

Discussion Paper No. 15-084

**Inequality of Income Acquisition:  
The Role of Childhood Circumstances**

Paul Hufe, Andreas Peichl, John Roemer,  
and Martin Ungerer

**ZEW**

Zentrum für Europäische  
Wirtschaftsforschung GmbH

Centre for European  
Economic Research

Discussion Paper No. 15-084

## **Inequality of Income Acquisition: The Role of Childhood Circumstances**

Paul Hufe, Andreas Peichl, John Roemer,  
and Martin Ungerer

Download this ZEW Discussion Paper from our ftp server:

**<http://ftp.zew.de/pub/zew-docs/dp/dp15084.pdf>**

Die Discussion Papers dienen einer möglichst schnellen Verbreitung von  
neueren Forschungsarbeiten des ZEW. Die Beiträge liegen in alleiniger Verantwortung  
der Autoren und stellen nicht notwendigerweise die Meinung des ZEW dar.

---

Discussion Papers are intended to make results of ZEW research promptly available to other  
economists in order to encourage discussion and suggestions for revisions. The authors are solely  
responsible for the contents which do not necessarily represent the opinion of the ZEW.

# Inequality of income acquisition: The role of childhood circumstances<sup>+</sup>

Paul Hufe<sup>1</sup>, Andreas Peichl<sup>2</sup>, John Roemer<sup>3</sup>, Martin Ungerer<sup>4</sup>

23 December 2015

**Abstract:** Many studies have estimated the effect of circumstances on income acquisition. Perhaps surprisingly, the fraction of inequality attributable to circumstances is usually quite small – in the advanced democracies, on the order of 20%. One reason for this is the lack of data on circumstance variables in empirical research. Here, we argue that all behaviors and accomplishments of children should be considered the consequence of circumstances: That is, an individual should not be considered to be responsible for her choices before an age of consent is reached. Using two data sets that contain data on childhood accomplishments, other environmental circumstances, and the income as an adult, we compute that the fraction of income inequality due to circumstances in the US is over 45%, and in the UK it is over 31%.

JEL Codes: D63, D3

**Keywords:** Equality of Opportunity; Earnings Inequality; Lower bounds; Early childhood achievements

---

<sup>+</sup> We are grateful to the Russell Sage Foundation for support. We thank participants of seminars in Mannheim and a workshop in London for helpful comments and suggestions.

<sup>1</sup> Paul Hufe ([hufe@zew.de](mailto:hufe@zew.de)) is affiliated to ZEW and University of Mannheim

<sup>2</sup> Andreas Peichl ([peichl@zew.de](mailto:peichl@zew.de)) is affiliated to ZEW, University of Mannheim, IZA and CESifo

<sup>3</sup> John Roemer ([john.roemer@yale.edu](mailto:john.roemer@yale.edu)) is affiliated to Yale University

<sup>4</sup> Martin Ungerer ([ungerer@zew.de](mailto:ungerer@zew.de)) is affiliated to ZEW and University of Cologne

## I. INTRODUCTION

There is now a large theoretical and empirical literature in economics on inequality of opportunity (IOp)<sup>1</sup>, which evolved out of writings in political philosophy, beginning with John Rawls and extending through the present time. In one prominent formulation (Roemer 1993, 1998), outcomes that individuals enjoy (such as income) are consequences of two sorts of factor: *Circumstances*, those characteristics of a person and her environment that are beyond her control or for which she should not be held responsible, and *effort*, which comprises those choices within her realm of control. Equality of opportunity is said to hold when the chances that individuals face for achieving the outcome in question are independent of their circumstances, and sensitive only to personal effort. The empirical literature measures the extent of IOp for various outcomes (income, wages, health) in many countries.<sup>2</sup>

Almost all empirical studies estimate that the extent to which inequality of income is due to circumstances is quite small. Since it is this part of inequality that is ethically disturbing, the implication might be drawn that existing inequality of income is ethically acceptable, being due mostly to differential effort. Indeed, Kanbur and Wagstaff (2014) have recently argued that equality-of-opportunity studies may be doing more harm than good, because they could be used by policymakers in developing countries to argue that most income inequality is ethically acceptable, and social policy need not be concerned with reducing it.

We believe that the equal-opportunity approach based on the dichotomy between circumstances and effort is valuable, as it is based upon sound philosophical principles. Moreover, surveys routinely find that most people agree with the principle that inequalities due to circumstances should be rectified by social policy. Indeed preferences for redistribution are systematically correlated with beliefs about the relative importance of effort and luck in the determination of outcomes (Alesina and Giuliano 2011). Individuals are more willing to accept income differences that are due to effort (or laziness/industriousness) rather than to exogenous circumstances (e.g., Fong 2001).

---

<sup>1</sup> See also the recent 'The Equality of Opportunity Project' for the US: <http://www.equality-of-opportunity.org/>.

<sup>2</sup> For recent survey articles on both the theoretical and empirical literature, see Fleurbaey and Peragine 2013, Roemer and Trannoy 2015, Ferreira and Peragine 2015, or Ramos and van de gaer (2016).

However, we also believe that previous measurements of IOp are inadequate. Many important circumstances that play a causal role in income determination have been ignored in the empirical literature. The effects of these circumstances appear statistically as effort, because effort is often measured as the residual cause of income variation after explicit circumstances have been accounted for. Hence, the measurement of IOp is biased downward, perhaps seriously so (see the simulations in Bourguignon et al., 2007 and the discussions in Ferreira and Gignoux, 2011 and Niehues and Peichl, 2014).

In this paper we make use of high-quality micro-panel data to correct this shortcoming. In particular we use the National Longitudinal Survey of Youth 1979 (NLSY79) and the 1970 British Cohort Study (BCS70) to construct fine-grained sets of circumstance that take account of both the social environment of children and their cognitive and non-cognitive achievements during childhood.

The central issue we must confront is what aspects of the child's environment and performance should be deemed as comprising, or due to, circumstances. We take what some might find to be a radical position: That all measurable achievements and behaviors of children before an age of consent is attained are due to circumstances. We believe that children should not be held responsible for any of their accomplishments before that age. Indeed, we could take a cue from the law and use the sexual age of consent, or the age at which a child is judged to be an adult in a court of law to be the age of consent for responsible choice. Ideally, if we had a complete biography of the child at the age of, say, sixteen, we would consider that to comprise the child's circumstances.

In particular, we need not distinguish between the effects of nature and nurture: A child's genetic and somatic make-up is certainly a circumstance. Some may object to this, viewing the child as deserving to benefit from his or her innate traits. We demur – at least we do not believe a person *deserves a higher income* because he or she has valuable personal traits. This does not mean we would begrudge talented people the satisfaction they enjoy from being beautiful, intelligent or charming. But our study here concerns equality of opportunity for income, not life satisfaction, and we do not countenance the view that those desirable traits should command a more generous material condition. Naturally, this implies that equalizing opportunities for incomes must – to some degree at least – conflict with the reward structure of market economies.

Our analysis shows a significant increase of IOp measures when we expand the set of circumstances to include the attributes of the individual and her environment as a child. We find that the fraction of income inequality explained by circumstances to be 45.7% using the NLSY79 and up to 31.8% in the BCS70. These baselines estimates increase further when allowing for heterogeneous effects of circumstances on income.

In section 2, we present our conceptual framework and methodology. Section 3 provides some intuition on the potential role of circumstances in explaining income determination, section 4 describes the data sets, section 5 displays our results, and section 6 concludes.

## II. CONCEPTUAL FRAMEWORK AND METHODOLOGY

The main outcome of interest in this study is individual primary income  $y$ . One measurement of the extent of income inequality due to circumstances is defined as follows. Consider the mean logarithmic deviation (MLD) of an income distribution  $F(y)$ .  $MLD(F)$  captures the total inequality of outcomes (IO). Let us suppose we have partitioned the population into types, each type corresponding to the set of individuals with given circumstances. Each type is characterized by its own income distribution. Let the type distributions be  $\{F^t(y), t \in T\}$  where  $T$  is the set of types, and let type  $t$  have frequency  $f_t$  in the population and mean income  $\mu^t$ , summarized by the vectors  $f = f^1, \dots, f^T$  and  $\mu = \mu^1, \dots, \mu^T$ . We can construct a hypothetical distribution, denoted by  $\Phi_{(\mu, f)}$ , in which all members of each type  $t$  receive the mean income  $\mu^t$  of that type.  $\Phi_{(\mu, f)}$  has a cumulative distribution function that is a step function, with as many steps as types; it is often called the ‘smoothed’ distribution of  $F$  associated with the typology  $(f, \mu)$ . If  $\Phi_{(\mu, f)}$  were the true income distribution, we would say that all inequality is due to circumstances, and none to effort. The MLD of total inequality IO is decomposable as follows:

$$MLD(F) = MLD(\Phi) + \sum_{t=1}^T f_t MLD(F_t) \quad (1)$$

It is this decomposition that accounts for the popularity of the MLD, for we can interpret the first term on the right-hand side of (1) as the inequality due to circumstances, and the second as the inequality due to differential effort. Therefore, the ratio

$$r = \frac{MLD(\Phi)}{MLD(F)}$$

is a measure of the extent to which income inequality is due to circumstances.

The disturbing result we mentioned earlier is that, using popular data sets, which record information on a limited set of circumstances, one finds that the measured value of  $r$  is quite small – far less than one-half – especially in the highly developed countries (Brunori et al. 2013). Is it, then, really the case that much more than half of income inequality is due to differential effort, as these results would suggest, or are we seriously underestimating the effect of circumstances, due to poor data sets?

The approach that we have just summarized, using the MLD decomposition, is non-parametric: It partitions the population into types, defined by their circumstances, and takes as data the type distributions and the aggregate distribution of the outcome of interest. This non-parametric estimation of the role of circumstances in generating inequality is of limited use, because it is only feasible if we have meaningful distributions of income by type, and that requires either a very large data set, or a small set of types. Suppose, for example, we had 20 circumstances, each of which could take on three values (low, middle, high). Then the set of types would have  $3^{20}$  elements. Even if one-third of these were empty, we would have  $3^{19}$  types. To get statistically meaningful distributions for all types, we would need, say,  $50 * 3^{19}$ , or about 84 billion observations. To circumvent this problem, practitioners use parametric estimation, in which regression analysis replaces the partition of the population into a typology. Using a parametric approach, we can estimate the impact of numerous circumstance variables even in the presence of small sample and cell sizes.

We follow Ferreira and Gignoux (2011) and Niehues and Peichl (2014) who use a parametric specification to estimate lower bounds of IOp. The empirical specification reads

$$\ln y_i = \alpha C_i + u_i, \quad (2)$$

and can be estimated by OLS to derive the fraction of variance that is explained by circumstances. In this reduced form, the estimates measure the overall effect of circumstances on earnings, including the indirect effect of type-specific effort heterogeneity. Based on this estimation, we can construct a parametric estimate of the smoothed distribution  $\Phi$  defined earlier by replacing earnings outcomes by their predictions:

$$\hat{y} = \exp(\hat{\alpha} C_i) \quad (3).$$

We then let  $\Phi$  be the distribution of these estimated incomes. In this counterfactual, all individuals with the same circumstances necessarily have the same incomes. Thus, in the case where all income differences are due to circumstances (and so the error terms in (2) are all zero), the ratio  $r$  would be unity. Thus  $r$  can be rewritten as:

$$r = \frac{MLD(\{y_i\})}{MLD(\{\hat{y}_i\})}$$

Practitioners recognize that this procedure leads to lower bound estimates of the true share of inequalities due to circumstances. The intuition for this is just like that of an  $R^2$ -measure (Ferreira and Gignoux 2011): Adding another circumstance variable to the analysis increases the explained variation (or at least does not decrease it in the case it is orthogonal), and hence the share of inequality due to circumstances cannot decrease although coefficients might be upward or downward biased. However, usually not all potential circumstances are observable in the data. Therefore, the extent of this underestimation bias is unclear.

Moreover, circumstances, taken from typical data sets, often appear to explain very little of the inequality in the aggregate distribution of income. Roemer (2015) attempts to explain this fact by raising the following question: Given an aggregate distribution of income  $F$ , and  $T$  types with frequencies  $f = f^1, \dots, f^T$  and mean incomes  $\mu = \mu^1, \dots, \mu^T$ , what is the *maximum* value of  $r$  that could attain, if we were able to choose the  $T$ -component (type) distributions  $F^t$  subject only to the conditions that  $F^t$  possess mean  $\mu^t$ , types  $t$  have frequency  $f^t$  in the population, and, of course, that the convex combination of the component distributions equals  $F$ ?

Roemer (2015) shows that *the supports of the type distributions* resulting in a maximal  $r$  are mutually disjoint. However, this is typically not the case in reality. Instead, the supports of the type distributions are very much overlapping – and very far from being disjoint.<sup>3</sup> This observation suggests that to get relatively large values of  $r$ , we need circumstances that define types with the property that there are many subsets of types that share very little income mass. Usually this is not the case when we use the common circumstances of parental education, occupation, race, or region of the country. Put another way, market economies do a pretty good job of equalizing opportunities for income acquisition, if we define the typology to be sufficiently coarse.

---

<sup>3</sup> See Assaad et al. (2015) for an example calculation using Egyptian data.



### III. DATA

For our empirical analysis, we use two data sets, NLSY79 and BCS70, which are described in turn after a short overview of the different circumstance sets that we are using.

#### A. Sets of circumstances

The empirical analysis comprises several scenarios including different sets of explanatory variables beyond the individual's control. We grouped the explanatory variables into meaningful subsets by topics. There is a base scenario and five further specifications. The base scenario is chosen to include the circumstances most commonly used in the literature (such as parental background and ethnic origin) whereas the other scenarios include more detailed childhood outcomes unique to the data at hand and novel to this literature. While scenarios one to five feature a certain degree of comparability between NLSY79 and BCS70, the sixth circumstance sets comprise variables unique to the respective data sets. Table 1 (see also Appendices 1 and 2 for more details on the respective variables) provides an overview of the circumstance sets we consider. Sets two to four contain information on attributes and achievements of the individual as a child, before the age of consent, while set five adds behavior of the child's parents.

| Scenario |       | Circumstance Set | Circumstance Var. |        |       |                           |  |
|----------|-------|------------------|-------------------|--------|-------|---------------------------|--|
| Sixth    | Fifth | Fourth           | Third             | Second | First | Base                      | Sex, Country of Birth, Ethnic Affiliation, Cohort, Age, Academic Achievement Mother, Occupation Code Mother, Rural/Urban, Height (16), Family Income |
|          |       |                  |                   |        |       | Ability                   | PIAT Math, PIAT Reading  |
|          |       |                  |                   |        |       | Behavioral Problems       | Behavioral Problems Index (BPI)  |
|          |       |                  |                   |        |       | Child-Parent Relationship | Play/Schoolwork w/ Parents, Perceived Quantity of Time w/ Parents, Parents Split, Parental Income  |
|          |       |                  |                   |        |       | Health-Related Behavior   | Smoking Habits Mother, Drinking Habits Mother, Health Restrictions Child   |
|          |       |                  |                   |        |       | Survey Specifics          | Specific to NLSY79 and BCS70. See text for more information.   |

**Table 1: Overview of Circumstance Scenarios**

*B. The National Longitudinal Survey of Youth 1979 (NLSY79)*

The NLSY79 is a longitudinal micro-study sponsored by the US Department of Labor, the first wave of which was collected from a nationally representative sample in 1979. It makes available a wealth of information on respondents' educational, income and employment biographies, family processes, health-related behaviors as well as psychological dispositions and cognitive abilities. At the time of the first round the sample consisted of 12,686 respondents covering the cohorts 1957-1964. This implies that respondents were aged between 47 and 56 in 2012, the year of the latest available survey round. As of 1986 the NLSY79 is complemented by the Child & Young Adults supplement (NLSY79 Child/YA), which tracks the lives of all biological children of female NLSY79 respondents. It thus greatly expands the scope of child information collected. Interviews are conducted on a biennial basis, where separate questionnaires are administered to children below the age of 15 and young adults senior in age. The former collects detailed information from both mothers and children on psychological and physiological child development, socio-economic background characteristics, family interactions and educational assessments. The latter is based on the NLSY79 questionnaire and provides a host of information on outcome variables such as income and educational achievements. As of the 2012 wave 11,512 descendants of the NLSY79 cohort have been interviewed covering the age range 14-41. The breadth of available information on parents originating from NLSY79 as well as the detailed records on living conditions and socialization processes of children before the age of consent originating from the child questionnaires, make this study particularly suitable to construct rich circumstance sets for the estimation of IOp.

Our base scenario comprises a set of basic demographic characteristics of the respondents. In particular, we included dummies for sex, race, country of birth, ethnic group and the respondent's cohort in order to take account of generational effects. Furthermore, we control for maternal educational achievement by including indicator variables for different academic degrees grouped in three categories of increasing rank. To be sure, we restrict ourselves to degrees attained before the child reaches the age of 16 in order to rule out feedback loops from child behavior beyond the responsibility cut-off. Similarly, we introduce a battery of occupation dummies for the mother, which are measured at the child age of 14. To further refine our account of the child's socio-economic background we employ net family income averaged over

all periods of observation from birth until the age of consent. Neighborhood characteristics are introduced by dummy variables for whether the child lives in a Metropolitan Statistical Area (MSA), i.e. a core urban area with a population of 50,000 or more, and if yes whether its residence is located in the center of such an area. Lastly, we include child height recorded at the age of 16.

The second scenario makes use of the availability of academic achievement tests in the NLSY79 Child/YA in order to proxy for the child's ability. Ability at age 16 or younger is assumed to be beyond personal control and hence can be interpreted as a circumstance. A common approximation of ability is the use of standardized test scores. Specifically, at this stage we include the standardized score of the Peabody Individual Achievement Test (PIAT) in the areas of mathematics and reading recognition. PIATs are widely-used measures of academic achievement credited with high test-retest reliability and concurrent validity. As we can draw on multiple rounds of PIATs for each observation we have averaged all respective scores over the age range 0-16 in order to cushion the influence of outlier test results. Note that we abstain from including further proxies of child ability at this stage in order to maintain comparability with the BCS70. They will, however, be integrated in a later scenario.

Thirdly, we include the total percentile score of the Behavioral Problems Index (BPI) among the circumstance variables. The BPI is an aggregate measure of child behavior and attitudes constructed from a series of 28 questions to mothers of children between four and 14 years of age. Again we make use of the availability of multiple test scores for each child by averaging over the relevant age range.

Scenario 4 extends the scope of circumstances to the child-parent relationship. In particular we take account of the family status of parents by controlling for the share of years parents did not live in the same household conditional on the fact that both were alive. Moreover, we took averages of the child's responses on whether parents spent time with them on schoolwork or engaged with them in activities of another kind. Lastly, a set of dummy variables is used to mirror the child's wish to spend more or less time with each of its parents at the age of 14.

Subsequently, we focus on health-related information for both children and parents in scenario 5. As regards the former, we make use of mothers' assessment on whether their child was restricted in school participation, schoolwork, or leisure activities due to a medical

condition. Again we average these reports over the relevant age range. With respect to parental behavior we are confined to maternal information. Therefore we include a dummy variable for whether the respective child has ever been exposed to a mother smoking on a daily basis. Additionally, we take account of the consumption of alcoholic beverages by including indicator variables for monthly drinking frequencies measured at the child age of eight. It is noteworthy that the integration of variables on maternal health behaviors yields important sample size reductions. Therefore, we will also consider an alternative reduced set of health-related information by exclusively focusing on the child's restrictions induced by medical conditions.

Lastly, we augment the circumstance set considered thus far by a host of variables that are specific to the NLSY79 and have no analogues in the BCS70. First, at the level of educational background we include a binary indicator for whether the child attends a public school. Moreover, four variables are introduced that measure the average number of persons with a certain educational level in the household over the relevant age range from birth to 16. The considered levels are "Less than 12 years", "12-13 years", "13-15 years" and ">15 years". We furthermore exploit the fact that all NLSY79 respondents were subjected to an Armed Forces Qualification Test (AFQT) at the beginning of the study. Thus, we are able to include the mother's AFQT-score as a proxy variable for maternal intelligence. Second, we aim to refine the ability measures of children by including the PIAT-score for Reading Comprehension and the Memory for Digit Span assessment. The digit span test is a measure of intelligence with particular focus on short-term memory. Third, we introduce a battery of average assessment scores on psychological dispositions of the children in the sample. These encompass the self-perception of children with respect to self-worth and scholastic performance as well as temperament assessments as regards child compliance, insecurity, and sociability. Similarly, NLSY79 conducted psychometric assessments with every respondent at the beginning of the study. As a result we are able to include the Pearlin Mastery Scale, Rotter's Locus of Control Scale and the Rosenberg Self-Esteem Scale at the level of mothers. The first two scores measure the extent to which respondents perceive themselves to be in control of forces that impact their lives. As its name suggests, the Rosenberg score can be interpreted as a measure of self-esteem. Lastly, socio-economic background variables of the circumstance set are enriched by a binary indicator on whether the mother ever was convicted of a crime and a home inventory score based on a test administered to children and mothers. Specifically, the latter

measure assesses the quality of a child's home environment with sub-scores focusing on the extent of cognitive stimulation and emotional support.

### *C. The 1970 British Cohort Study (BCS70)*

The BCS70 is a longitudinal survey funded by the Economic and Social Research Council and managed by the Centre for Longitudinal Studies. It follows the lives of more than 17,000 individuals born in England, Scotland and Wales in a single week in 1970. Since the first survey wave in 1970, there have been eight follow-up interviews of all cohort members at ages 5, 10, 16, 26, 30, 34, 38 and 42. The latest survey has been carried out in 2012. In addition to the main interviews, there have been five sub-studies where additional data has been collected for samples of cohort members selected for their particular characteristics or circumstances. The survey is accompanied by interviews of the parents and head-teachers, standardized test scores, health records, nutrition and activity diaries as well as labor market histories. Thereby, the BCS70 has collected information on health, physical, educational and social development, and economic circumstances. The data set contains detailed information on early childhood and parental background. In contrast to NLSY79, questionnaires are filled by both parents, revealing broader information on parental background. Moreover, similar and identical questions on family and social situation are addressed to parents and children.

The baseline scenario covers basic demographic and parental background variables. We include dummies for gender and foreign origin, defined by the birthplace of the father. Furthermore, we define four categories of academic achievement of the father, "no degree", "secondary", "intermediate", and "college", where the first category serves as the omitted variable. In the same way, we include six occupational categories for the father, with the blue-collar group used as the omitted baseline case. In order to account for the socio-economic background of the child, we use the childhood urbanization degree and the parental income over the childhood as explanatory variables. Urbanization is measured by three dummies, grouping rural, suburbs/towns, and inner urban areas at the age of ten. Parental income is measured at ages 10 and 16 and classified into seven to eleven groups. Finally, we use the body height of the individual at age 10.

Scenario 2 covers additional ability measures during childhood. The BCS70 provides information on standardized vocabulary test scores for multiple years as well as standardized

math test scores at the age of 10. The test used for the assessment of reading ability is the Edinburgh Reading Test, while the Friendly Math Test is used to account for ability in mathematics.

Within the third scenario, we extend the estimation to behavior related information. The Malaise Inventory test comprises a set of self-completion questions that are combined to measure levels of psychological distress or depression. The test was developed from the Cornell Medical Index Health Questionnaire (CMI) and covers questions on emotional disturbance and associated physical symptoms. It is designed to differentiate between individuals with and without psychiatric disorder. For our analysis, we use the standardized malaise test score at age 16. The BCS70 offers additional behavior tests for the parents, which we use in the data set specific scenario 6 in order to maintain comparability with the NLSY79 for scenario 3.

The fourth scenario deals with detailed information on family background. The time spent with the parents might affect the character of the individual. Therefore, we use the average time spent with the parents in a week as stated by the child at age 16. It is classified into 5 groups, from most days a week to little or never. In addition, we utilize the questions on common indoor or outdoor hobbies shared with the parents. It is noteworthy that one could further extend the circumstance set to the kind of common activities, household duties or personal relationships with the parents. As a final part of this scenario, we include the marital status of the parents at birth, categorized into “single”, “partnership”, “separated”, and “widowed”.

Scenario five deals with health and medical conditions of individuals and parents during childhood. The BCS70 contains detailed information on the smoking and drinking habits of parents, such that we can use both sets of information at age 16, categorized into six groups from “never” to “every or most days”. In addition, we include the medical condition of the child at age ten, stating whether the child experienced no, slight, or severe medical conditions up to this age.

Finally, scenario six consists of further variables available in the BCS70 but we refrained from using in the previous scenarios to maintain comparability with the NLSY79. As a starting point, we include further ability measures at age five for picture and reading scores, malaise and Rutter's behavior test scores for the parents as well as additional information on the

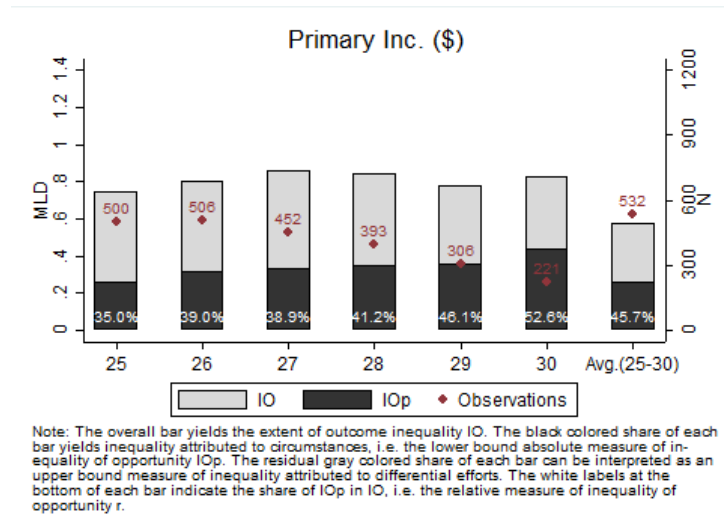
education of the father. In addition, we utilize further variables provided by the BCS70. These include information on the school type, especially whether a grammar school was attended. Moreover, we use information on the vaccination status of children at age five. Finally, we add a variable stating whether up to age five the child was raised primarily at home or in a childcare facility. However, the BCS70 offers a huge variety of variables that are of potential interest. In view of the small sample sizes, we refrain from using some of this information due to limitations in the available degrees of freedom.

## IV. RESULTS

### A. *NLSY79*

As outlined previously, the observational units covered by the NLSY79 span a wide time range. While the first children born to NLSY79 mothers were born as early as 1970, the latest birth we observe dates 1997. Therefore, a choice has to be made of how to treat the variance in age. In what follows, we offer three approaches. First, we implicitly control for age by measuring the outcome of interest at a specific age threshold. For instance, we use the primary income of person  $i$  at age 25 as the left-hand-side variable in equation (2), and then control for the respective cohort effects in the circumstance set. Figure 1 shows the results for different ages between 25 and 30 as well as for average income over this period for scenario 5. Depending on the age, circumstances account for between 35% and 53% of total inequality.

Figure 1: IOp at varying age thresholds (NLSY79), scenario 5



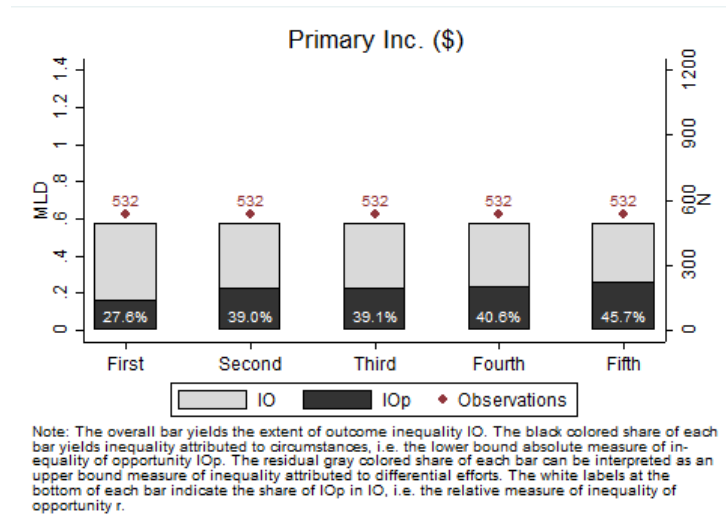
However, as suggested by Figure 1, this approach leads to very small sample sizes towards the age threshold of 30, which is commonly assumed to be a strong predictor of long-term earning potentials (Chetty et al. 2014). Therefore, we alternatively use the average income over the age range 25-30 (Figure 2). This approach strikes a balance between the ambition to maintain a reasonably sized sample and to cushion the effect of outlier incomes of younger cohorts, who are at the beginning of their careers. Lastly, we pool all observations and control for both age and cohort effects in the circumstance sets (Figure 4).

Using the averaged outcome variable, Figure 2 gives an overview of how our IOp estimates vary as we sequentially introduce the circumstance sets laid out in Table 1. First of all it is noteworthy, that inequality in the US is higher according to this sample than in comparison to other works. Pistolessi (2009) uses the PSID to calculate a Theil index of permanent labor earnings of 0.25 in 2001. Similarly, Niehues and Peichl (2014) rely on the same data source to calculate an MLD of 0.24 with respect to permanent gross earnings. The NLSY79 sample used in this study yields an MLD of approximately 0.57.

In terms of IOp, we find a value of  $r$  of 27.6% for the first scenario. This estimate is comparable to Niehues and Peichl's (2014) estimate of 29% and Pistolessi's (2009) estimate of about 20%, which is reassuring since we use a similar circumstance set than in the previous literature in this case. Adding more circumstances substantially increases the estimate of IOp by up to two thirds to 45.7% using the most comprehensive circumstance set.

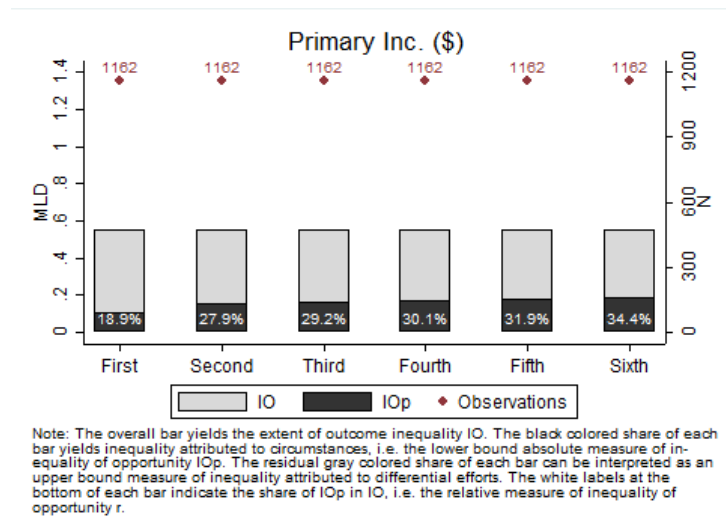


Figure 2: IOp with varying circumstance sets (NLSY79), comparable sample, average income



To examine whether this result is driven by the relatively small sample size, we retaylor the circumstance sets. In particular, we exclude all variables on drinking and smoking behaviors of the mother, which are sparsely populated in the NLSY79. Excluding these variables yields a doubling of the sample size. As health-related information is only introduced in scenario 5, we can compare the first four scenarios in Figure 2 and Figure 3 to observe a difference of around 10 percentage points depending on the sample under consideration.

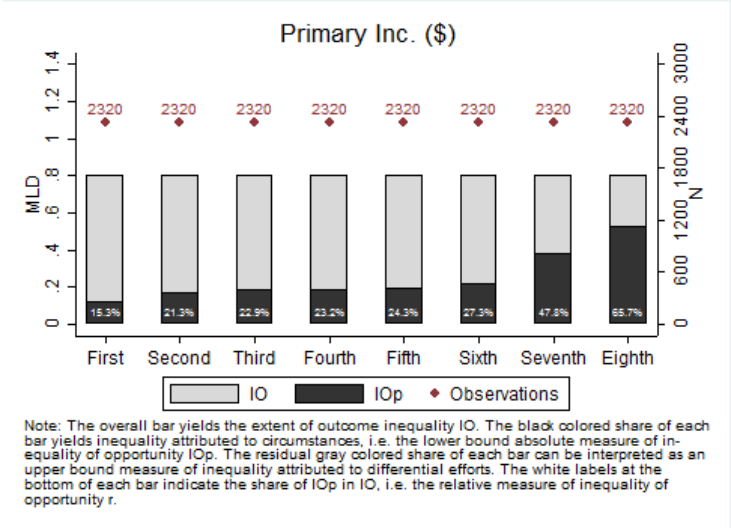
Figure 3: IOp with varying circumstance sets (NLSY79), survey-specific sample, average income



In the base scenario, we find an estimate for IOp of 18.9% which increases to 31.9% in the fifth scenario – again an increase of about two thirds. The estimate further increases to 34.4% in the sixth scenario which includes a host of circumstances that have no analogue in the BCS70, such as various test scores on psychological dispositions and ability. Clearly, these

variables provide an important contribution to IOp by driving up our result another 2.5 percentage points.

Figure 4: IOp with varying circumstance sets (NLSY79), survey-specific pooled sample



In a last step we pool all variables and control for age and cohort in the circumstance set. The sample is balanced on the sixth circumstance set and thus should be compared to Figure 3. As suggested by Figure 4, this approach drives the IOp shares further in the direction of previous results on IOp in the US. The downward shift of  $r$  is largely attributable to an increase in its denominator: The MLD of primary income. The previous approach averaged incomes over a six-year-period which lead to a smoothing of transitory income changes. Pooling all observations now increases intra-individual income inequality which is captured by an increase of IO from 0.57 to approximately 0.8. Notwithstanding this change in the relative share of IOp, the absolute measure of IOp remains constant.

So far, we have introduced the circumstances only linearly. However, the effect of circumstances on income could be heterogeneous across types. To analyze this, we introduce two additional scenarios. While the circumstance set remains the same as in scenario 6, we introduce additional type heterogeneity by allowing all circumstance coefficients to vary across age (Scenario 7) and combined age-sex cells (Scenario 8). This is implemented by interacting all circumstances with age, and age and sex respectively.<sup>4</sup>

<sup>4</sup> The ideal approach, were data sets sufficiently large, would be the non-parametric one, in which we create income distribution functions for each vector of circumstances. Clearly, in doing so, all the effects

The resulting increase of  $r$  to almost 66% is impressive and highlights the importance of taking seriously the differential impact of single circumstances across types.<sup>5</sup> Alternatively, we fully interact the circumstance variables on sex, country of birth, ethnic, the rural/urban divide and maternal academic achievement in the base scenario and then sequentially introduce all other circumstances as outlined in the previous discussion. The results (Appendix 3) indicate a doubling of  $r$  in the first scenario, while the subsequent increase by 10.1 percentage points until scenario six is only slightly lower in absolute terms than the increase portrayed in Figure 4 (12 percentage points). Thus, the introduction of heterogeneous effects across types does not merely capture variation that is reflected in other circumstances, but provides an independent upwards correction of lower bound IOp estimates.

### *B. The British Cohort Study*

In contrast to the NLSY79, the BCS70 only observes one cohort of individuals. Therefore, dealing with the variance in age is relatively straightforward. We observe individual net earnings from 2000 to 2012 (age 30 to 42), in 4-year steps. Using the annual information as well as average earnings over the whole period, we are able to cover a period of 12 years, including early employment time as well as individuals of age 30 to 40. The sample size is relatively constant for all years, indicating a small dropout rate and sound data quality. First, we run separate regressions for each year and a regression for average income using circumstance set specification five, which is still comparable to NLSY79.

Checchi et al. (2010) find a value of 20.5% for  $r$  as well as a MLD of 0.204 using EU-SILC data from 2005. Similarly, OECD data from 2010 indicate a MLD of 0.20 in the mid-2000s (OECD 2010). Generally, our measures for IO are somewhat higher than these estimates which may be attributed to the fact that we observe one cohort instead of a representative sample of the entire population. The IOp shares for net income take values from 17.3% in 2000 to 39.0% in 2008. Hence, our measure is up to 19 percentage points higher compared to previous studies.

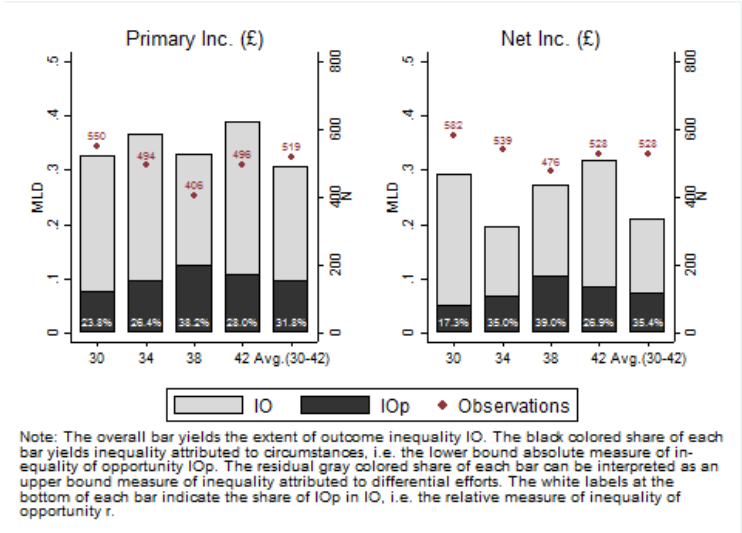
---

of interactions among the circumstances will be manifest in the income distribution of the type. The linear, parametric estimation of income, which is used in scenarios 1 through 6 is clearly capturing none of the interaction effects of the circumstances, and so scenarios 7 and 8 bring us somewhat closer to an approximation of the non-parametric approach (e.g. Checchi and Peragine 2010).

<sup>5</sup> See Hufe and Peichl (2015) for a more thorough discussion of the role of interaction terms in parametric lower bound estimations of IOp.

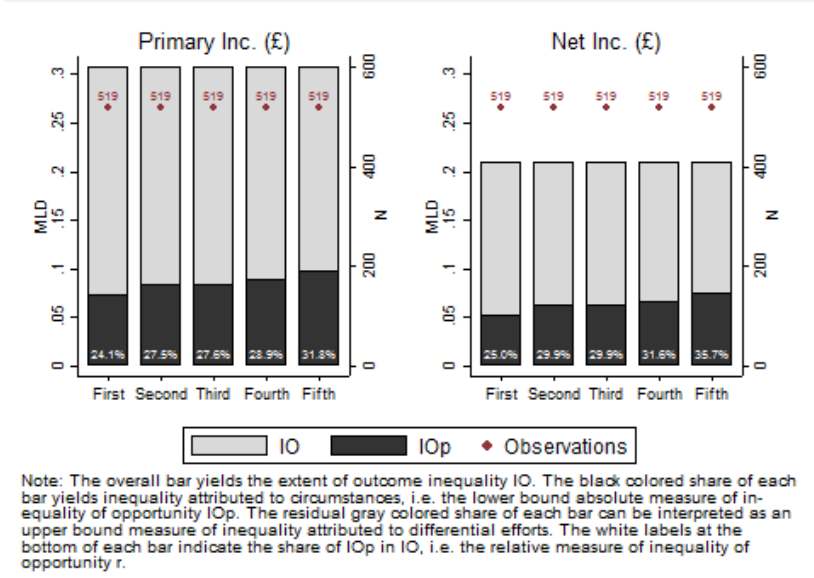
As the BCS70 carries information on primary and net income, we use these numbers to compare IOp before and after taxes and transfers.

Figure 5: IOp at varying age thresholds (BCS70), scenario 5



By restricting the sample to those individuals with valid data on both income definitions, we lose some observations but ensure comparability in the results. The main finding is that IO is higher in primary terms, but  $r$  is actually lower in primary income considering average income over ages 30 to 42. However, the level of IOp is higher in primary earnings.

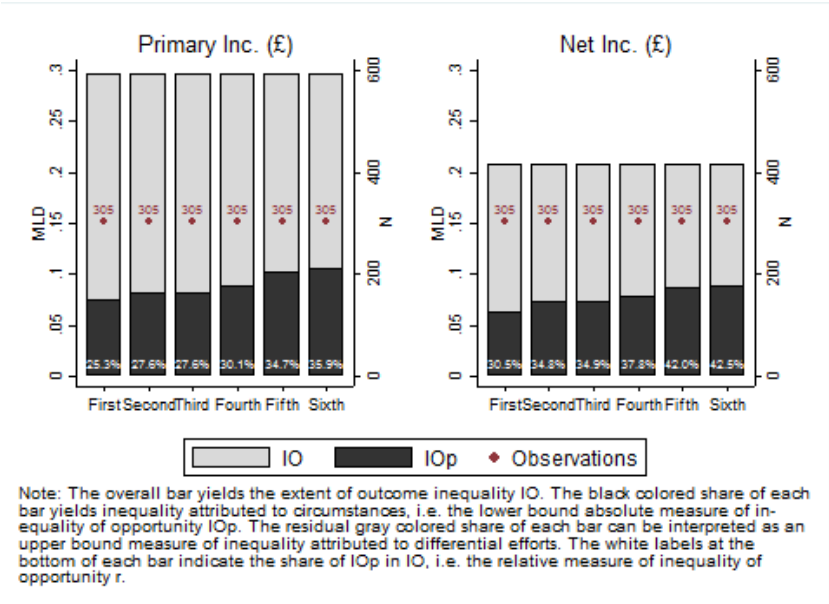
Figure 6: IOp with varying circumstance sets (BCS70), comparable sample, average income



In order to compare these figures with each other as well as with the NLSY79, we keep the number of observations constant at the composition of the fifth circumstance set and look at both, primary and net income. Thereby, we end up with 519 observations. A balanced panel

enables a comparison of the different circumstance sets. We find that the MLD of IO is quite comparable to previous studies in net terms and as usually somewhat higher in primary income. The relative share  $r$ , however, increases constantly with the number of explanatory variables, reaching from 24.1% (25.0%) in the first scenario to 31.8% (35.7%) in scenario five in primary (net) terms. Compared with the US results presented in the previous section, both the level of IOp and the impact of adding further circumstances are lower in the UK.

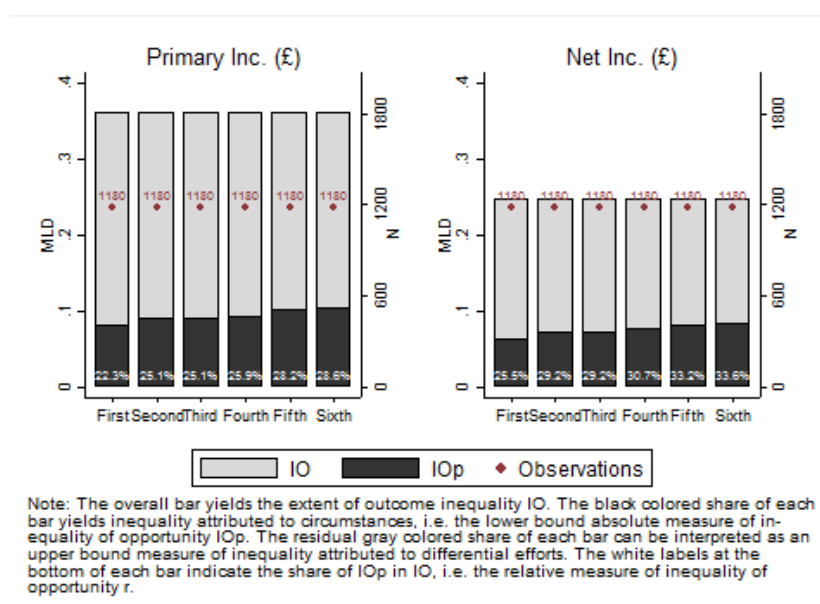
Figure 7: IOp with varying circumstance sets (BCS70), survey-specific sample, average income



Including the BCS70 specific variables and balancing on the sixth circumstance set reduces the number of observations and in addition decreases comparability with the NLSY79. Nevertheless, we find that adding the specific circumstance variables yields an even higher  $r$  of now 35.9% in primary and 42.5% in net income.

In a second analysis, we pool the data from different cohorts and explicitly control for age in the regression. Similarly to the analysis of the NLSY79, the relative share  $r$  declines for all scenarios due to the increase in IO, taking them closer to the results of previous studies. Nevertheless, with an  $r$  of 28.6% and 33.6% in the sixth circumstance set, they are still up to 8 (13) percentage points higher compared to the results of Checchi et al. (2010) in primary (net) terms.

Figure 8: IOp with varying circumstance sets (BCS70), survey-specific pooled sample, average income



## V. CONCLUSION

In this paper, we argue that important circumstances that play a causal role in income determination have been ignored in the empirical literature on IOp. From our perspective, all behaviors and accomplishments of children should be considered the consequence of circumstances: That is, an individual should not be held responsible for her choices before an age of consent is reached. In rich societies, it is appropriate to choose the age of consent as occurring no earlier than adolescence. Ideally, if we had a complete biography of the child at the age of, say, sixteen, we would consider that to comprise the child’s circumstances.

The credulous reader might well ask, “Well, if you take the complete biography of the child at the age of consent as comprising her circumstances, where can effort play a role? After all, the choices the adult makes will be strongly influenced by her ‘biography’ at age sixteen.” We agree, and that is why we believe circumstances account for a far larger fraction of outcome inequality than studies to date have calculated. Nevertheless, we resist the response that would decrease the age of consent to something like four or six years of age. Perhaps thirteen, as in the Jewish religion, would be acceptable – although one must ponder the fact that thirteen was acclaimed as the beginning of adulthood when life expectancies were barely one-third of what they are now, and the resources society had to allocate to children were far less.

In any case, our data sets are not sufficiently fine-grained to make these distinctions. Even if one prefers thirteen to sixteen, it is probably better to include achievements at age sixteen as circumstances when the complete biography at age thirteen is unavailable.

Using the NLSY79 and the BCS70, we construct fine-grained sets of circumstance that take account of both the social environment of children and their childhood accomplishments, to compute the fraction of income inequality due to circumstances in the US as well as the UK. Our analysis shows a significant increase of IOp measures when we expand the set of circumstances to include the attributes of the individual and her environment as a child. We find that the fraction of income inequality explained by circumstances to be 45.7% using the NLSY79 and up to 31.8% in the BCS70. These baseline estimates increase substantially when allowing for heterogeneous effects of circumstances on income.

Our findings suggest revisiting the sets of circumstances used in previous studies, as the results we obtained indicate substantially higher IOp when taking into account additional childhood information. In fact, extending circumstance sets by childhood achievements before an age of consent addresses some of the concerns on the policy relevance of the concept by providing substantial upward corrections of lower bound measures of IOp. Obviously, in many country contexts data limitations impose strong restrictions on the ability of applied researchers to conduct analyses as detailed as ours for the US and the UK. To address this problem, it could be one avenue for future research to combine different data sets for calculations of IOp. For instance, one might use a first data set with detailed information on circumstances to predict childhood accomplishments of different types. In a second step one could then use these intermediate types as circumstances in a second data set to calculate a measure of IOp. Such a procedure would be one promising route to overcome data limitations and enhance the data basis for analyses on IOp.

## REFERENCES

Alesina, A., and Giuliano, P. (2011). Preferences for redistribution. In: Benhabib, J., Jackson, M.O., and Bisin, A. (eds). *Handbook of social economics*. North-Holland, Amsterdam: pp 93–132.

Assaad, R., Krafft, C., and Roemer, J.E. (2015). Inequality of opportunity for income and consumption in Egypt. *mimeo*.

Bourguignon, F., Ferreira, F.H.G., and Menéndez, M. (2007). Inequality of opportunity in Brazil. *Rev Income Wealth* 53(4): pp 585–618.

Brunori, P., Ferreira, F.H.G., and Peragine, V. (2013). Inequality of opportunity, income inequality and economic mobility : some international comparisons. *World Bank Policy Research Working Paper Series* No. 6304.

Cecchi, D., Peragine, V., and Serlenga, L. (2010). Fair and unfair income inequalities in Europe. *IZA Discussion Paper* No. 5025.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *Quart. J. Econ.* 129: pp 1553-1623.

Cecchi, D., and Peragine, V. (2010). Inequality of opportunity in Italy. *The Journal of Economic Inequality* 8(4): pp 429-450.

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

Ferreira, F. H. G., and Peragine, V. (2015). Equality of opportunity: Theory and evidence. *IZA Discussion Paper* No. 8994.

Fleurbaey M., and Peragine, V. (2013). Ex ante versus ex post equality of opportunity. *Economica* 80(317): pp 118– 130.

Fong, C. (2001). Social preferences, self-interest, and the demand for redistribution. *J Public Econ* 82(2): pp 225– 246.



Hufe, P., and Peichl, A., (2015). Lower bounds and the linearity assumption in parametric estimations of inequality of opportunity. *IZA Discussion Paper*.

Kanbur, R., and Wagstaff, A. (2014). How useful is inequality of opportunity as a policy construct? *World Bank Policy Research Working Paper Series No. 6980*.

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

Pistolesi, N. (2009). Inequality of opportunity in the land of opportunities, 1968–2001. *The Journal of Economic Inequality* 7(4): pp 411-433.

Ramos, X., and Van de gaer, D. (2016). Empirical Approaches to Inequality of Opportunity: Principles, Measures, and Evidence. *Journal of Economic Surveys* (forthcoming).

Roemer, J.E. (1993). A pragmatic theory of responsibility for the egalitarian planner. *Philos Public Aff* 22(2): pp 146–166.

Roemer, J.E. (1998). *Equality of opportunity*. Harvard University Press, Cambridge.

Roemer, J.E. (2015). On the importance of circumstances in explaining income inequality. *Revue Economique* (forthcoming).

Roemer, J.E., and Trannoy, A. (2015). Equality of opportunity: Theory and measurement. *Journal of Economic Literature* (forthcoming).

### Appendix 1: Circumstance Variables NLSY79

| Scenario | Var. Name                | N     | Mean     | SD     | Min   | Max     | Study           | Ref. Age | Question Name   | Var. Type    | Question Text   |
|----------|--------------------------|-------|----------|--------|-------|---------|-----------------|----------|-----------------|--------------|---|
| 1        | Cohort                   | 7,999 | 1,986    | 5.611  | 1,970 | 1,997   | NLSY79 Child/YA | -        | CYRB            | Primary Var. | Enter correct birth date for Child: [Month] [Day] [Year]  |
| 1        | Male                     | 7,999 | 0.513    | 0.500  | 0     | 1       | NLSY79 Child/YA | -        | CSEX            | Primary Var. | Child Sex=Female?   |
| 1        | Hispanic                 | 7,999 | 0.221    | 0.415  | 0     | 1       | NLSY79 Child/YA | -        | CRACE           | Primary Var. | -   |
| 1        | Non-black/Non-Hispanic   | 7,999 | 0.447    | 0.497  | 0     | 1       | NLSY79 Child/YA | -        | CRACE           | Primary Var. | -   |
| 1        | Born in US               | 7,999 | 0.923    | 0.266  | 0     | 1       | NLSY79          | -        | FAM-2A          | Primary Var. | In what country were you born?  |
| 1        | Secondary (16)           | 7,815 | 0.566    | 0.496  | 0     | 1       | NLSY79          | 16       | Q3-10B          | Primary Var. | What is the name of the highest degree you have ever received?  |
| 1        | Intermediate (16)        | 7,815 | 0.102    | 0.302  | 0     | 1       | NLSY79          | 16       | Q3-10B          | Primary Var. | What is the name of the highest degree you have ever received?  |
| 1        | College (16)             | 7,815 | 0.172    | 0.378  | 0     | 1       | NLSY79          | 16       | Q3-10B          | Primary Var. | What is the name of the highest degree you have ever received?  |
| 1        | SMSA, Not center (16)    | 6,699 | 0.546    | 0.498  | 0     | 1       | NLSY79          | 16       | SMSARES         | Created      | -   |
| 1        | SMSA, Center (16)        | 6,699 | 0.282    | 0.450  | 0     | 1       | NLSY79          | 16       | SMSARES         | Created      | -   |
| 1        | Mom: Blue Collar (14)    | 3,650 | 0.585    | 0.493  | 0     | 1       | NLSY79          | 14       | Q6-56; CPSOCC70 | Primary Var. | Were you employed by...[Sector of occupation]; What kind of work were you doing for this [most recent] job?                               |
| 1        | Mom: Farmer (14)         | 3,650 | 0.000548 | 0.0234 | 0     | 1       | NLSY79          | 14       | Q6-56; CPSOCC70 | Primary Var. | Were you employed by...[Sector of occupation]; What kind of work were you doing for this [most recent] job?                               |
| 1        | Mom: White Collar (14)   | 3,650 | 0.0721   | 0.259  | 0     | 1       | NLSY79          | 14       | Q6-56; CPSOCC70 | Primary Var. | Were you employed by...[Sector of occupation]; What kind of work were you doing for this [most recent] job?                               |
| 1        | Mom: Professional (14)   | 3,650 | 0.0901   | 0.286  | 0     | 1       | NLSY79          | 14       | Q6-56; CPSOCC70 | Primary Var. | Were you employed by...[Sector of occupation]; What kind of work were you doing for this [most recent] job?                               |
| 1        | Mom: Self-Employed (14)  | 3,650 | 0.0704   | 0.256  | 0     | 1       | NLSY79          | 14       | Q6-56; CPSOCC70 | Primary Var. | Were you employed by...[Sector of occupation]; What kind of work were you doing for this [most recent] job?                               |
| 1        | Mom: Govt. Sctr. (14)    | 3,650 | 0.180    | 0.384  | 0     | 1       | NLSY79          | 14       | Q6-56; CPSOCC70 | Primary Var. | Were you employed by...[Sector of occupation]; What kind of work were you doing for this [most recent] job?                               |
| 1        | Height in inches (16)    | 6,452 | 66.31    | 3.819  | 48    | 83      | NLSY79 Child/YA | 16       | -               | Primary Var. | -   |
| 1        | Avg. Net Fam. Inc.       | 7,999 | 30,393   | 29,228 | 24.50 | 402,321 | NLSY79          | Avg.     | TNFI_TRUNC      | Created      | -   |
| 2        | Avg. PIAT: Math          | 7,701 | 99.96    | 12.20  | 43.33 | 135     | NLSY79 Child/YA | Avg.     | MATHZ[year]     | Created      | -   |
| 2        | Avg. PIAT: Reading       | 7,697 | 103.4    | 13.29  | 65    | 135     | NLSY79 Child/YA | Avg.     | RECOGZ[year]    | Created      | -   |
| 3        | Avg. BPI-Score           | 7,915 | 598.4    | 231.4  | 30    | 1,000   | NLSY79 Child/YA | Avg.     | BPIP[year]      | Primary Var. | -   |
| 3        | BPI-Score(14)            | 6,540 | 608.3    | 279.9  | 82    | 1,000   | NLSY79 Child/YA | 14       | BPIP[year]      | Primary Var. | -   |
| 4        | % years w/ 2 prnts       | 7,976 | 0.587    | 0.414  | 0     | 1       | NLSY79 Child/YA | Avg.     | DADHM[year]     | Primary Var. | Does father of child (living in HH) live in this HH?  |
| 4        | Avg. Activ.: Schlwrk     | 7,189 | 0.366    | 0.386  | 0     | 1       | NLSY79 Child/YA | Avg.     | -               | Primary Var. | Within the last week, have you and your parents: Worked on schoolwork together?   |
| 4        | Avg. Activ.: Games/Play  | 7,188 | 0.469    | 0.404  | 0     | 1       | NLSY79 Child/YA | Avg.     | -               | Primary Var. | Within the last week, have you and your parents: Played a game or sport?  |
| 4        | Mom (14): More Time      | 5,554 | 0.770    | 0.421  | 0     | 1       | NLSY79 Child/YA | 14       | -               | Primary Var. | Do you think your mother spends enough time with you?   |
| 4        | Mom (14): Less Time      | 5,554 | 0.0646   | 0.246  | 0     | 1       | NLSY79 Child/YA | 14       | -               | Primary Var. | Do you think your mother spends enough time with you?   |
| 4        | Dad (14): More Time      | 4,498 | 0.594    | 0.491  | 0     | 1       | NLSY79 Child/YA | 14       | -               | Primary Var. | Do you think your father spends enough time with you?   |
| 4        | Dad (14): Less Time      | 4,498 | 0.0382   | 0.192  | 0     | 1       | NLSY79 Child/YA | 14       | -               | Primary Var. | Do you think your father spends enough time with you?   |
| 5        | Avg. Health: School      | 7,910 | 0.0199   | 0.0930 | 0     | 1       | NLSY79 Child/YA | Avg.     | -               | Primary Var. | Does [Child Name] have any physical, emotional, or mental condition that limits or prevents [his/her] ability to attend school regularly? |
| 5        | Avg. Health: School Work | 7,909 | 0.0311   | 0.123  | 0     | 1       | NLSY79 Child/YA | Avg.     | -               | Primary Var. | Does [Child Name] have any physical, emotional, or mental   |

| Scenario | Var. Name                     | N     | Mean   | SD    | Min | Max   | Study           | Ref. Age | Question Name      | Var. Type    | Question Text   |
|----------|-------------------------------|-------|--------|-------|-----|-------|-----------------|----------|--------------------|--------------|---|
| 5        | Avg. Health: Play             | 7,983 | 0.0336 | 0.109 | 0   | 1     | NLSY79 Child/YA | Avg.     | -                  | Primary Var. | condition that limits or prevents [his/her] ability to do regular school work?  |
| 5        | Non-Smoker Mthr               | 4,566 | 0.217  | 0.412 | 0   | 1     | NLSY79          | 1-16     | DS5; DS6           | Primary Var. | Does [Child Name] have any physical, emotional, or mental condition that limits or prevents [his/her] ability to do usual childhood activities such as play, or participate in games or sports?               |
| 5        | Drinks 3-4 Times/Wk           | 4,426 | 0.0359 | 0.186 | 0   | 1     | NLSY79          | 8        | Q12-5              | Primary Var. | Do you now smoke daily, occasionally or not at all?;How long has it been since you last smoked cigarettes daily?  |
| 5        | Drinks 1-2 Times/Wk           | 4,426 | 0.365  | 0.482 | 0   | 1     | NLSY79          | 8        | Q12-5              | Primary Var. | During the last 30 days, on how many days did you drink any alcoholic beverages, including beer, wine, or liquor?   |
| 5        | Drinks 2-3 Times/Mth          | 4,426 | 0.262  | 0.440 | 0   | 1     | NLSY79          | 8        | Q12-5              | Primary Var. | During the last 30 days, on how many days did you drink any alcoholic beverages, including beer, wine, or liquor?   |
| 5        | Drinks Once/Mth               | 4,426 | 0.155  | 0.362 | 0   | 1     | NLSY79          | 8        | Q12-5              | Primary Var. | During the last 30 days, on how many days did you drink any alcoholic beverages, including beer, wine, or liquor?   |
| 5        | Drinks Never                  | 4,426 | 0.148  | 0.355 | 0   | 1     | NLSY79          | 8        | Q12-5              | Primary Var. | During the last 30 days, on how many days did you drink any alcoholic beverages, including beer, wine, or liquor?   |
| 6        | AFQT Score                    | 7,656 | 34.55  | 27.27 | 1   | 99    | NLSY79          | -        | AFQT-1             | Created      | -   |
| 6        | Avg. PIAT: Compreh.           | 7,582 | 99.87  | 12.12 | 65  | 135   | NLSY79 Child/YA | Avg.     | COMPZ[year]        | Created      | -   |
| 6        | No Pub. School (14)           | 6,708 | 0.129  | 0.335 | 0   | 1     | NLSY79 Child/YA | 14       | -                  | Primary Var. | Is the school your child usually attends public private or religious? Or, doesn't your child attend school? [For years <2000]; Which of the following describes the school Child attends:... [For years>2000] |
| 6        | Avg. # in HH w/ Educ <12      | 7,999 | 0.514  | 0.685 | 0   | 4.615 | NLSY79 Child/YA | Avg.     | NAHGC0[year]       | Created      | -   |
| 6        | Avg. # in HH w/ Educ 12-13    | 7,999 | 0.977  | 0.725 | 0   | 4.917 | NLSY79 Child/YA | Avg.     | NAHGC1[year]       | Created      | -   |
| 6        | Avg. # in HH w/ Educ 13-15    | 7,999 | 0.236  | 0.415 | 0   | 2.750 | NLSY79 Child/YA | Avg.     | NAHGC2[year]       | Created      | -   |
| 6        | Avg. # in HH w/ Educ >15      | 7,999 | 0.272  | 0.561 | 0   | 3     | NLSY79 Child/YA | Avg.     | NAHGC3[year]       | Created      | -   |
| 6        | No Convict.                   | 7,735 | 0.980  | 0.140 | 0   | 1     | NLSY79          | -        | POLICE_3           | Primary Var. | Have you been ever convicted of any charges other than a minor traffic violation?   |
| 6        | Avg. Emotional Supp.          | 7,888 | 47.06  | 21.21 | 0   | 99    | NLSY79 Child/YA | Avg.     | EMOTP[year]        | Created      | -   |
| 6        | Avg. Cognitive Stim.          | 7,915 | 46.04  | 22.72 | 0   | 97    | NLSY79 Child/YA | Avg.     | COGNP[year]        | Created      | -   |
| 6        | Avg. Compliance               | 6,870 | 22.73  | 3.841 | 6   | 30    | NLSY79 Child/YA | Avg.     | COMPLY[year]       | Created      | -   |
| 6        | Avg. Insecurity               | 6,874 | 18.91  | 3.878 | 7   | 35    | NLSY79 Child/YA | Avg.     | INSECUR[year]      | Created      | -   |
| 6        | Avg. Sociability              | 6,620 | 11.32  | 2.408 | 3   | 15    | NLSY79 Child/YA | Avg.     | SOCIAB[year]       | Created      | -   |
| 6        | Avg. Self-Worth               | 7,027 | 204.8  | 28.96 | 60  | 240   | NLSY79 Child/YA | Avg.     | SPPCG[year]        | Created      | -   |
| 6        | Avg. Self-Perception (School) | 6,988 | 177.3  | 35.86 | 19  | 240   | NLSY79 Child/YA | Avg.     | SPPCS[year]        | Created      | -   |
| 6        | Avg. Digital Memory           | 7,464 | 9.727  | 2.841 | 1   | 19    | NLSY79 Child/YA | Avg.     | DIGITZ[year]       | Created      | -   |
| 6        | Pearlin Scale (Mother)        | 7,609 | 493.8  | 92.37 | 51  | 891   | NLSY79          | -        | PEARLIN_ZSCORECW   | Created      | -   |
| 6        | Rotter Scale (Mother)         | 7,913 | 8.950  | 2.405 | 4   | 16    | NLSY79          | -        | ROTTER_SCORE       | Created      | -   |
| 6        | Rosenberg Scale (Mother)      | 7,736 | 488.8  | 104.3 | 14  | 941   | NLSY79          | -        | ROSENBERG_ZSCORECW | Created      | -   |

## Appendix 2: Circumstance Variables BCS70

| Scenario | Var. Name                        | N    | Mean   | Std. Dev. | Min | Max  | Study | Ref.Age | Question Name                    | Var. Type        | Question Text   |
|----------|----------------------------------|------|--------|-----------|-----|------|-------|---------|----------------------------------|------------------|---|
| 1        | Ethnic group father              | 6418 | 0,022  | 0,162     | 0,0 | 2,0  | BCS70 | 10      | e246b                            | Primary Var.     | Ethnic Group-Father   |
| 1        | Ethnic Group Mother              | 6620 | 0,018  | 0,144     | 0,0 | 2,0  | BCS70 | 10      | e246a                            | Primary Var.     | Ethnic Group-Mother   |
| 1        | Gender                           | 6679 | 0,570  | 0,495     | 0,0 | 1,0  | BCS70 | 0       | a0255                            | Primary Variable | Sex of the Baby   |
| 1        | Foreign origin                   | 6601 | 0,054  | 0,226     | 0,0 | 1,0  | BCS70 | 0       | a0007a                           | Primary Variable | Region of Birth of Father   |
| 1        | Body height                      | 6602 | 1,708  | 0,100     | 1,4 | 2,4  | BCS70 | 42      | BD9HGHTM                         | Created          | How tall are you without shoes?   |
| 1        | Parents income age 10            | 5908 | 4,193  | 1,232     | 1,0 | 7,0  | BCS70 | 10      | c9_1 - c9_7                      | Primary Variable | Total gross family income   |
| 1        | Parents income age 16            | 4395 | 5,213  | 2,469     | 1,0 | 11,0 | BCS70 | 16      | oe2                              | Primary Variable | Combined income of parents per wk/mth   |
| 1        | Childhood urbanization area      | 4889 | 0,775  | 0,542     | 0,0 | 2,0  | BCS70 | 10      | m304 - m307                      | Primary Variable | Neighbourhood description rural/COUNTRY-IN OR CLOSE TO VILLAGE/OUTSKIRTS OF TOWN OR CITY/INNER URBAN AREA   |
| 1        | Fathers education                | 5805 | 1,111  | 1,029     | 0,0 | 3,0  | BCS70 | 10      | c1_1 / c1_2 / c1_3 / c1_6 / c1_9 | Primary Variable | Father's qualifications   |
| 1        | Fathers occupation               | 6381 | 1,058  | 1,275     | 0,0 | 5,0  | BCS70 | 0       | a0013                            | Primary Variable | Socio Economic Group Father   |
| 2        | Stand. math test score age 16    | 5790 | 47,687 | 11,398    | 1,0 | 71,0 | BCS70 | 16      | BD3MATHS                         | Created          | Generated by ESRC Researcher Development Initiative (RDI) project ( <a href="http://www.cls.ioe.ac.uk/teachingresources">http://www.cls.ioe.ac.uk/teachingresources</a> )   |
| 2        | Stand. reading test score age 16 | 5704 | 10,462 | 1,403     | 5,4 | 13,1 | BCS70 | 16      | BD3RDAGE                         | Created          | Generated by ESRC Researcher Development Initiative (RDI) project ( <a href="http://www.cls.ioe.ac.uk/teachingresources">http://www.cls.ioe.ac.uk/teachingresources</a> )   |
| 2        | Highest level of education       | 6667 | 2,769  | 1,344     | 0,0 | 5,0  | BCS70 | 34      | bd7hq5                           | Primary Variable | What is the highest level of qualification you have obtained?<br>01 CSE<br>02 GCSE<br>03 GCE O Level<br>04 A/S Level<br>05 GCE A Level (or S Level)<br>06 Scottish School Certificate, Higher School Certificate or Scottish School Qualification<br>07 Diploma of Higher Education<br>08 Degree (e.g. BA, BSc)<br>09 Other degree level qualification such as graduate membership of professional institutes<br>10 Higher degree (e.g. PhD, MSc)<br>11 Nursing or other para-medical qualification<br>12 PGCE - Post-graduate Certificate of Education |
| 3        | Malese test score                | 3932 | 1,135  | 0,342     | 1,0 | 2,0  | BCS70 | 16      | BD4MALG                          | Created          | Generated by ESRC Researcher Development Initiative (RDI) project ( <a href="http://www.cls.ioe.ac.uk/teachingresources">http://www.cls.ioe.ac.uk/teachingresources</a> )   |
| 4        | Common activities with parents   | 3935 | 1,604  | 0,885     | 1,0 | 4,0  | BCS70 |         | c5r6 / c5r7                      | Primary Variable | Share outdoor hobby with parents/Share indoor hobby with parents  |

| Scenario | Var. Name                          | N    | Mean  | Std. Dev. | Min  | Max | Study | Ref.Age | Question Name | Var. Type        | Question Text   |
|----------|------------------------------------|------|-------|-----------|------|-----|-------|---------|---------------|------------------|---|
| 4        | Parents married                    | 6671 | 0,042 | 0,242     | 0,0  | 3,0 | BCS70 | 0       | a0012         | Primary Variable | Present Marital Status  |
| 4        | Time spent with parents            | 4128 | 3,173 | 1,416     | 1,0  | 5,0 | BCS70 | 16      | gb8_3         | Primary Variable | Do things with both parents together?   |
| 5        | Alcohol use father                 | 5135 | 3,631 | 1,806     | 0,0  | 6,0 | BCS70 | 16      | pg8_2         | Primary Variable | How often husband has alcoholic drink   |
| 5        | Alcohol use mother                 | 5445 | 2,794 | 1,898     | 0,0  | 6,0 | BCS70 | 16      | pg8_3         | Primary Variable | How often mother has alcoholic drink  |
| 5        | Smoking habit father               | 5884 | 2,819 | 2,707     | 0,0  | 6,0 | BCS70 | 16      | e263          | Primary Variable | Smoking Habits Since 1970:Father  |
| 5        | Smoking habit mother               | 5909 | 3,596 | 2,694     | 0,0  | 6,0 | BCS70 | 16      | e262          | Primary Variable | Smoking Habits Since 1970:Mother  |
| 6        | Test score picture recognition     | 5007 | 0,206 | 0,905     | -3,0 | 1,9 | BCS70 | 5       | BD2READ       | Created          | Generated by ESRC Researcher Development Initiative (RDI) project ( <a href="http://www.cls.ioe.ac.uk/teachingresources">http://www.cls.ioe.ac.uk/teachingresources</a> ) |
| 6        | Test score reading                 | 5007 | 5,206 | 0,905     | 2,1  | 6,9 | BCS70 | 5       | BD2RDAGE      | Created          | Generated by ESRC Researcher Development Initiative (RDI) project ( <a href="http://www.cls.ioe.ac.uk/teachingresources">http://www.cls.ioe.ac.uk/teachingresources</a> ) |
| 6        | Years of educ. Father              | 6074 | 0,629 | 1,418     | 0,0  | 6,0 | BCS70 | 10      | e196          | Primary Variable | Years of Ft Educ After Age 15-Father  |
| 6        | Years of educ. Mother              | 6416 | 0,430 | 1,016     | 0,0  | 6,0 | BCS70 | 10      | e195          | Primary Variable | Years of Ft Educ After Age 15-Mother  |
| 6        | Highest educ. Father               | 6152 | 3,007 | 2,258     | 1,0  | 8,0 | BCS70 | 10      | e189b         | Primary Variable | Fathers Highest Educ Qualification  |
| 6        | Highest educ. mother               | 6446 | 2,312 | 1,668     | 1,0  | 8,0 | BCS70 | 10      | e189a         | Primary Variable | Mothers Highest Educ Qualification  |
| 6        | Malese score mother                | 5385 | 1,176 | 0,453     | 1,0  | 3,0 | BCS70 | 16      | BD4MMALA      | Created          | Generated by ESRC Researcher Development Initiative (RDI) project ( <a href="http://www.cls.ioe.ac.uk/teachingresources">http://www.cls.ioe.ac.uk/teachingresources</a> ) |
| 6        | Medical condition during childhood | 5709 | 0,165 | 0,449     | 0,0  | 2,0 | BCS70 | 10      | m13 / m14     | Primary Variable | HAS CHILD ANY MEDICAL CONDITION / EFFECT ON DAILY LIFE HOME OR SCHOOL   |
| 6        | Ruthers test score mother          | 5003 | 1,161 | 0,428     | 1,0  | 3,0 | BCS70 | 16      | BD4MRUTG      | Created          | Generated by ESRC Researcher Development Initiative (RDI) project ( <a href="http://www.cls.ioe.ac.uk/teachingresources">http://www.cls.ioe.ac.uk/teachingresources</a> ) |
| 6        | Type of school                     | 2431 | 1,984 | 2,166     | 1,0  | 8,0 | BCS70 | 16      | stype         | Primary Variable | Type of School  |

### Appendix 3: Interactions in base circumstance set

Figure 9: IOp with varying circumstance sets (NLSY79), survey-specific pooled sample

