056	0.010	0.055	0.035
+ education father	0.055	0.250	0.005	0.049	0.005	0.059	0.011	0.058	0.037
+ education mother	0.055	0.250	0.005	0.049	0.005	0.060	0.011	0.059	0.037
+ birth order	0.056	0.252	0.005	0.049	0.005	0.060	0.011	0.059	0.037
+ age of parents at birth	0.056	0.254	0.005	0.050	0.005	0.061	0.011	0.060	0.038

	Gini	Gini(share)	GE0	GE0(share)	GE1	GE1(share)	CV2	CV2(share)	R2
Siblings' Income	0.034	0.153	0.002	0.022	0.002	0.026	0.004	0.024	0.018
+ Siblings' Income (sqrt)	0.043	0.193	0.003	0.029	0.003	0.035	0.006	0.035	0.022
+ Siblings' Income (3rd order)	0.044	0.197	0.003	0.031	0.003	0.039	0.007	0.039	0.023
+ Siblings' Income (4th order)	0.044	0.197	0.003	0.031	0.003	0.038	0.007	0.039	0.023
+ Municipality Effect	0.055	0.248	0.005	0.047	0.005	0.058	0.011	0.057	0.035
+ Municipality Effect (sqrt)	0.055	0.249	0.005	0.047	0.005	0.058	0.011	0.058	0.035
+ Municipality Effect (3rd order)	0.055	0.249	0.005	0.047	0.005	0.058	0.011	0.058	0.035
+ Municipality Effect (4th order)	0.055	0.249	0.005	0.047	0.005	0.058	0.011	0.058	0.035
+ family circumstances	0.067	0.304	0.008	0.070	0.008	0.086	0.016	0.085	0.053
+ individual circumstances	0.068	0.306	0.008	0.071	0.008	0.087	0.016	0.086	0.053
Siblings' Income + Municipality Effect	0.038	0.169	0.003	0.027	0.003	0.031	0.005	0.029	0.021

Notes: This Table shows Gini for long run income and estimated long run income $\hat{y}$, which has been estimated using a log model.

Source: All individuals for whom we observe at least 15 times their income after the age of 30, we keep only the observations from the age of 30 and for which income is strictly positive. To compute the family effect, we restrict the sample to families for which at least two individuals respect the sample criteria.

is explained by circumstances that are not observed (noted z_{mf} in section 4.3).

Accounting for our components a_m and b_{mf} in our prediction of income due to circumstances $\hat{y}$, siblings' income (b_{mf}) - net from municipality effect, as a continuous variables explains already 15,3% of the Gini index, and up to 19,3% when we include a quadratic term, when we include the average income at the level of the municipality of birth, this share goes up to 24,8%[9]. Since those variables are proxied for family and municipality fixed effects, we still miss a part of family circumstances[10] ; accounting for observed family circumstances, the share of the Gini explained by family and municipality circumstances reaches 30,4%. Accounting for birth order and age of parents at birth do not raise much this share. Comparing these estimates to the one found when only observed circumstances are controlled for (Panel A) indicates that accounting for siblings' income and municipality of birth fixed effects allow us to raise the lower bound of IOP by 5 percentage points where the Gini index is concerned. Using types and a Shapley decomposition, Hederos et al. (2017) find that 32,2% of income inequality is attributed to circumstances where brother's IQ and NCS are accounted for.

Turning to other indices of inequality, the share of other indices explained by circumstances are again smaller and does not exceed 6% when family and municipality effects are accounted for, and less than 9% when we add observed circumstances. Using types and a Shapley decomposition, this share is around 25% for other indices in Hederos et al. (2017). Using the mean-log deviation (GE(0)), our lower bound estimate is around 7%. As a matter of comparison, and considering the Roemerian definition of effort[11], Niehues & Peichl (2014) estimate the upper bound of inequality of opportunity to be around 60% for both men and women, in the US and Germany.

Including the siblings' income in our estimation, the R^2 is harder to interpret, see section A3 for further details.

[9]Note that we find similar results when we use centiles of these two components.

[10]Accounting for those variables as a fixed-effect model would lead to too many coefficients to estimate (one coefficient per family).

[11]According to the Roemerian approach (Roemer 1998), any indirect effect of circumstances on effort should be considered as a circumstance, while according to Fleurbaey's approach (Fleurbaey 1995; Fleurbaey et al. 2008), the effort affected by circumstances should be considered as effort

5.2.2 Results for Men

Table 5: IOP indicators $y_{mfi.}$, permanent income - Men

	Gini	Gini(share)	GE0	GE0(share)	GE1	GE1(share)	CV2	CV2(share)	R2
log income father	0.047	0.211	0.005	0.045	0.005	0.052	0.009	0.046	0.036
+log father income (sqrt)	0.067	0.300	0.008	0.070	0.008	0.089	0.018	0.089	0.048
+ log mother income	0.069	0.308	0.008	0.073	0.009	0.093	0.018	0.092	0.051
+log income mother (sqrt)	0.069	0.311	0.009	0.074	0.009	0.094	0.018	0.093	0.051
+ family size	0.072	0.320	0.009	0.076	0.009	0.097	0.019	0.095	0.054
+ education father	0.072	0.321	0.009	0.076	0.009	0.097	0.019	0.095	0.054
+ education mother	0.072	0.321	0.009	0.076	0.009	0.097	0.019	0.095	0.054
+ birth order	0.072	0.324	0.009	0.077	0.009	0.098	0.019	0.096	0.054
+ age of parents at birth	0.073	0.326	0.009	0.078	0.009	0.099	0.019	0.097	0.055
+ IQ	0.091	0.406	0.013	0.113	0.013	0.142	0.027	0.138	0.080
+ NCS	0.097	0.435	0.015	0.130	0.015	0.163	0.031	0.158	0.092
+NCS X IQ	0.098	0.439	0.015	0.132	0.016	0.167	0.032	0.162	0.093

	Gini	Gini(share)	GE0	GE0(share)	GE1	GE1(share)	CV2	CV2(share)	R2
Siblings' Income	0.040	0.178	0.003	0.028	0.003	0.033	0.006	0.030	0.021
+ Siblings' Income (sqrt)	0.049	0.220	0.004	0.035	0.004	0.045	0.008	0.043	0.025
+ Siblings' Income (3rd order)	0.050	0.224	0.004	0.038	0.005	0.049	0.010	0.050	0.026
+ Siblings' Income (4th order)	0.050	0.225	0.004	0.038	0.005	0.048	0.009	0.048	0.026
+ Municipality Effect	0.063	0.284	0.007	0.057	0.007	0.073	0.014	0.071	0.040
+ Municipality Effect (sqrt)	0.063	0.283	0.007	0.057	0.007	0.072	0.014	0.070	0.040
+ Municipality Effect (3rd order)	0.063	0.283	0.007	0.057	0.007	0.072	0.014	0.070	0.040
+ Municipality Effect (4th order)	0.063	0.283	0.007	0.057	0.007	0.073	0.014	0.070	0.041
+ all circumstances	0.104	0.467	0.017	0.149	0.018	0.188	0.036	0.183	0.105
Siblings' Income + Municipality Effect	0.044	0.197	0.004	0.034	0.004	0.040	0.007	0.036	0.026

Notes: This Table shows Gini for long run income and estimated long run income $\hat{y}$, which has been estimated using a log model.
Source: All individuals for whom we observe at least 15 times their income after the age of 30, we keep only the observations from the age of 30 and for which income is strictly positive. To compute the family effect, we restrict the sample to families for which at least two individuals respect the sample criteria.

Let's now turn to the sample of men[12]. The share of the Gini index explained by father's income and its quadratic term is around 30%, this is larger for the mens' sample than for the whole sample (21%), this is consistent with larger inter-generational correlation and elasticity between fathers and sons than fathers and daughters (Björklund & Jäntti 2012). Accounting for mother's income and its quadratic term raises the share of the Gini explained by circumstances by one point of percentage, this is the same for family size, and this is similar to results found for the whole sample. Parents' education however does not seem to explain a sizeable share of the Gini index in the sample of men. Overall, observed family and individual circumstances account for 32,6% of the gini index. However, accounting for the IQ and the NCS measured in early adulthood (around age 18) raises this bound by more than

[12] Siblings income and municipality effects are computed on the whole sample and also account for women's income.

10 points of percentage, the share of the Gini index explained by the whole set of circumstances reaches 43,9%. As a matter of comparison, Björklund et al. (2012) define types of individuals according to a set of observed circumstances similar to ours, they estimate the Gini share due to circumstances to be around 28,2% for Swedish men born between 1955 and 1967 ; and according to Björklund & Jäntti (2020) who use a log-model, this share is around 37,6% for Swedish men when they use a log model. Those differences may come from differences in the sample or in the method.

Again, the share of other indices explained by circumstances is smaller and does not go beyond 17%. Björklund et al. (2012) attribute 15% of the mean-log deviation to circumstances, this share is about 20% and 40% for GE(1) and CV^2, respectively. Björklund & Jäntti (2020) find that observed circumstances account for 8% of the CV^2 for Swedish men. Note that Björklund et al. (2012) show that the share of CV^2 is very different across cohorts. The R^2 of the regression including only observed family circumstances is around 5,4% for this sample of men, this is much smaller compared to the share of the variance of permanent income attributed to family background found in section 5.1 which is 19%, indicating that a large of the variance in permanent income attributed to family background is due to unobserved circumstances.

Let's turn to the IOP measures using our municipality and family effects, a_m and b_{mf}. Siblings' income (b_{mf}) - net from municipality effect, as a continuous variable explains already 17,8% of the Gini index, and up to 22% when we include a quadratic term, this is smaller than the share explained by father's income, since siblings' income also include sisters'[13]. This share goes up to 28% when we include the average income at the level of the municipality of birth[14]. Accounting for all circumstances, including abilities, this share reaches 46,3%, which is 3 points larger than when a_m and b_f are omitted.

Turning to other indices, the share explained by circumstances does not go beyond 18,5%. Using the mean log deviation, we find that at least 14,7% of IOP is explained by circumstances, while Niehues & Peichl (2014) estimate estimate the upper bound of inequality of opportunity to be around 60% for

[13]Using brother's income rather than siblings, this share is about 21%, and 26% when the quadratic term is accounted for, and goes up to 27% when using a polynome of fourth order.

[14]Note that we find similar results when we use centiles of these two components.

both men and women (using the Roemerian definition of effort), in the US and Germany.

6 Concluding Discussion

The table 6 summarizes our results about lower and upper bounds of estimation of IOp when comparing Niehues & Peichl (2014)'s methodology and the methodology in this paper. In the table, there are still many empty cells that correspond to pending estimations. The gap introduced by the new communal and family variables between the lower and upper bound for permanent income is about 5% with social circumstances alone but falls to 3% when the talent variables (IQ and socio-emotional skills) are added (focusing on men). This indicates that the levels of the family variable also reflects differences in talent between families. Overall circumstances, including socio-emotional skills, explain a small half of the inequality in permanent income for Sweden, with approximately two-thirds for social circumstances alone and the remaining third for inequalities related to socio-emotional skills. At this stage of the study, the findings do not seem to depend on whether one is studying permanent or current income.

There are several avenues for further research in the framework of this paper. In our first stage, we suppose that the current shock on incomes has no impact other than on the current period. This assumption is unduly too restrictive, and the effect of the shock can be described at least as an AR (1). Integrating this persistence perturbs the estimation of the permanent income, which has to be corrected appropriately. Another dimension that can be explored is whether our results are sensitive to introducing some circumstances heterogeneity among siblings. Birth order and age of parents at birth are regressors, but we can go further (birth spacing, for instance). Intermediate outcomes such as educational achievement are likely to be treated with the same methodology. We also plan to compare Sweden with the US and Germany, two countries studied by Niehues & Peichl (2014). Finally, we admit that the ethical statement about sibling joint effort can be challenged on the ground that it is also the result of sibling interaction and not only parental supervision. It remains to see whether a more fine-tune value judgment will have a significant impact on results.

Table 6: Summary of the results for IOP using the Gini index

Main sample		
	Permanent income (y_{mfi})	Current income (y_{mfit})
Lower Bound (obs. circumstances)	25,4%	23,3%
Lower Bound (obs. circumstances + a_m + b_f)	30,6 %	28%
Men sample (with IQ and socio-emotional skills as circumstances)		
	Permanent income (y_{mfi})	Current income (y_{mfit})
Lower Bound (obs. circumstances)	43,9%	40,5%
Lower Bound (obs. circumstances + a_m + b_f)	46,7%	43,1%

Notes: This table summarizes the share of the Gini explained by circumstances when only observed circumstances are included and when we add the family and the municipality components.

References

Autor, D., Figlio, D., Karbownik, K., Roth, J., & Wasserman, M. (2019). Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes. *American Economic Journal: Applied Economics*, *11*(3), 338–381. doi: 10.1257/app.20170571

Björklund, A., & Jäntti, M. (2012). How important is family background for labor-economic outcomes? *Labour Economics*, *19*(4), 465–474.

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*, 100455.

Björklund, A., Jäntti, M., & Lindquist, M. (2007). Family background and income during the rise of the welfare state: Brother correlations in income for swedish cohorts born 1932-1968.

Björklund, A., Jäntti, M., & Lindquist, M. J. (2009). Family background and income during the rise of the welfare state: brother correlations in income for swedish men born 1932–1968. *Journal of Public Economics*, *93*(5-6), 671–680.

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*(2-3), 675–696.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2011). Older and wiser? birth order and iq of young men. *CESifo Economic Studies*, *57*(1), 103–120.

Blau, F. D., Kahn, L. M., Brummund, P., Cook, J., & Larson-Koester, M. (2020). Is there still son preference in the united states? *Journal of Population Economics*, 1–42.

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), 879–896.

Briole, S., Le Forner, H., & Lepinteur, A. (2020). Children's socio-emotional skills: Is there a quantity–quality trade-off? *Labour Economics*, *64*, 101811.

Chetty, R., Hendren, N., Lin, F., Majerovitz, J., & Scuderi, B. (2016). Childhood Environment and Gender Gaps in Adulthood. *American Economic Review*, *106*(5), 282–288. doi: 10.1257/aer.p20161073

Clark, G. (2015). *The son also rises: Surnames and the history of social mobility.* Princeton University Press.

Cunha, F., Heckman, J. J., & Schennach, S. M. (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, *78*, 883–931.

Fleurbaey, M. (1995). Three solutions for the compensation problem. *Journal of Economic Theory*, *65*(2), 505–521.

Fleurbaey, M., et al. (2008). *Fairness, responsibility, and welfare.* Oxford University Press.

Hederos, K., Jäntti, M., & Lindahl, L. (2017). Gender and inequality of opportunity in sweden. *Social Choice and Welfare*, *49*(3), 605–635.

Jäntti, M., & Jenkins, S. P. (2015). Income mobility. In *Handbook of income distribution* (Vol. 2, pp. 807–935). Elsevier.

Jenkins, S. (1991). The measurement of income inequality. *Economic Inequality and Poverty*, 3–31.

Lefranc, A., Pistolesi, N., & Trannoy, A. (2009, December). Equality of opportunity and luck: Definitions and testable conditions, with an application to income in france (1979-2000). *Journal of Public Economics*, *93*(11-12), 1189-1207.

Lindahl, L. (2011). A comparison of family and neighborhood effects on grades, test scores, educational attainment and income—evidence from sweden. *The Journal of Economic Inequality*, *9*(2), 207–226.

Lindqvist, E., & Vestman, R. (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the swedish enlistment. *American Economic Journal: Applied Economics*, *3*(1), 101–28.

Niehues, J., & Peichl, A. (2014). Upper bounds of inequality of opportunity: theory and evidence for germany and the us. *Social Choice and Welfare*, *43*(1), 73–99.

Price, J. (2008). Parent-child quality time does birth order matter? *Journal of Human Resources*, *43*, 240–265.

Roemer, J. E. (1998). Equality of opportunity.

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

Solon, G. (1999). Intergenerational mobility in the labor market. In *Handbook of labor economics* (Vol. 3, pp. 1761–1800). Elsevier.

7 Appendix

A1 Decomposing the variance: accounting for persistence of shocks.

Following Björklund et al. (2009), we model this transitory component as a first order auto-regressive process

$$\varepsilon_{mfit} = \lambda\ \varepsilon_{mfit-1} + u_{mfit} \qquad \text{(Eq . 18)}$$

where u_{mfit} is a random shock to current income, with white noise properties, assumed to be uncorrelated across individuals born in a same municipality and across siblings. This AR(1) process accounts for persistence in transitory shocks, otherwise, those shocks would be caught in the individual's specific component c_{mfi}, while it is independent from the individual's non time-varying characteristics.

Table A.1: Variance decomposition

	All Sample	Men	Women
var(Municipality)	0.00324	0.00649	0.00290
	(0.000197)	(0.000412)	(0.000207)
var(Family)	0.0360	0.0640	0.0418
	(0.000602)	(0.00127)	(0.00108)
var(Individual)	0.134	0.153	0.0798
	(0.000827)	(0.00142)	(0.00131)
var(Residual)	0.466	0.413	0.519
	(0.000629)	(0.000769)	(0.000998)
lambda	0.768	0.763	0.771
	(0.000321)	(0.000451)	(0.000453)
N	9800751	5023373	4777378

Notes: This Table shows the decomposition of the variance, explained in section 4.2.
Source: All individuals for whom we observe at least 15 times their income after the age of 30, we keep only the observations from the age of 30 and for which income is strictly positive. To compute the family effect, we restrict the sample to families for which at least two individuals respect the sample criteria.

The variance attributed to municipality and family background is similar to the case when we do not consider the persistence of transitory shocks. But the variance attributed to the individual component is 0.10 points smaller, this is slightly more pronounced for women. Indeed, part of the individual component reflects persistence of transitory shock, we find a persistence coefficient (λ) close 0,77 and similar across genders. It turns that the variance due to transitory shocks (residual) is much larger when we account for their persistence, this is again particularly true for women, who experience more transitory shocks associated with their child's birth and parental leave, leading to more variability in their life-cycle income profile (Böhlmark & Lindquist 2006).

A2 Results for current income

A2.1 Main results

Table A.2: IOP indicators y_{mfit}, current income

	Gini	Gini(share)	GE0	GE0(share)	GE1	GE1(share)	CV2	CV2(share)	R2
log income father	0.037	0.155	0.003	0.020	0.003	0.026	0.006	0.022	0.012
+log father income (sqrt)	0.047	0.195	0.004	0.026	0.004	0.037	0.008	0.033	0.015
+ log mother income	0.050	0.209	0.005	0.029	0.005	0.041	0.009	0.036	0.017
+log income mother (sqrt)	0.052	0.213	0.005	0.030	0.005	0.042	0.010	0.037	0.018
+ family size	0.053	0.221	0.005	0.031	0.005	0.044	0.010	0.039	0.019
+ education father	0.055	0.229	0.005	0.033	0.005	0.047	0.011	0.041	0.019
+ education mother	0.055	0.229	0.005	0.033	0.005	0.047	0.011	0.042	0.019
+ birth order	0.056	0.230	0.005	0.033	0.005	0.048	0.011	0.042	0.019
+ age of parents at birth	0.056	0.233	0.005	0.034	0.005	0.048	0.011	0.043	0.020

	Gini	Gini(share)	GE0	GE0(share)	GE1	GE1(share)	CV2	CV2(share)	R2
Siblings' Income	0.034	0.140	0.002	0.015	0.002	0.020	0.004	0.017	0.009
+ Siblings' Income (sqrt)	0.043	0.177	0.003	0.019	0.003	0.028	0.006	0.024	0.011
+ Siblings' Income (3rd order)	0.044	0.180	0.003	0.021	0.003	0.030	0.007	0.028	0.012
+ Siblings' Income (4th order)	0.044	0.180	0.003	0.021	0.003	0.030	0.007	0.027	0.012
+ Municipality Effect	0.055	0.228	0.005	0.032	0.005	0.046	0.011	0.041	0.018
+ Municipality Effect (sqrt)	0.055	0.228	0.005	0.032	0.005	0.046	0.011	0.041	0.018
+ Municipality Effect (3rd order)	0.055	0.228	0.005	0.032	0.005	0.046	0.011	0.041	0.018
+ Municipality Effect (4th order)	0.055	0.228	0.005	0.032	0.005	0.046	0.011	0.041	0.018
+ family circumstances	0.067	0.278	0.008	0.047	0.008	0.068	0.016	0.060	0.027
+ individual circumstances	0.068	0.280	0.008	0.048	0.008	0.069	0.016	0.061	0.028
Siblings' Income + Municipality Effect	0.038	0.155	0.003	0.018	0.003	0.024	0.005	0.020	0.011

Notes: This Table shows Gini for long run income and estimated current income $\hat{y}$, which has been estimated using a log model.

Source: All individuals for whom we observe at least 15 times their income after the age of 30, we keep only the observations from the age of 30 and for which income is strictly positive. To compute the family effect, we restrict the sample to families for which at least two individuals respect the sample criteria.

A2.2 Results for Men

Table A.3: IOP indicators y_{mfit}, current income - Men

	Gini	Gini(share)	GE0	GE0(share)	GE1	GE1(share)	CV2	CV2(share)	R2
log income father	0.047	0.194	0.005	0.032	0.005	0.040	0.009	0.028	0.020
+log father income (sqrt)	0.067	0.276	0.008	0.049	0.008	0.070	0.018	0.054	0.028
+ log mother income	0.069	0.284	0.008	0.052	0.009	0.072	0.018	0.056	0.029
+log income mother (sqrt)	0.069	0.287	0.009	0.052	0.009	0.073	0.018	0.057	0.030
+ family size	0.072	0.296	0.009	0.054	0.009	0.076	0.019	0.059	0.031
+ education father	0.072	0.296	0.009	0.054	0.009	0.076	0.019	0.059	0.031
+ education mother	0.072	0.296	0.009	0.054	0.009	0.076	0.019	0.059	0.031
+ birth order	0.072	0.299	0.009	0.055	0.009	0.077	0.019	0.059	0.031
+ age of parents at birth	0.073	0.301	0.009	0.055	0.009	0.077	0.019	0.060	0.031
+ IQ	0.091	0.374	0.013	0.080	0.013	0.111	0.027	0.085	0.046
+ NCS	0.097	0.401	0.015	0.092	0.015	0.127	0.031	0.097	0.053
+NCS X IQ	0.098	0.405	0.015	0.094	0.016	0.130	0.032	0.100	0.053

	Gini	Gini(share)	GE0	GE0(share)	GE1	GE1(share)	CV2	CV2(share)	R2
Siblings' Income	0.040	0.164	0.003	0.020	0.003	0.026	0.006	0.019	0.012
+ Siblings' Income (sqrt)	0.049	0.203	0.004	0.025	0.004	0.035	0.008	0.026	0.014
+ Siblings' Income (3rd order)	0.050	0.207	0.004	0.027	0.005	0.039	0.010	0.031	0.015
+ Siblings' Income (4th order)	0.050	0.208	0.004	0.027	0.005	0.038	0.009	0.029	0.015
+ Municipality Effect	0.063	0.262	0.007	0.041	0.007	0.057	0.014	0.043	0.023
+ Municipality Effect (sqrt)	0.063	0.261	0.007	0.041	0.007	0.056	0.014	0.043	0.023
+ Municipality Effect (3rd order)	0.063	0.261	0.007	0.041	0.007	0.056	0.014	0.043	0.023
+ Municipality Effect (4th order)	0.063	0.261	0.007	0.041	0.007	0.056	0.014	0.043	0.023
+ all circumstances	0.104	0.431	0.017	0.106	0.018	0.146	0.036	0.112	0.060
Siblings' Income + Municipality Effect	0.044	0.181	0.004	0.024	0.004	0.031	0.007	0.022	0.015

Notes: This Table shows Gini for long run income and estimated current income $\hat{y}$, which has been estimated using a log model.

Source: All individuals for whom we observe at least 15 times their income after the age of 30, we keep only the observations from the age of 30 and for which income is strictly positive. To compute the family effect, we restrict the sample to families for which at least two individuals respect the sample criteria.

A3 Siblings correlations and inter-generational regression - statistical framework

Consider two siblings, $i = 1, 2$. To simplify, we neglect the decomposition between municipality and family effect, and note j the family f leaving in a municipality m, $\alpha_j = a_m + b_{mf}$. We focus on permanent income and neglect the effect of transitory shocks. Each siblings has an income (in log), noted y_{j1}, y_{j2}.

A3.1 Variance decomposition estimates

$$\begin{aligned} \text{cov}(y_{j1}, y_{j2}) &= \text{cov}(\alpha_j + c_{j1}, \alpha_j + c_{j2}) \\ &= \text{cov}(\alpha_j, \alpha_j) + 0 \\ &= \sigma_\alpha^2 \end{aligned}$$

Variance-share of family effect is equal to the correlation coefficient:

$$\begin{aligned}\rho &= \frac{\mathrm{cov}(y_{j1}, y_{j2})}{\sigma_{y_{j1}}\sigma_{y_{j2}}} \\ \rho &= \frac{\mathrm{cov}(y_{j1}, y_{j2})}{\mathrm{V}(y_{ji})} \\ &= \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_c^2}\end{aligned}$$

A3.2 Regression of y_{j1} on y_{j2}

Let's write:

$$y_{j1} = \gamma y_{j2} + \varepsilon_{i1} \qquad \text{(Eq . 19)}$$

- Under the iid assumption, $\gamma = \rho$:

$$\begin{aligned}\gamma &= \frac{\mathrm{cov}(y_{j1}, y_{j2})}{\mathrm{V}(y_{j2})} \\ &= \frac{\mathrm{cov}(y_{j1}, y_{j2})}{\mathrm{V}(y_{ji})} \\ &= \rho\end{aligned}$$

- Under the iid assumption, $\mathrm{R}^2 = \rho^2$:

$$\begin{aligned}\mathrm{R}^2 &= \frac{\gamma^2 \mathrm{V}(y_{j2})}{\mathrm{V}(y_{j1})} \\ &= \gamma^2 = \rho^2\end{aligned}$$

A3.2.1 Case with more than one sibling

When we have more than one sibling, let's write:

$$y_{j1} = \gamma \frac{\Sigma_{i=2}^S y_{ji'}}{S-1} + \varepsilon_{i1} \qquad \text{(Eq . 20)}$$

$$\begin{aligned}\mathrm{cov}(y_{j1}, \frac{\Sigma_{i=2}^S y_{ji'}}{S-1}) &= \mathrm{cov}(y_{j1}, \frac{\Sigma_{i=2}^S \alpha_j + c_{ji'}}{S-1}) \\ &= \mathrm{cov}(\alpha_j + c_{j1}, \alpha_j + \frac{\Sigma_{i=2}^S c_{ji'}}{S-1}) \\ &= \sigma_\alpha^2\end{aligned}$$

Under the iid assumption,

$$\begin{aligned}
\gamma &= \frac{\mathrm{cov}(y_{j1}, \frac{\Sigma_{i=2}^{S} y_{ji'}}{S-1})}{\mathrm{V}(\frac{\Sigma_{i=2}^{S} y_{ji'}}{S-1})} \\
&= \frac{\mathrm{cov}(y_{j1}, \frac{\Sigma_{i=2}^{S} y_{ji'}}{S-1})}{V(y_{ij})} \frac{V(y_{ij})}{\mathrm{V}(\frac{\Sigma_{i=2}^{S} y_{ji'}}{S-1})} \\
&= \rho \quad \frac{V(y_{ij})}{\mathrm{V}(\frac{\Sigma_{i=2}^{S} y_{ji'}}{S-1})}
\end{aligned}$$

Under the iid assumption, $\mathrm{R}^2 \neq \rho^2$:

$$\begin{aligned}
\mathrm{R}^2 &= \frac{\gamma^2 \mathrm{V}(\frac{\Sigma_{i=2}^{S} y_{ji'}}{S-1})}{\mathrm{V}(y_{j1})} \\
&= \rho^2 (\frac{V(y_{ij})}{\mathrm{V}(\frac{\Sigma_{i=2}^{S} y_{ji'}}{S-1})})^2 \frac{\mathrm{V}(\frac{\Sigma_{i=2}^{S} y_{ji'}}{S-1})}{\mathrm{V}(y_{j1})} \\
&= \rho^2 \quad \frac{V(y_{ij})}{\mathrm{V}(\frac{\Sigma_{i=2}^{S} y_{ji'}}{S-1})}
\end{aligned}$$

A3.3 Regression of y_{j1} on observable family characteristics W_i

- Write α_i as a function of observable family characteristics W_i (Solon 1999)):

$$\alpha_i = \beta W_i + z_i \quad \text{with } W_i \perp z_i \tag{Eq . 21}$$

- Individual's income can be expressed as :

$$y_{ji} = \beta W_i + z_i + c_{ji} \tag{Eq . 22}$$

- R^2 in the intergenerational regression Eq . 22 provides a lower bound estimate of ρ, which corresponds to the variance in α_i explained by W_i:
  - from equation Eq . 21 :
  $$\sigma_\alpha^2 = \beta^2 \sigma_W^2 + \sigma_z^2$$
  - from equation Eq . 22:
  $$\begin{aligned}
  \mathrm{R}^2 &= \frac{\beta^2 \mathrm{V}(W_i)}{\mathrm{V}(y_{j1})} = \frac{\beta^2 \sigma_W^2}{\sigma_\alpha^2 + \sigma_c^2} \\
  &< \frac{\beta^2 \sigma_W^2 + \sigma_z^2}{\sigma_\alpha^2 + \sigma_c^2} \\
  &= \rho
  \end{aligned}$$

– NB : if $\sigma_z = 0$, $\mathrm{R}^2 = \rho$

A3.4 Extended regression of y_{j1} on observable family characteristics W_i and sibling's income

A3.4.1 Computing coefficients

Let's write y_{j1} as a function of observable family characteristics W_i (Solon 1999) and of **sibling's income** y_{j2}:

$$y_{j1} = bW_i + cy_{j2} + e_{i1} \tag{Eq . 23}$$

Computing c :

- From partitioned regression, c is equal to the coefficient in the regression of $y_{j1}^{\perp W_i}$ on $y_{j2}^{\perp W_i}$ where $Z^{\perp U}$ is the residual of the orthogonal projection of Z on U;
- From Eq . 22, we have :
$$y_{ji}^{\perp W_i} = y_{ji} - \beta W_i = z_i + c_{ji}$$
- Thus

$$\begin{aligned} c &= \frac{\mathrm{cov}(z_i + c_{j1}, z_i + c_{j2})}{\mathrm{V}(z_i + c_{ji})} \\ &= \frac{\sigma_z^2}{\sigma_z^2 + \sigma_c^2} \end{aligned}$$

Computing b :

- From partitioned regression, b is equal to the coefficient in the regression of $y_{j1}^{\perp y_{j2}}$ on $W_i^{\perp y_{j2}}$ where $Z^{\perp U}$ is the residual of the orthogonal projection of Z on U;
- From Eq . 19, we have :
$$y_{ji}^{\perp y_{j2}} = y_{ji} - \rho y_{j2} = (1-\rho)\alpha_i + c_{j1} - \rho c_{j2}$$
- We can express $W_i^{\perp y_{j2}}$ as :

$$\begin{aligned} W_i^{\perp y_{j2}} &= W_i - \theta y_{j2} \\ \text{with} \quad \theta &= \frac{\mathrm{cov}(W_i, y_{j2})}{\mathrm{V}(y_{j2})} \\ &= \beta \frac{\sigma_W^2}{\sigma_\alpha^2 + \sigma_c^2} \end{aligned}$$

- Note that :

$$
\begin{aligned}
\mathrm{V}(W_i^{\perp y_{j2}}) &= \mathrm{V}(W_i - \theta y_{j2}) \\
&= \sigma_W^2 + \theta^2 \sigma_y^2 - 2\theta \mathrm{cov}(W_i, y_{j2}) \\
&= \sigma_W^2 + \theta\beta \frac{\sigma_W^2}{\sigma_\alpha^2 + \sigma_c^2}(\sigma_\alpha^2 + \sigma_c^2) - 2\theta\beta\sigma_W^2 \\
&= \sigma_W^2(1 - \beta\theta)
\end{aligned}
$$

- $\mathrm{cov}(y_{j1}^{\perp y_{j2}}, W_i^{\perp y_{j2}})$ is given by :

$$
\begin{aligned}
\mathrm{cov}(y_{j1}^{\perp y_{j2}}, W_i^{\perp y_{j2}}) &= \mathrm{cov}((1-\rho)\alpha_i + c_{j1} - \rho c_{j2}, W_i - \theta\alpha_i - \theta c_{j2}) \\
&= \mathrm{cov}((1-\rho)\alpha_i, W_i) + \mathrm{cov}((1-\rho)\alpha_i, -\theta\alpha_i) + \mathrm{cov}(-\rho c_{j2}, -\theta c_{j2}) \\
&= (1-\rho)\beta\sigma_W^2 - (1-\rho)\theta\sigma_\alpha^2 + \rho\theta\sigma_c^2
\end{aligned}
$$

- Using the expression for θ, we get :

$$
\begin{aligned}
\mathrm{cov}(y_{j1}^{\perp y_{j2}}, W_i^{\perp y_{j2}}) &= (1-\rho)\beta\sigma_W^2 - (1-\rho)\beta\sigma_W^2\rho + \rho\beta\sigma_W^2(1-\rho) \\
&= (1-\rho)\beta\sigma_W^2
\end{aligned}
$$

- b can thus be written as :

$$
\begin{aligned}
b &= \frac{\mathrm{cov}(y_{j1}^{\perp y_{j2}}, W_i^{\perp y_{j2}})}{\mathrm{V}(W_i^{\perp y_{j2}})} \\
&= \frac{\beta(1-\rho)}{1-\beta\theta} \\
Rq : b &= \frac{\beta\frac{\sigma_c^2}{\sigma_y^2}}{\frac{\sigma_c^2+\sigma_z^2}{\sigma_y^2}} \\
&= \beta\frac{\sigma_c^2}{\sigma_c^2+\sigma_z^2} \\
&= \beta\frac{\sigma_c^2+\sigma_z^2-\sigma_z^2}{\sigma_c^2+\sigma_z^2} \\
&= \beta(1-c) \\
&= \beta - \beta c
\end{aligned}
$$

Two ways two recover ρ from the estimation of equations Eq . 22 and Eq . 23 :

- denote $\mathrm{R}^2_{Y_{i1|W_i}}$ the R-square from equation Eq . 22

- using equation Eq . 21 and the expressions for c and $\mathrm{R}^2_{Y_{i1|W_i}}$, we have :

$$
\begin{aligned}
\rho &= \frac{\beta^2\sigma_W^2}{\sigma_\alpha^2+\sigma_c^2} + \frac{\sigma_z^2}{\sigma_\alpha^2+\sigma_c^2} \\
&= \frac{\beta^2\sigma_W^2}{\sigma_\alpha^2+\sigma_c^2} + \frac{\sigma_z^2}{\sigma_z^2+\sigma_c^2}\frac{\sigma_z^2+\sigma_c^2}{\sigma_\alpha^2+\sigma_c^2} \\
&= \frac{\beta^2\sigma_W^2}{\sigma_\alpha^2+\sigma_c^2} + c\ \ \frac{\sigma_z^2+\sigma_c^2}{\sigma_\alpha^2+\sigma_c^2}
\end{aligned}
$$

$$\sigma_z^2+\sigma_c^2 = \sigma_y^2 - \beta^2\sigma_W^2$$

$$
\begin{aligned}
\Rightarrow \rho &= \frac{\beta^2\sigma_W^2}{\sigma_\alpha^2+\sigma_c^2} + c\ \ (1-\frac{\beta^2\sigma_W^2}{\sigma_\alpha^2+\sigma_c^2}) \\
\Rightarrow \rho &= \frac{\beta^2\sigma_W^2}{\sigma_\alpha^2+\sigma_c^2} + c\ \ (1-\frac{\beta^2\sigma_W^2}{\sigma_\alpha^2+\sigma_c^2}) \\
\Rightarrow \rho &= \mathrm{R}^2_{Y_{i1|W_i}} + c\ \ (1-\mathrm{R}^2_{Y_{i1|W_i}})
\end{aligned}
$$

- alternatively, consider the expression for b, noting that $\beta\theta = \mathrm{R}^2_{Y_{i1|W_i}}$:

$$b = \frac{\beta(1-\rho)}{1-\mathrm{R}^2_{Y_{i1|W_i}}}$$

  solving for ρ yields:

$$\rho = 1 - b\frac{1-\mathrm{R}^2_{Y_{i1|W_i}}}{\beta} \qquad \text{(Eq . 24)}$$

A3.4.2 Predicting $\hat{y}_{j1}$ for the IOP

$$
\begin{aligned}
\hat{y}_{j1} &= b_0 + bW_i + cy_{j2} \\
&= b_0 + (\beta - \beta c)W_i + cy_{j2} \\
&= b_0 + \beta W_i + c(y_{j2} - \beta W_i) \\
&= b_0 + \beta W_i + c(z_i + c_{j2})
\end{aligned}
$$

Predicting y_{j1} from the siblings' income hence also accounts for siblings spillovers $(c.c_{j2})$.

A3.4.3 Computing the R^2

$$
\begin{aligned}
V(\hat{y}_{j1}) &= \beta^2 V(W_i) + c^2 V(z_i + c_{j2}) \\
&= \beta^2 V(W_i) + c^2 (V(z_i) + V(c_{j2})) \\
&= \beta^2 \sigma_W^2 + (\frac{\sigma_z^2}{\sigma_c^2 + \sigma_z^2})^2 (\sigma_z^2 + \sigma_c^2) \\
&= \beta^2 \sigma_W^2 + \frac{(\sigma_z^2)^2}{\sigma_c^2 + \sigma_z^2}
\end{aligned}
$$

$$
\begin{aligned}
\frac{V(\hat{y}_{j1})}{V(Y)} &= \frac{\beta^2 \sigma_W^2}{\sigma_y^2} + \frac{\sigma_z^2}{\sigma_y^2} - \frac{\sigma_z^2}{\sigma_y^2} + \frac{\sigma_z^2}{\sigma_y^2} \; \frac{\sigma_z^2}{\sigma_c^2 + \sigma_z^2} \\
&= \rho + \frac{\sigma_z^2}{\sigma_y^2} (\frac{\sigma_z^2}{\sigma_c^2 + \sigma_z^2} - 1) \\
&= \rho + \frac{\sigma_z^2}{\sigma_y^2} (\frac{-\sigma_c^2}{\sigma_c^2 + \sigma_z^2}) \\
&= \rho - c \; \frac{\sigma_c^2}{\sigma_y^2} \quad < \rho \\
&= \rho - c \; (1 - \rho) \quad < \rho
\end{aligned}
$$