Measuring Income Inequality of Opportunity

Accounting for Dynamic Complementarity

Aman Desai*

Job Market Paper

04 December, 2024

Abstract

The egalitarian principle of justice attributes life success to two main factors: circumstances beyond an individual's control and personal effort within it. Roemer's equality of opportunity concept proposes compensating individuals for inequalities arising from unequal circumstances. Dynamic complementarity in skill formation suggests that early childhood skill gaps often persist into adulthood, leading to unequal outcomes. Using PSID data, I classify all measurable factors before age 18 (the age of majority) as circumstances, creating sets based on critical childhood stages to account for dynamic complementarity. My findings show that over 40% of total income inequality can be attributed to inequality of opportunity before adulthood. Moreover, nearly one-third of total income inequality stems from circumstances faced by individuals at or before age five. Using only circumstances identified as important through a random forest—a supervised machine learning model—based on permutation-based importance scores, I estimate the lower bound of inequality of opportunity's share in total inequality before the age of majority to be about 31%. These results underscore the importance of considering childhood circumstances when measuring inequality of opportunity. This consideration is crucial for any public policy involving ex-post compensation or ex-ante investment in human capital to equalize opportunities.

Keywords: Inequality of opportunity, dynamic complementarity, random forest.
JEL Classification: C60 ; D31 ; D63; J13 ; J24

*PhD Program in Economics,The Graduate Center, City University of New York, 365 Fifth Avenue, New York NY 10016, United States. Website: https://amandesai.com/. Email: adesai@gradcenter.cuny.edu. I am grateful to my committee for all their invaluable guidance and support : Miles Corak (chair), Theodore Joyce, Wim Vijverberg, Lilia Maliar. I am also grateful to Noura Insolera for her support on the usage of PSID data and conference participants for their comments on the paper. Additionally, I appreciate Janet Gornick and the seminar participants at the Stone Center on Socio-Economic Inequality for their valuable input on the paper. Finally, I am indebted to the CUNY Graduate Center PhD Program in Economics dissertation seminar participants for their helpful comments. All errors and omissions are mine.

1 Introduction

Some forms of inequality in society are unjust, yet determining which types are fair presents an ethical dilemma. When addressing inequality, it is crucial to consider the mechanisms that enable individuals to succeed in life. Since Rawls (1971), the concept of egalitarianism has shifted from focusing on welfare derived from final outcomes to examining the processes leading to those outcomes. Economists now incorporate the idea of fairness in rewarding individual responsibility while acknowledging the existence of unfair inequalities in their analyses.

Roemer (1993) made a vital contribution by proposing that success in life is broadly determined by two elements: “circumstances,” over which individuals have no control and for which they should not be held responsible, and “effort,” which represents factors within an individual’s control. Equality of opportunity is achieved when the distribution of outcomes depends only on effort, not on circumstances. This formulation aligns with the concept of a “level playing field.” The literature on redistributive preferences explains how individual views on such policies correlate with beliefs about the impact of effort versus circumstances on outcomes (Alesina and Giuliano (2011)). Fong (2001) demonstrates that people are more accepting of inequality resulting from differential effort rather than unequal circumstances. From a behavioral perspective, Starmans, Sheskin, and Bloom (2017) uses laboratory studies, cross-cultural research, and experiments with infants and young children to show that humans naturally favor fair distribution over unequal distribution. When equality and fairness conflict, people prefer fair inequality to unfair equality.

The empirical literature measures the extent of inequality of opportunity (IOp hereafter) for various outcomes, including income, wages, and health, in many countries (Fleurbaey and

Peragine 2013; Roemer and Trannoy 2016; Ferreira and Peragine 2015; Ramos and Van de gaer 2016). I contribute to this literature by measuring income inequality due to unequal opportunities, creating age-based circumstance sets using the age of majority as a responsibility cutoff. Numerous studies have estimated the extent of income inequality due to circumstances. For the US, Pistolesi (2009) estimates IOp between 20% and 43% of earnings inequality. Using NLSY79 data, Hufe et al. (2017) estimate IOp shares in income inequality from 27.1% to 43.5%. The recent launch of the Global Estimates of Opportunity and Mobility (GEOM) database marks a significant step toward understanding global inequality of opportunity. This public data repository includes estimates from 72 countries representing 67% of the world's population, aiming to highlight how income inequality is influenced by circumstances beyond individual control, such as parental background and geographic location.

The lack of high-quality datasets reflecting all circumstances faced by individuals leads to partial observability, resulting in a downward bias in IOp estimates (Bourguignon, Ferreira, and Menéndez 2007; Ferreira and Gignoux 2011; Niehues and Peichl 2014). There's also an issue of arbitrary categorization of circumstances and effort variables. A factor considered a circumstance by one researcher may not be categorized as such by others. Since the distinction between "effort" and "circumstances" is a value judgment, measuring the role of circumstances accurately in predicting adult outcomes is challenging. I take a radical—but not unprecedented—position (Hufe et al. (2017)): all measurable factors, behavioral or otherwise, before the age of majority are considered circumstances. The law determines when a child becomes an adult and is ready to stand on their own (e.g., voting laws, drinking age). Therefore, I propose using this societal value judgment to categorize variables as either "effort" or "circumstances." Theoretically, if I had all information about a child before the age of majority, I would categorize that information as circumstances. Roemer's idea of equality of opportunity requires that a child should not be held responsible for factors affecting them before the age of majority, including their achievements.

Following this view, all measurable factors before the age of majority (e.g., 18 years) could be categorized as circumstances. The inequality in outcomes generated via these circumstances could be considered "unfair" and should be addressed. Roemer (1993) proposes that individuals affected by adverse circumstances warrant compensation. In this paper, I bring an insight—dynamic complementarity—from the literature surveyed by Heckman and Mosso (2014) to contribute to the inequality of opportunity literature. While measuring IOp, I account for dynamic complementarity in skill formation. Skills gaps that open early in childhood due to unequal circumstances tend to persist into adulthood. Any policy to address this inequality using compensation later in life may prove inefficient if early childhood skills gaps haven't been addressed. It is important to measure inequality of opportunity rigorously using these early childhood circumstances to better inform policy decisions. I measure inequality of opportunity using circumstances children face at critical stages in their development before the age of majority.

Recent empirical studies have used machine learning algorithms to create counterfactual distributions of outcomes and identify circumstances. Using representative survey data from 31 European countries, Brunori, Hufe, and Mahler (2023) show the superiority of tree-based models in creating counterfactual distributions using circumstance data. Machine learning algorithms such as decision trees and their ensemble random forest also allow interaction among circumstance factors. These algorithms offer flexibility in modeling non-linear relationships between circumstances and outcome variables. I follow this practice and utilize the random forest algorithm to calculate estimates of inequality of opportunity. To account for dynamic complementarity in skill formation, I create age-based opportunity sets using circumstances at or before critical childhood stages. I then use the random forest algorithm to create counterfactual distributions of adult income for these different circumstance sets and apply an inequality measure to obtain IOp estimates and their share in total income inequality. Additionally, machine learning techniques offer an advantage over subjective variable selection by researchers. These methods empirically identify useful variables from the set of circumstances with minimal human intervention. Leveraging this advantage, I use permutation-based variable importance scores to identify the circumstances that contribute most significantly to predicting adult income inequality. Using these key circumstances in the model, I obtain lower-bound estimates of inequality of opportunity.

Using Panel Study of Income Dynamics (PSID) data on both the Survey Research Center (SRC) sample and the full sample—which includes the Survey of Economic Opportunity sample—I estimate the share of “unfair inequality” in total income inequality to be about 40-45%, depending on the sample used. I argue these are upper-bound estimates of the IOp share in adult income inequality. Additionally, I perform the analysis using only circumstances identified by the permutation-based variable importance scores. For the full sample, I estimate the lower bound of the inequality of opportunity share in total income inequality to be about 31%. For the SRC sample, this estimate is about 29%. It is important to clarify that the nature of the problem examined is not causal. Rather, the objective is to determine the extent to which variations in adult incomes can be attributed to circumstances perceived as “unfair.” This approach classifies it as a prediction problem, best addressed using supervised machine learning techniques.

The paper is structured as follows: the next section briefly covers the theoretical framework, explaining the concepts of inequality of opportunity and dynamic complementarity. Section 3 describes the data. Section 4 details the measurement of IOp, section 5 presents the results, and section 6 concludes.

2 Theoretical Framework

2.1 Inequality of Opportunity

Consider a population $\mathcal{N} = \{1, 2, \dots, N\}$. Each individual in the population is characterized by a triple (y, C, e) where $C \in \Omega^c$, $e \in \Omega^e$, and $y = g(C, e)$, with $g : \Omega^c \times \Omega^e \Longrightarrow R$. The outcome vector $y = (y_1, \dots, y_N)$ represents the incomes of individuals, which depend on circumstances C and effort e. An individual in the population is identified by a *type* and a *tranch*. A *type* consists of individuals with the same circumstances beyond their control. If the population is divided into M mutually exclusive and exhaustive groups, called *types*, such that $\prod = \{\tau_1, \tau_2, \dots, \tau_M\}$, then all individuals belonging to the same *type* τ_m share the same circumstances: $\tau_1 \cup \tau_2 \cup \dots \cup \tau_M = \{1, \dots, N\}$, $\tau_m \cap \tau_k = \emptyset, \forall m, k$ and $C_i = C_j, \forall i, j \mid i, j \in \tau_m, \forall m$.A *tranch* consists of individuals with the same effort. According to Roemer, equality of opportunity is achieved when inequality generated due to differential circumstances is eliminated (between *types*), that is $F(y|C) = F(y)$. Inequality of opportunity is measured by the extent to which this principle is violated, that is $F(y|C) \neq F(y)$.

Following the egalitarian project, Roemer (1993) argues for the *ex-post compensation* principle, which calls for compensation after the effort is realized. The *ex-post compensation* principle requires that individuals exerting the same degree of effort receive the same outcomes, regardless of their circumstances. Roemer proposes a model of optimal taxation where the social planner's objective function incorporates an aversion to inequality caused by circumstances beyond an individual's control. Effort is typically unobservable. Roemer offers a solution to identify the effort predicated on some assumptions.

1. The circumstances faced by the individuals are fully observed.
2. Outcome is monotonically increasing in effort. Higher effort implies higher outcome.

$$y^m(e_i) \geq y^m(e_j) \Leftrightarrow e_i^m \geq e_j^m, \forall m = 1, ...M, \forall\ e_i, e_j \in R \tag{1}$$

3. Effort, by definition, is orthogonal to circumstances.

Roemer (2002) argues that when comparing the efforts of different individuals, we should take into account their specific effort distributions based on their *types*, and individuals should not be held solely responsible for these differences. Indeed, Roemer distinguishes between "level of effort" and "degree of effort", with latter being ethically relevant which can be identified with the quantile of the *type*-specific effort distribution of the individual. We denote distribution of effort in *type* m with $G^m(e)$ and its quantiles with $\pi \in [0, 1]$. For example, consider two individuals, A and B, born into a wealthy family and a poor family respectively. If they exert the same level of effort, the degree of effort is higher for child B due to her less advantaged circumstances. Instead of directly comparing their effort levels, Roemer suggests comparing their ranks (quantiles) on individual *type*-specific effort distributions. Since, effort distributions are mostly unobservable, Roemer suggests to identify the degree of effort exerted by the individual with the quantile of their *type*-specific outcome distribution. i.e. $y^m(G^m(e)) = y^m(\pi)$. Then the requirement for the same outcome due to same degree of effort exerted by the individuals is

$$y^m(\pi) = y^k(\pi) \Leftrightarrow F^m(y) = F^k(y); \forall \pi \in [0, 1], \forall m;\ k = 1, ..., M. \tag{2}$$

As explained, the implication of Roemer's adherence to the ex-post compensation principle is that society should compensate individuals for their unequal circumstances after individual effort is realized. This contrasts with the *ex-ante compensation* principle, where compensation is due before effort is realized by equalizing the opportunity sets available to everyone, regardless of their circumstances.

2.2 Importance of Skills in Early Childhood

Skills are multidimensional, covering cognition, personality, as well as mental and physical health. They reflect an individual's capacity to act. Borghans et al. (2008) and Almlund et al. (2011) offer comprehensive surveys of recent studies showing that both cognitive and non-cognitive skills have an impact on labor market outcomes. The literature on human capital and child development provides ample evidence of how early circumstances can predict adult outcomes. P. M. Carneiro and Heckman (2003) demonstrate that significant differences in children's skills, depending on their family backgrounds, emerge at an early age and persist over time. These skill differences impact success in the labor market and other life aspects. Cunha, Heckman, and Navarro-Lozano (2004) report that approximately 60% of the residual variance in log wages can be attributed to skills developed by late adolescence. Neal and Johnson (1996) link a significant portion of the black-white wage gap for men to cognitive skill disparities identified years before these individuals enter the job market. Furthermore, Heckman, Stixrud, and Urzua (2006) highlight that both cognitive and non-cognitive skills directly influence not only labor market outcomes but also a wide range of life experiences. These include the likelihood of unemployment, welfare usage, teenage pregnancy, criminal activity participation, and drug use.

Gaps in both cognitive and non-cognitive skills emerge early in childhood, across individuals and socio-economic groups. There is substantial evidence of early divergence in these skills even before schooling begins.[1] These skill gaps correspond to gaps in family investment and the environment in which individuals are brought up. Hart and Risley (1995) showed that children from professional families speak 50% more words than children from working-class families, and twice as many as children from welfare families. There is a substantial literature, summarized in Cunha et al. (2006), Lareau (2011), Kalil (2015) showing that disadvantaged children have compromised early environments as measured on a variety of dimensions. Moreover, various skills and abilities are critical at different stages of the life cycle. Early life disadvantages have a lasting impact on a range of outcomes in adulthood. Cunha et al. (2006) provide a review of studies that examine the significance of early childhood environments on socioeconomic outcomes in adolescence and adulthood. The empirical studies show that investing in disadvantaged young children yields higher economic returns.[2] Early interventions have been shown to be more effective than targeted interventions later in life, as high-quality interventions during the early years promote the development of skills in disadvantaged young children that lead to greater economic returns in the future. Non-cognitive skills foster cognitive skills and are an important product of successful families and successful interventions in disadvantaged families.

2.3 Technology of Skill Formation

Both cognitive and non-cognitive skills, the technology used for their development, and parental investment, which includes their own skills, are crucial in determining the dynamics of family

[1] Cunha et al. (2006), and Cunha and Heckman (2007).

[2] for comprehensive survey of empirical literature on human development and social mobility see Heckman and Mosso (2014).

influence. Cunha and Heckman (2007) model technology for skill formation, where formulation of skills is conceptualized as a law of motion. Let $\omega_{i,t}$ denote the human capital of child i at age t, $\omega_{i,t+1}$ the human capital at age $t+1$. Let parental investment for the child i be $x_{i,t}$ at age t and parental human capital be ω_i^p. $\epsilon_{i,t}$ is an independently and identically distributed unobserved individual component.

$$\omega_{i,t+1} = f(\omega_{i,t}, x_{i,t}, \omega_i^p, \epsilon_{i,t}) \tag{3}$$

The equation 3 captures the idea of static complementarity between investment in human capital in period t and skills in t. Children who are more intelligent, healthier, have better non cognitive skills acquire more capability from the same level of investment. $f(.)$ is assumed to be twice continuously differentiable, increasing in all arguments, and concave in $x_{i,t}$. Stock of skills $\omega_{i,t}$ and $\omega_{i,t+1}$ include both cognitive and non cognitive skills. The dimensions of $\omega_{i,t}$ and $f(.)$ are allowed to increase with the stage of the life cycle. The technology in the model is stage-specific and allows for critical and sensitive periods in the formation of capabilities and effectiveness of the investment. This formulation of technology has two implications.

First it implies that $\frac{\partial \omega_{i,t+1}}{\partial \omega_{i,t}} > 0$, that is, when higher stocks of skills in one period create higher stocks of skills in the next period. Second it implies that $\frac{\partial^2 \omega_{i,t+1}}{\partial \omega_{i,t} \partial x_{i,t}} > 0$, that is, when stocks of skills acquired by period t, $\omega_{i,t}$, make investment in period $t+1$, $x_{i,t}$ more productive. For the case of skill vectors, this includes own and cross effects. These generate dynamic complementarity between investment in period t and in period k where $k > t$. Higher investment in period t increases $\omega_{i,t+1}$ as $f(.)$ is increasing in $x_{i,t}$. This in turn raises $\omega_{i,k}$ because technology is increasing in $\omega_{i,m}$ for any m between t and k. This in turn leads to $\frac{\partial f}{\partial x_{i,k}} > 0$, since $\omega_{i,k}$ and

$x_{i,k}$ are complements. It follows that

$$\frac{\partial^2 \omega_{i,t+k+1}}{\partial x_{i,t} \partial x_{i,t+k}} > 0, \quad \forall k \geq 1. \tag{4}$$

Investment in period $t+k$ and investment in any prior years t are always complements as long as $\omega_{i,t+k}$ and $x_{i,t+k}$ are complements. These properties help explain why early investment in disadvantaged children can yield high productivity, which is both fair and economically efficient. Conversely, the return on investment tends to be lower at later stages for disadvantaged children, due to their lower skill base and hence reduced complementarity effect. While this may seem fair, it may be less economically efficient.

The concept of dynamic complementarity implies that early differences in skill investments can lead to enduring disparities in adult outcomes. I argue that a child encounters situations beyond their control before reaching the age of majority. By applying the principle of dynamic complementarity, I highlight the unequal opportunities arising from unequal circumstances in early life stages when measuring inequality of opportunity. By identifying specific age milestones in childhood, I can analyze the extent to which inequality in adult incomes can be attributed to circumstances before or during these stages. In the United States, for instance, children typically start speaking at age 2, attend kindergarten at age 5, begin high school at age 14, and transition into adulthood at age 18. By focusing on these significant stages of development, I can more accurately measure the opportunity gaps resulting from circumstances preceding these critical childhood stages.

3 Data Description

The data used in this study comes from the Michigan Panel Study of Income Dynamics (PSID)—the longest-running longitudinal survey in the United States—beginning in 1968 with a coverage of 4,800 households. The survey annually until 1997, and has since been conducted biennially. The genealogical design of PSID data allows to link individuals of interest to their parents and grandparents.

3.1 Analytical Sample

The PSID was originally created to study poverty. As a result, it disproportionately sampled individuals from poor households, which are included in the SEO (Survey of Economic Opportunity) sample. I construct two complementary analytical samples. The first is restricted to individuals from the SRC (Survey Research Center) sample, ensuring a representative sample of the US population. The second—full sample—includes individuals from both the SRC and SEO samples. To account for the inclusion of the SEO sample in the full sample, I apply appropriate weights provided by the PSID. This approach yields a larger sample size compared to limiting the analysis solely to the SRC sample.

The individuals of interest are the heads of the family and their spouses/partners who were present during the interviews conducted in the 2013-2019 waves. Since any measurable data on a child before the age of majority is considered circumstances, the goal is to use the data before the age of majority, which in the American context for certain rights and privileges I take to be 18. I use these to predict individual labor market incomes. I restrict my analytical samples to the family heads and their spouses/partners who participated in interviews during the 2013-2019 waves.

These individuals were born between 1978 and 1983. The data includes various characteristics about them and the families they grew up in during the first 18 years of their lives. I use the Family Relationship Matrix (FRM) to identify the family of the individual. I also use the Family Identification Mapping System (FIMS) provided by PSID to link these individuals to their parents in their first 18 years of life. The sample consists of data on family heads, their spouses, and in some cases, other family members. The head of the family can be a father, a step-father, a grandfather, or in some cases, a single mother. Therefore, some children at some point in time may or may not have their parents as the heads of their family. If the family head is a parent, then using FIMS ensures those parents are the individual's biological parents.

The data includes around 60 variables on 1022 individuals in the full sample and 639 individuals in the SRC sample across 18 years. I create wide datasets according to pre-determined age cut-offs where each row reflects a biography of an individual across their first 18 years. Since the data only includes variables before a child's majority age, all these factors should be considered circumstance variables. As mentioned earlier, various skills and abilities are critical at different stages of the life cycle. Dynamic complementarity suggests that gaps in skills attained at different critical and sensitive periods of childhood tend to persist in adulthood and lead to unequal outcomes. Moreover, dynamic complementarity and self productivity together suggest that lack of investments in skills at early stages lead to low returns to human capital investment in later stages of life. The Panel Study of Income Dynamics (PSID) includes data on measurable factors at various stages of childhood. This makes it suitable for accounting for critical stages before the age of majority, allowing for the creation of age-based circumstance sets.

Table 1: Selected Circumstances

Family/Demographic	**Market/Monetary**	**Government/Community**
Race, sex of the individual	Family income	Usage of foodstamps
Race of the family head, spouse	Childcare cost	Medicaid/Medicare usage in the family
Sex of the head	Homeownership	Help from family members, others, insiders
Education of the family head, spouse	Marginal tax rate on family income	Any outside dependents for head?
Occupation of the family head, spouse	Value of family home	Union membership of the family head, spouse
Number of children to father, mother		Availability of a car
Marital status of mother when individual was born		
Number of rooms in family home		
State of residence of family		
Birthweight		
Birthcohort		

Ideally, we would have a complete biography of individuals spanning their first 18 years of life. The selection of factors to include as circumstances is guided by economic literature. In the table 1, I present the circumstances I considered within the contexts of family, markets, and government. All of these variables are measured across the first 18 years of an individual's life. However, some factors are measured more frequently than others. The frequency of these factors depends on the availability of data in the PSID. For instance, family income is measured for all 18 years, while family wealth is measured less frequently. Using these factors, I construct circumstance sets based on age cutoffs corresponding to critical stages in childhood. For example, in the United States, children typically start speaking at age 2, attend kindergarten at age 5, enter high school at age 14, and transition into adulthood at age 18. By focusing on these developmental stages, we can more accurately assess the opportunity gaps resulting from conditions that occur before these critical points. As a child matures into an adult, the number of circumstances they encounter increases, which aligns with the evolving dimensions of factors in skill formation technology.

The main outcome of interest in this study is individual's permanent income. To proxy for an individual's permanent income, I use two measures. First, I use their labor income measured at age 35. This means that for individuals born in 1978, their labor income is measured in 2013. Similarly, for individuals from other birth cohorts, such as those born in 1983, their incomes are measured in 2018. For simplicity, the labor income at age 35 excludes farm and unincorporated business income. Additionally, I omit individuals with income below $500 from the analyses. The decision to measure labor income at age 35 provides a large enough sample size while ensuring it serves as a reasonable proxy for an individual's permanent income.

Labor income measured at a single point in time is susceptible to measurement error and can lead to attenuation bias (Solon 1992; Nybom and Stuhler 2017). To minimize this attenuation bias, I proxy the individual's permanent income using their labor income in adulthood, averaged over four years from 2013-2019. For individuals with missing income data in any wave, I calculate their average income using only the available years. For example, if an individual has income data for 2012, 2014, and 2018, but missing in 2016, I compute their average income using three years (2012, 2014, 2018). For the cohorts under consideration (born in 1978-1983), average incomes are measured at different ages. For example, in 2015, an individual born in 1978 is on average 37 years old, while someone born in 1982 is on average 33 years old.

The outcome variables are measured in logarithmic terms. Finally, I use urban CPI series (CUUR0000SA0) from Bureau of Labor Statistics for consumer price index to convert all monetary variables to 2018 dollars. The data is in wide format. Hence, I use individual cross-sectional weights from the survey waves in which individuals were identified. For example, for individuals born in 1978, their cross-sectional weights were pulled from the 2013 survey wave. For those born in 1979 and 1980, their cross-sectional weights were taken from the 2015 survey wave. This same approach applies to individuals in the remaining cohorts. I perform analyses using both outcome variables 1) individual labor income at age 35 and 2) age adjusted averaged labor income across four years, for both the full sample and the SRC sample.

4 Measuring Inequality of Opportunity

(Ramos and Van de gaer 2016) provide a comprehensive survey of various measures proposed based on differing normative views. I use a widely adopted ex-ante utilitarian measure of inequality of opportunity (Van de Gaer 1993; Checchi and Peragine 2010.). The idea is to construct a counterfactual smoothed distribution of outcomes obtained by removing inequality within types (circumstances) from the original outcome distribution. Measuring inequality of opportunity involves two steps. First, I form a counterfactual smoothed distribution of outcomes based on individual types, or circumstances. Then, I apply a standard measure of inequality that satisfies anonymity, the principle of transfers, population replication, and scale invariance to the counterfactual distribution conditional solely on circumstances.[3] I use what is referred to as parametric specification in the literature for estimation of lower bounds of IOp (Bourguignon, Ferreira, and Menéndez 2007; Ferreira and Gignoux 2011; Niehues and Peichl 2014).[4]

$$ln(y_i) = \alpha_0 + \sum_{l=1}^{L}(\alpha_l C^s_{i,l}) + u_i \tag{5}$$

where y is the adult income, C is the collection of factors that are categorized as circumstance belonging to a finite set Ω^c. $s \in \{2, 5, 14, 18\}$ reflecting four different sets of circumstances

[3] See Cowell (2016) for more information

[4] In the US, discussions typically revolve around intergenerational income mobility (Corak 2013; Chetty et al. 2014), which is indeed a special case of equation 5 where parental income enters as the sole circumstance. In the appendix, I provide IGE estimates using the age cutoffs employed in this study to measure inequality of opportunity. The directional patterns are consistent with the results of this study.

based on chosen age cutoffs. The smoothed distribution of $\hat{y}$ is then obtained using equation 5

$$\hat{y}_i = exp\left[\alpha_0 + \sum_{l=1}^{L}(\hat{\alpha}_l C^s_{i,l})\right] \tag{6}$$

If any measurable data in a child's life before the age of majority is considered part of the circumstance set Ω^c, then the data that ideally includes a biography of a child across the first 18 years will form a circumstance set. Although incomplete, the PSID offers an extensive list of factors across the first 18 years that make up the circumstance set. To account for dynamic complementarity, I construct four circumstance sets. I use age-based circumstances to create opportunity structures based on critical stages of development in childhood.

$$C^2 \subseteq C^5 \subseteq C^{14} \subseteq C^{18} \subseteq \Omega^c \tag{7}$$

This formulation allows us to expand the circumstance set with age to account for additional circumstances a child faces at different stages of childhood before she becomes an adult. For instance $C^2 \subseteq \Omega^c$ includes the circumstances the child faces prior to or at age 2. $C^{18} \subseteq \Omega^c$ will make use of full set of circumstances in the data that include factors across first 18 years of the child's life. Similar interpretation holds for all other circumstance sets.

Obtaining $\hat{y}$ is a prediction problem where the relationship between circumstances and outcome is unknown a priori. Researchers have proposed various methods to obtain $\hat{y}$ in the existing literature. Economists are increasingly turning to machine learning techniques to solve such prediction problems. Notably, supervised machine learning methods outperform traditional OLS regression in generating out-of-sample predictions (Mullainathan and Spiess 2017). Brunori, Hufe, and Mahler (2023) demonstrate the superiority of tree-based supervised machine learning models over existing estimation methods. They use conditional inference trees and forests to generate $E(y|C)$ predictions. These machine learning models outperform the standard OLS as well as latent class models proposed by Donni, Rodríguez, and Dias (2015), especially when the potential number of types exceeds the available degrees of freedom. I adopt their estimation strategy, but unlike Brunori, Hufe, and Mahler (2023), I use random forests to generate predictions. Random forests are an interactive function class and hence allow for non-linearity among the "types", that is circumstances. After obtaining the adult income predictions based on circumstances, I apply an inequality measure, mean logarithmic deviation to the predicted income distribution to calculate absolute inequality of opportunity.

$$Absolute\ IOp = I(\hat{y}_{EA}) \tag{8}$$

where $I(\hat{y}_{EA})$ is the ex-ante measure of inequality of opportunity. I also report relative inequality of opportunity as a ratio of inequality in predicted income distribution to inequality in adult income. It can be interpreted as the share of inequality in adult income that is attributed to inequality of opportunity. The value of relative IOp ranges from 0 to 1. If all income differences are solely due to circumstances, relative IOp will be 1.

$$Relative\ IOp = \frac{I(\hat{y}_{EA})}{I(y)} \tag{9}$$

4.1 Regression Trees

Machine learning techniques extract information from data, identify patterns, and make statistical decisions with minimal human intervention. Instead of relying on subjective variable selection by researchers, these techniques allow us to empirically identify useful variables from the set of circumstances. An algorithm used in the literature to identify types (circumstances) is a regression tree. Similar to a linear regression function, a regression tree also predicts an outcome value for each feature vector. The prediction function takes the form of a tree that splits the feature space into two at every node. At each node, a single variable determines whether the left or right child node is considered next. When a terminal node, or “leaf,” is reached, a prediction is returned. Trees are thus a highly interactive function class. and allow to create “types”, that is, circumstances from the data.

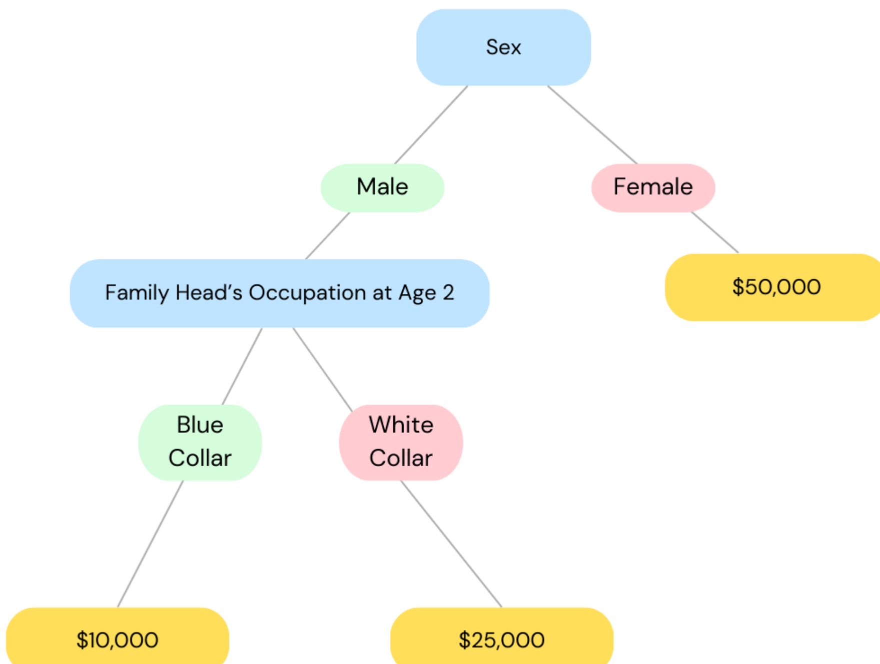

Figure 1: An example of a tree

A regression tree algorithm makes predictions by stratifying the feature space through a process called *recursive binary splitting*. This top-down, greedy approach starts at the top of the tree and splits the predictor space into two new branches further down the tree at each split.

During each step of the tree-building process, the best split is created at that step, rather than looking ahead and selecting a split that will lead to a better tree in a future step. The goal is to minimize the loss function

$$\sum_{j=1}^{|T|} \sum_{i:x_i \in C_j} (y_i - \hat{y}_{C_j})^2 + \alpha|T| \tag{10}$$

where, $|T|$ is the number of terminal nodes of the tree, C_j is the region corresponding to j^{th} terminal node, and $\hat{y}_{C_j}$ the predicted value of the outcome variable in the region C_j, which the mean value of the observations in the training data in that region. The hyper parameter, α, controls a trade-off between the subtree's complexity and its fit to the training data.

The algorithms works as follows :

1. To grow a large tree on the training data using recursive binary splitting, continue splitting until each terminal node has fewer than a specified minimum number of observations.

2. To obtain a sequence of best subtrees as a function of α, apply cost complexity pruningto the large tree.[5]

3. To tune the cost complexity hyper parameter α, use the k-fold cross validation or bootstrap resampling to obtain validation set results as function of α. Then, pick the value of α that minimizes the root mean squared error (rmse).

4. For the chosen value of α obtain the subtree fitted in step 2.

[5] A strategy here is to grow a very large tree and then prune it back to a smaller simple subtrees that can perform better on test data. In broad terms, without the cost complexity parameter the algorithm provides the biggest tree as it only reduces the first term of the loss function. As α increases, the price to be paid for a tree with many terminal nodes increases and hence the loss function above minimizes for a small enough sub tree.

4.2 Random Forest Construction

Although regression trees are easy to interpret and understand, they have low bias but high variance, making them prone to overfitting. To reduce overfitting, I use a tree ensemble algorithm called Random Forest. The idea is to create B bootstrap samples of training data and fit a regression tree for each dataset, resulting in B regression tree predictions. Finally, these B sets of predictions are averaged to reduce the variance.

The process of tree construction is similar to a single decision tree, with some modifications. In each iteration, a tree is constructed using a random subsample. The number of features in these subsamples is determined through hyperparameter tuning. Random sampling in each iteration ensures less correlation among the regression trees constructed. The prediction function in my case becomes

$$\hat{y} = F(C) = \frac{1}{K}\sum_{k=1}^{K} h_k(C) \tag{11}$$

where C stands for circumstances, which is a subset of the full set of circumstances in consideration. C is chosen randomly before constructing each tree. K is the total number of trees. $h_k(C)$ denotes predictions from each tree. Averaging predictions from K trees reduces the overall variance.

4.3 Procedure

The data is in wide format with as many as 396 circumstance variables as features for the biggest circumstance set in the model. The number of individuals in the data is 639 for the SRC sample and 1022 for the full sample that also includes individuals from the SEO sample. Each row reflects an incomplete biography of an individual across the first 18 years of their life. To train the model effectively, I use as much data as possible. As such, I do not create a separate test data set. Instead, I employ the full dataset with a cross-validation process to minimize overfitting and tune the hyper parameters. The goal is to calculate the shares of inequality of income opportunity as shown in equation 9. To that extent, I obtain predictions using the final model on the whole data to obtain absolute and relative IOp estimates for all age cutoffs. I fit the models on training data, tune the hyper parameters on validation data, and then use the best model(with the lowest rmse) on the full data set. The algorithm runs as follows:

- Execute the random forest algorithm and use 5-fold cross validation for hyperparameter tuning. Select the models with hyperparameters that yield the lowest $rmse$. In each fold, the data is divided into $N_{train} = \frac{4}{5}N$ and $N_{validation} = \frac{1}{5}N$. This 5-fold cross-validation process helps minimize overfitting. To increase efficiency, I repeat this cross-validation process twice.[6]
- Store the prediction functions $\widehat{f}_{train}(\widehat{\Omega}^c)$.
- Obtain final predictions using the full data $\widehat{y}_{EA} = \widehat{f}_{train}(\widehat{\Omega}^c_{fulldata})$.

This procedure is repeated for all circumstance sets in consideration based on cut-offs at age 2, 5, 14, and 18 for both the full sample as well as the SRC sample using individual labor income at age 35 and average age adjusted income across 2013-2019 waves as proxy for permanent income.

[6] Cross validation process and results of hyperparameter tunning are provided in the appendix.

5 Results

5.1 Selected Descriptive Statistics

Table 2 presents unweighted summary statistics for the adult incomes and selected variables comprising the circumstance sets. These represent the standard set of circumstances commonly used in the literature to estimate Inequality of Opportunity. I provide descriptive statistics for my analytical sample, which includes family heads and their spouses/partners present during interviews from 2013 to 2019. The statistics are reported for both the Survey Research Center (SRC) sample and the full sample, which also includes the Survey of Economic Opportunity (SEO) sample.

The mean individual labor income at age 35, measured in natural logarithm, is 10.5 in the full sample and 10.7 in the SRC sample. The male-to-female ratio is similar in both samples. However, in the full sample, individuals from the Black population are overrepresented, comprising 44% compared to 11% in the SRC sample. During the child's first year, family heads had, on average, about 12 years of education in the full sample, while it exceeded 13 in the SRC sample. For the spouses of family heads, the average years of education was just over 10 years in the full sample and 12 years in the SRC sample. The standard deviation was higher for spouses, indicating substantial variation in their educational attainment during the child's first year. The occupational composition of family heads during the child's first year is fairly similar for both samples. Throughout the analysis, I use appropriate PSID cross-sectional survey weights for the year when individuals' adult incomes are measured. Note that in table 2, "head" refers to the head of the family in which the child grew up during childhood, while "spouse" refers to the spouse of the family head.[7]

Table 2: Selected Descriptive Statistics

Characteristic	**Full Sample** N = 1,022	**SRC Sample** N = 639
Individual labor income at age 35 (in natural logarithms)	10.5 (0.9)	10.7 (0.9)
Family income during the child's first year (in natural logarithms)	10.8 (0.8)	11.0 (0.8)
Sex		
Male	474 (46%)	311 (49%)
Female	548 (54%)	328 (51%)
Race		
White	559 (55%)	554 (87%)
Black	446 (44%)	72 (11%)
Other	17 (1.7%)	13 (2.0%)
Occupation of the head during the child's first year		
Other	178 (17%)	60 (9.4%)
Professional, Technical, and Kindred Workers	168 (16%)	157 (25%)
Managers and Administrators, except Farm	72 (7.0%)	62 (9.7%)
Sales Workers	22 (2.2%)	20 (3.1%)
Clerical and Kindred Workers	51 (5.0%)	28 (4.4%)
Craftsman and Kindred Workers	219 (21%)	151 (24%)
Operatives, except Transport	128 (13%)	72 (11%)
Transport Equipment Operatives	45 (4.4%)	23 (3.6%)
Laborers, except Farm	41 (4.0%)	24 (3.8%)
Farmers and Farm Managers	13 (1.3%)	12 (1.9%)
Farm Laborers and Farm Foremen	5 (0.5%)	2 (0.3%)
Service Workers, except Private Household	79 (7.7%)	28 (4.4%)
Private Household Workers	1 (<0.1%)	
Years of education of the head during the child's first year	12.4 (2.6)	13.1 (2.5)
Years of education of the spouse during the child's first year	10.4 (5.1)	12.0 (4.1)

[1] Mean (SD); n (%)

5.2 IOp Estimates Across Critical Stages in Childhood

Table 3 presents estimates of absolute and relative income inequality of opportunity. Also reported in the table are the estimates of total inequality (IO) in adult income. The adult income is proxied by individual labor income at age 35 as well as age adjusted individual labor incomes averaged over survey waves from 2013-2019. As a measure of inequality, mean logarithmic deviation is used.[8] The results are reported from analysis performed on both full sample and SRC

[7] A complete list of summary statistics is provided in the appendix.

[8] $MLD(x) = ln(\bar{x}) - \overline{ln(x)}$ MLD of 0 reflects everyone has the same income, i.e. perfect equality.

sample. For labor income measured at age 35, total income inequality is 0.368 for the full sample and 0.337 for the SRC sample. For an age-adjusted distribution of average incomes across four years, total income inequality is 0.327 for the full sample and 0.308 for the SRC sample.

I start with a baseline model which uses an OLS specification using standard set of circumstances displayed in table 3. These circumstances includes individual's race, gender, occupation of the head of the family in which the child grew up as well as completed years of education of the head and their spouse in the family during the child's first year. For the full sample, absolute IOp is 0.083 when adult income is proxied by individual labor income at age 35 and 0.078 when the adult income is calculated as an average income of incomes across waves from 2013-2019. In case of SRC sample, the absolute IOp is 0.065 for the baseline specification when individual's labor income at age 35 is considered. Averaging incomes across aforementioned waves does not affect the absolute IOp measure with estimate being 0.063. These small estimates represent overall less inequality in the distribution of incomes conditional on standard circumstances used in the literature. In addition to the absolute IOp estimates using the baseline circumstances, estimates for the circumstance sets created using age cutoffs based on critical stages in childhood are also reported. The estimates obtained using random forest model are greater than those obtained using standard set of circumstances and OLS specification. It is also apparent that these estimates based on age-based circumstance sets increase with age given that circumstance sets expand with age. In case of adult incomes proxied by labor income at age 35 the absolute IOp estimates range from 0.104 to 0.158 for age cut off for circumstances sets at age 2 and 18 respectively for the full sample. For the SRC sample the estimates are smaller. Considering average incomes across four waves as a measure of adult incomes, the absolute IOp estimates are lower compared to the case above, as they range from 0.078 to 0.146 for age cutoffs at 2 years and 18 years for for the full sample. For the SRC sample, these estimates are 0.063 to 0.123 for the same age cutoffs.

Table 3: IOp Estimates for Different Circumstance Sets

	Full Sample (N = 1022)		SRC sample (N = 639)	
	Absolute IOp	Relative IOp	Absolute IOp	Relative IOp
Outcome : Labor Income at age 35 (Full sample IO = 0.368; SRC sample IO = 0.337)				
Baseline	0.083	0.225	0.065	0.192
Age cutoff at 2 years	0.104	0.282	0.093	0.276
Age cutoff at 5 years	0.126	0.342	0.100	0.297
Age cutoff at 14 years	0.146	0.396	0.115	0.340
Age cutoff at 18 years	0.158	0.429	0.128	0.380
Outcome : Age-adjusted Labor Income (Full sample IO = 0.327; SRC sample IO = 0.308)				
Baseline	0.078	0.239	0.063	0.206
Age cutoff at 2 years	0.101	0.310	0.085	0.277
Age cutoff at 5 years	0.109	0.335	0.097	0.316
Age cutoff at 14 years	0.129	0.394	0.111	0.360
Age cutoff at 18 years	0.146	0.448	0.123	0.400

Also displayed in table 3 are the estimates of relative IOp, the share of inequality of opportunity in total income inequality for all age-based circumstance sets. These shares are calculated using labor income at age 35 as an outcome variable to proxy individual income in adulthood. While considering individual labor incomes at age 35 as a proxy for adult income, about 34% of income inequality in adult income can be ascribed to unequal circumstances faced by an individual up to age five while performing the analysis on the full sample. For the SRC sample, this relative IOp is 29.7%. Consistent with expectations, the role of unequal circumstances, and consequently the share of inequality of opportunity in income inequality, increases with age as the circumstance sets expand. This share of inequality increases up to 43% considering all circumstances encountered before or at the age of majority at 18 for the full sample and 38% for the SRC sample. These estimates of relative IOp are greater than those obtained from OLS regressions using a standard set of circumstances. For example, the relative IOp estimates obtained from OLS specification using standard set of circumstances are 22.5% for the full sample and 19.2% for the SRC sample when adult income is proxied by labor income at age 35.

I also conduct the analysis using an age-adjusted average income of the individuals from 2013-2019 PSID waves as a proxy for adult income. In this analysis, since incomes are averaged at different ages for individuals, I also include their birth year in the equation. An individual cannot choose their birth year, so it enters the equation as a circumstance and helps account for average incomes measured at different ages in adulthood. Table 3 shows about 34% of income inequality in adult income can be attributed to unequal circumstances faced by individuals up to age five when analyzing the full sample. For the SRC sample, the relative IOp is approximately 32%. These estimates are similar to those found using individual labor income as a proxy for adult income. The share of inequality increases up to 45% considering all circumstances encountered up to age of majority at 18 for the full sample and 40% for the SRC sample. Once again, these estimates of relative IOp are greater than those obtained from OLS regressions using a standard set of circumstances. For example, the relative IOp estimates are about 24% for the full sample and 20.6% for the SRC sample when adult income is proxied by labor incomes averaged over four survey waves from 2013-2019. By definition, the share of IOp in income inequality that stems from unequal circumstances is deemed to be unfair inequality.

Figure 2 illustrates the profile of relative IOp—the proportion of inequality of opportunity in total income inequality—across all age cut-offs used in this study. I also estimate relative IOp using circumstances based on age cutoffs at 10, 12, and 16 years. These profiles are created for both the full and SRC samples, using two income measures as outcome variables to calculate relative IOp. The shares of IOp consistently increase with age across all cases. As expected, the highest IOp share stems from considering all measurable data on a child before the age of consent at 18. It is evident that relative IOp estimates increase at a decreasing rate with age.

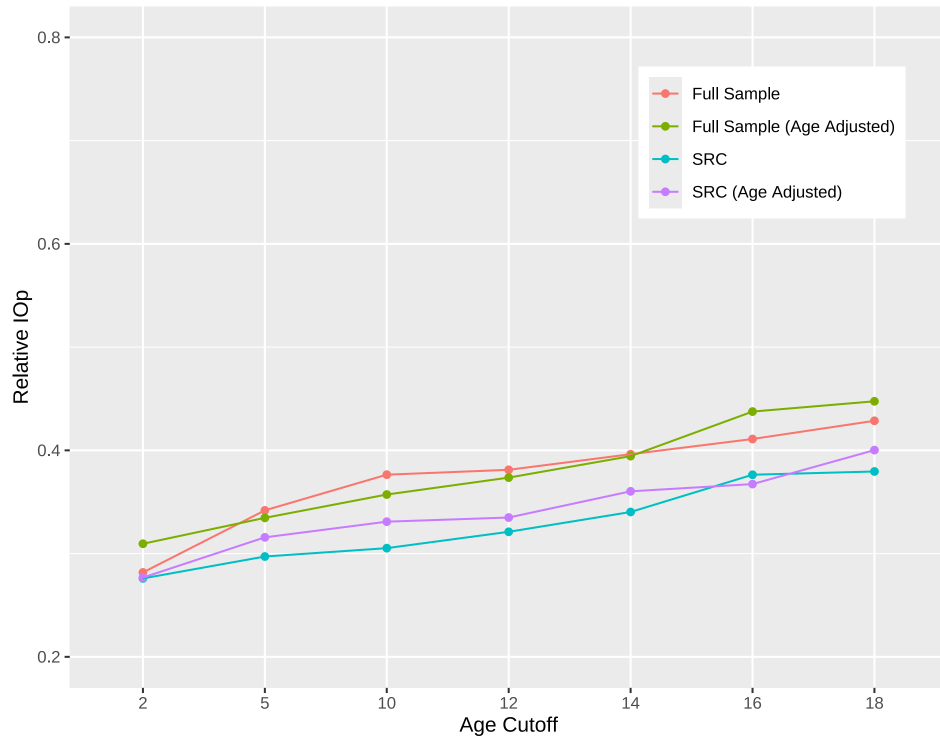

Figure 2: Relative IOp Profiles Across All Age Cutoffs

The results of this study align with findings from other research estimating inequality of opportunity (IOp) in the United States. For example, Pistolesi (2009) found IOp to account for 20 to 43 percent of earnings inequality in the US. I also compare my results to those of Hufe et al. (2017), whose study most closely resembles the analysis conducted here. They estimated IOp using circumstances at birth, age 12, and age 16 in the US. While they measured adult income in various years, they also reported results using average adult income for 2008–2012 from NLSY79 data. Their findings show a higher proportion of income inequality due to circumstances for age cutoffs at 12 and 16 years, at 44.6% and 58.8% respectively. I compare these estimates with obtained in this study. Using labor income at age 35, I obtained relative IOp 38% and 41% for age cut offs at 12 and 16 respectively for the full sample as well as 32% and 38% for the same cutoff using the SRC sample. When average age adjusted incomes are used as a proxy for income in adulthood, for the full sample relative IOp estimates are 37% and 44% for age cutoffs at 12 and 16 respectively. For the SRC sample, these estimates are 34% and 38%.

It is important to note that I do not account for ability variables such as IQ or other test scores explicitly. One implication of dynamic complementarity is that early investments in cultivating non-cognitive skills can promote cognitive skills. A lack of early investments in disadvantaged children may lead to a lower stock of skills in subsequent years. Since children do not have control over their circumstances, these missed opportunities early in life may lead to lower stocks of skills in the future. Consequently, the inequalities generated due to these factors in outcomes should be accounted for in the measurement of IOp. Hufe et al. (2017) argue that ability is a circumstance and categorize it as such, which may have led to estimates of relative IOp as high as 58.8 % for age cutoff at 16 years. However, I report results from circumstance set created at an age cutoff of 10, as it is well documented that IQ is rank stable after age 10 (Mackintosh 2011). For the full sample using average age adjusted incomes, about 36%of IOp could be attributable to the circumstances faced by the child before or at age 10. This estimate is 33% for the analysis performed on the SRC sample. When individual labor income at age 35 is used as the outcome, about 38% of income inequality could be attributed to unequal circumstances up to age 10. This estimate is 30.5% for the SRC sample. In addition to that, I permit the set of circumstances to expand with age, consistent with the formulation of the skill formation technology. Therefore, unlike Hufe et al. (2017), circumstances may reappear as the set enlarges. A set of circumstances that includes data on a child up to age 14 will be a superset of a set that contains data on a child up to age 5. The biggest set will be the set of circumstances including data on the child up to the age of majority at 18 years.

5.3 Lower and Upper Bound Estimates

Displaying results from a random forest algorithm using a single tree is challenging due to the construction of multiple trees during the model fitting process. Instead, I can explore feature importance scores to understand the "importance" of different variables in constructing the trees and predicting adult income (Breiman 2001; Fisher, Rudin, and Dominici 2019). The idea is to calculate the increase in model's prediction error after permuting a feature. A feature is "important" if shuffling its values increases the model error as it implies that the model relied on the feature for prediction. If prediction error of the model does not change by much while shuffling the feature values, the feature is considered unimportant in predicting the outcome, adult income.

Let $x_1, x_2, ..., x_j$ be the features of interest and let $rmse_{base}$ be the baseline performance metric for the trained model. The permutation-based importance scores can be computed as follows:

1. For $i = 1, 2, ..., j$:
    1. Permute the values of feature x_i in the training data.
    2. Recompute the performance metric on the permuted data $rmse_{perm}$.
    3. Record the difference from baseline using $vi(x_i) = \frac{rmse_{perm}}{rmse_{base}}$.
2. Return the vi scores $vi(x_1), vi(x_2), ..., vi(x_j)$.

Table 4 presents the top ten circumstances that were most important in predicting adult incomes, as measured by age-adjusted average individual incomes across the 2013–2019 PSID waves. The table displays these key circumstances for all age cutoffs, showcasing results for both the full sample and the SRC sample. This permutation approach introduces randomness into the procedure. I repeat the procedure for 50 simulations to obtain mean importance scores and their corresponding standard deviations.

The table 4 shows that an individual's sex and the family's use of food stamps during the child's first year of life are the most crucial circumstances in the data, as evidenced by the highest variable importance scores derived from the model. This finding holds true for analyses of both the full and SRC samples. Notably, these circumstances maintain high importance scores across all age cutoffs used to create circumstance sets. It is important to note that these scores do not indicate causality. However, they offer insight into how various circumstances, measured at different life stages, can influence adult income predictions and contribute to inequality of opportunity. For instance, recent evidence suggests that the timing of food stamp receipt can have long-term implications (Bond et al. 2022). Depending on the sample—full or SRC—and age cutoff for circumstance sets, other important predictors of adult income include the family head's and spouse's occupation, their education levels, the marginal tax rate on family income, and the value of the house in which the child grew up.

Table 4: Important Circumstances

Full Sample	SRC Sample
Age cutoff at 2	
Usage of food stamps at age 1	Usage of food stamps at age 1
Sex of the individual	Sex of the individual
Marginal tax rate on family income at age 1	Occupation of the head at age 2
Race of the individual	Birthcohort
Race of the head at age 2	Usage of food stamps at age 2
Birthcohort	House value at age 1
House value at age 1	Occupation of the spouse at age 2
Occupation of the head at age 2	Region where the head grew up at age 2
Usage of food stamps at age 2	Area where the head grew up at age 1
Race of the head at age 1	Area where the head grew up at age 2
Age cutoff at 5	
Usage of food stamps at age 1	Usage of food stamps at age 1
Sex of the individual	Sex of the individual
Usage of food stamps at age 3	Occupation of the head at age 5
Occupation of the head at age 5	House value at age 1
House value at age 4	Marginal tax rate on family income at age 1
Usage of food stamps at age 4	House value at age 4
Marginal tax rate on family income at age 1	Marginal tax rate on family income at age 3
Race of the head at age 4	Usage of food stamps at age 3
House value at age 1	Usage of food stamps at age 4
Occupation of the head at age 4	Occupation of the head at age 2
Age cutoff at 14	
Usage of food stamps at age 1	Usage of food stamps at age 1
Sex of the individual	Sex of the individual
House value at age 4	Marginal tax rate on family income at age 6
Childcare cost at age 14	Education of the spouse at age 13
Education of the spouse at age 14	Occupation of the head at age 6
Birthcohort	Number of children to father
Total family income at age 14	Number of children to mother
Occupation of the head at age 11	House value at age 1
Usage of food stamps at age 4	Occupation of the head at age 8
Union membership of the head at age 11	Occupation of the head at age 12
Age cutoff at 18	
Usage of food stamps at age 1	Sex of the individual
Sex of the individual	Marginal tax rate on family income at age 6
Education of the spouse at age 15	Usage of food stamps at age 1
Occupation of the head at age 16	House value at age 10
Union membership of the head at age 11	Education of the spouse at age 13
Childcare cost at age 14	Occupation of the head at age 8
House value at age 16	Number of children to mother
Occupation of the head at age 15	House value at age 1
Race of the head at age 16	Union membership of the head at age 11
Occupation of the head at age 11	Occupation of the head at age 15

These circumstances are ranked by importance based on permutation-based variable importance scores, with the most important at the top. *Age* denotes the individual's age during childhood. *Head* refers to the family head under whom the child grew up before reaching the age of majority (18), while *spouse* indicates that family head's partner.As explained in the data section, while most family heads and their spouses are the individuals' parents, it is possible for non-parents to occupy these roles.

The advantage of using supervised machine learning to predict adult income based on circumstances and subsequently estimate IOp is that we can analyze variable importance scores and utilize the most "important" circumstances—as determined by the procedure described earlier. This approach is particularly helpful when dealing with high-dimensional data, such as in this study, where there is a need to reduce the dimensionality of the feature space. This method partially circumvents the need to categorize circumstances based on researchers' subjective judgments when measuring IOp. Table 4 presents the top ten circumstances from all circumstance sets used, organized by age cutoffs. These circumstances have the highest variable importance scores, calculated using age-adjusted average incomes as the outcome variable for both the full and SRC samples. I then use these algorithm-selected circumstances to repeat the entire analysis and obtain new relative IOp estimates.

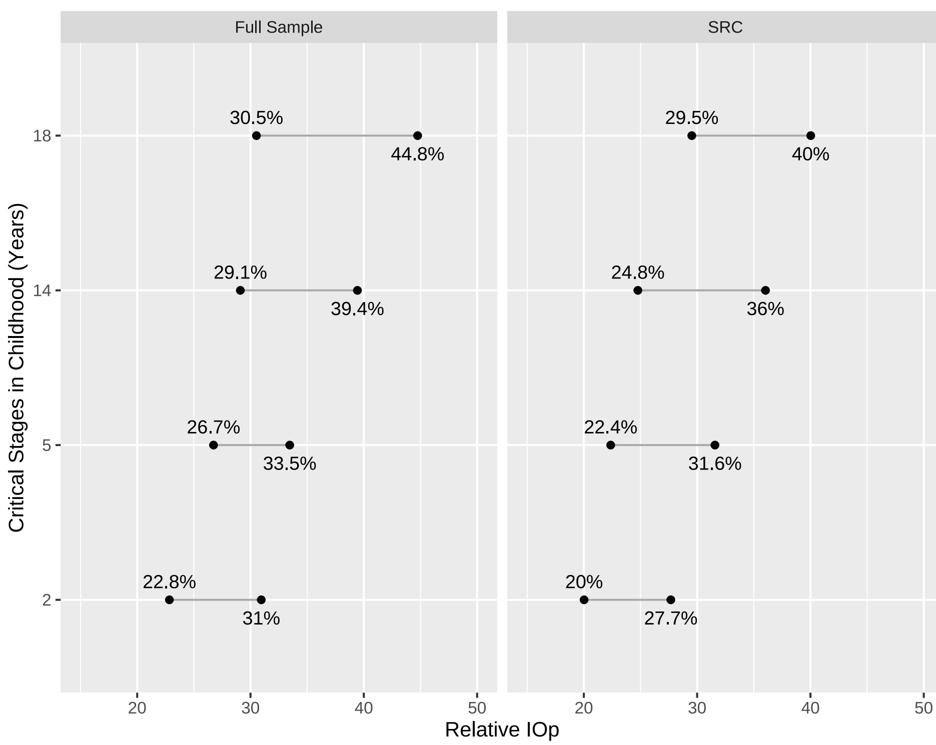

Figure 3: Lower and Upper Bounds of Relative IOp Estimates (Age-adjusted Average Incomes)

Although the literature typically provides lower bound estimates of IOp due to incomplete circumstance sets, Figure 3 suggests we may have reached an upper bound given the number of circumstances used in this analysis. The figure displays the upper and lower bounds of relative IOp, using age-adjusted average incomes across PSID waves 2013–2019 as the outcome variable. The lower bound estimates are derived from the ten most "important" circumstances, determined by permutation-based variable importance scores calculated as described earlier. As we can see most of the relative IOp could be attributed to just top ten circumstances selected by the model for all age cutoffs.

Considering the circumstance set created using the age cutoff at the age of majority at 18, about 31% of total adult income inequality could be attributed to IOp measured using circumstances deemed "important"—as reported in Table 4—by the model for the full sample. This share increases to almost 45% when all available circumstances before the age of majority are considered, suggesting an upper bound may have been obtained. When performing the same analysis on the SRC sample, the lower and upper bound estimates of IOp were approximately 30% and 40%, respectively. For the circumstance set created using age cutoff at five years, figure 3 shows that for full sample, the lower bound of IOp—obtained using only "important" circumstances—is estimated to be about 27%, while the upper bound is about 34%. For the SRC sample, estimates are 22.4% and 31.6% respectively.

6 Conclusion

In this paper, I measure income inequality of opportunity—unfair inequality—using age-based circumstance sets while accounting for the dynamic complementarity across the first 18 years of individuals' lives. I categorize any measurable data on an individual up to the age of majority at 18 years as circumstances. If a child is not considered an adult before the age of majority, they should not be held responsible for inequalities in adult incomes due to unequal childhood circumstances. I use random forest—a supervised machine learning algorithm—to create a counterfactual distribution of adult incomes that depends only on circumstances faced by individuals before the age of majority. This approach allows for empirical identification of relevant circumstance variables, reducing reliance on arbitrary value judgement of the researcher. Using mean logarithmic deviation to measure inequality in the counterfactual distribution of adult income, I obtain inequality that is dependent solely on circumstances and thus reflects only unfair inequality.

Using Panel Study of Income Dynamics (PSID) data on both the Survey Research Center (SRC) sample and the full sample—which also includes the Survey of Economic Opportunity sample—I found that relative inequality of opportunity (IOp), the share of income inequality in adult income attributable to unequal circumstances faced by an individual up to age five, is about 34% for the full sample and 29% for the SRC sample. Adult income was proxied as individual labor income at age 35. When using average age-adjusted incomes across four waves from 2013–2019, these estimates remained unchanged. These figures are higher than those obtained using a standard set of circumstances in OLS regression. I also calculated permutation-based variable importance scores for each circumstance in all circumstance sets based on age cutoffs, using age-adjusted average incomes across four waves. Using the top ten circumstances based on importance scores, I obtained relative IOp estimates for all circumstance sets under consideration. I argue that these estimates are lower bound estimates of relative IOp, unlike the upper bound estimates I obtained using the full—albeit incomplete—biography of a child across the first 18 years of their life. For the circumstance set created using the top 10 circumstances selected by the model (using variable importance scores for a cutoff at the age of majority, 18 years), the relative IOp is estimated to be about 30–33%. Adding more circumstances to the model only increased the relative IOp estimates to about 40–43%, depending on whether the full or the SRC sample was used.

These findings have limitations. First, the boundary between circumstances and effort factors at age 18 could be seen as arbitrary, given the persistent effects of circumstances even after this age. I acknowledge this and report relative IOp estimates for age cutoff at 16 years as well in the paper. The estimates remained consistent. Although the boundary between circumstance and effort is debatable, it's difficult to argue that a 5-year-old has any control over their circumstances. Additionally, the purpose of using the age of majority is not to provide a strong value judgment but to open up the possibility of categorizing circumstance and effort variables based on societal judgment rather than researchers' value judgments. The age-based boundary between circumstances and effort variables could differ based on culture and societal norms, allowing us to move away from standard circumstance sets used in the literature. Finally, I recognize that the estimation of inequality of opportunity using machine learning algorithms is only as good as the next best algorithm.

Rigorously measuring unfair income inequality is crucial if equalizing opportunities is a public policy goal. This study identifies "unfair" inequality by incorporating early childhood circumstances into the measurement of inequality of opportunity. This approach to measuring IOp could inform the development of public policies aimed at equalizing opportunities in early childhood, consistent with ex-ante investments in human capital. Additionally, Roemer (1993) proposes ex-post compensation for individuals who experience outcome inequalities due to their unequal circumstances. Measuring IOp by accounting for unequal childhood circumstances provides better estimates, as it considers the persistent effects of limited opportunities during childhood. Public policies aimed at eliminating the effects of unequal opportunities through compensation are better informed when they account for childhood circumstances. While this research focuses on childhood circumstances in measuring inequality of opportunity (IOp), it would be valuable to examine cross-country differences in the share of unequal opportunities using age-based circumstance sets. This approach could complement the current practice of using fixed sets of circumstances. Furthermore, this analysis could be extended to other outcomes such as health and education, as well as exploring geographical heterogeneity in IOp estimates.

Appendix

6.1 Family Relationship Matrix

The Family Relationship Matrix File (FRM), spanning from 1968 to 2019, is designed to consolidate all existing relationship data gathered during these years as part of the Panel Study of Income Dynamics (PSID). This file outlines the relationship between each individual in a PSID Family Unit (FU) and all other members within the same FU for each wave from 1968 to 2019. Each record set for a specific individual shows their relationship to all other FU members during that wave. Only individuals who resided in the FU at the time of the interview are included in the file for each wave.I use this FRM file to identify the family heads of the individuals of interest during their first 18 years of life. As the individuals were born between 1978-1983, the interview waves are limited to the period from 1978-2001. This matrix provides information on whether the family head was the individual's parent before they reached the age of 18. It's also possible for the family head to be a non-parent. Moreover, the family head can change over the course of an individual's first 18 years.

6.2 Family Identification Mapping System

FIMS relies on information stored in the Parent Identification File (PID), a cumulative file compiled from several PSID data sources. The PID summarizes information about parent-child relationships collected from various sources since the 1983 wave of the Panel Study of Income Dynamics (PSID). This file includes identifier variables that link children to their birth and adoptive parents, and it also indicates the source of the information. I use FIMS to identify whether the parent-head of a child is a mother or a father.

6.3 Handling Missing Data

The analysis uses wide data, where each observation is a joint distribution of all measurable data on individuals. All of these are considered circumstances over which individuals have no control. The sample suffers from missing data issues.

When the training sample is moderate in size, one effective method to impute missing data is the K nearest neighbors algorithm (Eskelson et al. 2009; Tutz and Ramzan 2015). I use this algorithm to impute missing data in my sample.

The procedure finds a sample with one or more missing values and then identifies the K most similar samples in the complete training data with no missing values. Similarity of samples is defined by a distance metric. After computing this distance metric, the nearest K samples to the sample with the missing value are identified, and the mean value is calculated. This mean value is then used to replace the missing value in the sample.

Usually, Euclidean distance is used as a sample similarity metric when all the features are numeric. In this study, I have 396 features (for a full dataset) with missing values for both numeric and categorical features. Instead of using Euclidean distance, I use a good alternative called Gower's distance (Gower (1971)). This distance metric uses separate measures for both numeric and categorical features. For a categorical feature, the distance between two samples is 0 if the samples have the same value and 1 otherwise. For a numeric feature, the sample distance between two observations is defined as

$$d(x_i, x_j) = 1 - \frac{|x_i - x_j|}{R_x} \tag{12}$$

where R_x is the range of the feature for which the missing values are being imputed using KNN. This measure is computed for each feature, and the average distance is used as an overall distance. Once the K neighbors are found, their mean values are used to impute the missing data. For categorical features, the mode is used, while an average or a median can be used for numeric data. I use the average in my analyses. I explain the feature engineering steps in next section.

6.4 Data Preprocessing

While converting long data to wide data based on an individual's age in childhood, some columns have all values missing. I start by removing these columns. Next, I exclude features where more than 50% of the values are missing from my analysis.

I then run the KNN algorithm to impute the missing values in the rest of the data for all quantitative and qualitative features using Gower's distance to measure the distance between neighbors. A rule of thumb is to use $k = \sqrt{n}$, where k is the number of neighbors and n is the number of observations in the sample. I settled on k = 25 while using the KNN algorithm for missing data imputation.

For the remaining features, I remove numerical features with near-zero variance. Finally, to address multicollinearity, I remove numeric features with a correlation greater than 0.8 with other features.

6.5 Tuned Hyperparameters

Table 5: Tuned Hyperparameters

		Full Sample (N = 1022)		SRC sample (N = 639)	
Age Cutoffs	Trees	min_n	mtry	min_n	mtry
Outcome : Labor Income at age 35					
2	500	30	9	30	14
5	500	25	24	25	27
10	500	25	43	25	57
12	500	25	57	30	111
14	500	25	74	30	152
16	500	20	69	25	218
18	500	20	84	20	110
Outcome : Age-adjusted Labor Income					
2	500	25	8	30	14
5	500	30	21	35	38
10	500	30	51	30	48
12	500	30	73	35	96
14	500	30	72	25	93
16	500	25	80	25	97
18	500	25	89	20	130

Table 5 in the appendix lists the hyperparameter values obtained through a 5-fold cross-validation process for each circumstance set, based on their respective age cut-offs. I utilize three essential hyperparameters for building a random forest model.

- *mtry*: An integer representing the number of predictors that will be randomly selected at each split during the tree model creation.
- *n_trees*: An integer representing the number of trees in the ensemble.
- *min_n*: An integer representing the minimum number of data points a node must contain before it can be split further.

To reduce the complexity and run time of the code, I only tune *mtry* parameter using 5-fold cross validation. The number of trees are chosen following the standard practice in the literature of machine learning. I keep the number of trees arbitrarily high and do not tune that parameter Oshiro, Perez, and Baranauskas (2012). Following the same practice, I do not tune *min_n* hyperparameter. I choose high enough number for this hyperparameter instead of tuning it to reduce the run time as well as the model complexity.

When *mtry* is set to 1, the split variable is chosen at random, which can lead to biased results. When *mtry* is set to the total number of predictors, the split is optimized along all possible directions. Each value of *mtry* in the table 5 is obtained through a 5-fold cross-validation process, repeated twice. The model with the lowest root mean square error (rmse) - as indicated by the hyperparameters in the table 5 - is selected. This model is then fitted on the entire dataset to generate a counterfactual distribution of predictions, based on the factors in the respective circumstance sets.

This process is repeated for both full and SRC samples, as well as for labor income at age 35 and averaged age-adjusted income as outcome variables. All values for tuned hyper parameters are reported in the table 5 above.

6.6 IOp Estimates for All Age Cutoffs

Table 6: IOp Estimates for All Age Cutoffs

	Full Sample (N = 1022)			SRC sample (N = 639)		
	Income Inequality	Absolute IOp	Relative IOp	Income Inequality	Absolute IOp	Relative IOp
Outcome : Labor Income at age 35						
Cutoff at age 2	0.368	0.104	0.282	0.337	0.093	0.276
Cutoff at age 5	0.368	0.126	0.342	0.337	0.100	0.297
Cutoff at age 10	0.368	0.138	0.376	0.337	0.103	0.305
Cutoff at age 12	0.368	0.140	0.381	0.337	0.108	0.321
Cutoff at age 14	0.368	0.146	0.396	0.337	0.115	0.340
Cutoff at age 16	0.368	0.151	0.411	0.337	0.127	0.376
Cutoff at age 18	0.368	0.158	0.429	0.337	0.128	0.380
Outcome : Age-adjusted Labor Income						
Cutoff at age 2	0.327	0.101	0.310	0.308	0.085	0.277
Cutoff at age 5	0.327	0.109	0.335	0.308	0.097	0.316
Cutoff at age 10	0.327	0.117	0.357	0.308	0.102	0.331
Cutoff at age 12	0.327	0.122	0.374	0.308	0.103	0.335
Cutoff at age 14	0.327	0.129	0.394	0.308	0.111	0.360
Cutoff at age 16	0.327	0.143	0.438	0.308	0.113	0.367
Cutoff at age 18	0.327	0.146	0.448	0.308	0.123	0.400

6.7 IOp Estimates Using Gini

In my main study, I use mean logarithmic deviation (MLD). Any standard inequality measure that satisfies anonymity, the principle of transfers, population replication, and scale invariance could be used. Here, I present the absolute and relative IOp estimates along with their contributions to total inequality, as measured by the Gini coefficient.

Table 7: IOp Estimates for Different Circumstance Sets (Using Gini)

	Full Sample (N = 1022)			SRC sample (N = 639)		
	Income Inequality	Absolute IOp	Relative IOp	Income Inequality	Absolute IOp	Relative IOp
Outcome : Labor Income at age 35						
Cutoff at age 2	0.425	0.258	0.607	0.408	0.241	0.591
Cutoff at age 5	0.425	0.281	0.661	0.408	0.249	0.609
Cutoff at age 10	0.425	0.294	0.692	0.408	0.253	0.619
Cutoff at age 12	0.425	0.296	0.695	0.408	0.258	0.632
Cutoff at age 14	0.425	0.301	0.709	0.408	0.265	0.649
Cutoff at age 16	0.425	0.306	0.720	0.408	0.277	0.677
Cutoff at age 18	0.425	0.312	0.735	0.408	0.278	0.680
Outcome : Age-adjusted Labor Income						
Cutoff at age 2	0.410	0.254	0.620	0.394	0.230	0.584
Cutoff at age 5	0.410	0.262	0.640	0.394	0.242	0.615
Cutoff at age 10	0.410	0.271	0.661	0.394	0.250	0.633
Cutoff at age 12	0.410	0.275	0.671	0.394	0.250	0.633
Cutoff at age 14	0.410	0.282	0.689	0.394	0.258	0.654
Cutoff at age 16	0.410	0.296	0.721	0.394	0.259	0.657
Cutoff at age 18	0.410	0.299	0.730	0.394	0.270	0.685

Table 7 shows the shares IOp in total income inequality using Gini coefficient as the inequality measure. Despite the shares being higher, the upward trend until the age of majority at 18 aligns with what is observed when using MLD as the inequality measure. Most of the income inequality, approximately 60%, attributed to the inequality of opportunity, stems from circumstances at or before the age of 2.

6.8 Intergenerational Income Elasticity

Policy discussions have shifted from inequality of outcome to inequality of opportunity, informed by intergenerational mobility (Corak 2013; Chetty et al. 2014). The literature on intergenerational income mobility offers variety of measures, the most popular being Intergenerational Income Elasticity (IGE). IGE is measured as a coefficient in a Galtonian regression of a child's income on parental income.[9] Indeed, the IGE measure is a special case of IOp estimated using equation 5, where parental income is the sole circumstance variable (Brunori, Ferreira, and Salas-Rojo n.d. for a theoretical framework on inherited inequalities.) Usually, in the measurement of IGE, the child's income and the parent's income are averaged over multiple years to address the attenuation bias (Solon 1992; Mazumder 2005).

$$ln(y_{child}) = \alpha + \beta_{IGE}\, ln(y_{parent}) + u \tag{13}$$

Evidence suggest that the timing of parental income measured may be as or more important than a single measure of parental income (P. Carneiro et al. (2021)). I show the IGE estimates to compare them with the IOp estimates obtained in the study, considering the parental income averaged over years before and at critical stages of childhood. For parental incomes, I use family incomes from the first 18 years of the child's life. I proxy a child's permanent income using their labor income in adulthood, averaged over four years from 2013-2019. For individuals with missing income data in any wave, I calculate their average income using only the available years. For example, if an individual has income data for 2012, 2014, and 2018, but missing in 2016, I compute their average income using three years (2012, 2014, 2018). For the cohorts under

[9] for a comprehensive discussion on full set of measures, see Deutscher and Mazumder (2023).

consideration (born in 1978-1983), average incomes are measured at different ages. For example, in 2015, an individual born in 1978 is on average 37 years old, while someone born in 1982 is on average 33 years old. To account for these age differences when measuring income, I include both age and age-squared terms in the equation 13. This approach follows standard practice in the intergenerational income mobility literature.

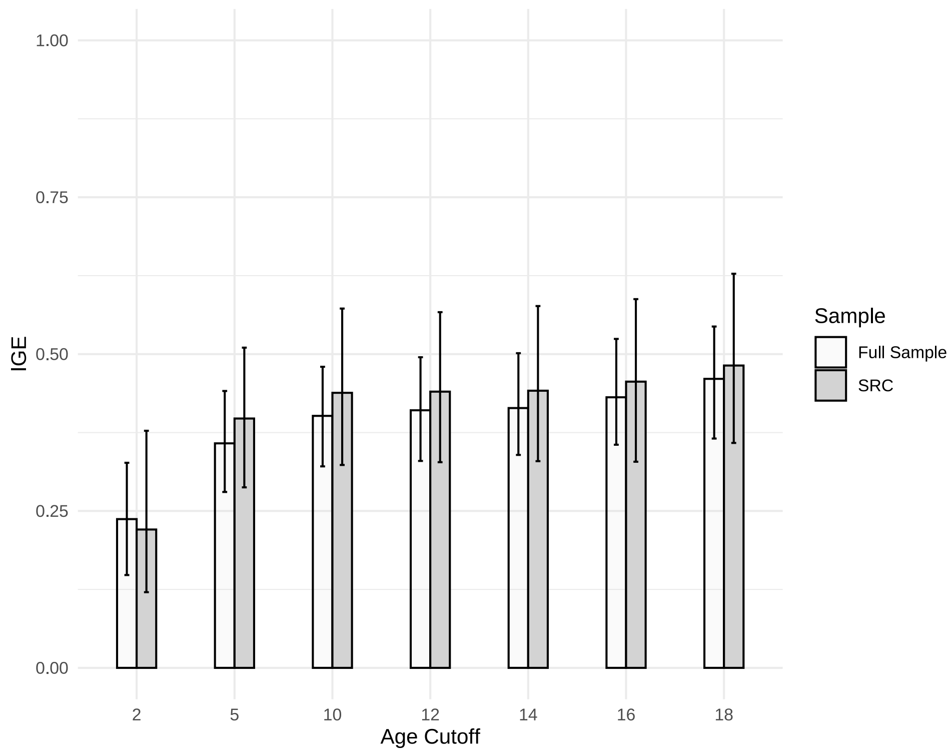

Figure 4: IGE Estimates Based on Age Cutoffs

Figure 4 displays the coefficients obtained from the IGE equation 13, accounting for age differences at which the incomes of children are measured, with family incomes averaged over the years before and at the age cutoffs considered in the study. Once again, I report IGE coefficients using both the full sample as well as the SRC sample. For instance, IGE is 0.237 when computed using the equation with family income averaged over the first two years of the child's life for the full sample. For the SRC sample, this estimate is 0.22. Similarly, using family income averaged over the first 18 years of the child's life, I compute IGE at around 0.461 for the full sample and 0.482 for the SRC sample. We can see the IGE measures estimated in the full and SRC samples using age cutoff at 5 is 0.358 and 0.397 respectively. The estimates do increase with age but at a decreasing rate. These measures are not causal but are comparable to IOp measures obtained previously and follow the same pattern directionally. To obtain the intervals, I use 95% bootstrap confidence intervals with 500 iterations.

6.9 Summary Statistics

Table 8 presents summary statistics for the full set of numerical variables used as circumstances. All monetary circumstances are measured in 2018 dollars. Education is measured in total completed years of schooling. "Head" refers to the head of the family in which the child grew up during childhood. "Spouse" refers to the spouse of the family head. "Age" refers to the child's age during the first 18 years of their life.

Table 8: Summary Statistics (Numerical Variables)

	Full Sample (N = 1022)		SRC sample (N = 639)	
Circumstance	Mean	StDev	Mean	StDev
Individual labor income at age 35	10.5	0.9	10.7	0.9
Individual age-adjusted average income	10.6	0.9	10.5	0.9
Number of children to mother	3.3	5.4	2.8	1.1
Order of birth to mother	1.9	1.1	1.8	1.0
Number of children to father	3.2	3.7	2.9	1.2
Order of birth to father	2.0	1.2	1.8	1.0
House value at age 1	75766.3	113445.2	99124.9	121315.1
Marginal Tax Rate on family Income at age 1	18.7	12.1	21.9	11.2
Number of rooms in family home at age 1	5.3	1.6	5.6	1.5
Education of the spouse at age 1	10.7	5.2	12.2	4.1
Education of the head at age 1	12.5	2.7	13.2	2.6
Total family income at age 1	63207.0	45902.2	72793.9	48512.0
House value at age 2	78166.4	114427.3	106266.2	128250.7
Marginal Tax Rate on family income at age 2	18.3	12.0	21.5	11.3
Number of rooms in family home at age 2	5.6	1.7	6.0	1.6
Education of the spouse at age 2	10.6	5.3	11.8	4.5
Education of the head at age 2	12.5	2.7	13.2	2.5
Total family income at age 2	64932.1	48502.9	75588.6	51389.0

Table 8: Summary Statistics (Numerical Variables) *(continued)*

Circumstance	Mean	StDev	Mean	StDev
House value at age 3	82589.4	119312.5	113523.0	134074.2
Marginal Tax Rate on family income at age 3	18.4	12.2	21.7	11.2
Number of rooms in family home at age 3	5.6	1.7	6.0	1.6
Education of the spouse at age 3	10.4	5.3	11.8	4.5
Education of the head at age 3	12.6	2.5	13.2	2.3
Total family income at age 3	65918.5	53180.0	78663.9	57639.7
House value at age 4	93229.9	147422.8	128216.9	169609.8
Marginal Tax Rate on family income at age 4	18.2	11.9	21.5	11.2
Number of rooms in family home at age 4	5.8	1.7	6.2	1.7
Total family income at age 4	68582.1	60236.8	81922.3	66365.3
House value at age 5	96096.9	147218.2	133370.8	167438.9
Marginal Tax Rate on family income at age 5	17.5	11.3	20.8	10.5
Number of rooms in family home at age 5	5.9	1.8	6.3	1.7
Total family income at age 5	71662.9	63795.7	86023.3	70626.6
House value at age 6	104290.6	171957.9	146195.4	198184.9
Marginal Tax Rate on family income at age 6	16.8	11.0	20.0	10.2
Number of rooms in family home at age 6	5.9	1.9	6.4	1.9
Total family income at age 6	74857.3	68588.2	90205.9	76914.4
House value at age 7	111343.8	174441.2	153680.9	199708.9
Marginal Tax Rate on family income at age 7	16.7	10.8	19.6	10.0
Number of rooms in family home at age 7	6.1	1.9	6.5	1.9
Total family income at age 7	78568.7	78416.6	94461.9	89213.4
Cost of childcare at age 7	1388.3	4701.0	1314.2	2630.3
House value at age 8	114360.4	177323.6	156757.9	202887.6
Marginal Tax Rate on family income at age 8	16.2	10.9	18.9	10.3
Number of rooms in family home at age 8	6.2	1.9	6.6	2.0
Total family income at age 8	81748.1	83635.2	99011.7	95425.8
Cost of childcare at age 8	1752.7	8798.8	2052.6	10787.2
House value at age 9	117130.6	179763.7	158777.8	205128.2
Marginal Tax Rate on family income at age 9	16.2	10.3	18.8	9.7
Number of rooms in family home at age 9	6.3	1.9	6.8	1.9
Total family income at age 9	84181.8	91749.6	101977.0	104948.7
Cost of childcare at age 9	1341.2	4209.8	1249.1	2164.5
House value at age 10	116977.6	185902.2	159834.8	209476.9
Marginal Tax Rate on family income at age 10	15.9	10.0	18.5	9.4
Number of rooms in family home at age 10	6.3	1.9	6.8	1.9
Total family income at age 10	86204.7	109920.9	105746.7	129411.3
Cost of childcare at age 10	1178.4	4838.7	1320.8	5835.3
House value at age 11	114875.4	169942.3	158573.9	194575.0
Number of rooms in family home at age 11	6.4	1.9	6.9	1.9
Education of the spouse at age 11	9.4	6.4	11.2	5.7
Education of the head at age 11	13.1	2.4	13.7	2.4
Total family income at age 11	87953.5	97973.9	108426.9	113477.3
Cost of childcare at age 11	832.6	1976.2	880.9	2144.1
House value at age 12	115524.6	166700.3	159964.9	190108.8
Number of rooms in family home at age 12	6.5	2.0	7.0	1.9
Education of the spouse at age 12	9.3	6.4	10.9	5.8
Education of the head at age 12	13.1	2.4	13.6	2.4
Total family income at age 12	90084.8	110566.5	112219.8	129975.1
Cost of childcare at age 12	624.4	1726.8	676.6	1820.8
House value at age 13	115607.1	160334.0	158919.8	181807.1
Number of rooms in family home at age 13	6.6	1.9	7.0	2.0
Education of the spouse at age 13	9.0	6.5	10.5	5.9
Education of the head at age 13	13.0	2.4	13.6	2.4

Table 8: Summary Statistics (Numerical Variables) *(continued)*

Circumstance	Mean	StDev	Mean	StDev
Total family income at age 13	90864.9	99000.3	111337.0	114319.0
Cost of childcare at age 13	555.6	2255.7	640.5	2669.3
House value at age 14	120869.8	162043.0	166084.6	183006.8
Number of rooms in family home at age 14	6.6	1.9	7.1	2.0
Education of the spouse at age 14	8.9	6.5	10.3	6.1
Education of the head at age 14	13.0	2.4	13.6	2.4
Total family income at age 14	92786.8	99029.7	113792.9	112581.9
Cost of childcare at age 14	727.0	5041.2	779.4	5042.8
House value at age 15	120170.1	161521.0	163067.8	182266.9
Number of rooms in family home at age 15	6.7	1.9	7.1	2.0
Education of the spouse at age 15	8.9	6.5	10.2	6.1
Education of the head at age 15	13.0	2.4	13.6	2.3
Total family income at age 15	97655.1	147777.8	119594.8	177201.3
Cost of childcare at age 15	291.9	1065.0	286.8	1060.7
House value at age 16	128595.1	168289.4	170459.5	190005.3
Number of rooms in family home at age 16	6.8	2.0	7.2	2.0
Education of the spouse at age 16	8.8	6.6	10.3	6.2
Education of the head at age 16	13.0	2.4	13.5	2.4
Total family income at age 16	101085.3	101390.5	122777.6	116181.7
Education of the individual at age 16	2.7	5.3	2.3	4.1
Cost of childcare at age 16	293.4	1146.2	303.0	1207.6
House value at age 17	122109.9	156730.5	166456.0	176930.1
Number of rooms in family home at age 17	6.7	1.9	7.1	1.9
Education of the spouse at age 17	8.8	6.4	10.2	6.1
Education of the head at age 17	13.0	2.4	13.5	2.3
Total family income at age 17	64766.6	93027.1	78858.8	112091.3
Education of the individual at age 17	9.7	1.1	9.7	1.2
Cost of childcare at age 17	795.4	9039.5	669.3	8085.3
House value at age 18	140888.1	181874.6	188368.3	205024.4
Number of rooms in family home at age 18	6.8	2.0	7.1	2.1
Education of the spouse at age 18	9.1	6.4	10.6	5.9
Education of the head at age 18	12.9	2.5	13.5	2.5
Total family income at age 18	32683.2	25860.5	37481.4	27571.0
Education of the individual at age 18	10.8	1.0	10.8	1.0
Cost of childcare at age 18	454.1	6167.3	543.8	7635.8

Table 9 presents summary statistics for the full set of categorical variables used as circumstances. Head" refers to the head of the family in which the child grew up during childhood. “Spouse” refers to the spouse of the family head. “Age” refers to the child’s age during the first 18 years of their life.

Table 9: Summary Statistics (Categorical Variables)

		Full Sample (N = 1022)		SRC sample (N = 639)	
Circumstance	Category	Obs	Percent	Obs	Percent
Birthcohort	1978	153	15.0	99	15.5
	1979	169	16.5	104	16.3
	1980	172	16.8	108	16.9
	1981	187	18.3	131	20.5
	1982	145	14.2	83	13.0
	1983	196	19.2	114	17.8
Sex of the individual	Male	474	46.4	311	48.7
	Female	548	53.6	328	51.3
Birthweight	high	772	75.5	511	80.0
	low	55	5.4	27	4.2
	Other	30	2.9	12	1.9
Marital status of the mother when child was born	Married	734	71.8	547	85.6
	NeverMarried	215	21.0	56	8.8
	Other	29	2.8	16	2.5
Race of the individual	White	556	54.4	0	0.0
	Black	440	43.1	639	100.0
	Other	17	1.7	551	86.2
State where the family lived at age 1	CA	50	4.9	70	11.0
	IN	43	4.2	13	2.0
	MI	45	4.4	36	5.6
	MS	49	4.8	34	5.3
	OH	43	4.2	25	3.9
	TX	55	5.4	8	1.3
	Other	575	56.3	31	4.9
Sex of the family head at age 1	Male	725	70.9	35	5.5
	Female	135	13.2	404	63.2
Car availability in the family at age 1	Yes	737	72.1	529	82.8
	No	123	12.0	44	6.9
Type of parent at age 1	Father	682	66.7	542	84.8
	Mother	96	9.4	31	4.9
	NP	82	8.0	515	80.6
Type of the family head at age 1	H	82	8.0	34	5.3
	PH	778	76.1	24	3.8
Homeownership of the head at age 1	Own	433	42.4	24	3.8
	Rent	372	36.4	549	85.9
	Neither	55	5.4	342	53.5
Occupation of the head at age 1	Not applicable	98	9.6	198	31.0
	Professional, Technical, and Kindred Workers	107	10.5	33	5.2
	Managers and Administrators, except Farm	69	6.8	40	6.3
	Clerical and Kindred Workers	49	4.8	99	15.5
	Craftsman and Kindred Workers	124	12.1	59	9.2
	Operatives, except Transport	86	8.4	27	4.2
	Transport Equipment Operatives	43	4.2	93	14.6

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
	Laborers, except Farm	39	3.8	54	8.5
	Service Workers, except Private Household	65	6.4	23	3.6
	Other	41	4.0	24	3.8
Union membership of the head at age 1	Not applicable	208	20.4	25	3.9
	Yes	198	19.4	34	5.3
	No	451	44.1	121	18.9
Occupation of the spouse at age 1	Not applicable	369	36.1	133	20.8
	Professional, Technical, and Kindred Workers	77	7.5	316	49.5
	Clerical and Kindred Workers	86	8.4	231	36.2
	Service Workers, except Private Household	55	5.4	62	9.7
	Other	71	6.9	61	9.5
Household help from inside at age 1	Yes	206	20.2	39	6.1
	No	493	48.2	47	7.4
	Other	3	0.3	100	15.6
Other dependent for the head at age 1	Yes	64	6.3	358	56.0
	No	795	77.8	1	0.2
Area where the head grew up at age 1	Rural	187	18.3	35	5.5
	Town	303	29.6	537	84.0
	City	336	32.9	124	19.4
	Other	19	1.9	225	35.2
State where the head grew up at age 1	IL	44	4.3	200	31.3
	MI	43	4.2	17	2.7
	MS	59	5.8	34	5.3
	NY	43	4.2	28	4.4
	NC	48	4.7	10	1.6
	OH	51	5.0	35	5.5
	SC	43	4.2	16	2.5
	Other	514	50.3	38	5.9
Head's income class while growing up at age 1	Poor	269	26.3	10	1.6
	Average	337	33.0	396	62.0
	Rich	199	19.5	116	18.2
Region where the family lived at age 1	Northeast	114	11.2	278	43.5
	NorthCentral	275	26.9	147	23.0
	South	354	34.6	95	14.9
	West	108	10.6	208	32.6
	Other	9	0.9	171	26.8
Region where the head grew up at age 1	Northeast	121	11.8	91	14.2
	NorthCentral	274	26.8	8	1.3
	South	354	34.6	104	16.3
	West	88	8.6	220	34.4
	Other	8	0.8	158	24.7
Did head live at different state from childhood at age 1	SameState	633	61.9	77	12.1
	SameReg	84	8.2	8	1.3
	DiffReg	128	12.5	408	63.8
Race of the head at age 1	White	508	49.7	61	9.5
	Black	339	33.2	98	15.3

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
	Other	13	1.3	506	79.2
Union membership of the spouse at age 1	Not applicable	588	57.5	55	8.6
	No	215	21.0	12	1.9
	Other	38	3.7	383	59.9
Medicare/Medicaid usage in the family at age 1	Yes	124	12.1	155	24.3
	No	561	54.9	22	3.4
	Other	0	0.0	43	6.7
Usage of food stamps by the head at age 1	no	655	64.1	402	62.9
	yes	190	18.6	0	0.0
State where the family lived at age 2	CA	50	4.9	487	76.2
	IN	44	4.3	74	11.6
	MI	48	4.7	36	5.6
	MS	55	5.4	35	5.5
	TX	53	5.2	25	3.9
	Other	624	61.1	9	1.4
Sex of the family head at age 2	Male	718	70.3	34	5.3
	Female	156	15.3	429	67.1
Car availability in the family at age 2	Yes	737	72.1	514	80.4
	No	137	13.4	54	8.5
Type of parent at age 2	Father	685	67.0	529	82.8
	Mother	125	12.2	39	6.1
	NP	64	6.3	505	79.0
Type of the family head at age 2	H	64	6.3	47	7.4
	PH	810	79.3	16	2.5
Homeownership of the head at age 2	Own	451	44.1	16	2.5
	Rent	364	35.6	552	86.4
	Neither	59	5.8	354	55.4
Occupation of the head at age 2	Not applicable	140	13.7	177	27.7
	Professional, Technical, and Kindred Workers	117	11.4	37	5.8
	Managers and Administrators, except Farm	77	7.5	53	8.3
	Clerical and Kindred Workers	54	5.3	107	16.7
	Craftsman and Kindred Workers	139	13.6	68	10.6
	Operatives, except Transport	89	8.7	29	4.5
	Transport Equipment Operatives	55	5.4	95	14.9
	Service Workers, except Private Household	67	6.6	53	8.3
	Other	88	8.6	31	4.9
Union membership of the head at age 2	Not applicable	238	23.3	29	4.5
	Yes	182	17.8	66	10.3
	No	450	44.0	131	20.5
Occupation of the spouse at age 2	Not applicable	451	44.1	115	18.0
	Professional, Technical, and Kindred Workers	83	8.1	319	49.9

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
	Clerical and Kindred Workers	90	8.8	278	43.5
	Service Workers, except Private Household	65	6.4	67	10.5
	Other	83	8.1	61	9.5
Household help from inside at age 2	Yes	243	23.8	47	7.4
	No	482	47.2	53	8.3
	Other	1	0.1	131	20.5
Other dependent for the head at age 2	Yes	75	7.3	337	52.7
	No	797	78.0	1	0.2
Area where the head grew up at age 2	Rural	184	18.0	41	6.4
	Town	299	29.3	526	82.3
	City	357	34.9	120	18.8
	Other	20	2.0	220	34.4
State where the head grew up at age 2	IL	44	4.3	203	31.8
	MI	47	4.6	18	2.8
	MS	64	6.3	34	5.3
	NY	46	4.5	29	4.5
	NC	50	4.9	11	1.7
	OH	49	4.8	36	5.6
	SC	47	4.6	17	2.7
	Other	513	50.2	36	5.6
Head's income class while growing up at age 2	Poor	273	26.7	10	1.6
	Average	346	33.9	389	60.9
	Rich	202	19.8	110	17.2
Region where the family lived at age 2	Northeast	109	10.7	276	43.2
	NorthCentral	274	26.8	149	23.3
	South	368	36.0	89	13.9
	West	110	10.8	204	31.9
	Other	13	1.3	170	26.6
Region where the head grew up at age 2	Northeast	123	12.0	94	14.7
	NorthCentral	277	27.1	11	1.7
	South	361	35.3	102	16.0
	West	90	8.8	217	34.0
	Other	9	0.9	157	24.6
Did head live at different state from childhood at age 2	SameState	636	62.2	77	12.1
	SameReg	90	8.8	9	1.4
	DiffReg	134	13.1	396	62.0
Race of the head at age 2	White	504	49.3	63	9.9
	Black	360	35.2	103	16.1
	Other	10	1.0	499	78.1
Union membership of the spouse at age 2	Not applicable	574	56.2	60	9.4
	Yes	45	4.4	9	1.4
	No	250	24.5	364	57.0
Medicare/Medicaid usage in the family at age 2	Yes	109	10.7	25	3.9
	No	618	60.5	176	27.5
	Other	0	0.0	34	5.3
Usage of food stamps by the head at age 2	no	683	66.8	435	68.1
	yes	191	18.7	0	0.0
State where the family lived at age 3	CA	51	5.0	500	78.2
	MI	53	5.2	68	10.6
	MS	56	5.5	35	5.5

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
	TX	52	5.1	27	4.2
	Other	670	65.6	9	1.4
Sex of the family head at age 3	Male	709	69.4	34	5.3
	Female	173	16.9	460	72.0
Car availability in the family at age 3	Yes	730	71.4	506	79.2
	No	152	14.9	59	9.2
Type of parent at age 3	Father	683	66.8	525	82.2
	Mother	148	14.5	40	6.3
	NP	51	5.0	498	77.9
Type of the family head at age 3	H	51	5.0	55	8.6
	PH	831	81.3	12	1.9
Homeownership of the head at age 3	Own	481	47.1	12	1.9
	Rent	354	34.6	553	86.5
	Neither	47	4.6	371	58.1
Occupation of the head at age 3	Not applicable	158	15.5	167	26.1
	Professional, Technical, and Kindred Workers	124	12.1	27	4.2
	Managers and Administrators, except Farm	95	9.3	57	8.9
	Clerical and Kindred Workers	49	4.8	112	17.5
	Craftsman and Kindred Workers	150	14.7	77	12.1
	Operatives, except Transport	91	8.9	29	4.5
	Transport Equipment Operatives	54	5.3	105	16.4
	Service Workers, except Private Household	62	6.1	56	8.8
	Other	98	9.6	30	4.7
Union membership of the head at age 3	Not applicable	245	24.0	28	4.4
	Yes	184	18.0	70	11.0
	No	448	43.8	135	21.1
Occupation of the spouse at age 3	Not applicable	534	52.3	115	18.0
	Professional, Technical, and Kindred Workers	94	9.2	312	48.8
	Clerical and Kindred Workers	100	9.8	315	49.3
	Service Workers, except Private Household	63	6.2	75	11.7
	Other	90	8.8	70	11.0
Household help from inside at age 3	Yes	279	27.3	45	7.0
	No	453	44.3	60	9.4
	Other	1	0.1	151	23.6
Other dependent for the head at age 3	Yes	84	8.2	318	49.8
	No	797	78.0	0	0.0
Area where the head grew up at age 3	Rural	176	17.2	44	6.9
	Town	301	29.5	520	81.4
	City	369	36.1	117	18.3

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
	Other	19	1.9	222	34.7
State where the head grew up at age 3	IL	44	4.3	202	31.6
	MI	51	5.0	17	2.7
	MS	65	6.4	33	5.2
	NY	49	4.8	29	4.5
	NC	47	4.6	11	1.7
	OH	49	4.8	39	6.1
	SC	47	4.6	15	2.3
	Other	514	50.3	36	5.6
Head's income class while growing up at age 3	Poor	276	27.0	10	1.6
	Average	337	33.0	386	60.4
	Rich	213	20.8	111	17.4
Region where the family lived at age 3	Northeast	114	11.2	267	41.8
	NorthCentral	269	26.3	155	24.3
	South	376	36.8	92	14.4
	West	112	11.0	198	31.0
	Other	11	1.1	171	26.8
Region where the head grew up at age 3	Northeast	123	12.0	95	14.9
	NorthCentral	278	27.2	9	1.4
	South	362	35.4	103	16.1
	West	94	9.2	213	33.3
	Other	9	0.9	155	24.3
Did head live at different state from childhood at age 3	SameState	648	63.4	79	12.4
	SameReg	85	8.3	9	1.4
	DiffReg	133	13.0	400	62.6
Race of the head at age 3	White	500	48.9	55	8.6
	Black	371	36.3	104	16.3
	Other	11	1.1	495	77.5
Union membership of the spouse at age 3	Not applicable	567	55.5	60	9.4
	Yes	45	4.4	10	1.6
	No	256	25.0	346	54.1
Medicare/Medicaid usage in the family at age 3	Yes	115	11.3	27	4.2
	No	618	60.5	183	28.6
	Other	0	0.0	36	5.6
Usage of food stamps by the head at age 3	no	679	66.4	434	67.9
	yes	203	19.9	0	0.0
State where the family lived at age 4	CA	50	4.9	495	77.5
	MI	53	5.2	70	11.0
	MS	54	5.3	34	5.3
	TX	57	5.6	27	4.2
	Other	666	65.2	7	1.1
Sex of the family head at age 4	Male	702	68.7	39	6.1
	Female	178	17.4	458	71.7
Car availability in the family at age 4	Yes	603	59.0	506	79.2
	No	117	11.4	59	9.2
Type of parent at age 4	Father	683	66.8	434	67.9
	Mother	155	15.2	32	5.0
	Other	42	4.1	499	78.1
Type of the family head at age 4	PH	838	82.0	55	8.6
	Other	42	4.1	11	1.7
Homeownership of the head at age 4	Own	491	48.0	554	86.7
	Rent	343	33.6	11	1.7
	Neither	46	4.5	377	59.0

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
Occupation of the head at age 4	Not applicable	150	14.7	163	25.5
	Professional, Technical, and Kindred Workers	121	11.8	25	3.9
	Managers and Administrators, except Farm	105	10.3	55	8.6
	Craftsman and Kindred Workers	159	15.6	106	16.6
	Operatives, except Transport	92	9.0	94	14.7
	Transport Equipment Operatives	49	4.8	105	16.4
	Service Workers, except Private Household	68	6.7	57	8.9
	Other	136	13.3	31	4.9
Union membership of the head at age 4	Not applicable	235	23.0	29	4.5
	Yes	180	17.6	88	13.8
	No	453	44.3	133	20.8
Occupation of the spouse at age 4	Not applicable	531	52.0	117	18.3
	Professional, Technical, and Kindred Workers	94	9.2	307	48.0
	Clerical and Kindred Workers	97	9.5	314	49.1
	Service Workers, except Private Household	70	6.8	78	12.2
	Other	88	8.6	65	10.2
Household help from inside at age 4	Yes	269	26.3	49	7.7
	No	315	30.8	59	9.2
	Other	0	0.0	156	24.4
Other dependent for the head at age 4	Yes	75	7.3	222	34.7
	No	803	78.6	0	0.0
Area where the head grew up at age 4	Rural	174	17.0	39	6.1
	Town	302	29.5	524	82.0
	City	368	36.0	115	18.0
	Other	19	1.9	226	35.4
State where the head grew up at age 4	IL	46	4.5	200	31.3
	MI	50	4.9	17	2.7
	MS	68	6.7	35	5.5
	NY	49	4.8	27	4.2
	NC	46	4.5	11	1.7
	OH	49	4.8	39	6.1
	SC	48	4.7	15	2.3
	Other	510	49.9	35	5.5
Head's income class while growing up at age 4	Poor	267	26.1	10	1.6
	Average	341	33.4	387	60.6
	Rich	216	21.1	108	16.9
Region where the family lived at age 4	Northeast	112	11.0	274	42.9
	NorthCentral	271	26.5	152	23.8
	South	379	37.1	91	14.2
	West	113	11.1	198	31.0

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
	Other	5	0.5	176	27.5
Region where the head grew up at age 4	Northeast	125	12.2	95	14.9
	NorthCentral	278	27.2	5	0.8
	South	363	35.5	104	16.3
	West	92	9.0	210	32.9
	Other	8	0.8	158	24.7
Did head live at different state from childhood at age 4	SameState	640	62.6	79	12.4
	SameReg	95	9.3	8	1.3
	DiffReg	131	12.8	395	61.8
Race of the head at age 4	White	500	48.9	65	10.2
	Black	372	36.4	99	15.5
	Other	8	0.8	496	77.6
Union membership of the spouse at age 4	Not applicable	568	55.6	62	9.7
	Yes	50	4.9	7	1.1
	No	253	24.8	347	54.3
Medicare/Medicaid usage in the family at age 4	Yes	112	11.0	31	4.9
	No	630	61.6	182	28.5
	Other	0	0.0	39	6.1
Usage of food stamps by the head at age 4	no	686	67.1	436	68.2
	yes	194	19.0	0	0.0
State where the family lived at age 5	CA	50	4.9	500	78.2
	MI	51	5.0	65	10.2
	MS	56	5.5	33	5.2
	TX	51	5.0	27	4.2
	Other	652	63.8	8	1.3
Sex of the family head at age 5	Male	684	66.9	34	5.3
	Female	176	17.2	450	70.4
Car availability in the family at age 5	Yes	499	48.8	490	76.7
	No	82	8.0	62	9.7
Type of parent at age 5	Father	666	65.2	362	56.7
	Mother	161	15.8	21	3.3
	Other	33	3.2	483	75.6
Type of the family head at age 5	PH	827	80.9	61	9.5
	Other	33	3.2	8	1.3
Homeownership of the head at age 5	Own	486	47.6	544	85.1
	Rent	326	31.9	8	1.3
	Neither	48	4.7	380	59.5
Occupation of the head at age 5	Not applicable	139	13.6	149	23.3
	Professional, Technical, and Kindred Workers	129	12.6	23	3.6
	Managers and Administrators, except Farm	107	10.5	49	7.7
	Clerical and Kindred Workers	55	5.4	118	18.5
	Craftsman and Kindred Workers	160	15.7	93	14.6
	Operatives, except Transport	85	8.3	26	4.1
	Transport Equipment Operatives	46	4.5	105	16.4

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
	Service Workers, except Private Household	58	5.7	51	8.0
	Other	81	7.9	29	4.5
Union membership of the head at age 5	Not applicable	239	23.4	22	3.4
	Yes	167	16.3	59	9.2
	No	436	42.7	136	21.3
Occupation of the spouse at age 5	Not applicable	478	46.8	108	16.9
	Professional, Technical, and Kindred Workers	108	10.6	299	46.8
	Clerical and Kindred Workers	96	9.4	282	44.1
	Service Workers, except Private Household	84	8.2	84	13.1
	Other	92	9.0	68	10.6
Household help from inside at age 5	Yes	303	29.6	56	8.8
	No	275	26.9	62	9.7
	Other	0	0.0	184	28.8
Other dependent for the head at age 5	Yes	79	7.7	196	30.7
	No	775	75.8	0	0.0
Area where the head grew up at age 5	Rural	168	16.4	42	6.6
	Town	296	29.0	507	79.3
	City	362	35.4	112	17.5
	Other	19	1.9	221	34.6
State where the head grew up at age 5	IL	47	4.6	195	30.5
	MI	50	4.9	17	2.7
	MS	65	6.4	36	5.6
	NY	50	4.9	27	4.2
	NC	45	4.4	10	1.6
	OH	50	4.9	40	6.3
	SC	47	4.6	15	2.3
	Other	492	48.1	35	5.5
Head's income class while growing up at age 5	Poor	260	25.4	9	1.4
	Average	336	32.9	374	58.5
	Rich	208	20.4	102	16.0
Region where the family lived at age 5	Northeast	111	10.9	273	42.7
	NorthCentral	264	25.8	146	22.8
	South	370	36.2	91	14.2
	West	111	10.9	193	30.2
	Other	4	0.4	171	26.8
Region where the head grew up at age 5	Northeast	125	12.2	93	14.6
	NorthCentral	278	27.2	4	0.6
	South	351	34.3	104	16.3
	West	86	8.4	208	32.6
	Other	6	0.6	153	23.9
Did head live at different state from childhood at age 5	SameState	634	62.0	75	11.7
	SameReg	89	8.7	6	0.9
	DiffReg	123	12.0	389	60.9
Race of the head at age 5	White	490	47.9	61	9.5
	Black	362	35.4	96	15.0
	Other	8	0.8	486	76.1
Union membership of the spouse at age 5	Not applicable	519	50.8	59	9.2

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
	Yes	48	4.7	7	1.1
	No	281	27.5	320	50.1
Medicare/Medicaid usage in the family at age 5	Yes	114	11.2	27	4.2
	No	743	72.7	198	31.0
	Other	0	0.0	31	4.9
Usage of food stamps by the head at age 5	no	670	65.6	519	81.2
	yes	190	18.6	0	0.0
Race of the spouse at age 5	Not applicable	131	12.8	486	76.1
	White	314	30.7	66	10.3
	Black	135	13.2	46	7.2
	Other	2	0.2	311	48.7
State where the family lived at age 6	CA	49	4.8	18	2.8
	MI	52	5.1	1	0.2
	MS	53	5.2	32	5.0
	TX	48	4.7	29	4.5
	Other	646	63.2	7	1.1
Sex of the family head at age 6	Male	659	64.5	32	5.0
	Female	189	18.5	445	69.6
Type of parent at age 6	Father	643	62.9	475	74.3
	Mother	177	17.3	70	11.0
	Other	28	2.7	467	73.1
Type of the family head at age 6	PH	820	80.2	69	10.8
	Other	28	2.7	9	1.4
Homeownership of the head at age 6	Own	488	47.7	536	83.9
	Rent	319	31.2	9	1.4
	Other	41	4.0	383	59.9
Occupation of the head at age 6	Not applicable	131	12.8	141	22.1
	Professional, Technical, and Kindred Workers	122	11.9	21	3.3
	Managers and Administrators, except Farm	112	11.0	47	7.4
	Clerical and Kindred Workers	58	5.7	112	17.5
	Craftsman and Kindred Workers	143	14.0	98	15.3
	Operatives, except Transport	76	7.4	32	5.0
	Transport Equipment Operatives	55	5.4	90	14.1
	Service Workers, except Private Household	65	6.4	44	6.9
	Other	86	8.4	33	5.2
Union membership of the head at age 6	Not applicable	227	22.2	30	4.7
	Yes	170	16.6	59	9.2
	No	430	42.1	131	20.5
Occupation of the spouse at age 6	Not applicable	451	44.1	114	17.8
	Professional, Technical, and Kindred Workers	106	10.4	284	44.4
	Clerical and Kindred Workers	106	10.4	253	39.6

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
	Operatives, except Transport	44	4.3	84	13.1
	Service Workers, except Private Household	89	8.7	78	12.2
	Other	52	5.1	24	3.8
Other dependent for the head at age 6	Yes	78	7.6	65	10.2
	No	769	75.2	41	6.4
Area where the head grew up at age 6	Rural	162	15.9	42	6.6
	Town	290	28.4	503	78.7
	City	364	35.6	110	17.2
	Other	17	1.7	218	34.1
State where the head grew up at age 6	IL	47	4.6	195	30.5
	MI	50	4.9	15	2.3
	MS	64	6.3	36	5.6
	NY	48	4.7	28	4.4
	NC	42	4.1	10	1.6
	OH	47	4.6	39	6.1
	SC	46	4.5	13	2.0
	Other	489	47.8	34	5.3
Head's income class while growing up at age 6	Poor	256	25.0	9	1.4
	Average	328	32.1	369	57.7
	Rich	208	20.4	102	16.0
Region where the family lived at age 6	Northeast	112	11.0	267	41.8
	NorthCentral	262	25.6	144	22.5
	South	358	35.0	93	14.6
	West	110	10.8	195	30.5
	Other	6	0.6	162	25.4
Region where the head grew up at age 6	Northeast	120	11.7	92	14.4
	NorthCentral	273	26.7	3	0.5
	South	347	34.0	100	15.6
	West	88	8.6	206	32.2
	Other	5	0.5	150	23.5
Did head live at different state from childhood at age 6	SameState	630	61.6	77	12.1
	SameReg	84	8.2	5	0.8
	DiffReg	119	11.6	386	60.4
Race of the head at age 6	White	490	47.9	60	9.4
	Black	351	34.3	92	14.4
	Other	7	0.7	482	75.4
Union membership of the spouse at age 6	Not applicable	507	49.6	57	8.9
	Yes	60	5.9	6	0.9
	No	267	26.1	305	47.7
Medicare/Medicaid usage in the family at age 6	Yes	122	11.9	36	5.6
	No	726	71.0	194	30.4
	Other	0	0.0	33	5.2
Usage of food stamps by the head at age 6	no	683	66.8	512	80.1
	yes	165	16.1	0	0.0
Race of the spouse at age 6	Not applicable	166	16.2	490	76.7
	White	377	36.9	55	8.6
	Black	170	16.6	64	10.0
	Other	2	0.2	373	58.4
State where the family lived at age 7	CA	48	4.7	22	3.4
	MI	50	4.9	1	0.2
	MS	54	5.3	32	5.0

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Percent	Obs	Percent
	TX	48	4.7	28	4.4
	Other	639	62.5	8	1.3
Sex of the family head at age 7	Male	655	64.1	31	4.9
	Female	184	18.0	445	69.6
Type of parent at age 7	Father	641	62.7	474	74.2
	Mother	173	16.9	70	11.0
	Other	25	2.4	467	73.1
Type of the family head at age 7	PH	814	79.6	68	10.6
	Other	25	2.4	9	1.4
Homeownership of the head at age 7	Own	508	49.7	535	83.7
	Rent	283	27.7	9	1.4
	Neither	48	4.7	395	61.8
Occupation of the head at age 7	Not applicable	136	13.3	124	19.4
	Professional, Technical, and Kindred Workers	138	13.5	25	3.9
	Managers and Administrators, except Farm	114	11.2	51	8.0
	Clerical and Kindred Workers	51	5.0	122	19.1
	Craftsman and Kindred Workers	142	13.9	98	15.3
	Operatives, except Transport	65	6.4	27	4.2
	Transport Equipment Operatives	54	5.3	93	14.6
	Service Workers, except Private Household	61	6.0	43	6.7
	Other	78	7.6	30	4.7
Union membership of the head at age 7	Not applicable	226	22.1	24	3.8
	Yes	159	15.6	56	8.8
	No	431	42.2	132	20.7
Occupation of the spouse at age 7	Not applicable	413	40.4	99	15.5
	Professional, Technical, and Kindred Workers	114	11.2	298	46.6
	Clerical and Kindred Workers	112	11.0	230	36.0
	Operatives, except Transport	50	4.9	92	14.4
	Service Workers, except Private Household	77	7.5	87	13.6
	Other	73	7.1	21	3.3
Other dependent for the head at age 7	Yes	69	6.8	57	8.9
	No	763	74.7	57	8.9
Area where the head grew up at age 7	Rural	167	16.3	39	6.1
	Town	283	27.7	501	78.4
	City	355	34.7	113	17.7
	Other	18	1.8	215	33.6
State where the head grew up at age 7	IL	45	4.4	192	30.0
	MI	49	4.8	16	2.5

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
	MS	63	6.2	35	5.5
	NY	49	4.8	28	4.4
	NC	44	4.3	10	1.6
	OH	45	4.4	40	6.3
	SC	42	4.1	14	2.2
	Other	487	47.7	32	5.0
Head's income class while growing up at age 7	Poor	256	25.0	9	1.4
	Average	321	31.4	368	57.6
	Rich	208	20.4	105	16.4
Region where the family lived at age 7	Northeast	110	10.8	263	41.2
	NorthCentral	259	25.3	143	22.4
	South	359	35.1	92	14.4
	West	108	10.6	194	30.4
	Other	3	0.3	164	25.7
Region where the head grew up at age 7	Northeast	122	11.9	91	14.2
	NorthCentral	269	26.3	3	0.5
	South	340	33.3	100	15.6
	West	88	8.6	204	31.9
	Other	5	0.5	150	23.5
Did head live at different state from childhood at age 7	SameState	622	60.9	77	12.1
	SameReg	89	8.7	5	0.8
	DiffReg	113	11.1	384	60.1
Race of the head at age 7	White	490	47.9	63	9.9
	Black	344	33.7	89	13.9
	Other	5	0.5	482	75.4
Union membership of the spouse at age 7	Not applicable	461	45.1	59	9.2
	Yes	46	4.5	3	0.5
	No	321	31.4	274	42.9
Medicare/Medicaid usage in the family at age 7	Yes	114	11.2	33	5.2
	No	722	70.6	228	35.7
	Other	0	0.0	34	5.3
Usage of food stamps by the head at age 7	no	672	65.8	507	79.3
	yes	167	16.3	0	0.0
Race of the spouse at age 7	Not applicable	189	18.5	492	77.0
	White	448	43.8	52	8.1
	Black	196	19.2	73	11.4
	Other	3	0.3	443	69.3
State where the family lived at age 8	CA	45	4.4	25	3.9
	MI	50	4.9	2	0.3
	MS	54	5.3	33	5.2
	SC	42	4.1	26	4.1
	TX	48	4.7	8	1.3
	Other	583	57.0	11	1.7
Sex of the family head at age 8	Male	626	61.3	31	4.9
	Female	196	19.2	425	66.5
Type of parent at age 8	Father	616	60.3	458	71.7
	Mother	185	18.1	76	11.9
	Other	21	2.1	453	70.9
Type of the family head at age 8	PH	801	78.4	73	11.4
	Other	21	2.1	8	1.3
Homeownership of the head at age 8	Own	497	48.6	526	82.3
	Rent	281	27.5	8	1.3
	Neither	44	4.3	381	59.6

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
Occupation of the head at age 8	Not applicable	123	12.0	125	19.6
	Professional, Technical, and Kindred Workers	137	13.4	28	4.4
	Managers and Administrators, except Farm	109	10.7	43	6.7
	Clerical and Kindred Workers	55	5.4	120	18.8
	Craftsman and Kindred Workers	141	13.8	95	14.9
	Operatives, except Transport	68	6.7	31	4.9
	Transport Equipment Operatives	52	5.1	96	15.0
	Service Workers, except Private Household	65	6.4	42	6.6
	Other	72	7.0	29	4.5
Union membership of the head at age 8	Not applicable	215	21.0	29	4.5
	Yes	144	14.1	49	7.7
	No	438	42.9	124	19.4
Occupation of the spouse at age 8	Not applicable	403	39.4	88	13.8
	Professional, Technical, and Kindred Workers	103	10.1	306	47.9
	Clerical and Kindred Workers	122	11.9	225	35.2
	Service Workers, except Private Household	80	7.8	86	13.5
	Other	113	11.1	86	13.5
Other dependent for the head at age 8	Yes	59	5.8	60	9.4
	No	753	73.7	76	11.9
Area where the head grew up at age 8	Rural	165	16.1	37	5.8
	Town	277	27.1	491	76.8
	City	349	34.1	111	17.4
	Other	15	1.5	209	32.7
State where the head grew up at age 8	IL	46	4.5	193	30.2
	MI	51	5.0	13	2.0
	MS	61	6.0	36	5.6
	NY	45	4.4	27	4.2
	NC	41	4.0	9	1.4
	OH	45	4.4	39	6.1
	SC	45	4.4	12	1.9
	Other	471	46.1	33	5.2
Head's income class while growing up at age 8	Poor	256	25.0	10	1.6
	Average	312	30.5	359	56.2
	Rich	200	19.6	109	17.1
Region where the family lived at age 8	Northeast	110	10.8	256	40.1
	NorthCentral	254	24.9	138	21.6
	South	349	34.1	93	14.6
	West	106	10.4	186	29.1
	Other	3	0.3	161	25.2

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per- cent	Obs	Percent
Region where the head grew up at age 8	Northeast	115	11.3	92	14.4
	NorthCentral	267	26.1	2	0.3
	South	334	32.7	97	15.2
	West	84	8.2	201	31.5
	Other	5	0.5	148	23.2
Did head live at different state from childhood at age 8	SameState	609	59.6	74	11.6
	SameReg	85	8.3	5	0.8
	DiffReg	111	10.9	376	58.8
Race of the head at age 8	White	477	46.7	62	9.7
	Black	339	33.2	87	13.6
	Other	5	0.5	470	73.6
Union membership of the spouse at age 8	Not applicable	452	44.2	60	9.4
	Yes	48	4.7	4	0.6
	No	300	29.4	270	42.3
Medicare/Medicaid usage in the family at age 8	Yes	116	11.4	31	4.9
	No	703	68.8	216	33.8
	Other	0	0.0	38	5.9
Usage of food stamps by the head at age 8	no	666	65.2	494	77.3
	yes	156	15.3	0	0.0
Race of the spouse at age 8	Not applicable	205	20.1	483	75.6
	White	429	42.0	51	8.0
	Black	182	17.8	83	13.0
	Other	3	0.3	423	66.2
State where the family lived at age 9	CA	45	4.4	25	3.9
	MI	49	4.8	2	0.3
	MS	54	5.3	32	5.0
	TX	49	4.8	25	3.9
	Other	614	60.1	8	1.3
Sex of the family head at age 9	Male	618	60.5	33	5.2
	Female	193	18.9	435	68.1
Type of parent at age 9	Father	610	59.7	458	71.7
	Mother	184	18.0	75	11.7
	Other	17	1.7	453	70.9
Type of the family head at age 9	PH	794	77.7	73	11.4
	Other	17	1.7	7	1.1
Homeownership of the head at age 9	Own	501	49.0	526	82.3
	Rent	280	27.4	7	1.1
	Other	30	2.9	385	60.3
Occupation of the head at age 9	Not applicable	116	11.4	130	20.3
	Professional, Technical, and Kindred Workers	135	13.2	18	2.8
	Managers and Administrators, except Farm	97	9.5	44	6.9
	Clerical and Kindred Workers	57	5.6	118	18.5
	Craftsman and Kindred Workers	137	13.4	85	13.3
	Operatives, except Transport	65	6.4	32	5.0
	Transport Equipment Operatives	59	5.8	94	14.7

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
	Service Workers, except Private Household	65	6.4	40	6.3
	Other	80	7.8	36	5.6
Union membership of the head at age 9	Not applicable	206	20.2	27	4.2
	Yes	147	14.4	57	8.9
	No	434	42.5	125	19.6
Occupation of the spouse at age 9	Not applicable	394	38.6	90	14.1
	Professional, Technical, and Kindred Workers	108	10.6	303	47.4
	Clerical and Kindred Workers	111	10.9	215	33.6
	Service Workers, except Private Household	91	8.9	89	13.9
	Other	107	10.5	81	12.7
Other dependent for the head at age 9	Yes	71	6.9	71	11.1
	No	730	71.4	77	12.1
Area where the head grew up at age 9	Rural	165	16.1	43	6.7
	Town	273	26.7	484	75.7
	City	345	33.8	113	17.7
	Other	14	1.4	205	32.1
State where the head grew up at age 9	IL	43	4.2	195	30.5
	MI	50	4.9	12	1.9
	MS	62	6.1	33	5.2
	NY	43	4.2	27	4.2
	OH	42	4.1	9	1.4
	SC	41	4.0	38	5.9
	Other	514	50.3	31	4.9
Head's income class while growing up at age 9	Poor	259	25.3	9	1.4
	Average	308	30.1	377	59.0
	Rich	195	19.1	111	17.4
Region where the family lived at age 9	Northeast	105	10.3	254	39.7
	NorthCentral	254	24.9	137	21.4
	South	343	33.6	90	14.1
	West	106	10.4	189	29.6
	Other	3	0.3	161	25.2
Region where the head grew up at age 9	Northeast	113	11.1	90	14.1
	NorthCentral	262	25.6	3	0.5
	South	331	32.4	97	15.2
	West	84	8.2	198	31.0
	Other	5	0.5	151	23.6
Did head live at different state from childhood at age 9	SameState	597	58.4	73	11.4
	SameReg	88	8.6	5	0.8
	DiffReg	110	10.8	370	57.9
Race of the head at age 9	White	477	46.7	67	10.5
	Black	329	32.2	87	13.6
	Other	5	0.5	471	73.7
Union membership of the spouse at age 9	Not applicable	449	43.9	59	9.2
	Yes	52	5.1	3	0.5
	No	295	28.9	266	41.6
Medicare/Medicaid usage in the family at age 9	Yes	116	11.4	34	5.3
	No	694	67.9	219	34.3

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Percent	Obs	Percent
	Other	0	0.0	42	6.6
Usage of food stamps by the head at age 9	no	655	64.1	490	76.7
	yes	156	15.3	0	0.0
Race of the spouse at age 9	Not applicable	205	20.1	488	76.4
	White	427	41.8	45	7.0
	Black	173	16.9	84	13.1
	Other	3	0.3	422	66.0
State where the family lived at age 10	CA	44	4.3	24	3.8
	MI	51	5.0	2	0.3
	MS	54	5.3	31	4.9
	SC	41	4.0	26	4.1
	TX	50	4.9	8	1.3
	Other	578	56.6	11	1.7
Sex of the family head at age 10	Male	622	60.9	32	5.0
	Female	196	19.2	420	65.7
Type of parent at age 10	Father	615	60.2	450	70.4
	Mother	187	18.3	78	12.2
	Other	16	1.6	446	69.8
Type of the family head at age 10	PH	802	78.5	76	11.9
	Other	16	1.6	6	0.9
Homeownership of the head at age 10	Own	513	50.2	522	81.7
	Rent	271	26.5	6	0.9
	Other	34	3.3	392	61.3
Occupation of the head at age 10	Not applicable	114	11.2	118	18.5
	Professional, Technical, and Kindred Workers	137	13.4	18	2.8
	Managers and Administrators, except Farm	106	10.4	42	6.6
	Clerical and Kindred Workers	69	6.8	119	18.6
	Craftsman and Kindred Workers	137	13.4	95	14.9
	Operatives, except Transport	65	6.4	36	5.6
	Transport Equipment Operatives	55	5.4	93	14.6
	Service Workers, except Private Household	59	5.8	33	5.2
	Other	74	7.2	29	4.5
Union membership of the head at age 10	Not applicable	213	20.8	27	4.2
	Yes	150	14.7	52	8.1
	No	439	43.0	130	20.3
Occupation of the spouse at age 10	Not applicable	385	37.7	86	13.5
	Professional, Technical, and Kindred Workers	103	10.1	303	47.4
	Clerical and Kindred Workers	129	12.6	208	32.6
	Service Workers, except Private Household	93	9.1	83	13.0

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
	Other	107	10.5	103	16.1
Other dependent for the head at age 10	Yes	70	6.8	65	10.2
	No	736	72.0	68	10.6
Area where the head grew up at age 10	Rural	168	16.4	38	5.9
	Town	275	26.9	480	75.1
	City	344	33.7	114	17.8
	Other	16	1.6	205	32.1
State where the head grew up at age 10	IL	43	4.2	188	29.4
	MI	50	4.9	14	2.2
	MS	62	6.1	32	5.0
	NY	46	4.5	27	4.2
	NC	42	4.1	9	1.4
	OH	42	4.1	38	5.9
	SC	43	4.2	11	1.7
	Other	475	46.5	31	4.9
Head’s income class while growing up at age 10	Poor	260	25.4	11	1.7
	Average	309	30.2	363	56.8
	Rich	200	19.6	111	17.4
Region where the family lived at age 10	Northeast	104	10.2	251	39.3
	NorthCentral	254	24.9	138	21.6
	South	351	34.3	88	13.8
	West	106	10.4	186	29.1
	Other	3	0.3	161	25.2
Region where the head grew up at age 10	Northeast	118	11.5	91	14.2
	NorthCentral	262	25.6	2	0.3
	South	338	33.1	100	15.6
	West	81	7.9	197	30.8
	Other	4	0.4	152	23.8
Did head live at different state from childhood at age 10	SameState	604	59.1	69	10.8
	SameReg	89	8.7	4	0.6
	DiffReg	110	10.8	372	58.2
Race of the head at age 10	White	470	46.0	67	10.5
	Black	341	33.4	83	13.0
	Other	6	0.6	465	72.8
Union membership of the spouse at age 10	Not applicable	438	42.9	59	9.2
	Yes	53	5.2	4	0.6
	No	315	30.8	253	39.6
Medicare/Medicaid usage in the family at age 10	Yes	116	11.4	35	5.5
	No	701	68.6	231	36.2
	Other	0	0.0	41	6.4
Usage of food stamps by the head at age 10	no	661	64.7	486	76.1
	yes	157	15.4	0	0.0
Race of the spouse at age 10	Not applicable	210	20.5	485	75.9
	White	418	40.9	43	6.7
	Black	183	17.9	87	13.6
	Other	3	0.3	413	64.6
State where the family lived at age 11	CA	49	4.8	25	3.9
	MI	53	5.2	2	0.3
	MS	55	5.4	35	5.5
	NC	43	4.2	28	4.4
	TX	52	5.1	8	1.3
	Other	591	57.8	12	1.9
Sex of the family head at age 11	Male	631	61.7	34	5.3
	Female	212	20.7	424	66.4

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
Type of parent at age 11	Father	623	61.0	459	71.8
	Mother	202	19.8	82	12.8
	Other	18	1.8	454	71.0
Type of the family head at age 11	PH	825	80.7	80	12.5
	Other	18	1.8	7	1.1
Homeownership of the head at age 11	Own	531	52.0	534	83.6
	Rent	280	27.4	7	1.1
	Other	32	3.1	404	63.2
Occupation of the head at age 11	Not applicable	120	11.7	119	18.6
	Professional, Technical, and Kindred Workers	140	13.7	18	2.8
	Managers and Administrators, except Farm	112	11.0	41	6.4
	Clerical and Kindred Workers	63	6.2	122	19.1
	Craftsman and Kindred Workers	131	12.8	99	15.5
	Operatives, except Transport	66	6.5	34	5.3
	Transport Equipment Operatives	63	6.2	92	14.4
	Service Workers, except Private Household	66	6.5	31	4.9
	Other	82	8.0	32	5.0
Union membership of the head at age 11	Not applicable	213	20.8	29	4.5
	Yes	150	14.7	61	9.5
	No	470	46.0	126	19.7
Occupation of the spouse at age 11	Not applicable	379	37.1	90	14.1
	Professional, Technical, and Kindred Workers	125	12.2	320	50.1
	Clerical and Kindred Workers	127	12.4	193	30.2
	Service Workers, except Private Household	89	8.7	102	16.0
	Other	123	12.0	98	15.3
Other dependent for the head at age 11	Yes	79	7.7	65	10.2
	No	747	73.1	83	13.0
Area where the head grew up at age 11	Rural	169	16.5	46	7.2
	Town	289	28.3	483	75.6
	City	353	34.5	114	17.8
	Other	17	1.7	213	33.3
State where the head grew up at age 11	MI	51	5.0	193	30.2
	MS	65	6.4	14	2.2
	NY	46	4.5	28	4.4
	NC	44	4.3	9	1.4
	OH	44	4.3	38	5.9
	SC	43	4.2	12	1.9
	Other	535	52.3	33	5.2
Head’s income class while growing up at age 11	Poor	277	27.1	10	1.6

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per- cent	Obs	Percent
	Average	311	30.4	405	63.4
	Rich	204	20.0	120	18.8
Region where the family lived at age 11	Northeast	105	10.3	253	39.6
	NorthCentral	257	25.1	138	21.6
	South	363	35.5	88	13.8
	West	113	11.1	188	29.4
	Other	5	0.5	165	25.8
Region where the head grew up at age 11	Northeast	116	11.4	97	15.2
	NorthCentral	267	26.1	3	0.5
	South	355	34.7	99	15.5
	West	85	8.3	202	31.6
	Other	5	0.5	157	24.6
Did head live at different state from childhood at age 11	SameState	619	60.6	72	11.3
	SameReg	89	8.7	5	0.8
	DiffReg	120	11.7	377	59.0
Race of the head at age 11	White	477	46.7	65	10.2
	Black	359	35.1	93	14.6
	Other	6	0.6	474	74.2
Union membership of the spouse at age 11	Not applicable	430	42.1	62	9.7
	Yes	63	6.2	5	0.8
	No	340	33.3	238	37.2
Medicare/Medicaid usage in the family at age 11	Yes	121	11.8	41	6.4
	No	718	70.3	253	39.6
	Other	0	0.0	40	6.3
Usage of food stamps by the head at age 11	no	688	67.3	498	77.9
	yes	155	15.2	0	0.0
Race of the spouse at age 11	Not applicable	228	22.3	492	77.0
	White	426	41.7	49	7.7
	Black	181	17.7	90	14.1
	Other	3	0.3	421	65.9
State where the family lived at age 12	CA	49	4.8	26	4.1
	MI	53	5.2	2	0.3
	MS	59	5.8	35	5.5
	NC	45	4.4	29	4.5
	TX	54	5.3	8	1.3
	Other	619	60.6	14	2.2
Sex of the family head at age 12	Male	640	62.6	34	5.3
	Female	239	23.4	440	68.9
Type of parent at age 12	Father	633	61.9	466	72.9
	Mother	226	22.1	94	14.7
	Other	20	2.0	461	72.1
Type of the family head at age 12	PH	859	84.1	92	14.4
	Other	20	2.0	7	1.1
Homeownership of the head at age 12	Own	562	55.0	553	86.5
	Rent	283	27.7	7	1.1
	Other	34	3.3	425	66.5
Occupation of the head at age 12	Not applicable	119	11.6	115	18.0
	Professional, Technical, and Kindred Workers	154	15.1	20	3.1
	Managers and Administrators, except Farm	118	11.5	36	5.6

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
	Clerical and Kindred Workers	75	7.3	135	21.1
	Craftsman and Kindred Workers	124	12.1	105	16.4
	Operatives, except Transport	69	6.8	40	6.3
	Transport Equipment Operatives	64	6.3	90	14.1
	Service Workers, except Private Household	73	7.1	37	5.8
	Other	81	7.9	34	5.3
Union membership of the head at age 12	Not applicable	217	21.2	30	4.7
	Yes	157	15.4	53	8.3
	No	495	48.4	117	18.3
Occupation of the spouse at age 12	Not applicable	402	39.3	96	15.0
	Professional, Technical, and Kindred Workers	126	12.3	342	53.5
	Managers and Administrators, except Farm	46	4.5	199	31.1
	Clerical and Kindred Workers	138	13.5	104	16.3
	Service Workers, except Private Household	79	7.7	37	5.8
	Other	87	8.5	109	17.1
Other dependent for the head at age 12	Yes	79	7.7	54	8.5
	No	787	77.0	56	8.8
Area where the head grew up at age 12	Rural	179	17.5	48	7.5
	Town	305	29.8	503	78.7
	City	360	35.2	121	18.9
	Other	20	2.0	223	34.9
State where the head grew up at age 12	MI	51	5.0	192	30.0
	MS	65	6.4	16	2.5
	NY	51	5.0	29	4.5
	NC	44	4.3	10	1.6
	OH	49	4.8	40	6.3
	Other	604	59.1	12	1.9
Head's income class while growing up at age 12	Poor	289	28.3	36	5.6
	Average	332	32.5	426	66.7
	Rich	205	20.1	126	19.7
Region where the family lived at age 12	Northeast	114	11.2	263	41.2
	NorthCentral	267	26.1	140	21.9
	South	378	37.0	93	14.6
	West	115	11.3	197	30.8
	Other	5	0.5	168	26.3
Region where the head grew up at age 12	Northeast	128	12.5	98	15.3
	NorthCentral	275	26.9	4	0.6
	South	370	36.2	106	16.6
	West	87	8.5	208	32.6
	Other	4	0.4	161	25.2
Did head live at different state from childhood at age 12	SameState	640	62.6	74	11.6

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Percent	Obs	Percent
	SameReg	95	9.3	4	0.6
	DiffReg	129	12.6	386	60.4
Race of the head at age 12	White	497	48.6	68	10.6
	Black	373	36.5	99	15.5
	Other	7	0.7	493	77.2
Union membership of the spouse at age 12	Not applicable	453	44.3	62	9.7
	Yes	64	6.3	5	0.8
	No	358	35.0	242	37.9
Medicare/Medicaid usage in the family at age 12	Yes	134	13.1	45	7.0
	No	741	72.5	269	42.1
	Other	0	0.0	48	7.5
Usage of food stamps by the head at age 12	no	727	71.1	509	79.7
	yes	152	14.9	0	0.0
Race of the spouse at age 12	Not applicable	256	25.0	519	81.2
	White	429	42.0	41	6.4
	Black	183	17.9	105	16.4
	Other	3	0.3	425	66.5
State where the family lived at age 13	CA	50	4.9	23	3.6
	MI	52	5.1	2	0.3
	MS	59	5.8	36	5.6
	NC	46	4.5	28	4.4
	TX	58	5.7	9	1.4
	Other	646	63.2	13	2.0
Sex of the family head at age 13	Male	646	63.2	36	5.6
	Female	266	26.0	458	71.7
Type of parent at age 13	Father	636	62.2	469	73.4
	Mother	251	24.6	112	17.5
	Other	25	2.4	462	72.3
Type of the family head at age 13	PH	887	86.8	109	17.1
	Other	25	2.4	10	1.6
Homeownership of the head at age 13	Own	596	58.3	571	89.4
	Rent	291	28.5	10	1.6
	Other	25	2.4	447	70.0
Occupation of the head at age 13	Not applicable	139	13.6	119	18.6
	Professional, Technical, and Kindred Workers	144	14.1	15	2.3
	Managers and Administrators, except Farm	128	12.5	50	7.8
	Clerical and Kindred Workers	84	8.2	126	19.7
	Craftsman and Kindred Workers	126	12.3	113	17.7
	Operatives, except Transport	70	6.8	47	7.4
	Transport Equipment Operatives	58	5.7	88	13.8
	Service Workers, except Private Household	75	7.3	34	5.3
	Other	88	8.6	30	4.7
Union membership of the head at age 13	Not applicable	241	23.6	32	5.0

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Percent	Obs	Percent
	Yes	164	16.0	61	9.5
	No	492	48.1	140	21.9
Occupation of the spouse at age 13	Not applicable	430	42.1	98	15.3
	Professional, Technical, and Kindred Workers	135	13.2	338	52.9
	Clerical and Kindred Workers	139	13.6	223	34.9
	Service Workers, except Private Household	78	7.6	113	17.7
	Other	130	12.7	108	16.9
Other dependent for the head at age 13	Yes	92	9.0	51	8.0
	No	819	80.1	86	13.5
Area where the head grew up at age 13	Rural	178	17.4	58	9.1
	Town	320	31.3	523	81.8
	City	378	37.0	122	19.1
	Other	18	1.8	232	36.3
State where the head grew up at age 13	MI	52	5.1	203	31.8
	MS	64	6.3	15	2.3
	NY	52	5.1	30	4.7
	NC	45	4.4	10	1.6
	OH	50	4.9	42	6.6
	SC	45	4.4	12	1.9
	Other	585	57.2	36	5.6
Head’s income class while growing up at age 13	Poor	296	29.0	9	1.4
	Average	347	34.0	433	67.8
	Rich	214	20.9	128	20.0
Region where the family lived at age 13	Northeast	117	11.4	275	43.0
	NorthCentral	276	27.0	146	22.8
	South	394	38.6	96	15.0
	West	120	11.7	205	32.1
	Other	5	0.5	173	27.1
Region where the head grew up at age 13	Northeast	131	12.8	103	16.1
	NorthCentral	288	28.2	4	0.6
	South	379	37.1	110	17.2
	West	91	8.9	218	34.1
	Other	4	0.4	161	25.2
Did head live at different state from childhood at age 13	SameState	658	64.4	79	12.4
	SameReg	98	9.6	4	0.6
	DiffReg	137	13.4	397	62.1
Race of the head at age 13	White	512	50.1	70	11.0
	Black	386	37.8	105	16.4
	Other	8	0.8	508	79.5
Union membership of the spouse at age 13	Not applicable	475	46.5	64	10.0
	Yes	79	7.7	7	1.1
	No	355	34.7	261	40.8
Medicare/Medicaid usage in the family at age 13	Wildcode	121	11.8	59	9.2
	Yes	133	13.0	259	40.5
	No	658	64.4	87	13.6
Usage of food stamps by the head at age 13	no	755	73.9	36	5.6
	yes	154	15.1	458	71.7
Race of the spouse at age 13	Not applicable	289	28.3	536	83.9
	White	428	41.9	44	6.9

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
	Black	182	17.8	127	19.9
	Other	4	0.4	424	66.4
State where the family lived at age 14	CA	49	4.8	23	3.6
	MI	55	5.4	3	0.5
	MS	60	5.9	36	5.6
	NC	48	4.7	29	4.5
	TX	62	6.1	10	1.6
	Other	649	63.5	13	2.0
Sex of the family head at age 14	Male	645	63.1	39	6.1
	Female	280	27.4	460	72.0
Type of parent at age 14	Father	635	62.1	465	72.8
	Mother	263	25.7	124	19.4
	Other	27	2.6	460	72.0
Type of the family head at age 14	PH	898	87.9	119	18.6
	Other	27	2.6	10	1.6
Homeownership of the head at age 14	Own	622	60.9	579	90.6
	Rent	279	27.3	10	1.6
	Other	24	2.3	464	72.6
Occupation of the head at age 14	Not applicable	125	12.2	114	17.8
	Professional, Technical, and Kindred Workers	166	16.2	11	1.7
	Managers and Administrators, except Farm	125	12.2	45	7.0
	Clerical and Kindred Workers	93	9.1	139	21.8
	Craftsman and Kindred Workers	130	12.7	105	16.4
	Operatives, except Transport	60	5.9	55	8.6
	Transport Equipment Operatives	63	6.2	89	13.9
	Service Workers, except Private Household	75	7.3	32	5.0
	Other	87	8.5	30	4.7
Union membership of the head at age 14	Not applicable	234	22.9	33	5.2
	Yes	154	15.1	61	9.5
	No	533	52.2	141	22.1
Occupation of the spouse at age 14	Not applicable	438	42.9	93	14.6
	Professional, Technical, and Kindred Workers	140	13.7	353	55.2
	Managers and Administrators, except Farm	53	5.2	219	34.3
	Clerical and Kindred Workers	147	14.4	116	18.2
	Service Workers, except Private Household	70	6.8	43	6.7
	Other	76	7.4	113	17.7
Other dependent for the head at age 14	Yes	106	10.4	47	7.4

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per- cent	Obs	Percent
	No	818	80.0	50	7.8
Area where the head grew up at age 14	Rural	181	17.7	66	10.3
	Town	330	32.3	523	81.8
	City	374	36.6	123	19.2
	Other	19	1.9	240	37.6
State where the head grew up at age 14	MI	41	4.0	200	31.3
	MS	53	5.2	16	2.5
	NY	47	4.6	23	3.6
	OH	40	3.9	9	1.4
	SC	41	4.0	40	6.3
	Other	505	49.4	29	4.5
Head's income class while growing up at age 14	Poor	301	29.5	9	1.4
	Average	354	34.6	362	56.7
	Rich	212	20.7	133	20.8
Region where the family lived at age 14	Northeast	114	11.2	279	43.7
	NorthCentral	273	26.7	144	22.5
	South	409	40.0	95	14.9
	West	124	12.1	204	31.9
	Other	5	0.5	179	28.0
Region where the head grew up at age 14	Northeast	129	12.6	107	16.7
	NorthCentral	288	28.2	4	0.6
	South	390	38.2	109	17.1
	West	89	8.7	221	34.6
	Other	5	0.5	164	25.7
Did head live at different state from childhood at age 14	SameState	663	64.9	78	12.2
	SameReg	99	9.7	5	0.8
	DiffReg	139	13.6	397	62.1
Race of the head at age 14	White	515	50.4	74	11.6
	Black	387	37.9	106	16.6
	Other	12	1.2	512	80.1
Union membership of the spouse at age 14	Not applicable	485	47.5	64	10.0
	Yes	70	6.8	10	1.6
	No	365	35.7	262	41.0
Medicare/Medicaid usage in the family at age 14	Wildcode	89	8.7	48	7.5
	Yes	87	8.5	275	43.0
	No	570	55.8	57	8.9
Usage of food stamps by the head at age 14	no	791	77.4	36	5.6
	yes	133	13.0	390	61.0
Race of the spouse at age 14	Not applicable	300	29.4	548	85.8
	White	427	41.8	41	6.4
	Black	184	18.0	139	21.8
	Other	6	0.6	423	66.2
State where the family lived at age 15	CA	40	3.9	20	3.1
	MI	42	4.1	5	0.8
	MS	48	4.7	31	4.9
	NC	41	4.0	22	3.4
	SC	42	4.1	10	1.6
	TX	55	5.4	13	2.0
	Other	506	49.5	9	1.4
Sex of the family head at age 15	Male	556	54.4	32	5.0
	Female	222	21.7	383	59.9
Type of parent at age 15	Father	545	53.3	398	62.3
	Mother	211	20.6	106	16.6

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
	Other	22	2.2	394	61.7
Type of the family head at age 15	PH	756	74.0	102	16.0
	Other	22	2.2	8	1.3
Homeownership of the head at age 15	Own	524	51.3	496	77.6
	Rent	234	22.9	8	1.3
	Other	20	2.0	391	61.2
Occupation of the head at age 15	Not applicable	106	10.4	101	15.8
	Professional, Technical, and Kindred Workers	126	12.3	12	1.9
	Managers and Administrators, except Farm	122	11.9	42	6.6
	Clerical and Kindred Workers	78	7.6	103	16.1
	Craftsman and Kindred Workers	115	11.3	107	16.7
	Operatives, except Transport	46	4.5	50	7.8
	Transport Equipment Operatives	44	4.3	82	12.8
	Service Workers, except Private Household	65	6.4	19	3.0
	Other	74	7.2	22	3.4
Union membership of the head at age 15	Not applicable	203	19.9	23	3.6
	Yes	123	12.0	55	8.6
	No	447	43.7	130	20.3
Occupation of the spouse at age 15	Not applicable	350	34.2	75	11.7
	Professional, Technical, and Kindred Workers	111	10.9	297	46.5
	Managers and Administrators, except Farm	51	5.0	183	28.6
	Clerical and Kindred Workers	133	13.0	93	14.6
	Service Workers, except Private Household	65	6.4	43	6.7
	Other	66	6.5	102	16.0
Other dependent for the head at age 15	Yes	87	8.5	41	6.4
	No	689	67.4	42	6.6
Area where the head grew up at age 15	Rural	150	14.7	61	9.5
	Town	283	27.7	443	69.3
	City	303	29.6	102	16.0
	Other	18	1.8	202	31.6
State where the head grew up at age 15	MI	33	3.2	172	26.9
	MS	34	3.3	16	2.5
	NY	41	4.0	18	2.8
	OH	37	3.6	7	1.1
	SC	37	3.6	33	5.2
	Other	441	43.2	27	4.2
Head's income class while growing up at age 15	Poor	242	23.7	9	1.4

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
	Average	298	29.2	319	49.9
	Rich	181	17.7	106	16.6
Region where the family lived at age 15	Northeast	102	10.0	236	36.9
	NorthCentral	217	21.2	130	20.3
	South	347	34.0	86	13.5
	West	105	10.3	166	26.0
	Other	7	0.7	153	23.9
Region where the head grew up at age 15	Northeast	113	11.1	93	14.6
	NorthCentral	234	22.9	6	0.9
	South	329	32.2	96	15.0
	West	73	7.1	184	28.8
	Other	5	0.5	139	21.8
Did head live at different state from childhood at age 15	SameState	547	53.5	68	10.6
	SameReg	83	8.1	4	0.6
	DiffReg	124	12.1	335	52.4
Race of the head at age 15	White	444	43.4	60	9.4
	Black	314	30.7	96	15.0
	Other	10	1.0	439	68.7
Union membership of the spouse at age 15	Not applicable	396	38.7	53	8.3
	Yes	71	6.9	7	1.1
	No	308	30.1	221	34.6
Medicare/Medicaid usage in the family at age 15	Wildcode	126	12.3	47	7.4
	Yes	92	9.0	235	36.8
	No	423	41.4	95	14.9
Usage of food stamps by the head at age 15	no	683	66.8	32	5.0
	yes	94	9.2	298	46.6
Race of the spouse at age 15	Not applicable	249	24.4	477	74.6
	White	360	35.2	27	4.2
	Black	157	15.4	123	19.2
	Other	7	0.7	356	55.7
State where the family lived at age 16	CA	46	4.5	18	2.8
	MI	51	5.0	5	0.8
	MS	44	4.3	37	5.8
	NC	45	4.4	24	3.8
	TX	56	5.5	8	1.3
	Other	585	57.2	13	2.0
Sex of the family head at age 16	Male	577	56.5	34	5.3
	Female	252	24.7	417	65.3
Type of parent at age 16	Father	568	55.6	420	65.7
	Mother	238	23.3	115	18.0
	Other	23	2.3	417	65.3
Type of the family head at age 16	PH	806	78.9	111	17.4
	Other	23	2.3	7	1.1
Homeownership of the head at age 16	Own	572	56.0	528	82.6
	Rent	238	23.3	7	1.1
	Other	19	1.9	420	65.7
Occupation of the head at age 16	Not applicable	113	11.1	104	16.3
	Professional, Technical, and Kindred Workers	137	13.4	11	1.7
	Managers and Administrators, except Farm	126	12.3	49	7.7

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
	Clerical and Kindred Workers	74	7.2	106	16.6
	Craftsman and Kindred Workers	115	11.3	110	17.2
	Operatives, except Transport	53	5.2	45	7.0
	Transport Equipment Operatives	53	5.2	74	11.6
	Service Workers, except Private Household	76	7.4	27	4.2
	Other	80	7.8	34	5.3
Union membership of the head at age 16	Not applicable	215	21.0	35	5.5
	Yes	145	14.2	55	8.6
	No	464	45.4	137	21.4
Occupation of the spouse at age 16	Not applicable	371	36.3	84	13.1
	Professional, Technical, and Kindred Workers	133	13.0	310	48.5
	Managers and Administrators, except Farm	55	5.4	194	30.4
	Clerical and Kindred Workers	132	12.9	109	17.1
	Service Workers, except Private Household	76	7.4	43	6.7
	Other	62	6.1	102	16.0
Other dependent for the head at age 16	Yes	98	9.6	46	7.2
	No	730	71.4	41	6.4
Area where the head grew up at age 16	Rural	166	16.2	73	11.4
	Town	306	29.9	461	72.1
	City	318	31.1	113	17.7
	Other	17	1.7	218	34.1
Head’s income class while growing up at age 16	Poor	273	26.7	176	27.5
	Average	318	31.1	15	2.3
	Rich	180	17.6	120	18.8
Region where the family lived at age 16	Northeast	102	10.0	247	38.7
	NorthCentral	248	24.3	132	20.7
	South	355	34.7	84	13.1
	West	116	11.4	186	29.1
	Other	7	0.7	154	24.1
Region where the head grew up at age 16	Northeast	117	11.4	105	16.4
	NorthCentral	258	25.2	6	0.9
	South	348	34.1	96	15.0
	West	80	7.8	201	31.5
	Other	4	0.4	148	23.2
Did head live at different state from childhood at age 16	SameState	594	58.1	73	11.4
	SameReg	84	8.2	4	0.6
	DiffReg	129	12.6	367	57.4
Race of the head at age 16	White	459	44.9	58	9.1
	Black	343	33.6	97	15.2
	Other	17	1.7	455	71.2
Union membership of the spouse at age 16	Not applicable	410	40.1	63	9.9

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
	Yes	88	8.6	13	2.0
	No	329	32.2	229	35.8
Usage of food stamps by the head at age 16	no	725	70.9	60	9.4
	yes	102	10.0	244	38.2
Race of the spouse at age 16	Not applicable	272	26.6	502	78.6
	White	382	37.4	32	5.0
	Black	160	15.7	128	20.0
	Other	8	0.8	379	59.3
State where the family lived at age 17	MI	40	3.9	18	2.8
	MS	41	4.0	6	0.9
	NY	31	3.0	18	2.8
	NC	33	3.2	9	1.4
	TX	38	3.7	25	3.9
	Other	424	41.5	9	1.4
Sex of the family head at age 17	Male	431	42.2	22	3.4
	Female	176	17.2	297	46.5
Type of parent at age 17	Father	420	41.1	300	46.9
	Mother	166	16.2	80	12.5
	Other	21	2.1	296	46.3
Type of the family head at age 17	PH	586	57.3	75	11.7
	Other	21	2.1	9	1.4
Homeownership of the head at age 17	Own	424	41.5	371	58.1
	Rent	171	16.7	9	1.4
	Other	12	1.2	307	48.0
Occupation of the head at age 17	Not applicable	75	7.3	66	10.3
	Professional, Technical, and Kindred Workers	89	8.7	7	1.1
	Managers and Administrators, except Farm	91	8.9	22	3.4
	Clerical and Kindred Workers	52	5.1	73	11.4
	Craftsman and Kindred Workers	93	9.1	76	11.9
	Operatives, except Transport	33	3.2	36	5.6
	Transport Equipment Operatives	50	4.9	63	9.9
	Service Workers, except Private Household	59	5.8	14	2.2
	Other	61	6.0	25	3.9
Union membership of the head at age 17	Not applicable	159	15.6	22	3.4
	Yes	117	11.4	47	7.4
	No	327	32.0	95	14.9
Occupation of the spouse at age 17	Not applicable	271	26.5	69	10.8
	Professional, Technical, and Kindred Workers	92	9.0	215	33.6
	Managers and Administrators, except Farm	39	3.8	134	21.0

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per- cent	Obs	Percent
	Clerical and Kindred Workers	105	10.3	69	10.8
	Service Workers, except Private Household	56	5.5	35	5.5
	Other	44	4.3	78	12.2
Other dependent for the head at age 17	Yes	65	6.4	35	5.5
	No	542	53.0	29	4.5
Area where the head grew up at age 17	Rural	125	12.2	38	5.9
	Town	212	20.7	342	53.5
	City	235	23.0	87	13.6
	Other	14	1.4	147	23.0
Head's income class while growing up at age 17	Poor	200	19.6	125	19.6
	Average	223	21.8	13	2.0
	Rich	136	13.3	92	14.4
Region where the family lived at age 17	Northeast	78	7.6	171	26.8
	NorthCentral	178	17.4	95	14.9
	South	264	25.8	65	10.2
	West	81	7.9	126	19.7
	Other	6	0.6	115	18.0
Region where the head grew up at age 17	Northeast	88	8.6	69	10.8
	NorthCentral	183	17.9	5	0.8
	South	254	24.9	74	11.6
	West	59	5.8	136	21.3
	Other	3	0.3	105	16.4
Did head live at different state from childhood at age 17	SameState	433	42.4	54	8.5
	SameReg	64	6.3	3	0.5
	DiffReg	90	8.8	261	40.8
Race of the head at age 17	White	330	32.3	44	6.9
	Black	258	25.2	67	10.5
	Other	12	1.2	325	50.9
Union membership of the spouse at age 17	Not applicable	306	29.9	44	6.9
	Yes	56	5.5	9	1.4
	No	245	24.0	162	25.4
Usage of food stamps by the head at age 17	no	454	44.4	34	5.3
	yes	105	10.3	184	28.8
Race of the spouse at age 17	Not applicable	195	19.1	313	49.0
	White	268	26.2	38	5.9
	Black	128	12.5	91	14.2
	Other	7	0.7	263	41.2
State where the family lived at age 18	CA	40	3.9	16	2.5
	MI	39	3.8	5	0.8
	MS	34	3.3	33	5.2
	NC	40	3.9	20	3.1
	TX	48	4.7	5	0.8
	Other	463	45.3	11	1.7
Sex of the family head at age 18	Male	484	47.4	29	4.5
	Female	180	17.6	329	51.5
Type of parent at age 18	Father	463	45.3	353	55.2
	Mother	165	16.1	74	11.6
	NP	36	3.5	341	53.4
Type of the family head at age 18	H	36	3.5	68	10.6
	PH	628	61.4	18	2.8
Homeownership of the head at age 18	Own	479	46.9	18	2.8

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
	Rent	169	16.5	409	64.0
	Other	16	1.6	349	54.6
Occupation of the head at age 18	Not applicable	96	9.4	68	10.6
	Professional, Technical, and Kindred Workers	104	10.2	10	1.6
	Managers and Administrators, except Farm	94	9.2	38	5.9
	Clerical and Kindred Workers	60	5.9	85	13.3
	Craftsman and Kindred Workers	103	10.1	83	13.0
	Operatives, except Transport	39	3.8	32	5.0
	Transport Equipment Operatives	43	4.2	71	11.1
	Service Workers, except Private Household	59	5.8	17	2.7
	Other	64	6.3	23	3.6
Union membership of the head at age 18	Not applicable	170	16.6	26	4.1
	Yes	112	11.0	51	8.0
	No	376	36.8	102	16.0
Occupation of the spouse at age 18	Not applicable	297	29.1	68	10.6
	Professional, Technical, and Kindred Workers	98	9.6	253	39.6
	Managers and Administrators, except Farm	51	5.0	152	23.8
	Clerical and Kindred Workers	100	9.8	79	12.4
	Service Workers, except Private Household	57	5.6	44	6.9
	Other	56	5.5	81	12.7
Other dependent for the head at age 18	Yes	82	8.0	31	4.9
	No	582	56.9	38	5.9
Area where the head grew up at age 18	Rural	126	12.3	50	7.8
	Town	242	23.7	377	59.0
	City	262	25.6	85	13.3
	Other	13	1.3	173	27.1
Head’s income class while growing up at age 18	Poor	219	21.4	147	23.0
	Average	254	24.9	11	1.7
	Rich	134	13.1	97	15.2
Region where the family lived at age 18	Northeast	82	8.0	193	30.2
	NorthCentral	189	18.5	103	16.1
	South	295	28.9	66	10.3
	West	94	9.2	145	22.7
	Other	4	0.4	127	19.9
Region where the head grew up at age 18	Northeast	97	9.5	85	13.3
	NorthCentral	200	19.6	4	0.6
	South	280	27.4	77	12.1

Table 9: Summary Statistics (Categorical Variables) *(continued)*

Circumstance	Category	Obs	Per-cent	Obs	Percent
	West	60	5.9	157	24.6
	Other	5	0.5	121	18.9
Did head live at different state from childhood at age 18	SameState	472	46.2	55	8.6
	SameReg	72	7.0	5	0.8
	DiffReg	98	9.6	290	45.4
Race of the head at age 18	White	360	35.2	50	7.8
	Black	274	26.8	75	11.7
	Other	19	1.9	356	55.7
Union membership of the spouse at age 18	Not applicable	326	31.9	50	7.8
	Yes	71	6.9	14	2.2
	No	266	26.0	177	27.7
Usage of food stamps by the head at age 18	no	430	42.1	51	8.0
	yes	129	12.6	199	31.1
Race of the spouse at age 18	Not applicable	201	19.7	300	46.9
	White	307	30.0	46	7.2
	Black	134	13.1	89	13.9
	Other	13	1.3	304	47.6
				20	3.1
				8	1.3

7 References

Alesina, Alberto, and Paola Giuliano. 2011. "Chapter 4 - Preferences for Redistribution." In *Handbook of Social Economics*, edited by Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, 1:93–131. North-Holland. https://doi.org/10.1016/B978-0-444-53187-2.00004-8.

Almlund, Mathilde, Angela Lee Duckworth, James Heckman, and Tim Kautz. 2011. "Personality Psychology and Economics." In *Handbook of the Economics of Education*, 4:1–181. Elsevier. https://doi.org/10.1016/B978-0-444-53444-6.00001-8.

Bond, Timothy N., Jillian B. Carr, Analisa Packham, and Jonathan Smith. 2022. "Hungry for Success? SNAP Timing, High-Stakes Exam Performance, and College Attendance." *American Economic Journal: Economic Policy* 14 (4): 51–79. https://doi.org/10.1257/pol.20210026.

Borghans, Lex, Angela Duckworth, James Heckman, and Bas Weel. 2008. "The Economics and Psychology of Personal Traits." *The Journal of Human Resources* 43 (February). https://doi.org/10.1353/jhr.2008.0017.

Bourguignon, François, Francisco H. G. Ferreira, and Marta Menéndez. 2007. "Inequality of Opportunity in Brazil." *Review of Income and Wealth* 53 (4): 585–618. https://doi.org/10.1111/j.1475-4991.2007.00247.x.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.

Brunori, Paolo, Francisco H. G. Ferreira, and Pedro Salas-Rojo. n.d. "Inherited Inequality: A General Framework and an Application to South Africa." *Working Papers*. Accessed July 30, 2024. https://ideas.repec.org//p/inq/inqwps/ecineq2023-658.html.

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. https://doi.org/10.1111/sjoe.12530.

Carneiro, Pedro Manuel, and James J. Heckman. 2003. "Human Capital Policy." {SSRN} {Scholarly} {Paper}. Rochester, NY. https://doi.org/10.2139/ssrn.434544.

Carneiro, Pedro, Italo López García, Kjell G. Salvanes, and Emma Tominey. 2021. "Intergenerational Mobility and the Timing of Parental Income." *Journal of Political Economy*, March. https://doi.org/10.1086/712443.

Checchi, Daniele, and Vito Peragine. 2010. "Inequality of Opportunity in Italy." *J Econ Inequal* 8 (4): 429–50. https://doi.org/10.1007/s10888-009-9118-3.

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." *Quarterly Journal of Economics* 129 (4): 1553–623.

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

Cowell, Frank. 2016. "Inequality and Poverty Measures." *Oxford Handbooks Online*, January. https://www.academia.edu/69017672/Inequality_and_Poverty_Measures.

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2): 31–47. https://doi.org/10.1257/aer.97.2.31.

Cunha, Flavio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov. 2006. "Chapter 12 Interpreting the Evidence on Life Cycle Skill Formation." In *Handbook of the Economics of Education*, 1:697–812. Elsevier. https://doi.org/10.1016/S1574-0692(06)01012-9.

Cunha, Flavio, James J. Heckman, and Salvador Navarro-Lozano. 2004. "Separating Uncertainty from Heterogeneity in Life Cycle Earnings." {SSRN} {Scholarly} {Paper}. Rochester, NY. https://doi.org/10.2139/ssrn.643622.

Deutscher, Nathan, and Bhashkar Mazumder. 2023. "Measuring Intergenerational Income Mobility: A Synthesis of Approaches." *Journal of Economic Literature* 61 (3): 988–1036. https://doi.org/10.1257/jel.20211413.

Donni, Paolo Li, Juan Gabriel Rodríguez, and Pedro Rosa Dias. 2015. "Empirical Definition of Social Types in the Analysis of Inequality of Opportunity: A Latent Classes Approach." *Social Choice and Welfare* 44 (3): 673–701. https://www.jstor.org/stable/43662611.

Eskelson, Bianca, Hailemariam Temesgen, Valerie Lemay, Tara Barrett, Nicholas Crookston, and Andrew Hudak. 2009. "The Roles of Nearest Neighbor Methods in Imputing Missing Data in Forest Inventory and Monitoring Databases." *United States Department of Agriculture, Forest Service / University of Nebraska-Lincoln: Faculty Publications*, January. https://digitalcommons.unl.edu/usdafsfacpub/217.

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): 622–57. https://doi.org/10.1111/j.1475-4991.2011.00467.x.

Ferreira, Francisco H. G., and Vito Peragine. 2015. "Equality of Opportunity: Theory and Evidence." {SSRN} {Scholarly} {Paper}. Rochester, NY. https://papers.ssrn.com/abstract=2584375.

Fisher, Aaron, Cynthia Rudin, and Francesca Dominici. 2019. "All Models Are Wrong, but Many Are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously." *J Mach Learn Res* 20: 177. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8323609/.

Fleurbaey, Marc, and Vito Peragine. 2013. "Ex Ante Versus Ex Post Equality of Opportunity." *Economica* 80 (317): 118–30. https://doi.org/10.1111/j.1468-0335.2012.00941.x.

Fong, Christina. 2001. "Social Preferences, Self-Interest, and the Demand for Redistribution." *Journal of Public Economics* 82 (2): 225–46. https://doi.org/10.1016/S0047-2727(00)00141-9.

Gower, J. C. 1971. "A General Coefficient of Similarity and Some of Its Properties." *Biometrics* 27 (4): 857–71. https://doi.org/10.2307/2528823.

Hart, Betty, and Todd R. Risley. 1995. *Meaningful Differences in the Everyday Experience of Young American Children*. Brookes Publishing Company, Inc.

Heckman, James J., and Stefano Mosso. 2014. "The Economics of Human Development and Social Mobility." *Annu. Rev. Econ.* 6 (1): 689–733. https://doi.org/10.1146/annurev-economics-080213-040753.

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." {SSRN} {Scholarly} {Paper}. Rochester, NY. https://papers.ssrn.com/abstract=881240.

Hufe, Paul, Andreas Peichl, John Roemer, and Martin Ungerer. 2017. "Inequality of Income Acquisition: The Role of Childhood Circumstances." *Soc Choice Welf* 49 (3-4): 499–544. https://doi.org/10.1007/s00355-017-1044-x.

Kalil, Ariel. 2015. "Inequality Begins at Home: The Role of Parenting in the Diverging Destinies of Rich and Poor Children." In *Families in an Era of Increasing Inequality: Diverging Destinies*, 63–82. National Symposium on Family Issues. Cham, Switzerland: Springer International Publishing/Springer Nature. https://doi.org/10.1007/978-3-319-08308-7_5.

Lareau, Annette. 2011. "Unequal Childhoods: Class, Race, and Family Life." In *Unequal Childhoods*. University of California Press. https://doi.org/10.1525/9780520949904.

Mackintosh, Nicholas. 2011. *IQ and Human Intelligence*. Second Edition, Second Edition. Oxford, New York: Oxford University Press.

Mazumder, Bhashkar. 2005. "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data." *The Review of Economics and Statistics* 87 (2): 235–55. https://doi.org/10.1162/0034653053970249.

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31 (2): 87–106. https://doi.org/10.1257/jep.31.2.87.

Neal, Derek A., and William R. Johnson. 1996. "The Role of Premarket Factors in Black-White Wage Differences." *Journal of Political Economy* 104 (5): 869–95. https://doi.org/10.1086/262045.

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. https://www.jstor.org/stable/43662521.

Nybom, Martin, and Jan Stuhler. 2017. "Biases in Standard Measures of Intergenerational Income

Dependence." *Journal of Human Resources* 52 (3): 800–825.

Oshiro, Thais Mayumi, Pedro Santoro Perez, and José Augusto Baranauskas. 2012. "How Many Trees in a Random Forest?" In *Machine Learning and Data Mining in Pattern Recognition*, edited by David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, et al., 7376:154–68. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-31537-4_13.

Pistolesi, Nicolas. 2009. "Inequality of Opportunity in the Land of Opportunities, 1968–2001." *J Econ Inequal* 7 (4): 411–33. https://doi.org/10.1007/s10888-008-9099-7.

Ramos, Xavier, and Dirk Van de gaer. 2016. "Approaches to Inequality of Opportunity: Principles, Measures and Evidence." *Journal of Economic Surveys* 30 (5): 855–83. https://doi.org/10.1111/joes.12121.

Rawls, John. 1971. *A Theory of Justice: Original Edition*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9z6v.

Roemer, John E. 1993. "A Pragmatic Theory of Responsibility for the Egalitarian Planner." *Philosophy & Public Affairs* 22 (2): 146–66. https://www.jstor.org/stable/2265444.

———. 2002. "Equality of Opportunity: A Progress Report." *Social Choice and Welfare* 19 (2): 455–71. https://www.jstor.org/stable/41106460.

Roemer, John E., and Alain Trannoy. 2016. "Equality of Opportunity: Theory and Measurement." *Journal of Economic Literature* 54 (4): 1288–1332. https://doi.org/10.1257/jel.20151206.

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

Starmans, Christina, Mark Sheskin, and Paul Bloom. 2017. "Why People Prefer Unequal Societies." *Nat Hum Behav* 1 (4): 1–7. https://doi.org/10.1038/s41562-017-0082.

Tutz, Gerhard, and Shahla Ramzan. 2015. "Improved Methods for the Imputation of Missing Data by Nearest Neighbor Methods." *Computational Statistics & Data Analysis* 90 (C): 84–99. https://ideas.repec.org//a/eee/csdana/v90y2015icp84-99.html.

Van de Gaer, Dirk. 1993. *Equality of Opportunity and Investment in Human Capital*. Reeks van de Faculteit Der Economische En Toegepaste Economische Wetenschappen, Katholieke Universiteit Te Leuven.