WORKING PAPER 2025:2

The relationship between intergenerational mobility and equality of opportunity

Adrian Adermon
Gunnar Brandén
Martin Nybom

IFAU
INSTITUTE FOR EVALUATION OF LABOUR MARKET AND EDUCATION POLICY

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala.

IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

More information about IFAU and the institute's publications can be found on the website www.ifau.se

ISSN 1651-1166

The relationship between intergenerational mobility and equality of opportunity*

Adrian Adermon[†] Gunnar Brandén [‡] Martin Nybom[§]

March 3, 2025

Abstract

Among economists, empirical analysis of social mobility and the role of parental background is largely carried out in two separate strands of research. The *intergenerational mobility* literature estimates parent-child persistence in a certain outcome of interest, such as income. In contrast, the *equality of opportunity* literature is rooted in a normative framework, and has only more recently started generating empirical evidence. Intergenerational mobility regressions are relatively straightforward to estimate, but their normative implications are less obvious. In contrast, measures of equality of opportunity have a policy-relelvant interpretation, but are demanding in terms of data, requiring the researcher to observe a large set of determinants of socioeconomic status for large samples. But maybe the two approaches capture similar underlying dynamics? We compare the two approaches by estimating both equality of opportunity and intergenerational mobility measures — as well as sibling correlations — across 16 birth cohorts within 126 Swedish local labor markets. Using these estimates, we test to what extent the different measures correlate, resulting in insights on the plausibility of interpreting intergenerational mobility measures as informative about equality of opportunity.

Keywords: Equality of opportunity, Intergenerational mobility, Sibling correlations

JEL Classification: D31, J62, D63

1 Introduction

The study of intergenerational mobility has experienced growing interest in both research and policy circles. Common measures of intergenerational as-

*First draft: April 2017. We are grateful for comments from Anders Björklund, Anna Sjögren, Paul Hufe, Arnaud Lefranc, Andreas Peichl, Alain Trannoy, and seminar participants at the DIAL conference in Turku, IMCHILD project workshop in Stockholm, EQINEQ conference in Paris, and IFAU in Uppsala.

[†]Institute for Evaluation of Labour Market and Education Policy (IFAU), Box 513, SE-751 20 Uppsala, Sweden (Email: adrian.adermon@ifau.uu.se), UCLS, and UCFS.

[‡]Center for Epidemiology and Community Medicine, Stockholm, Sweden (Email: gunnar.branden@regionstockholm.se), and Department of Global Public Health, Global and Sexual Health, Karolinska Institutet, Stockholm, Sweden

[§]Institute for Evaluation of Labour Market and Education Policy (IFAU), Box 513, SE-751 20 Uppsala, Sweden (Email: martin.nybom@ifau.uu.se), SOFI, and UCLS.

sociation, such as the intergenerational elasticity of income, are comparably straightforward to estimate and provide an appealing (statistical) interpretation of overall regression-to-the-mean. By comparing such estimates across time or places, scholars often conjecture that a lack of mobility indicates inequality of opportunity. For example, in their influential work on mobility variation within the United States, Chetty, Hendren, Kline, and Saez (2014) characterize areas with high rates of income mobility across generations as "lands of opportunity".[1] The importance of family background has also been studied by estimating sibling correlations, an alternative approach which has emerged parallel to the empirical mobility literature (Corcoran et al. 1990; Solon 1999). Again, these estimates are often interpreted as being informative about the level equality of opportunity in a society.[2] Theoretically, however, rates of intergenerational mobility or sibling correlations are not necessarily good indicators of the degree of equality of opportunity — as Björklund and Jäntti (2020) put it, the intergenerational-mobility approach captures a "special case" of equality of opportunity. They also point out that the sibling correlation excludes all background factors that are not shared between siblings. While thus appealing to practitioners, it is unclear to what extent mobility and sibling correlation estimates are informative about issues of public interest, such as disparities in opportunity.

An alternative approach is to start from a theoretically defined concept of equality of opportunity (EOp). The dominant view in this tradition is that a society achieves equality of opportunity if individuals' accomplishments, with respect to some outcome of interest, are determined wholly by their personal choices and effort, rather than by *circumstances* beyond their control (Arneson 1989; Cohen 1989; Roemer 1993, 1998). Roemer (1998) formalized this concept, and a growing empirical literature seeks to derive indices of EOp by removing the influence of such circumstances (e.g., race, parental schooling, etc) on income or other outcomes of interest (see Roemer and Trannoy 2016, for a survey). However, empirical implementation faces substantial challenges (see, e.g., Fishkin 2014). First, it is difficult to distinguish between circumstances and effort, and some factors may be seen as a combination of both. Second, analyses suffer from an inherent omitted-variables problem: to isolate the influence of circumstances all such factors have to be perfectly observed, which is never the case.[3]

Our goal is to empirically connect these two literatures. Using population-wide data from Sweden, we examine the statistical relationship between measures of intergenerational mobility, sibling correlations, and indices of equality of opportunity. In a nutshell, we exploit variation in estimates of these measures across geographical areas and examine to what extent they correlate with each other. Our objective is to provide an empirical test of whether a lack of intergenerational mobility, or strong sibling similarities, indeed implies inequality of opportunity, and vice versa.

[1]Other influential papers that draw a link between intergenerational mobility and equality of opportunity (though often with some form of qualification) include, for example, Alesina, Stantcheva, and Teso (2018), Chetty, Hendren, Kline, Saez, and Turner (2014), Corak (2013), and Ward (2023).

[2]See, e.g., Björklund, Eriksson, et al. (2002), Björklund, Hederos, and Jäntti (2010), Björklund and Jäntti (2012), Hällsten (2014), Raaum, Salvanes, and Sørensen (2006), and Solon et al. (1991).

[3]Kanbur and Wagstaff (2016) argue that this problem complicates comparisons of EOp across space or time, since estimates will either have to be based on the lowest common denominator in terms of observed circumstances, or estimates will not be comparable.

The literature on intergenerational mobility, or its inverse, intergenerational *persistence*, goes back to Galton's (1886) seminal study of the intergenerational correlation in height. While sociologists have long studied class and occupational mobility, economic research on income and earnings mobility gained interest first in the 1980s and early 1990s following the theoretical work of Becker and Tomes (1979, 1986) and an increasing access to intergenerational data sets (e.g., Solon 1992; Zimmerman 1992). Subsequently, mobility estimates have been produced for a large number of countries, including some developing countries, and in several developed countries researchers have moved on to using administrative data sources to address new questions within the literature (for reviews, see Black and Devereux 2011; Jäntti and Jenkins 2015; Mogstad and Torsvik 2023). Noteworthy findings include that mobility is much lower than previously thought (e.g., Mazumder 2005), and also negatively correlated with cross-sectional inequality both within and across countries (Björklund and Jäntti 1997; Chetty, Hendren, Kline, and Saez 2014; Corak 2013). While we might be concerned about the intergenerational persistence of inequality in specific outcomes (such as income) per se, the normative implication of such outcome-specific transmission is somewhat unclear. There is no consensus view on whether having more mobility necessarily represents a social improvement or if it is undesirable.

The empirical literature on EOp is more recent. In a nutshell, its aim is to estimate how much inequality can be explained by *circumstances*, i.e., various measurable background factors such as parental education and income, race/ethnicity, family structure, etc. The earliest applications were concerned with estimating the extent to which tax-and-transfer systems equalize opportunities (Page and Roemer 2001; Roemer, Aaberge, et al. 2003). Subsequent studies have mostly tried to estimate the overall level of inequality of opportunity (IOp) in different countries.[4] Lefranc, Pistolesi, and Trannoy (2008) find that West Germany, Norway, and Sweden come quite close to achieving equality of opportunity, the U.S. and Italy are far from it, and Belgium, the Netherlands, France, and Great Britain lie in-between these two extremes. Using a large set of circumstance variables, Björklund, Jäntti, and Roemer (2012) estimate that around 70 percent of long-run income inequality in Sweden is due to individual effort, and that parental income and IQ differences are the most important barriers to equal opportunities. Ferreira and Gignoux (2011) estimate IOp indices for six Latin American countries, finding that between one quarter and one half of consumption inequality is due to circumstances. Hufe, Kanbur, and Peichl (2022) build on the concept of EOp to measure the extent of "unfair inequality" in income across countries and over time within the US.[5] For recent surveys of both the theoretical and empirical literature, see Ferreira and Peragine (2016) and Roemer and Trannoy (2015, 2016).

While the EOp approach builds on formal definitions and welfare axioms, the challenges facing applied work is an obstacle. As such, the approach has

[4] In the following, we will refer to the theoretical construct as *equality* of opportunity, or *EOp*, and the empirical estimates as *in*equality of opportunity, or *IOp*. IOp can be thought of as the inverse of EOp — when IOp is high, EOp is low, and vice versa.

[5] It is important to note that almost all existing studies estimate lower bounds on the role of circumstances, and that using a wider set of circumstances will results in larger estimates of the inequality of opportunity. This makes it hard to directly compare estimates of IOp between different studies.

probably had less impact compared to the mobility literature. A simplified view of these literatures could thus be that the mobility approach is practically appealing but lacks in conceptual clarity, while the EOp approach builds on a rigorous framework but is harder to empirically operationalize. But maybe they largely capture the same dynamics?

In our analysis, we divide Sweden into 126 local labor markets (similar to US commuting zones), and estimate *regional* measures of mobility and IOp in permanent income. We estimate two intergenerational measures: the intergenerational income elasticity (IGE) and rank persistence. We also consider the sibling correlation as an alternative measure of the role of family background. We refer to these three indices collectively as measures of *social mobility.* To estimate inequality of opportunity (IOp) indices, we follow the machine-learning approach of Brunori, Hufe, and Mahler (2023). This approach amounts to using conditional inference forests to estimate the extent to which existing income inequality is due to circumstances as opposed to "effort".[6] We compute regional measures of inequality, and merge several administrative data sources to form a large set of circumstance variables.

Our contributions are threefold. First, we link the different literatures by providing empirical evidence on the statistical relationship between measures of social mobility and IOp indices. Knowledge of the degree to which they correlate is of immediate value: a strong correlation would indicate that differences over time or place in social mobility are likely to capture differences also in equality of opportunity. Second, by distinguishing between different mobility measures, we are able to conjecture which measures are more or less correlated with IOp. Third, through a number of robustness tests, we provide evidence on which factors determine the extent of correlation and in what contexts the correlations become weaker and stronger.

We begin by documenting national-level estimates that are similar in magnitude to those in prior studies for Sweden, with the IGE and rank-slope of about 0.23–0.25, a sibling correlation of 0.27, and a (relative) inequality of opportunity (IOp) index of 0.39. But the core of our analysis concerns the regional variation in these measures. In particular, we find a strong positive relationship between measures of IOp and intergenerational persistence across Swedish regions. First, the intergenerational measures (IGE and rank correlation) are strongly associated with inequality of opportunity (IOp) indices, with cross-region correlations above 0.8. The sibling correlation, however, tends to be more moderately correlated with both the intergenerational and the IOp measures, which would indicate that these measures capture partly different aspects of intergenerational transmission. However, we also find that sampling variation across smaller regions depresses our estimated correlations between the different measures, and especially so for the sibling correlation. When weighting the estimates by region size or excluding smaller regions, the pattern for the sibling correlation is more similar to that of the other measures.

Second, the strength of the correlation between IOp and social mobility is not due to a mechanical role of parental income in the IOp indices. In fact, the different measures remain nearly as strongly correlated when parental income is excluded from the circumstances underlying the IOp index. Finally, we emphasize

[6]The estimation of IOp is essentially a prediction problem, which makes it eminently suited for applying machine learning techniques. Brunori, Hufe, and Mahler (2023) show that their approach substantially outperforms earlier methods.

that while the *variation* in the measures co-move strongly across regions and cohorts, their *levels* in terms of explained income variation differ. The sibling correlation or (relative) IOp indices attribute much more of the variance in income to family-background factors compared to the (squared) IGE or rank correlation (see, e.g., Björklund and Jäntti 2020; Solon 1999).

1.1 Related literature

The question of whether intergenerational measures are informative about EOp is contested. It has been argued that perfect equality of opportunity does not imply eliminating all resemblance between parents and children, because differences due to inherited ability and values will persist even in a perfectly fair society (Swift 2004). For this reason, Jencks and Tach (2006) argue that measures of intergenerational mobility are unreliable indicators of equality of opportunity. On the contrary Torche (2015) argues that these sources of transmission are likely to be both small, and similar in magnitude across space and time, so that differences in intergenerational mobility can be used to infer differences in EOp. We contribute to the discussion by testing this empirically.

Our paper also relates to a set of recent empirical papers. Deutscher and Mazumder (2023) compare the ranking of Australian regions across different measures of relative and absolute income mobility. While their focus is on providing a comprehensive framework for different measures of intergenerational mobility, they do include a measure of relative inequality of opportunity. However, their cross-region correlations are notably lower than ours, and our papers differ in several ways. First, rather than using only a single measure of IOp, we provide estimates for both absolute and relative IOp; lower and upper bounds; and using different inequality indices. Second, we have access to a very rich set of circumstances underlying our IOp indices. Furthermore, we probe the sensitivity of our results to varying the set of circumstances, provide estimates separately by gender, and address the impact of sampling variation across regions on the various measures.

Brunori, Ferreira, and Peragine (2013) correlate existing estimates across countries, finding correlations of around 0.6 between IOp and both IGEs (across 16 countries) and intergenerational schooling correlations (24 countries).

In another recent paper, Brunori, Hufe, and Mahler (2023) use machine-learning methods to estimate IOp. They provide cross-country correlations between IOp indices and IGEs for 10 countries, finding correlations ranging between 0.44 and 0.66 for traditional IOp estimates, while their new IOp estimates are much more strongly related to IGEs, with correlations of almost 0.9.

Blundell and Risa (2019) use machine-learning methods to predict child incomes in Norway using a rich set of family-background characteristics. They then compare the predictive power of this rich model, measured by R^2, to that of a simple intergenerational rank-rank regression in income. Across 40 Norwegian regions, the R^2 of the rank-rank model is highly correlated (0.87) with that of the full machine-learning model.

Naturally, we also build on the large number of prior studies measuring the role of family background for incomes in Sweden, using various measures and study populations.[7] In particular, Björklund and Jäntti (2020) provide a

[7] See e.g. Björklund and Jäntti (2009), Björklund, Jäntti, and Roemer (2012), Breen,

conceptual discussion of the intergenerational mobility, intergenerational causal effects, sibling correlations, and equality of opportunity approaches. They conclude that all four approaches are informative about important questions, but that using only one of the approaches in isolation could lead to mistaken conclusions. While they focus on national-level estimates of the various measures, we instead study the joint variation in such estimates across regions and cohorts.

The paper is structured as follows. We discuss how intergenerational mobility, sibling correlations, and IOp are measured and estimated in more detail in the next Section, where we also provide some discussion of the conceptual relationship between these approaches. We discuss our data in Section 3. Section 4 presents the results and some sensitivity analyses, and Section 5 concludes.

2 Concepts, Measurement, and Implementation

In this Section we outline our set of measures of intergenerational mobility, sibling correlation, and inequality of opportunity. We briefly discuss their measurement and conceptual nature.

2.1 Intergenerational Mobility

Most empirical studies of intergenerational income persistence (or inversely mobility) characterize the joint distribution of adult children's and their parents' lifetime incomes using various linear summary measures.[8] Each of these measures provides a specific perspective on intergenerational mobility within a population.

The most established measure is the *intergenerational elasticity* (IGE), commonly estimated as the slope coefficient in a bivariate regression of offspring on parent log income:

$$y_t = \beta y_{t-1} + \gamma X + \varepsilon_t, \tag{1}$$

where t indicates generation and X is a set of cohort and gender dummies. The elasticity, β, is a measure of persistence, and the lower it is the higher is the expected rate at which incomes regress to the mean between generations. The IGE has long been the most popular mobility measure among economists — partly due to the appeal of its regression-to-the-mean interpretation, and partly due to its derivability as a reduced-form relationship from models of parental investments in children (Becker and Tomes 1979; Solon 2004).

However, there are practical challenges related to the measurement of lifetime income affecting the estimation of both the IGE and other related measures.[9] Partly for this reason, much recent work estimates rank-based mobility measures.

The *rank-rank regression* (or rank persistence) is estimated by regressing child's percentile ranked incomes on parents' percentile ranked incomes, where

Mood, and Jonsson (2016), and Hederos, Jäntti, and Lindahl (2017). For regional variation in intergenerational mobility, see Heidrich 2017.

[8]We do not address non-linear mobility measures in this paper. See further discussion in Deutscher and Mazumder (2023).

[9]The standard concerns are that attenuation (e.g., Mazumder 2005) and life-cycle biases (Haider and Solon 2006; Nybom and Stuhler 2016) arising from the approximation of lifetime income using short-run incomes make consistent estimation of the IGE more demanding.

the rank is computed within each birth cohort and gender.[10] It measures the extent to which offspring income tends to increase with parental income, without requiring that relationship to be log-linear and abstracting from *any* distributional differences between generations. While the rank-rank regression provides a summary measure of positional mobility, its scale-invariant interpretation can also be unappealing. As an example, moving ten percentiles in the income distribution in a high-inequality country is significantly more meaningful in terms of changes in living standard than a comparable shift in a low-inequality country.

The usage of rank-based measures is often motivated by practical features. For example, Chetty, Hendren, Kline, and Saez (2014) argue that the approximate linearity of the conditional expectation of child income rank on parent income rank make them well suited for analyzing mobility differences across subgroups of a population, and Nybom and Stuhler (2017) show that rank-based measures suffer less from measurement-error biases when lifetime incomes are unobserved. We estimate these two intergenerational measures separately by region and following conventional procedures, using father's income as the parental variable.

2.2 Sibling Correlations

An alternative measure of the importance of family background is the sibling correlation (see, e.g., Björklund and Jäntti 2009; Corcoran et al. 1990; Solon 1999). A common motivation for its usage is that it captures a broader scope of family influences than intergenerational mobility measures. For example, Jäntti and Jenkins (2015) argue that if we would like to understand how important family background is for the distribution of economic status, a focus on parent-child associations captures only one specific dimension of the family. The sibling correlation instead captures the importance of all factors that siblings share in terms of some outcome. While part of what siblings share is parental income, a large part is not.

We can write log earnings for sibling j in family i as

$$y_{ij} = a_i + b_{ij}, \tag{2}$$

where a_i is a family component that captures everything that is common between siblings, including parental characteristics, place of birth, and neighborhood, while b_{ij} is an individual component which is taken to be orthogonal to the shared component. Thus, the variance of log earnings can be decomposed as $\sigma_y^2 = \sigma_a^2 + \sigma_b^2$. The correlation between two siblings within family i is then

$$\text{Corr}(y_{ij}, y_{ij'}) = \frac{\text{Cov}(y_{ij}, y_{ij'})}{\sigma_y^2} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2}, \tag{3}$$

and captures the share of the variance in log earnings due to shared factors.

Following Mazumder (2008), we estimate the multi-level model

$$y_{ij} = X_{ij}'\beta + a_i + b_{ij},$$

[10]When the ranks are computed within the full population of interest, a regression of child on parent income rank gives an estimate of the (Spearman) rank correlation. In our application, it is more correct to talk about a rank-rank slope, as in Chetty, Hendren, Kline, and Saez (2014).

where X_{ij} is a set of cohort and gender dummies, using restricted maximum likelihood (REML). This provides estimates of the variance components $\hat{\sigma}_a^2$ and $\hat{\sigma}_b^2$ which can be plugged into Equation (3) to get an estimate of the sibling correlation.

It is instructive to connect the sibling correlation to the IGE. Adding generation indices, we can write the family component as $a_i = \beta y_{i,t-1} + z_i$, where $y_{i,t-1}$ is parental log earnings and z_i captures all shared factors orthogonal to parental earnings. Substituting this expression into Equation (3)[11], and assuming that both generations have the same earnings variance, we get

$$\mathrm{Corr}(y_{ij}, y_{ij'}) = \beta^2 + \frac{\sigma_z^2}{\sigma_y^2}.$$

The sibling correlation thus equals the squared IGE plus an additional term (Bingley and Cappellari 2018; Solon 1999). We also note that by necessity the sibling correlation is generally estimated for a slightly different population (siblings) than mobility or IOp measures (all children).

2.3 Inequality of Opportunity

Roemer (2004) points out that intergenerational associations are direct measures of inequality of opportunity only if two specific conditions apply. First, the advantages associated with parental background are entirely summarized by parental income (including its correlates). Second, the concept of equality of opportunity (EOp) that is employed views as unacceptable any income differences in the child generation that are attributable to differences in innate talents (which might be partly genetically inherited).

In the concept of EOp proposed by Roemer (1993, 1998), the population is partitioned into *types*, where each type comprises the set of individuals with the same *circumstances*. The set of circumstances includes all factors beyond the child's control, which theoretically could include both typically observable (e.g. parental education) and unobservable (e.g. genetic makeup) factors. In empirical studies, this set is by necessity restricted to a host of observable background factors such as parental income and education, place of birth, race/ethnicity, etc. Each individual chooses their level of *effort*, which together with their circumstances results in a certain level of *advantage*. EOp obtains when individuals are rewarded for their effort, but not for their circumstances. Recognizing the potential for type-effort correlations, Roemer takes this further, arguing that EOp obtains only when the *distribution* of advantage is independent of type. In principle, this can be tested by forming types from groups of individuals with similar circumstances and comparing the empirical cumulative distribution functions of, e.g., earnings between types.

Let earnings Y be a function of circumstances C, efforts E, and unobserved random factors u:

$$Y = f(C, E, u). \tag{4}$$

Effort is partly influenced by circumstances, so we can rewrite this expression as $Y = f[C, E(C, w), u]$. Since we are only interested in the total impact of circumstances on earnings, we can work with the reduced form $Y = g(C, \varepsilon)$.

[11] Substituting the expression for a_i into Equation (2), we get $y_{ijt} = \beta y_{i,t-1} + z_i + b_{ij}$, which is identical to Equation (1) if $\varepsilon_t = z_i + b_{ij}$.

Define the counterfactual earnings distribution $Y^C = \mathrm{E}(Y \mid C)$, which captures expected earnings for an individual with circumstances C. A measure of *absolute* inequality of opportunity is then given by the level of inequality due to circumstances:

$$IOL = I(Y^C), \tag{5}$$

where $I()$ is an inequality index.[12] Alternatively, *relative* IOp measures the share of overall inequality that is unfair:

$$IOR = \frac{I(Y^C)}{I(Y)}. \tag{6}$$

The empirical challenge is to estimate the counterfactual distribution Y^C. Popular approaches include the *parametric* approach (Bourguignon, Ferreira, and Menéndez 2007; Ferreira and Gignoux 2011), typically using predicted values from a log-linear regression of earnings on circumstance variables to estimate Y^C, and the *nonparametric* approach (Checchi and Peragine 2010), which partitions the sample into a set of types based on observed circumstances, and estimates Y^C as average incomes within types. Both approaches face challenges: if the models are too restrictive,[13] they run the risk of underestimating inequality of opportunity, while conversely they run the risk of overfitting if made too flexible.[14] Brunori, Hufe, and Mahler (2023) propose a principled solution to this problem through the use of machine learning methods, in particular the conditional inference trees proposed by Hothorn, Hornik, and Zeileis (2006).

2.3.1 Conditional inference trees and forests

Conditional inference trees (CIT), like other tree-based methods such as CART (Breiman et al. 1984), use recursive binary splitting to form predictions for an outcome variable. In a first step, the sample is split in two by selecting a variable and cut-off value for that variable. Each sub-sample is then split in the same way, and so on until a stopping rule is reached. For each variable and split, the algorithm performs a statistical test of the null hypothesis that the distribution of the outcome is independent of the variable. If the test fails to reject the null for each variable, the algorithm terminates, and the tree is finished; if not, the variable with the lowest p-value is chosen to split on. To find the cut-off value, a new test of independence is performed for each potential value, and the one with the lowest p-value is chosen.[15]

Conditional inference forests (CIF) apply the random forest approach of Breiman (2001) to CITs. To construct a CIF, we draw 200 bootstrap samples with a random subset of circumstances from the original sample, and estimate a CIT in each bootstrap sample.[16] We then form $\hat{Y}^C$ by averaging predictions

[12] We use the Gini coefficient as our inequality index in the main analysis, but also present robustness checks using the mean logarithmic deviation.

[13] Due to a simple linear functional form in the parametric case, or a small and coarsened set of circumstances in the non-parametric case.

[14] Due to including interactions and polynomial terms in the parametric case, or dividing the sample into too finegrained a set of types in the non-parametric case.

[15] We use a size of 0.05 for the hypothesis, and adjust for multiple testing using the Bonferroni correction.

[16] Each bootstrap sample uses 60 percent of the sample, and is drawn without replacement. For each boostrap draw, we use $\lceil\sqrt{k}\rceil$ circumstances, where k is the total number of circumstance variables in the sample.

across the bootstrap samples for each individual.

A further advantage of CIF is that they can use *surrogate splits* (Rieger, Hothorn, and Strobl 2010), which enable us to retain individuals even if they have missing values on some circumstances.[17] We use the *party* R package (Hothorn, Hornik, Strobl, et al. 2023) to estimate conditional inference forests and calculate absolute and relative IOp for each local labor market.

Because we are unable to observe all circumstances, conventional IOp estimates are typically viewed as *lower bounds* on the true level of IOp. Niehues and Peichl (2014) propose an alternative estimator which produces an *upper bound* on IOp. We discuss their estimator in Appendix B.

3 Data and descriptive statistics

We combine several administrative registers maintained by Statistics Sweden. Our source data cover the universe of the Swedish population aged 0–74 from 1965–2020 and their biological parents. All individuals are linked to population registers containing information on incomes, education, family relationships, and demographic events such as civil status, residency, and death. These include the national censuses (FoB) 1960–1990, the education register 1985–2020, and the income and tax register for the years 1968–2020.[18]

3.1 Sample restrictions

We first select all 1,727,599 children born in Sweden between 1965 and 1980. We then restrict the sample to children whose mother and father were also born in Sweden, and were between 18 and 40 years old when their child was born, leaving 1,534,031 children in the sample. We further restrict the sample to children with at least three annual incomes above a minimum level in adulthood (as described in Section 3.2), reducing the sample by 190,844 observations, and to those who lived at least six consecutive years in the same local labor market during ages 2–12, further reducing the sample by 158,283 observations. These restrictions result in a core sample of 1,184,904 children.[19]

To maximize sample size and to retain comparability across the various measures, our main samples pool sons and daughters. For this reason, we adjust our income measures for mean differences by gender (see below). While gender might be seen as an important circumstance variable from an EOp perspective (Hederos, Jäntti, and Lindahl 2017), its role will play out in radically different ways for intergenerational measures and the sibling correlation. We thus proceed with pooled samples and gender-adjusted income measures, but also present gender-specific estimates as robustness tests.

For the intergenerational and IOp analyses, we construct our *main analysis sample* by restricting the core sample to the 1,077,046 children whose fathers

[17]For observations with missing data on a selected circumstance, the algorithm instead uses a surrogate variable which is selected to best predict the split in the originally chosen variable. In our application, we allow for up to three surrogate splits.

[18]Data from the income and taxation register up until 1985 is only available for the years 1968, 1971, 1973, 1976, 1979 and 1982.

[19]This latter restriction implies that we omit those from very mobile families who lack a stable region of residence in childhood.

have non-missing incomes. For the sibling correlations, we construct a *sibling sample* of 767,005 children by dropping all singletons from the core sample.

3.2 Variable definitions

We create pre-tax income panels spanning 1968–2020. All incomes are deflated to 2020 SEK. We use two income measures (see Appendix A for details). First, *labor income* includes labor earnings, business income, taxable benefits and some labor-related benefits such as short-term sick pay and parental benefits. Capital income, pensions and long-term sickness and parental leave benefits are not included. We observe labor incomes at least every third year between 1968 and 1985, and yearly thereafter. Second, *disposable income* is calculated as the individual's (consumption-weighted) share of household disposable income, which includes after-tax labor earnings, business income, capital income, and transfers including unemployment, parental, sickness benefits, means-tested income support, pensions, study grants, and housing grants. Disposable incomes are only available starting in 1990.

In our main analyses, we use labor incomes for the parental generation and individualized disposable household incomes for the child generation. The latter reduces the risk of underestimating consumption opportunities for women, who often spend a larger share of their time in unpaid household work. As we show below, the results are robust to using labor incomes for the child generation.[20]

To obtain a more time-consistent permanent income measure we drop all annual incomes below a threshold corresponding to two "price base amounts", which in 2020 amounted to around 44 percent of the lowest full-time entry wage in the collective agreements (Swedish National Mediation Office 2021).[21] We then approximate permanent incomes by averaging annual incomes between ages 30–40 for the child and 35–55 for the parental generation. We exclude individuals with fewer than three annual income observations within the relevant age range. We use these averages untransformed for the IOp estimates, take logs for the IGE and sibling correlations, and calculate national-level percentile ranks within cohort and gender for the rank regressions.

We use local labor markets as geographical units, following Chetty, Hendren, Kline, and Saez (2014). We observe residency at birth and then annually from 1969. We recode the residency data to map into the 1985 municipality division before aggregating the municipalities into a total of 126 local labor markets. Each individual is assigned the region where they resided for the most years up until age 12.

In addition to parental income, we define a set of circumstance variables for the IOp analyses. Parental education is reported in levels which we convert into the corresponding years of schooling. We define one-digit (ten categories) parental occupation from the census closest in time to when the child was ten years old. We also include family size and both parents' age when the child was born, as well as indicators of family stability during childhood. For the

[20]From the perspective of theories of parental investment in child human capital (e.g., Becker and Tomes 1979), one could argue that using disposable income among parents would be more appropriate. However, data restrictions prevent us from doing so, as data on disposable income is only available from 1990, while labor income is available from 1968.

[21]The price base amount is used across the Swedish social insurance system to price adjust transfers, pensions, and fines. In 2020, the amount was SEK 47 300 (around EUR 4,500).

latter, we use indicators for whether the child lived in the same parish as both biological parents at age 14; whether either of the parents completed a divorce (not necessarily from each other) before the child turned 21; and whether either parent died before age 55 (also acting as a coarse measure of a poor health endowment). Finally, in a robustness test we include data on adolescent cognitive and non-cognitive skills from military enlistment tests (for men only).

3.3 Summary statistics and national-level estimates

We show summary statistics for the main analysis sample in Table 1, and for the sibling sample in Table C.1. Panel E shows summary statistics of our various measures across regions, as well as their national-level counterparts (col. 5).[22] Reassuringly, our national-level estimates of the different measures are largely in line with prior evidence, despite some differences in either income concepts and/or cohorts studied.[23] Table 1 further highlights two important patterns. First, the means across regions (Panel E, col. 1) are consistently lower than their national-level counterparts (col. 5). A possible reason for this is that regional-level income differences are suppressed in the former case but not the latter, which needs to be kept in mind throughout our discussion.[24]

Second, the national-level estimates also highlight that the share of total inequality that is attributed to family-background factors is generally substantially higher for the sibling correlation and the IOp indices than what is revealed by intergenerational mobility estimates. Note that to get this inequality share for the IGE or rank correlation, we need to square those estimates (see also, e.g., Björklund and Jäntti 2020). However, while recognizing these differences in what the *levels* of the measures imply, the focus of our analysis is how differences in the various measures across regions (or cohorts) *correlate*.

4 Results

Figure 1 plots the different persistence measures against absolute and relative IOp for each local labor market. There is a clear positive association between IOp and persistence in all panels. There is some visual evidence of a nonlinear relationship, with stronger correlations for regions with higher intergenerational persistence — particularly for the absolute IOp measure in the left column.

This pattern is also reflected in the regression lines: the solid line shows the fit from a weighted regression, where larger regions (shown as larger circles) are given more weight than smaller ones, while the dashed line shows the unweighted regression. Larger regions tend to cluster in the upper right parts of the graphs (with higher intergenerational persistence as well as higher IOp), where the

[22] Table C.2 shows further IOp and inequality estimates at the national level (Panel A) and averaged across local labor markets (Panel B). In addition to what's shown in Table 1, we vary the included circumstances and show estimates using the mean logarithmic deviation (MLD) as inequality index. Estimates using the MLD are uniformly lower than our Gini-based estimates.

[23] See e.g. Björklund and Jäntti (2009, 2020), Björklund, Jäntti, and Roemer (2012), Breen, Mood, and Jonsson (2016), and Nybom and Stuhler (2017).

[24] Also worth keeping in mind is that our regional-level analyses allow for region-specific coefficients when performing the predictions underlying the IOp indices. Further, the means in col. 1 are region-weighted while those in col. 5 are person-weighted (larger regions have more influence).

Table 1: Summary statistics and national-level estimates

	Mean	Std. dev.	Min	Max	All
Panel A. Child					
Birth year	1972.5	4.5	1965.0	1980.0	
Income	199	146	94	77,074	
Share women	48%				
Panel B. Mother					
Birth year	1945.9	6.0	1925.0	1962.0	
Income	218	72	94	3,444	
Years of schooling	11.1	2.7	7.0	20.0	
Age at birth	26.5	4.5	18.0	40.0	
Panel C. Father					
Birth year	1943.5	6.3	1925.0	1962.0	
Income	322	150	97	20,642	
Years of schooling	10.8	3.0	7.0	20.0	
Age at birth	28.9	4.7	18.0	40.0	
Panel D. Family					
Family size	1.8	0.8	1.0	10.0	
Parents divorced	22%				
Parent died	11%				
Same parish	86%				
Panel E. Local labor markets					
IGE	0.18	0.04	0.04	0.34	0.23
Rank persistence	0.19	0.04	0.06	0.26	0.25
Sibling correlation	0.21	0.04	0.06	0.32	0.27
Absolute IOp	0.04	0.01	0.01	0.09	0.08
Relative IOp	0.24	0.06	0.05	0.38	0.39
Inequality (Gini)	0.18	0.01	0.15	0.24	0.20
N, main sample	8,548	19,419	110	182,857	1,077,046
N, sibling sample	6,087	13,712	80	129,044	767,005

Note: Panels A–D show summary statistics for the individual-level data, while Panel E shows summary statistics for the 126 local labor markets. The *All* column shows estimates and sample sizes for the full sample.

association is steeper. This results in substantially steeper slopes (stronger correlations) for the weighted compared to the unweighted regressions. The sibling correlations are noisier, showing more dispersion at the upper end than the other measures.

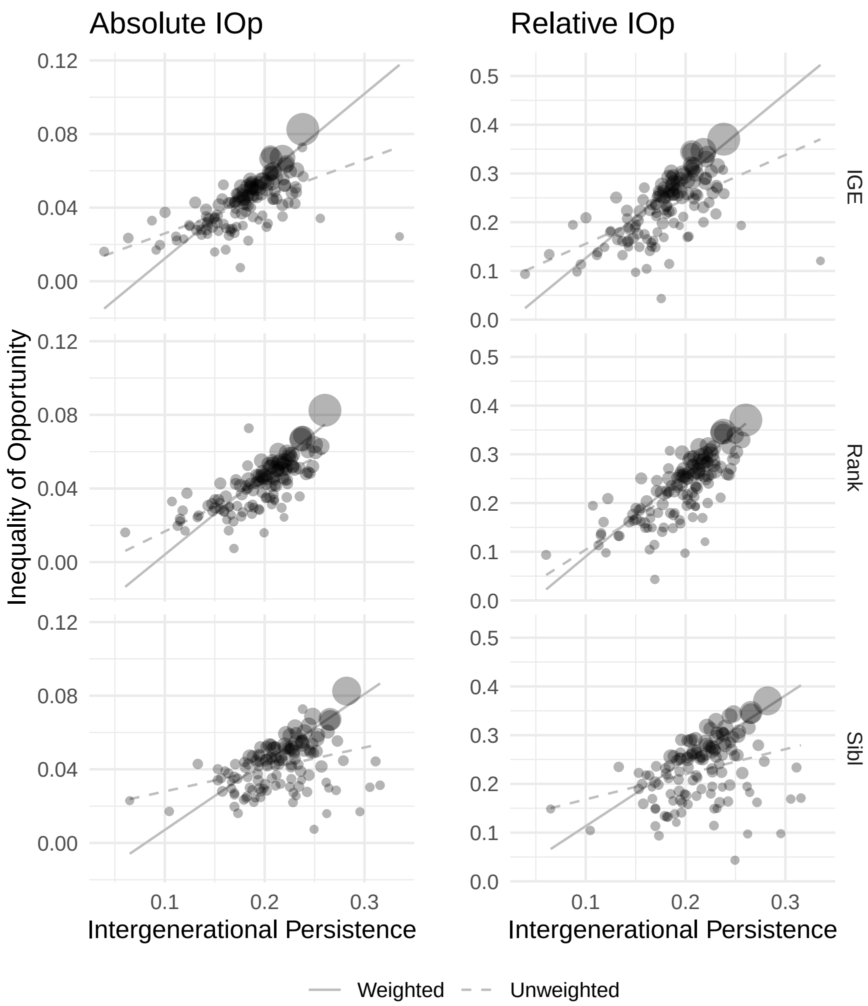

Figure 1: Relationship between inequality of opportunity and mobility measures

Notes: Each circle represents a local labor market. Circle size corresponds to the IGE sample size in each local labor market. Solid lines show OLS regressions weighted by the sample size, while dashed lines show unweighted regressions lines.

Panel A of Table 2 confirms these results. Each table entry shows the Pearson correlation coefficient between the estimated IOp and one of the intergenerational persistence measures (or sibling correlation), as listed in the column headings. The first three columns show absolute IOp, while the last three show relative IOp (i.e., the share of total inequality accounted for by observed circumstances). We present unweighted correlations in the first row, and correlations weighted by sample size in the second row.

Table 2: Main results

	Absolute IOp			Relative IOp		
	IGE	Rank	Sibl	IGE	Rank	Sibl
Panel A. All						
Unweighted	0.59	0.71	0.33	0.55	0.72	0.29
	(0.07)	(0.06)	(0.08)	(0.07)	(0.06)	(0.09)
Weighted	0.84	0.87	0.80	0.83	0.89	0.76
	(0.05)	(0.04)	(0.05)	(0.05)	(0.04)	(0.06)
Panel B. Men						
Unweighted	0.49	0.66	0.13	0.46	0.66	0.09
	(0.08)	(0.07)	(0.09)	(0.08)	(0.07)	(0.09)
Weighted	0.75	0.82	0.70	0.75	0.83	0.63
	(0.06)	(0.05)	(0.06)	(0.06)	(0.05)	(0.07)
Panel C. Women						
Unweighted	0.64	0.59	0.05	0.62	0.60	0.02
	(0.07)	(0.07)	(0.09)	(0.07)	(0.07)	(0.09)
Weighted	0.84	0.84	0.47	0.80	0.82	0.44
	(0.05)	(0.05)	(0.08)	(0.05)	(0.05)	(0.08)

Note: Each cell shows the correlation across 126 local labor markets between the measures of IOp and mobility indicated in the header, with standard errors in parentheses. "Weighted" rows show correlations weighted by the no. of observations of the local labor market. The IOp, IGE, and rank estimators use a larger sample than the sibling correlation: no. of observations is 1,077,046 and 767,005 respectively in Panel A.; 559,831 and 232,340 in Panel B.; and 517,215 and 201,137 in Panel C.

Overall, the two sets of measures are strongly correlated. The results are very similar for absolute and relative IOp. Further, the results are quite similar across the two intergenerational measures (IGE and rank-rank regression), while the sibling correlations are less correlated with IOp in the unweighted specifications.

The largest variation comes from whether we weight by region size or not. Weighed correlations, ranging from 0.8 to 0.9, are markedly higher than unweighted, ranging from 0.3 (for the sibling correlation) to 0.7. This is likely explained by the non-linear patterns observed Figure 1, where correlations appear stronger for larger regions. In Section 4.1, we show that this pattern arises due to a larger influence of sampling variation in smaller regions, rather than any true heterogeneity in the underlying processes. For this reason, we focus most of our remaining analyses on the weighted estimates.

Panels B and C of Table 2 show results separately for men and women (i.e., sons and daughters), respectively. The IOp-IGE correlations are slightly larger for women, while the associations between IOp and the sibling correlation are larger among men.

4.1 Region size and sampling variation

We already noted that the correlations between IOp and intergenerational measures differ substantially depending on whether we use weights or not when estimating our correlations. This observation would be consistent with the correlations being stronger among larger compared to smaller local labor markets. In this section, we examine whether this pattern is a true reflection of heterogeneity in processes of mobility and opportunity across large and small regions, or whether it is an artifact of noisier estimates in smaller samples.

In Table C.4, we split the regions at the median size, and estimate separate correlations for the larger and smaller regions. We find that *both* weighted and unweighted correlations are substantially lower for smaller regions. This large-vs-small difference is the largest for the sibling correlation, which is close to uncorrelated with IOp in small regions. Figure C.1 shows the relationship between local labor market size and our measures of intergenerational persistence and IOp. There appears to be a pattern with larger regions having lower intergenerational mobility and less equality of opportunity. Furthermore, the intergenerational and sibling measures are much more dispersed for smaller regions. This elevated dispersion could reflect sampling variation which introduces measurement-error induced attenuation bias in the correlations, and is consistent with the fact that the unweighted correlations were found to be considerably lower in our main analysis.

To probe this hypothesis further, we perform a set of analyses where we enforce small samples sizes for all regions. We sample 100 observations (with replacement) from each region, and estimate the correlations for this sample. The procedure is repeated 500 times. Table C.5 shows means and bootstrap standard errors for the correlations. We show unweighted and weighted (using sample sizes from the full sample) correlations, as well as correlations separately by larger and smaller regions (again, split using the full sample). Strikingly, these correlations are all much lower than our main estimates, on the order of 0.4–0.6. While the boostrap-based correlations for larger regions remain somewhat larger, the difference is small. Moreover, using our original weights based on labor-market size has virtually no impact on the bootstrap-based correlations.

We view these results as strong support for the hypothesis that the lower correlations for smaller regions are driven primarily by sampling variation, rather than reflecting a true feature of the structure of social mobility and opportunity across regions. The analysis also strengthens the case for focusing on weighted rather than unweighted correlations, as in the latter case the correlations will be more strongly attenuated by sampling variation among smaller regions.

4.2 Robustness to alternative specifications

Table 3 presents results from a set of alternative specifications. All panels show correlations weighted by sample size.

In our main specification, we use individualized *household income* as the outcome for the child generation. Panel A shows correlations when we instead use individual *labor income*.[25] The correlations fall somewhat across all measures, but remain high, around 0.6–0.8.

The main analyses use the Gini coefficient as inequality index underlying the IOp estimation. The results are basically unchanged if we instead use the mean logarithmic deviation (Panel B). Further, we use the Pearson correlation to measure the degree of association between the different measures of IOp and

[25] See Appendix A for precise definitions of these income measures.

Table 3: Alternative specifications

	Absolute IOp			Relative IOp		
	IGE	Rank	Sibl	IGE	Rank	Sibl
A. Individual incomes	0.69	0.79	0.60	0.71	0.81	0.59
	(0.07)	(0.05)	(0.07)	(0.06)	(0.05)	(0.07)
B. Mean Log Deviation	0.81	0.83	0.81	0.83	0.88	0.80
	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.05)
C. Spearman rank corr.	0.83	0.86	0.71	0.81	0.86	0.70
	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)
D. Excl. three largest cities	0.73	0.80	0.53	0.72	0.80	0.49
	(0.06)	(0.05)	(0.08)	(0.06)	(0.05)	(0.08)
E. Balanced sample	0.82	0.87	0.82	0.79	0.85	0.77
	(0.05)	(0.04)	(0.05)	(0.06)	(0.05)	(0.06)
F. No income restriction	0.79	0.82	0.53	0.72	0.83	0.48
	(0.06)	(0.05)	(0.08)	(0.06)	(0.05)	(0.08)
G. Income > one basic amt.	0.85	0.86	0.77	0.83	0.87	0.72
	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)
H. Cohorts, levels	0.64	0.79	0.70	0.58	0.76	0.69
	(0.21)	(0.16)	(0.19)	(0.22)	(0.17)	(0.19)
I. Cohorts, first differences	0.58	0.45	0.67	0.52	0.48	0.61
	(0.23)	(0.25)	(0.21)	(0.24)	(0.24)	(0.22)
J. County $\times$ cohort	0.73	0.72	0.69	0.72	0.73	0.65
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)

Note: Each cell shows the weighted correlation across local labor markets between the measures of IOp and mobility indicated in the header, with standard errors in parentheses. A. uses individual incomes as the outcome measure; B. uses the Mean Log Deviation instead of the Gini as the index of inequality; C. uses the Spearman rank correlation instead of the Pearson to estimate the correlations between IOp and mobility measures; D. excludes the three largest metropolitan areas: Stockholm, Gothenburg, and Malmö; E. uses a balanced sample with 706,589 obs. for each measure; F. removes the restriction on small incomes, while G. sets it at one basic amount; H. uses variation across 16 birth cohorts; I. uses first-differenced variation across birth cohorts; J. uses 384 county $\times$ cohort groups instead of local labor markets.

intergenerational persistence, but this estimator works best for linear associations. Given the non-linearities seen in Figure 1, a less restrictive estimator might better capture the relationship. To test this, we instead use the Spearman rank correlation in Panel C, with results essentially unchanged.

We might worry that the high correlations (especially in the weighted case, see Figure 1) are driven by the three major metropolitan areas in Sweden (Stockholm, Gothenburg, and Malmö). The correlations do fall slightly when we exclude them from the analysis (particularly for the sibling correlation), although the results are quite similar overall (Panel D).

To estimate sibling correlations, we require families with at least two children, and thus have to exclude singletons. This results in an unbalanced sample, where the IOp and intergenerational measures are estimated on a different (and larger) sample than the sibling correlation. Panel E reports estimates using a balanced sample which imposes all sample restrictions from both the main and sibling samples. Again, the results are basically unchanged.

The main sample excludes individuals with annual incomes below two price base amounts (see Appendix A for details). As Panels F–G show, the results are generally robust to dropping this low-income cut-off or setting the cut-off to one price base amount instead of two. The exception is the association between IOp measures and the sibling correlation, which falls to around 0.5 when no restriction is used.

Since parental income is included as a circumstance in the IOp calculations, we might worry that the correlations are driven by this factor alone. To test this, Panel H removes parental income from the set of circumstances, yielding, perhaps surprisingly, very similar results to the baseline specification.

The IOp estimates used so far can be viewed as *lower bounds* on the true level of IOp. In Appendix B, we show results using the *upper bound* estimator of Niehues and Peichl (2014). This approach yields very similar results for absolute IOp, whereas the *relative* upper bound IOp is weakly negatively correlated with both the mobility and the lower bound IOp estimates. However, given the strong assumptions that this approach is based on, we treat these results with a large amount of caution.

4.2.1 Different circumstance variables

It is interesting to study how different sets of circumstances impact the correlations. Table C.3 shows correlations between our mobility measures and IOp, estimated using different sets of circumstances. Column (4) is our baseline specification, and thus reproduces Panel A of Table 2. The results are robust to varying the set of circumstances (Columns (1)–(4)), including adding gender as a circumstance (Column (5)). Columns (7)–(8) show estimates using only men for whom we can observe cognitive and non-cognitive skills measured at military enlistment tests. Column (7) shows the baseline specification for this subsample, while column (8) adds skills to the set of circumstances in the IOp estimation.[26] Adding these skill measures changes the correlations only marginally.

[26]It is debatable whether one should see these measures as circumstances, effort, or a combination of both. See for example Björklund, Jäntti, and Roemer (2012), who provide arguments for their inclusion as circumstances.

4.2.2 Using variation over time

As an alternative to spatial variation, we explore variation over time in the associations between estimates of IOp and intergenerational persistence. Figure 2 shows time series plots of our measures, estimated at the national level. The measures appear to co-move over cohorts, with a non-trivial increase (less mobility, EOp, etc.) over the first five birth cohorts and a small subsequent reversion (more mobility, EOp, etc.) starting from the 1970 birth cohort.

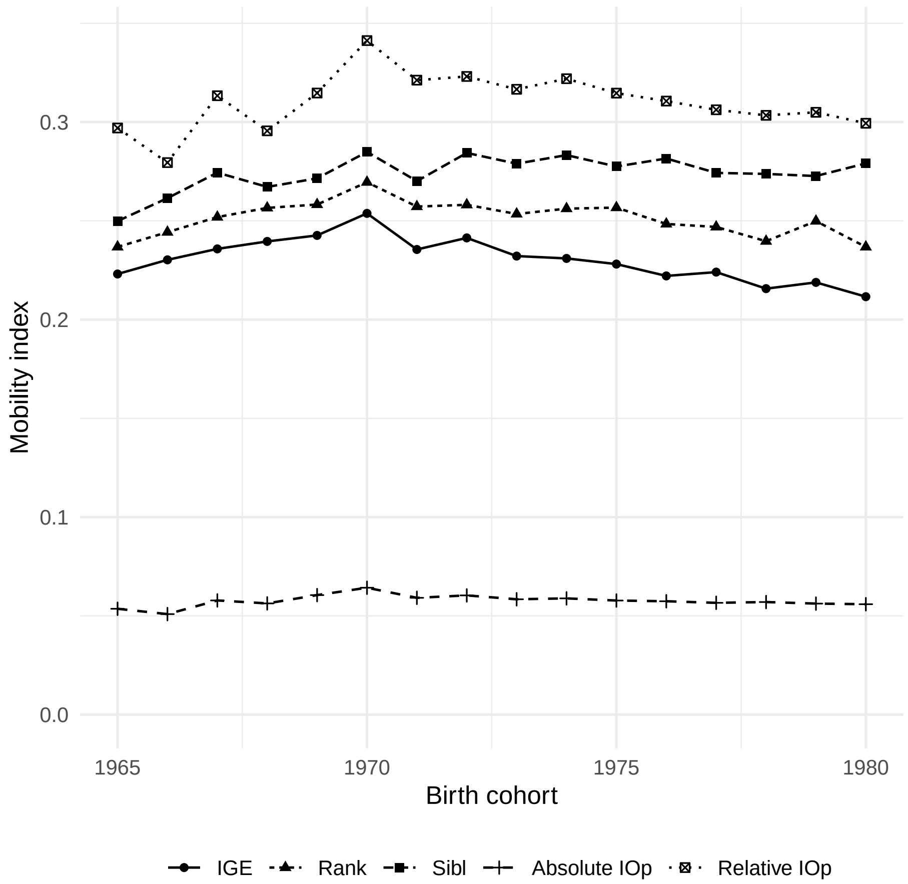

Figure 2: Relationship between inequality of opportunity and mobility measures

Notes: Each point shows the given measure estimated using the national cross-section for a given birth cohort.

In Panel I of Table 3, we show correlations between the measures using variation between cohorts. Since there is much less variation in cohort sizes than in region sizes, we only show unweighted estimates. Across cohorts, both absolute and relative IOp measures are highly correlated with the intergenerational measures, with correlations ranging from 0.6–0.8 (although it should be noted that these estimates are considerably less precise than those based on variation across regions).

These correlations might be driven by linear trends. A more stringent test is to estimate correlations in changes (by taking first differences of all variables). We show such results in Panel J of Table 3. This reduces the correlations somewhat (and increases the standard errors), but all correlations remain substantially positive, around 0.5–0.6.

Finally, in Panel K we replace the 126 local labor markets with 384 groups formed by interacting Sweden's 24 counties at the time with the 16 birth cohorts in our data, and perform our analyses across these groups. This results in a larger number of groups while reducing variation in sample sizes across groups.[27] Correlations drop slightly, but remain substantial at around 0.7.

5 Conclusions

The study of social mobility is often motivated with reference to the normative concept of *equality of opportunity* (EOp). However, it is not clear a priori how well EOp is actually captured by estimates of social mobility. The purpose of this paper is to provide a bridge between the different concepts. To this end, we estimate a set of intergenerational measures (IGEs, rank correlations) and sibling correlations, along with indices of inequality of opportunity, for each of 126 Swedish local labor markets as well as over cohorts. We then calculate how strongly these different measures correlate across regions and over time.

Our findings suggest that the *variation* in measures of IOp and intergenerational persistence is intimately related. First we show that, using income as the outcome of interest, the intergenerational measures (elasticity and rank correlation) correlate very strongly with inequality of opportunity (IOp) indices across Swedish regions, while the sibling correlation is only slightly less strongly correlated with IOp measures.

Moreover, the strong associations between IOp and intergenerational persistence (in income) is not driven by a mechanical role of parental income in the IOp indices. As we show, the various measures remain strongly correlated also when parental income is excluded from the set of circumstances underlying the IOp, as well as in a number of different robustness analyses. Finally, we study correlations across birth cohorts nationally to use variation over time instead of across space. While the correlations are somewhat smaller and less precisely estimated in these analyses, they are still substantial.

We want to emphasize, however, that the various measures we study provide quite different answers to the key question of what share of total inequality that can be attributed to family-background factors. This implied share is substantially higher for the sibling correlation and the IOp indices than what is implied by intergenerational estimates. But while the *levels* of the measures thus can have vastly different interpretations, our analysis emphasizes that *differences* in the various measures across regions (or cohorts) correlate strongly. Because the literatures in question are primarily comparative, studying variation across countries or over time, this is an important insight.

Taken together, our findings suggest that the estimators of intergenerational mobility typically used in the empirical literature can indeed be used to say something about variation in equality of opportunity, as the two concepts are

[27]Sample sizes for the county $\times$ cohort groups vary between 477 and 12,055, with a mean of 2,805. For the local labor markets, sample sizes vary between 110 and 182,857, with a mean of 8,548.

strongly correlated across both space and time. However, more research on this question would be valuable. We recognize that the landscape of Swedish local labor markets constitute a quite specific context, and that the patterns might differ across countries or over longer time periods.

References

Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso (2018). "Intergenerational Mobility and Preferences for Redistribution". *American Economic Review* 108(2), 521–554.

Arneson, Richard J. (1989). "Equality and Equal Opportunity for Welfare". *Philosophical Studies* 56(1), 77–93.

Becker, Gary S. and Nigel Tomes (1979). "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility". *Journal of Political Economy* 87(6), 1153–1189.

— (1986). "Human Capital and the Rise and Fall of Families". *Journal of Labor Economics* 4(3), S1–S39.

Bingley, Paul and Lorenzo Cappellari (2018). "Correlation of Brothers' Earnings and Intergenerational Transmission". *The Review of Economics and Statistics*.

Björklund, Anders, Tor Eriksson, et al. (2002). "Brother Correlations in Earnings in Denmark, Finland, Norway and Sweden Compared to the United States". *Journal of Population Economics* 15(4), 757–772.

Björklund, Anders, Eriksson Karin Hederos, and Markus Jäntti (2010). "IQ and Family Background: Are Associations Strong or Weak?" *The B.E. Journal of Economic Analysis & Policy* 10(1).

Björklund, Anders and Markus Jäntti (1997). "Intergenerational Income Mobility in Sweden Compared to the United States". *The American Economic Review* 87(5), 1009–1018.

— (2009). "Intergenerational Income Mobility and the Role of Family Background". In: *Oxford Handbook of Economic Inequality*. Ed. by Wiemer Salverda, Brian Nolan, and Timothy M. Smeeding. Oxford: Oxford University Press, 491–521.

— (2012). "How Important Is Family Background for Labor-Economic Outcomes?" *Labour Economics* 19(4), 465–474.

— (2020). "Intergenerational Mobility, Intergenerational Effects, Sibling Correlations, and Equality of Opportunity: A Comparison of Four Approaches". *Research in Social Stratification and Mobility*, 100455.

Björklund, Anders, Markus Jäntti, and John E. Roemer (2012). "Equality of Opportunity and the Distribution of Long-Run Income in Sweden". *Social Choice and Welfare* 39(2), 675–696.

Black, Sandra E. and Paul J. Devereux (2011). "Recent Developments in Intergenerational Mobility". In: *Handbook of Labor Economics*. Ed. by Orley Ashenfelter and David Card. Vol. 4B. Amsterdam: Elsevier, 1487–1541.

Blundell, Jack and Erling Risa (2019). "Income and Family Background: Are We Using the Right Models?" SSRN Scholarly Paper 3269576.

Bourguignon, François, Francisco H. G. Ferreira, and Marta Menéndez (2007). "Inequality of Opportunity in Brazil". *Review of Income and Wealth* 53(4), 585–618.

Breen, Richard, Carina Mood, and Jan O. Jonsson (2016). "How Much Scope for a Mobility Paradox? The Relationship between Social and Income Mobility in Sweden". *Sociological Science* 3, 39–60.

Breiman, Leo (2001). "Random Forests". *Machine Learning* 45(1), 5–32.

Breiman, Leo et al. (1984). *Classification and Regression Trees*. The Wadsworth Statistics/Probability Series. New York: Chapman and Hall/CRC.

Brunori, Paolo, Francisco H. G. Ferreira, and Vito Peragine (2013). “Inequality of Opportunity, Income Inequality, and Economic Mobility: Some International Comparisons”. In: *Getting Development Right.* Ed. by Eva Paus. Palgrave Macmillan US, 85–115.

Brunori, Paolo, Paul Hufe, and Daniel Mahler (2023). “The Roots of Inequality: Estimating Inequality of Opportunity from Regression Trees and Forests”. *The Scandinavian Journal of Economics* 125(4), 900–932.

Carranza, Rafael (2023). “Upper and Lower Bound Estimates of Inequality of Opportunity: A Cross-National Comparison for Europe”. *Review of Income and Wealth* 69(4), 838–860.

Checchi, Daniele and Vito Peragine (2010). “Inequality of Opportunity in Italy”. *The Journal of Economic Inequality* 8(4), 429–450.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez (2014). “Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States”. *The Quarterly Journal of Economics* 129(4), 1553–1623.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner (2014). “Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility”. Working Paper 19844. National Bureau of Economic Research.

Cohen, G. A. (1989). “On the Currency of Egalitarian Justice”. *Ethics* 99(4), 906–944.

Corak, Miles (2013). “Income Inequality, Equality of Opportunity, and Intergenerational Mobility”. *The Journal of Economic Perspectives* 27(3), 79–102.

Corcoran, Mary et al. (1990). “Effects of Family and Community Background on Economic Status”. *The American Economic Review* 80(2), 362–366.

Deutscher, Nathan and Bhashkar Mazumder (2023). “Measuring Intergenerational Income Mobility: A Synthesis of Approaches”. *Journal of Economic Literature* 61(3), 988–1036.

Ferreira, Francisco H. G. and Jérémie Gignoux (2011). “The Measurement of Inequality of Opportunity: Theory and an Application to Latin America”. *Review of Income and Wealth* 57(4), 622–657.

Ferreira, Francisco H. G. and Vito Peragine (2016). “Individual Responsibility and Equality of Opportunity”. In: *The Oxford Handbook of Well-Being and Public Policy.* Ed. by Matthew D. Adler and Marc Fleurbaey. Oxford: Oxford University Press, 746–784.

Fishkin, Joseph (2014). *Bottlenecks: A New Theory of Equal Opportunity.* Oxford: Oxford University Press.

Galton, Francis (1886). “Regression Towards Mediocrity in Hereditary Stature.” *The Journal of the Anthropological Institute of Great Britain and Ireland* 15, 246–263.

Haider, Steven and Gary Solon (2006). “Life-Cycle Variation in the Association between Current and Lifetime Earnings”. *American Economic Review* 96(4), 1308–1320.

Hällsten, Martin (2014). “Inequality Across Three and Four Generations in Egalitarian Sweden: 1st and 2nd Cousin Correlations in Socio-Economic Outcomes”. *Research in Social Stratification and Mobility.* Inequality Across Multiple Generations 35, 19–33.

Hederos, Karin, Markus Jäntti, and Lena Lindahl (2017). "Gender and Inequality of Opportunity in Sweden". *Social Choice and Welfare* 49(3), 605–635.

Heidrich, Stefanie (2017). "Intergenerational Mobility in Sweden: A Regional Perspective". *Journal of Population Economics* 30(4), 1241–1280.

Hothorn, Torsten, Kurt Hornik, Carolin Strobl, et al. (2023). *Party: A Laboratory for Recursive Partytioning.*

Hothorn, Torsten, Kurt Hornik, and Achim Zeileis (2006). "Unbiased Recursive Partitioning: A Conditional Inference Framework". *Journal of Computational and Graphical Statistics* 15(3), 651–674.

Hufe, Paul, Ravi Kanbur, and Andreas Peichl (2022). "Measuring Unfair Inequality: Reconciling Equality of Opportunity and Freedom from Poverty". *The Review of Economic Studies* 89(6), 3345–3380.

Hufe, Paul, Andreas Peichl, and Daniel Weishaar (2022). "Lower and Upper Bound Estimates of Inequality of Opportunity for Emerging Economies". *Social Choice and Welfare* 58(3), 395–427.

Jäntti, Markus and Stephen P. Jenkins (2015). "Income Mobility". In: *Handbook of Income Distribution.* Ed. by François Bourguignon and Anthony B. Atkinson. Vol. 2. Handbook of Income Distribution. Elsevier, 807–935.

Jencks, Christopher and Laura Tach (2006). "Would Equal Opportunity Mean More Mobility?" In: *Mobility and Inequality: Frontiers of Research in Sociology and Economics.* Ed. by Stephen L Morgan, David B Grusky, and Gary S Fields. Studies in Social Inequality. Stanford: Stanford University Press, 23–58.

Kanbur, Ravi and Adam Wagstaff (2016). "How Useful Is Inequality of Opportunity as a Policy Construct?" In: *Inequality and Growth: Patterns and Policy: Volume I: Concepts and Analysis.* Ed. by Kaushik Basu and Joseph E. Stiglitz. London: Palgrave Macmillan UK, 131–150.

Lefranc, Arnaud, Nicolas Pistolesi, and Alain Trannoy (2008). "Inequality of Opportunities Vs. Inequality of Outcomes: Are Western Societies All Alike?" *Review of Income and Wealth* 54(4), 513–546.

Mazumder, Bhashkar (2005). "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data". *Review of Economics and Statistics* 87(2), 235–255.

— (2008). "Sibling Similarities and Economic Inequality in the US". *Journal of Population Economics* 21(3), 685–701.

Mogstad, Magne and Gaute Torsvik (2023). "Family Background, Neighborhoods, and Intergenerational Mobility". In: *Handbook of the Economics of the Family.* Ed. by Shelly Lundberg and Alessandra Voena. Vol. 1. North-Holland, 327–387.

Niehues, Judith and Andreas Peichl (2014). "Upper Bounds of Inequality of Opportunity: Theory and Evidence for Germany and the US". *Social Choice and Welfare* 43(1), 73–99.

Nybom, Martin and Jan Stuhler (2016). "Heterogeneous Income Profiles and Lifecycle Bias in Intergenerational Mobility Estimation". *Journal of Human Resources* 51(1), 239–268.

— (2017). "Biases in Standard Measures of Intergenerational Income Dependence". *Journal of Human Resources* 52(3), 800–825.

Page, Marianne and John E. Roemer (2001). "The U.S. Fiscal System as an Opportunity-Equalizing Device". In: *Inequality and Tax Policy.* Ed. by Kevin A. Hassett and R. Glenn Hubbard. Washington DC: The AEI Press, 134–159.

Raaum, Oddbjørn, Kjell G. Salvanes, and Erik Ø. Sørensen (2006). "The Neighbourhood Is Not What It Used to Be". *The Economic Journal* 116(508), 200–222.

Rieger, Anna, Torsten Hothorn, and Carolin Strobl (2010). "Random Forests with Missing Values in the Covariates". Technical Report 79. Munich: Department of Statistics, University of Munich.

Roemer, John E. (1993). "A Pragmatic Theory of Responsibility for the Egalitarian Planner". *Philosophy & Public Affairs* 22(2), 146–166.

— (1998). *Equality of Opportunity.* Cambridge, MA, USA: Harvard University Press.

— (2004). "Equal Opportunity and Intergenerational Mobility: Going beyond Intergenerational Income Transition Matrices". In: *Generational Income Mobility in North America and Europe.* Cambridge University Press.

Roemer, John E., Rolf Aaberge, et al. (2003). "To What Extent Do Fiscal Regimes Equalize Opportunities for Income Acquisition among Citizens?" *Journal of Public Economics* 87(3), 539–565.

Roemer, John E. and Alain Trannoy (2015). "Equality of Opportunity". In: *Handbook of Income Distribution.* Ed. by Anthony B. Atkinson and François Bourguignon. Vol. 2. Handbook of Income Distribution. Elsevier, 217–300.

— (2016). "Equality of Opportunity: Theory and Measurement". *Journal of Economic Literature* 54(4), 1288–1332.

Solon, Gary (1992). "Intergenerational Income Mobility in the United States". *The American Economic Review* 82(3), 393–408.

— (1999). "Intergenerational Mobility in the Labor Market". In: *Handbook of Labor Economics.* Ed. by Orley C. Ashenfelter and David Card. Vol. 3, Part A. Elsevier, 1761–1800.

— (2004). "A Model of Intergenerational Mobility Variation Over Time and Place". In: *Generational Income Mobility in North America and Europe.* Cambridge University Press.

Solon, Gary et al. (1991). "A Longitudinal Analysis of Sibling Correlations in Economic Status". *The Journal of Human Resources* 26(3), 509–534.

Statistics Sweden (2010). "Construction and Use of Labour Market Areas in Sweden". Report. Örebro: Enterprise- and Register-based Employment Statistics Unit.

Swedish National Mediation Office (2021). "Avtalsrörelsen Och Lönebildningen 2020". Medlingsinstitutets Årsrapport. Stockholm.

Swift, Adam (2004). "Would Perfect Mobility Be Perfect?" *European Sociological Review* 20(1), 1–11.

Torche, Florencia (2015). "Intergenerational Mobility and Equality of Opportunity". *European Journal of Sociology / Archives Européennes de Sociologie* 56(03), 343–371.

Ward, Zachary (2023). "Intergenerational Mobility in American History: Accounting for Race and Measurement Error". *American Economic Review* 113(12), 3213–3248.

Zimmerman, David J. (1992). "Regression Toward Mediocrity in Economic Stature". *The American Economic Review* 82(3), 409–429.

A Variable definition details

Individualized disposable incomes are calculated by multiplying household disposable income by an adjustment factor, which takes two slightly different forms. For the period 1990–2004, this factor is calculated by dividing the individual's consumption weight by the sum of each family member's consumption weights; and for the (partly overlapping) period 1998–2021, it is calculated by dividing one by the sum of family consumption weights. We use the *dispinkpersf* variable for the earlier period, and the *dispinkke* variable for the later, both from the LISA register. For the overlapping period 1998–2004, when both definitions are available, we take the average.

The consumption weights are defined as follows: 1.16 for one adult, 1.92 for two adults, and 0.96 for each additional adult; 0.56 for children aged 0–3, 0.66 for children aged 4–10, and 0.76 for children aged 11–17. To illustrate, in a family of two adults and two children aged 3 and 5, the sum of family consumption weights is $1.92 + 0.56 + 0.66 = 3.14$. The earlier definition then gives an adjustment factor $0.96/3.14 = 0.31$ (where the numerator comes from dividing 1.92 by two), while the newer definition gives $1/3.14 = 0.32$.

Households are defined through individuals with a family relationship (married, registered partners, cohabiting with common children, parents, and guardians) who are registered as residents of the same property. Cohabiting unmarried couples with no children cannot be linked, and so appear as single households in our data.

For individual labor income, we use the following registers and variables: from the income and taxation (IoT) register, we use the sum of *injo*, *inro*, *intj*, and *sjoin* for 1968; the sum of *ainjo*, *ainro*, *aintj*, and *sjoin* for 1971, 73, and 76; and the *arbink* variable for 1979 and 82. We also use *arbink* from the 1970, 75, and 80 censuses (FoB). For 1985–1989, we take the sum of *loneink*, *fink*, and *arbers* from the LOUISE register; and for 1990–2021, we use *forvers* from the LISA register.

Information on highest level of completed education comes from the LISA register for the years 1990–2021. We translate this into years of schooling as follows: old primary school = 7 years; new primary school = 9 years; short high school = 11 years; long high school = 12 years; short tertiary education = 14 years; long tertiary education = 16 years; and Ph.D. = 20 years.

We also use data on highest completed level of education from the 1960 and 70 censuses. The 1970 census has a clearly defined coding scheme that we translate directly to years of schooling. For the 1960 census there is a variable for level of education, but due to lacking documentation it cannot be directly translated into years of schooling. To circumvent this, we exploit the panel structure of our data to impute years of schooling as follows: for each level in the 1960 variable, we calculate the modal value from the 1970 years of schooling variable using all individuals who were observed in the given category in 1960.

We use local labor markets for 1985 as the geographic units of observation in the main analyses. These are defined by grouping municipalities according to observed commuting patterns. The local labor markets were created by Statistics Sweden in a conscious effort to form local labor market regions suitable for economic analysis (Statistics Sweden 2010).

Family size is measured as the mother's total number of biological children. We observe this in the 2022 multigenerational register, when all mothers in our sample are at least 60 years old.

B Upper bounds on IOp

In this section, we implement the *upper bound* IOp estimator of Niehues and Peichl (2014) (see also Carranza 2023; Hufe, Peichl, and Weishaar 2022, for cross-country applications). Their estimator relies on the argument that an individual's circumstances are time-invariant, and can therefore be captured by individual fixed effects estimated from income panels. This approach relies on (at least) two key features/assumptions. The first is that circumstances do not change over time, which would for example be violated if the *effect of* fixed circumstances vary over time or due to genuinely time-varying factors, such as macro-economic shocks. We deal with this concern by including time fixed effects in our regressions.

The second feature is that efforts tend to *not be fixed*, so that the fixed effects primarily capture the influence of circumstances. This is clearly a strong "assumption" and the reason for why the approach merely yields upper bounds. If the assumption fails, then the estimated upper bounds become rather uninformative. Furthermore, in our case there is a risk that the role of such fixed efforts varies systematically across local labor markets, which would bias our correlations in unknown ways. We thus treat these analyses with a great deal of caution.

In a first step, we regress yearly log incomes for years $t \neq s$ on a set of individual and time fixed effects:

$$y_{it} = c_i + u_t + \eta_{it}. \tag{7}$$

In a second step, we then regress log incomes in year s on the estimated individual fixed effects:

$$y_{is} = \psi \hat{c}_i + \nu_{it}, \tag{8}$$

and finally we form predictions from this regression as our measures of the counterfactual incomes due to circumstances: $\hat{Y}_i^C = \exp(\hat{\psi}\hat{c}_i)$.

In our application, we use incomes at ages 30–39 to estimate Equation (7), and incomes at age 40 to estimate Equation (8). We then form estimates of absolute and relative IOp as in Equations (5) and (6).

Table B.1 shows correlations between the intergenerational measures and lower and upper bounds of IOp, where the lower bounds are the conditional inference forest estimates from the main specification. For absolute IOp, the results are remarkably similar whether we use lower or upper bounds. This breaks down, however, when we use relative IOp. Now the IOp upper bound is negatively correlated with the intergenerational measures. Table B.2 further shows that the estimates from the different IOp measures (upper vs lower bound, relative vs absolute) are all positively associated, apart from the upper-bound estimates of relative IOp. Surprisingly, these are negatively correlated not only with the intergenerational measures but also the other estimates of IOp.

C Additional results

Table B.1: IOp bounds

	Absolute IOp		Relative IOp	
	Lower	Upper	Lower	Upper
IGE	0.84	0.75	0.83	-0.29
	(0.05)	(0.06)	(0.05)	(0.09)
Rank	0.87	0.76	0.89	-0.27
	(0.04)	(0.06)	(0.04)	(0.09)
Sibl	0.80	0.76	0.76	-0.29
	(0.05)	(0.06)	(0.06)	(0.09)

Note: Each cell shows the correlation across 126 local labor markets between the measures of IOp and mobility indicated in the header, with standard errors in parentheses. Correlations are weighted by the no. of observations of the local labor market. The IOp, IGE, and rank estimators use a larger sample than the sibling correlation: no. of observations is 1,077,046 and 767,005 respectively.

Table B.2: Correlations, IOp measures

	Lower		Upper
	Absolute IOp	Relative IOp	Absolute IOp
Lower			
Relative IOp	0.97		
	(0.02)		
Upper			
Absolute IOp	0.92	0.79	
	(0.04)	(0.05)	
Relative IOp	-0.35	-0.29	-0.34
	(0.08)	(0.09)	(0.08)

Note: Each cell shows the correlation across 126 local labor markets between the different measures of IOp, with standard errors in parentheses. Correlations are weighted by the no. of observations of the local labor market. Total no. of observations is 1,077,046 .

Table C.1: Summary statistics, sibling sample

	Mean	Std. dev.	Min	Max
Panel A. Child				
Birth year	1972.6	4.2	1965.0	1980.0
Income	199	123	94	37,230
Share women	48%			
Panel B. Mother				
Birth year	1946.3	4.9	1925.0	1962.0
Income	217	70	94	2,040
Years of schooling	11.1	2.7	7.0	20.0
Age at birth	26.2	4.3	18.0	40.0
Panel C. Father				
Birth year	1943.9	5.2	1925.0	1962.0
Income	322	151	97	20,642
Years of schooling	10.8	3.0	7.0	20.0
Age at birth	28.6	4.5	18.0	40.0
Panel D. Family				
Family size	2.3	0.5	2.0	10.0
Parents divorced	22%			
Parent died	12%			
Same parish	88%			

Note: Table shows summary statistics for the individual-level data in the siblings sample.

Table C.2: IOp estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. National IOp								
Gini								
Absolute IOp	0.069	0.074	0.075	0.077	0.080	0.071	0.076	0.083
Relative IOp	0.353	0.376	0.383	0.393	0.409	0.362	0.385	0.421
Inequality	0.197	0.197	0.197	0.197	0.197	0.197	0.196	0.196
Mean log deviation								
Absolute IOp	0.008	0.009	0.009	0.009	0.010	0.008	0.009	0.011
Relative IOp	0.129	0.136	0.144	0.147	0.157	0.120	0.143	0.166
Inequality	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.065
Panel B. Mean IOp								
Gini								
Absolute IOp	0.048	0.050	0.050	0.042	0.043	0.036	0.036	0.042
Relative IOp	0.264	0.275	0.273	0.227	0.236	0.196	0.195	0.231
Inequality	0.182	0.182	0.182	0.182	0.182	0.182	0.181	0.181
Mean log deviation								
Absolute IOp	0.004	0.004	0.004	0.003	0.003	0.002	0.002	0.003
Relative IOp	0.078	0.079	0.077	0.055	0.058	0.042	0.044	0.057
Inequality	0.054	0.054	0.054	0.054	0.054	0.054	0.054	0.054
Circumstances								
Parental income	✓	✓	✓	✓	✓		✓	✓
Parental education		✓	✓	✓	✓	✓	✓	✓
Parental occupation			✓	✓	✓	✓	✓	✓
Family characteristics				✓	✓	✓	✓	✓
Gender					✓			
Skills								✓
No. of observations	1,077,046	1,077,046	1,077,046	1,077,046	1,077,046	1,077,046	501,591	501,591

Note: The table shows IOp and inequality estimates, using the *Gini* or *mean logarithmic deviation* as the inequality index. Panel A. shows estimates at the national level, while Panel B. shows averages across local labor markets. *Parental income*, *education*, and *occupation* includes the variable for both parents separately. *Family characteristics* includes family size, both parents' year of birth and age when the child was born, and indicators for early parental death, divorce during childhood, and living in the same parish as both parents during childhood.

Table C.3: Circumstances

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Absolute IOp								
IGE	0.86	0.84	0.84	0.84	0.83	0.80	0.73	0.72
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)
Rank persistence	0.86	0.86	0.86	0.88	0.87	0.87	0.81	0.80
	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
Sibling correlation	0.77	0.79	0.79	0.81	0.81	0.79	0.68	0.69
	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.07)	(0.07)
Panel B. Relative IOp								
IGE	0.89	0.84	0.84	0.82	0.81	0.76	0.73	0.70
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)
Rank persistence	0.90	0.89	0.89	0.88	0.87	0.85	0.81	0.80
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)
Sibling correlation	0.73	0.74	0.76	0.77	0.76	0.73	0.62	0.62
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)
Circumstances								
Parental income	✓	✓	✓	✓	✓		✓	✓
Parental education		✓	✓	✓	✓	✓	✓	✓
Parental occupation			✓	✓	✓	✓	✓	✓
Family characteristics				✓	✓	✓	✓	✓
Gender					✓			
Skills								✓
No. of observations	1,077,046	1,077,046	1,077,046	1,077,046	1,077,046	1,077,046	501,591	501,591

Note: *Parental income*, *education*, and *occupation* includes the variable for both parents separately. *Family characteristics* includes family size, both parents' year of birth and age when the child was born, and indicators for early parental death, divorce during childhood, and living in the same parish as both parents during childhood.

Table C.4: By sample size

	Absolute IOp			Relative IOp		
	IGE	Rank	Sibl	IGE	Rank	Sibl
Panel A. N $\leq$ 3,396						
Unweighted	0.41	0.44	0.10	0.34	0.46	0.05
	(0.12)	(0.12)	(0.12)	(0.12)	(0.11)	(0.12)
Weighted	0.56	0.53	0.22	0.51	0.55	0.19
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Panel B. N $>$ 3,396						
Unweighted	0.71	0.77	0.59	0.70	0.79	0.50
	(0.09)	(0.08)	(0.12)	(0.09)	(0.08)	(0.12)
Weighted	0.86	0.88	0.85	0.85	0.89	0.83
	(0.07)	(0.06)	(0.08)	(0.07)	(0.06)	(0.08)

Note: Each cell shows the correlation across local labor markets (LLM) between the measures of IOp and mobility indicated in the header, with standard errors in parentheses. "Weighted" rows show correlations weighted by the no. of observations of the LLM. The sample has been split at the median LLM size in two parts, with 63 LLMs each. The smaller LLMs are shown in Panel A., and the larger in Panel B. The IOp, IGE, and rank estimators use a different sample than the sibling correlation: no. of observations is 99,826 and 109,031 respectively in Panel A.; and 977,220 and 657,974 in Panel B.

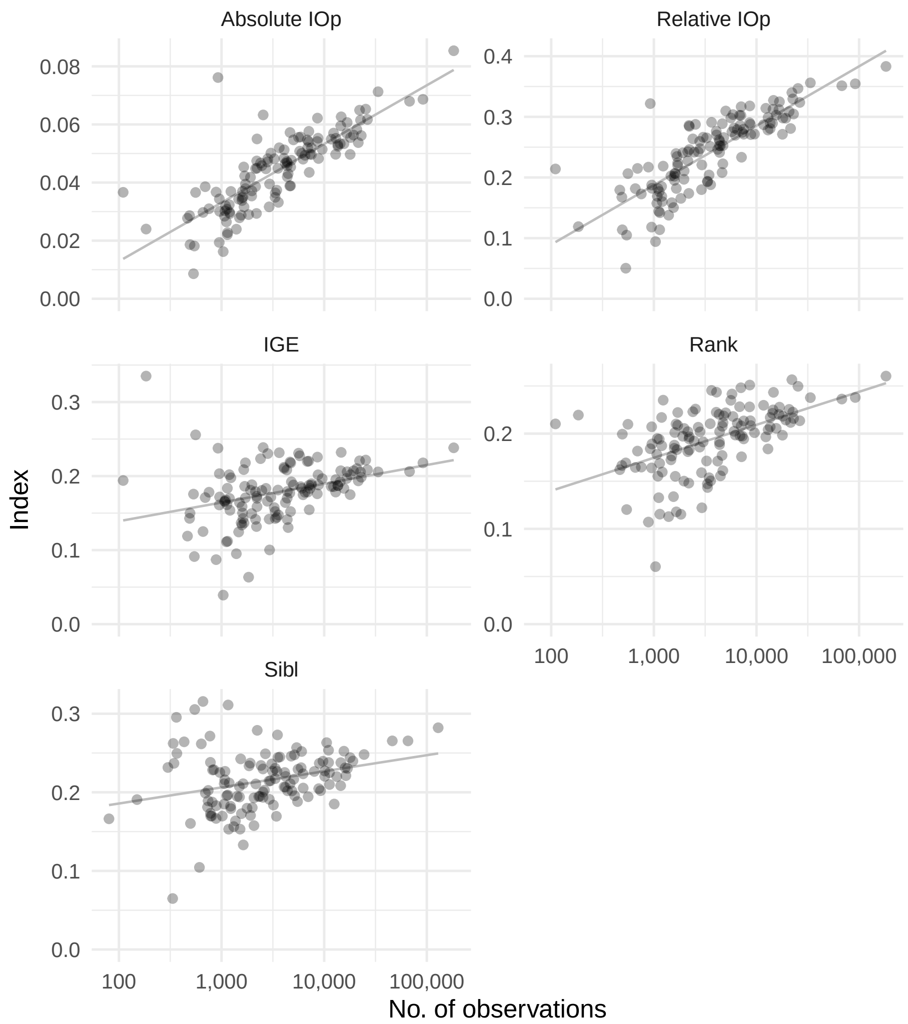

Figure C.1: Relationship between IOp and mobility indices and region size

Notes: Each circle represents a local labor market. Lines show OLS regressions of the indicated persistence or IOp measures on sample size.

Table C.5: Bootstrap simulation, 100 obs./LLM

	Absolute IOp		Relative IOp	
	IGE	Rank	IGE	Rank
A. Unweighted	0.54	0.42	0.52	0.45
	(0.08)	(0.09)	(0.07)	(0.07)
B. Weighted	0.56	0.45	0.53	0.45
	(0.14)	(0.17)	(0.14)	(0.16)
C. N $\leq$ 3,396	0.52	0.40	0.49	0.42
	(0.11)	(0.11)	(0.10)	(0.10)
D. N $>$ 3,396	0.56	0.45	0.53	0.47
	(0.10)	(0.10)	(0.10)	(0.09)

Note: The table shows versions of the main results where we enforce a uniform sample size of 100 observations from each local labor market by sampling with replacement. The procedure is repeated 500 times, and each cell shows mean correlations and standard errors from the bootstrap distributions. Panel A. shows unweighted correlations, while Panel B. shows correlations weighted using sample sizes from the full data. Panels C. and D. show correlations separately for smaller and larger local labor markets (in the original data), as in Table C.4.