Measuring Income Inequality of Opportunity
Accounting for Dynamic Complementarity

Aman Desai*
Quinnipiac University

August 07, 2025

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 21%. 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.

*Department of Economics, Quinnipiac University, 275 Mt Carmel Ave, Hamden, CT 06518, United States. Website: https://amandesai.com/. Email: aman.desai@quinnipiac.edu. I am grateful to Miles Corak, Theodore Joyce, Wim Vijverberg, Lilia Maliar for their invaluable guidance and support. 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—I estimate the share of "unfair inequality" in total income inequality to be about 41% based on age cutoff at 18 years. Additionally, I perform the analysis using only 10 most important circumstances identified by the permutation-based variable importance scores to estimate the inequality of opportunity share in total income inequality to be about 21%, which is nearly half of estimated inequality of opportunity. Hence, I argue to have reached upper-bound estimates of the IOp share in adult income inequality. I also calculate the IOp estimates for Survey of Economic Opportunity sample from the PSID which includes disproportionately higher number of poor households. The estimated share of inequality of opportunity in total income inequality is about 23% for this sample. 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 $g : \Omega^c \times \Omega^e \implies R$. 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. 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 influence. Cunha and Heckman (2007) model technology for skill formation, where formulation

[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).

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 ran 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 Research 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. PSID also has a Survey Research Center sample which is more representative of the US population. My research sample includes individuals from both samples to ensure adequate sample size. I refer to this as the “full sample” hereafter. I also provide estimates of inequality of opportunity for individuals from the SEO sample focusing on poor households.

The individuals of interest are the heads of the family and their spouses/partners born in 1978-1983 who were present during any of the survey waves from 2005 to 2015. 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. 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 PSID-SHELF file to identify the family of the individual during their childhood and obtain data on family heads and their spouses (Fabian T. Pfeffer, Davis Daumler, and Esther Friedman 2025). 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. The data includes around 25 distinct circumstances on 1374 individuals in total. I create wide datasets according to based on age of the individual in childhood 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: 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	
Sex of the head	Homeownership	
Education of the family head, spouse	Marginal tax rate on family income	
Occupation of the family head, spouse	Availability of a car	
Number of children to father, mother		
Marital status of mother when individual was born		
Marital status of the family head		
Size of the family		
Region of residence of family		
Birthweight		
Birthcohort		

Ideally, one would have a complete biography of individuals spanning their first 18 years of life, which is difficult to obtain due to lack of data availability. My decision to choose 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 circumstances 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 which lags the survey wave by a year. Using these factors, I construct circumstance sets based on age cutoffs corresponding to critical stages in childhood. By focusing on the developmental stages, I 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 for which I use their labor income. Labor income measured at a single point in time is susceptible to measurement error and leads 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 when they were aged 27-33 while present in the survey waves from 2007-2019.[3] For simplicity, the labor income excludes farm and unincorporated business income. Additionally, I omit individuals with zero incomes from the analyses. For individuals with missing income data in any wave, I calculate their average income using only the available years between the age of 27-33. For example, if an individual has income data for age 27 and age 31 while missing data for 29, I use an average of incomes at 27 and 31. It is important to note that the PSID reports labor income from the tax year one year before the survey year. The question asks individuals to report their income from the

[3]While it would be ideal to analyze each cohort individually, doing so would lead to very small sample sizes, especially around the critical age of 30, which is widely recognized in the literature as a key predictor of long-term earnings potential (Chetty et al. 2014). Therefore, I chose to group data into six consecutive cohorts, which provides enough cases for robust analysis while reducing the influence of unusually high or low incomes found among younger individuals who are just starting their careers.

previous year. So, the incomes are measured with a lag, although majority of individual incomes are averaged over at least two years.[4] The average age at which individuals incomes are measured ranges from 28 to 30 years. To account for this variation I include birth cohort of an individual as a circumstance in my analyses. The adult labor incomes are measured in logarithms. Finally, all currency variables are reported in real U.S. dollars in 2024 using the Personal Consumption Expenditures Price Index (PCEPI) to adjust for inflation. The data is in wide format. Hence, I use appropriate individual cross-sectional weights from the survey waves in which individuals were last observed while aged between 27-33.

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.[5] 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

[4] a tiny percent of individuals have incomes from only one year.

[5] See Cowell (2016) for more information.

Gignoux 2011; Niehues and Peichl 2014).[6]

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

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 based on chosen age cutoffs. The smoothed distribution of $\hat{y}$ is then obtained using equation 1.6

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

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.

[6] 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.

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.

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

Equation 9 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.

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.

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

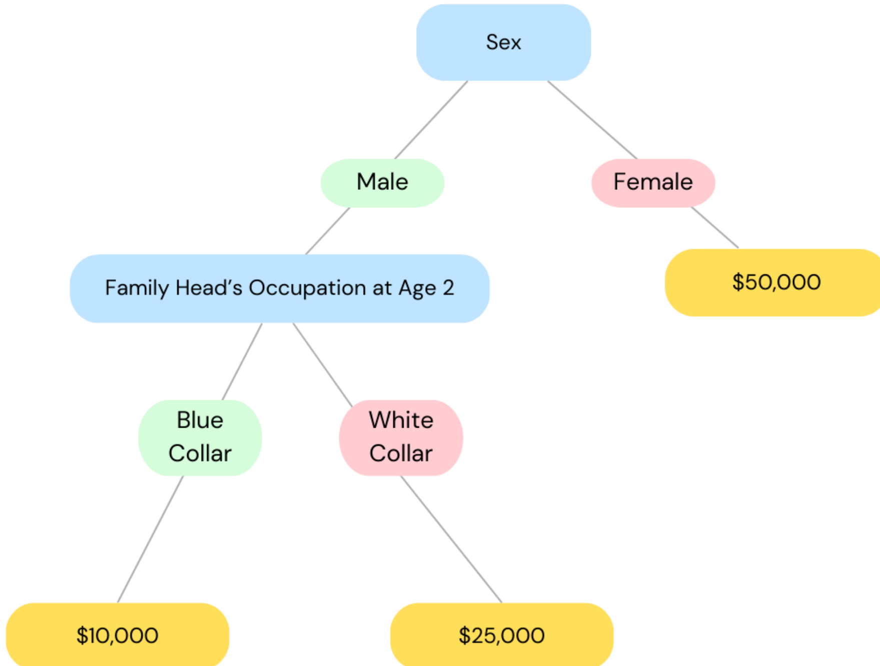

Figure 1: An example of a tree

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.[7]

3. To tune the cost complexity hyper parameter α, use the k-fold cross validation or bootstrap

[7] 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.

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.

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.

The data is in wide format with almost 175 circumstances as features for the biggest circumstance set for age cutoff at 18. The number of individuals in the data is 1374 and each row reflects an incomplete biography of an individual across the first 18 years of their life. I fit the models on training data, tune the hyperparameters on validation data using 5-fold cross-validation, and evaluate performance on test data. The goal is to calculate the shares of inequality of opportunity in total income inequality as shown in equation 9. To that extent, I run the final evaluated model (chosen based on the lowest rmse) on the full dataset and obtained a counterfactual distribution of adult incomes based on circumstances to obtain absolute and relative IOp estimates for all age cutoffs. The algorithm runs as follows:

- Initiate with splitting the data into training and test set, with $N_{train} = \frac{3}{4}N$ and $N_{test} = \frac{1}{4}N$.
- Execute an algorithm and use 5-fold cross validation for hyperparameter tuning. Select the model with hyperparameters that yield the lowest rmse during the cross-validation process.
- Store the prediction functions $\hat{f}_{train}(\hat{\Omega}^c)$.
- Obtain final predictions using the full data $\hat{y}_{EA} = \hat{f}_{train}(\hat{\Omega}^c_{fulldata})$.

This procedure is repeated for all circumstance sets in consideration based on cut-offs at age 2, 5, 14, and 18. I also report results for age cutoffs at 10,12, and 16 years to provide comparisons with the similar study by Hufe et al. (2017).

5 Results

5.1 Selected Descriptive Statistics

Table 2 presents weighted summary statistics for adult labor incomes and selected circumstances. Complete summary statistics for all circumstances are provided in table 6 and 7 in the appendix. The research sample consists of 1,374 individuals representing more than 14 million individuals in

Table 2: Selected weighted summary statistics

Circumstance	N = 14,074,440
Adult labor income in 2024 dollars (in natural logarithms)	10 (1.1)
Family income during the child's first year in 2024 dollars	72,572 (43,514.9)
Years of education of the head during the child's first year	13 (2.4)
Sex	
Male	6,979,487 (50%)
Female	7,094,953 (50%)
Race	
White	11,061,503 (79%)
Black	2,208,954 (16%)
Other race	159,772 (1.1%)
Hispanic	644,211 (4.6%)
Region of family residence during the child's first year	
Northeast	2,853,626 (20%)
Midwest	4,016,751 (29%)
South	4,869,462 (35%)
West	2,093,271 (15%)
Alaska or Hawaii	63,752 (0.5%)
Country outside the United States	177,578 (1.3%)

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

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. Monetary circumstances are reported from the previous tax year. For example, family income reported in the 1982 survey wave reflects data from the 1981 tax year.

the US with up to 175 circumstances in the largest circumstance set, which was created using the age 18 cutoff.

Average labor income is 10, measured at age 28-30 and in natural logarithms. Average family income in the previous tax year when the child was one year old is just over 70,000 in 2024 US dollars. Family heads have an average of 13 years of education in the sample, with a standard deviation of 2.4 years during the child's first year. The table also includes proportions of demographic circumstances such as sex and race, which reflect the US population distribution. Regarding region of residence during the child's first year, 35% of families lived in the South, followed by 29% in the Midwest, 20% in the Northeast, and 15% in the West.

Table 3: IOp estimates for different circumstance sets

	Outcome : adult labor income			
	Full sample (total inequality = 0.39)		SEO sample (total inequality = 0.372)	
	Absolute IOp	Relative IOp	Absolute IOp	Relative IOp
Baseline circumstances	0.09	0.23	0.06	0.15
Age cutoff at 2 years	0.08	0.21	0.04	0.10
Age cutoff at 5 years	0.12	0.30	0.05	0.13
Age cutoff at 14 years	0.14	0.36	0.08	0.21
Age cutoff at 18 years	0.16	0.41	0.08	0.23

5.2 IOp Estimates Across Critical Stages in Childhood

Table 3 presents estimates of absolute and relative income inequality of opportunity as well as total inequality (IO) in adult labor income. The adult labor income is measured at the age around 28-30. As a measure of inequality, mean logarithmic deviation is used.[8] The results are reported from analysis performed on both the full sample and the SEO sample. Total inequality in adult labor income in the full sample is 0.39, while it is 0.37 in the case of the Survey of Economic Opportunity sample.

I begin with a baseline model that uses an OLS specification with a fixed set of circumstances often used in the literature. These circumstances include family income, family head's education, and family's residence—all measured during the individual's first year. Also included are demographic characteristics such as race and sex of the individual. Using this simple specification, the absolute IOp calculated using equation 8 is 0.09. The estimated relative IOp is 0.23, meaning 23% of total income inequality is attributable to these circumstances used in the specification. For the Survey of Economic Opportunity sample, that estimate is 15%. In addition to the absolute and relative 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 generated by

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

the random forest model exceed those produced using the standard set of circumstances and OLS specification, with the sole exception of the circumstance set with age cutoff at 2. It is also apparent that these estimates based on age-based circumstance sets increase with age given that circumstance sets expand with age—consistent with the technology of skill formation during childhood. For the full sample, the relative IOp (share of inequality of opportunity in total income inequality) is 21% when accounting for circumstances up to age 2. This estimate rises to 30% with the age cutoff at 5 years. When considering circumstances up to the age of majority at 18, the relative IOp reaches 41%, demonstrating the growing influence of unequal circumstances throughout childhood on total income inequality. It is evident that relative IOp estimates increase at a decreasing rate with age, with about one-third of total income inequality attributable to unequal circumstances up to age 5. The numbers for these estimates in case of survey of economic opportunity sample are lower overall. The share of inequality of opportunity in total income inequality is 23% for the age cutoff at 18.

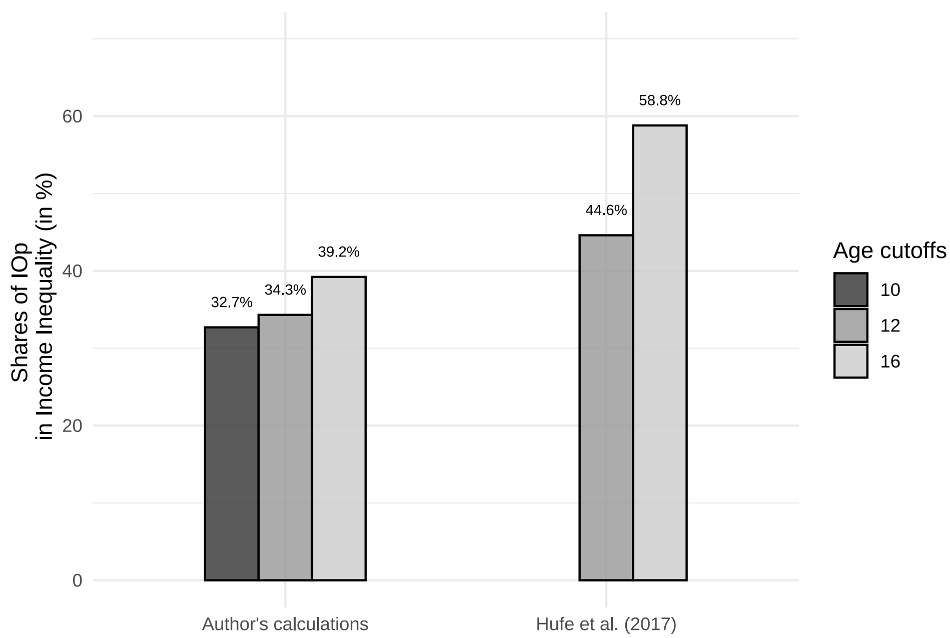

Figure 2: Comparison with Hufe et al. (2017)

I also estimate relative IOp using circumstances based on age cutoffs at 10, 12, and 16 years.

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 adult labor income as outcome, I obtained relative IOp of about 34% and 39% for age cut offs at 12 and 16 respectively as shown in figure 2.

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). The estimated share of inequality of opportunity in total income inequality using an age cutoff at age 10 is almost one-third. 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 labor 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 labor 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 $vi(x_1), vi(x_2), ..., vi(x_j)$.
3. Repeat steps 1 and 2 for 50 times to obtain mean variable importance and standard deviation for each predictor.

Figure 3 presents the top ten circumstances that were most important in predicting adult labor incomes. The figure displays important circumstances as identified using permutation-based variable importance scores as explained above. For example, sex of the individual and home ownership of the family head when the child is 2 years old are the top 2 important circumstances in cases where the age cutoff is set at 2 years, as evidenced by the graph in the top-left panel. Important predictors for other circumstance sets based on different age cutoffs are also shown in figure 3. Sex of the individual and usage of food stamps by the family head maintain high importance across all childhood stages defined in this study. 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 labor 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). Tree-based machine learning models can help identify these important circumstances that can be studied in detail to help with a causal analysis. I run the algorithm explained above for 50 iterations. The graphs in the figure display variable importance scores, with the bars representing one standard deviation above and below the mean importance scores.

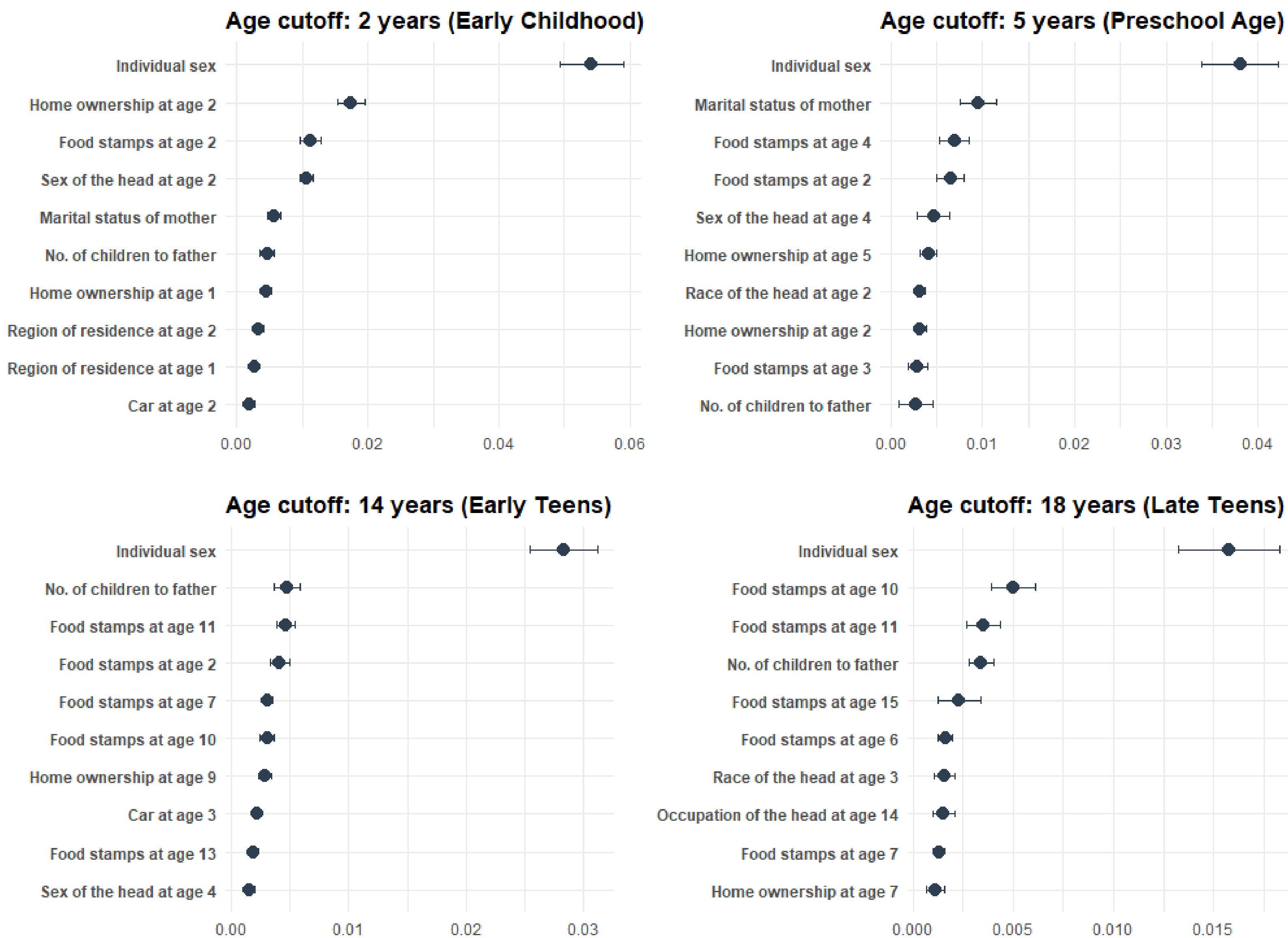

Figure 3: Important circumstances across different age cutoffs (age in the graph refers to the age of the individual in childhood)

As discussed earlier, one of the advantages of using supervised machine learning models 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. With that approach, I use these algorithm-chosen circumstances to redo all the analyses and generate relative IOp estimates.

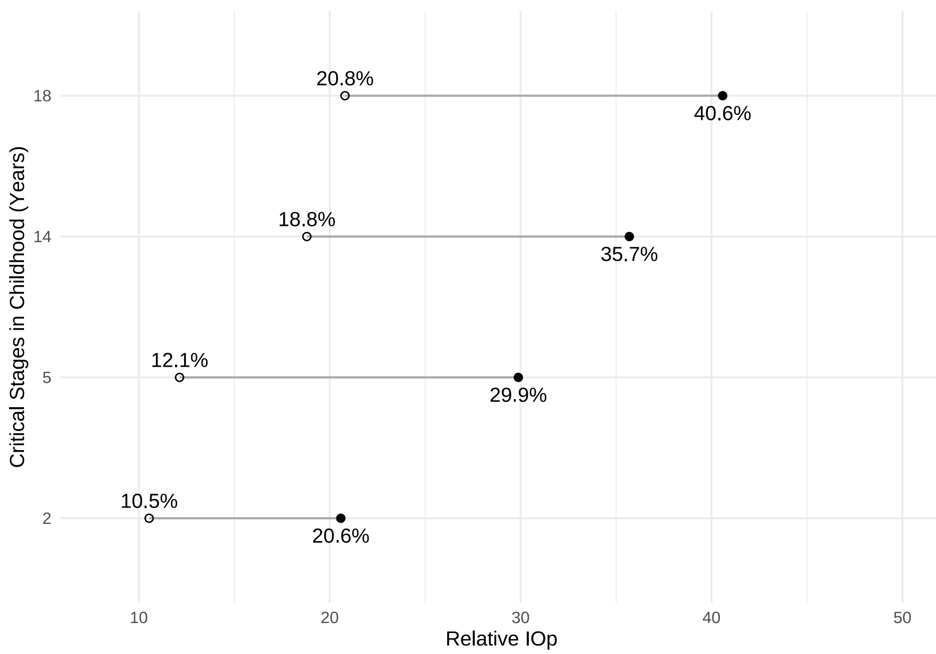

Figure 4: Lower and upper bounds of relative IOp estimates

Although the literature typically provides lower bound estimates of IOp due to incomplete circumstance sets, figure 4 suggests we may have reached an upper bound given the number of circumstances used in this analysis. Figure 4 displays the upper and lower bounds of relative IOp. 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 21% of total adult income inequality could be attributed to IOp measured using circumstances deemed "important". This share increases to almost 41% when all available circumstances before the age of majority are considered, suggesting an upper bound may have been obtained.

6 Conclusion

In this essay, 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 individuals born in 1978-1983 with their adult labor incomes measured at around age 28-30, I find that relative inequality of opportunity (IOp), the share of income inequality in adult labor income attributable to unequal circumstances faced by an individual up to the age of majority at 18, is just over 40%. This estimate is 23% for those in the Survey of Economic Opportunity (SEO) sample, which includes a disproportionately higher number of individuals from poor households. These figures are higher than those obtained using a standard set of circumstances in OLS regression. Using an OLS regression without interactions that includes individual's race, sex as well as education of the family head, family income, region of residence (all measured during child's first year), I estimate the share of inequality of opportunity in total income inequality to be 23% for the full sample. For the SEO sample, this number is 15%. Moreover, around 30% of the total inequality in adult labor incomes could be attributable to unequal circumstances up to age 5.

I also calculate permutation-based variable importance scores for each circumstance in all circumstance sets based on age cutoffs. Using the top ten circumstances based on these importance scores, I obtain 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 obtain 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 21%. Adding more circumstances (as many as 174) to the model only increases the relative IOp estimates to about 41%. Although it lacks causal interpretation, this exercise could be useful in identifying circumstances for causal analysis to better inform policy discourse.

These findings have limitations. One might contest using age 18 as the threshold for majority, which is reasonable. Different societies establish different standards. For instance, in France, the age of consent is 15 years, which could serve as a distinction between circumstance and effort in that context. Cultural norms also offer alternative thresholds. In Judaism, a boy becomes a bar mitzvah at age 13, marking when he is considered a Jewish adult responsible for his own actions and observing Jewish law. For girls, this transition occurs at age 12. The purpose of this essay is to use the age of majority as a responsibility cutoff between circumstances and effort based on societal value judgments that may vary depending on cultural norms while incorporating dynamic complementarity in skill formation in measurement of inequality of opportunity. 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 childhood circumstances into the measurement of inequality of opportunity. This approach to measuring IOp could inform the causal studies leading to 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 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.

6.2 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 25% 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 = 30 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.

Table 4: Tuned hyperparameters

		Full Sample (N = 1374)	
Age Cutoff	Trees	min_n	mtry
2	200	25	2
5	200	35	7
10	200	35	11
12	200	35	13
14	200	30	15
16	200	30	22
18	200	25	17

6.3 Tuned Hyperparameters

Table 4 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.

Table 5: IOp estimates for different circumstance sets (using Gini)

	Outcome : adult labor income			
	Full sample (total inequality = 0.402)		SEO sample (total inequality = 0.4)	
	Absolute IOp	Relative IOp	Absolute IOp	Relative IOp
2	0.22	0.55	0.15	0.37
5	0.27	0.66	0.17	0.43
10	0.28	0.69	0.21	0.52
12	0.28	0.70	0.22	0.54
14	0.29	0.71	0.22	0.55
16	0.30	0.75	0.23	0.57
18	0.30	0.76	0.23	0.58

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 4 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 4—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.

6.4 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 report absolute and relative IOp estimates along with their contributions to total inequality, as measured by the Gini coefficient, for all age cutoffs.

Table 5 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, 66%, attributed to the inequality of opportunity, stems from circumstances up to age 5. For only individuals in the SEO sample, that estimate is 43%.

6.5 Summary Statistics

Table 6 presents summary statistics for the full set of continuous variables used as circumstances. All monetary circumstances are measured in 2024 US dollars. Moreover, family income, childcare costs, marginal tax rates (in percentages) are from previous tax year to the survey wave. 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 6: Weighted summary statistics (complete set of continuous circumstances)

Circumstance	Mean	Std.Dev
Adult real labor income in 2024 US dollars (in natural lograithm)	10.4	1.1
Number of children to individual's mother	2.8	1.1
Number of children to individual's father	2.9	1.3
Years of education of the head at age 1	13.1	2.5
Years of education of the head at age 2	13.1	2.6
Years of education of the head at age 3	13.2	2.4
Years of education of the head at age 4	13.3	2.5
Years of education of the head at age 5	13.4	2.5
Years of education of the head at age 6	13.5	2.4
Years of education of the head at age 7	13.5	2.5
Years of education of the head at age 8	13.5	2.5
Years of education of the head at age 9	13.5	2.5
Years of education of the head at age 10	13.5	2.5
Years of education of the head at age 11	13.5	2.5
Years of education of the head at age 12	13.6	2.5
Years of education of the head at age 13	13.6	2.4
Years of education of the head at age 14	13.6	2.4
Years of education of the head at age 15	13.5	2.4
Years of education of the head at age 16	13.5	2.5
Years of education of the head at age 17	13.6	2.5
Years of education of the head at age 18	13.5	2.5
Years of education of the spouse at age 1	13.2	2.2
Years of education of the spouse at age 2	13.3	2.2
Years of education of the spouse at age 3	13.3	2.1
Years of education of the spouse at age 4	13.3	2.2
Years of education of the spouse at age 5	13.4	2.2
Years of education of the spouse at age 6	13.4	2.2
Years of education of the spouse at age 7	13.5	2.2
Years of education of the spouse at age 8	13.6	2.2
Years of education of the spouse at age 9	13.6	2.2
Years of education of the spouse at age 10	13.6	2.2
Years of education of the spouse at age 11	13.6	2.2
Years of education of the spouse at age 12	13.6	2.2
Years of education of the spouse at age 13	13.6	2.2
Years of education of the spouse at age 14	13.7	2.2

Table 6: Weighted summary statistics (complete set of continuous circumstances) *(continued)*

Circumstance	Mean	Std.Dev
Years of education of the spouse at age 15	13.6	2.1
Years of education of the spouse at age 16	13.6	2.2
Years of education of the spouse at age 17	13.5	2.2
Years of education of the spouse at age 18	13.6	2.3
family size at age 1	3.9	1.2
family size at age 2	4.0	1.2
family size at age 3	4.1	1.1
family size at age 4	4.2	1.1
family size at age 5	4.4	1.2
family size at age 6	4.4	1.1
family size at age 7	4.4	1.1
family size at age 8	4.4	1.1
family size at age 9	4.4	1.1
family size at age 10	4.4	1.1
family size at age 11	4.4	1.1
family size at age 12	4.4	1.2
family size at age 13	4.4	1.2
family size at age 14	4.3	1.2
family size at age 15	4.2	1.2
family size at age 16	4.2	1.2
family size at age 17	4.1	1.2
family size at age 18	4.0	1.3
family income at age 1	73739.6	45332.5
family income at age 2	75262.3	52478.2
family income at age 3	77078.8	54881.7
family income at age 4	80824.0	59121.4
family income at age 5	84537.9	67330.5
family income at age 6	86809.7	69382.9
family income at age 7	93392.6	75990.1
family income at age 8	97373.5	87818.0
family income at age 9	102733.2	94912.4
family income at age 10	106327.8	101978.3
family income at age 11	111144.7	120988.3
family income at age 12	112386.2	110238.4
family income at age 13	116463.6	125612.9
family income at age 14	119442.3	118945.0
family income at age 15	116367.6	106357.7
family income at age 16	126459.6	132186.5
family income at age 17	136042.9	204502.2
family income at age 18	143432.0	160872.0
Real childcare cost at age 1	1086.2	3937.5
Real childcare cost at age 2	71.6	414.7
Real childcare cost at age 3	1148.4	14212.0
Real childcare cost at age 4	1102.2	15117.2
Real childcare cost at age 5	3065.0	19548.6
Real childcare cost at age 6	3105.7	17777.3
Real childcare cost at age 7	3149.6	20177.9
Real childcare cost at age 8	2039.3	9191.7
Real childcare cost at age 9	2268.3	13730.5
Real childcare cost at age 10	3943.4	23311.2
Real childcare cost at age 11	2974.2	20320.4
Real childcare cost at age 12	4514.3	27198.4
Real childcare cost at age 13	3158.2	22015.0
Real childcare cost at age 14	1695.2	15461.2
Real childcare cost at age 15	3494.3	24997.7

Table 6: Weighted summary statistics (complete set of continuous circumstances) *(continued)*

Circumstance	Mean	Std.Dev
Real childcare cost at age 16	1113.9	18341.9
Real childcare cost at age 17	1448.6	14784.9
Real childcare cost at age 18	576.1	7629.2
Marginal tax rate on family head and spouse’s combined income at age 1	21.1	12.1
Marginal tax rate on family head and spouse’s combined income at age 2	20.8	12.0
Marginal tax rate on family head and spouse’s combined income at age 3	20.8	12.1
Marginal tax rate on family head and spouse’s combined income at age 4	21.1	11.7
Marginal tax rate on family head and spouse’s combined income at age 5	20.2	11.4
Marginal tax rate on family head and spouse’s combined income at age 6	19.1	11.1
Marginal tax rate on family head and spouse’s combined income at age 7	19.2	10.8
Marginal tax rate on family head and spouse’s combined income at age 8	18.8	10.8
Marginal tax rate on family head and spouse’s combined income at age 9	18.3	10.3
Marginal tax rate on family head and spouse’s combined income at age 10	17.6	10.2
Marginal tax rate on family head and spouse’s combined income at age 11	17.6	10.0
Marginal tax rate on family head and spouse’s combined income at age 12	17.6	9.7
Marginal tax rate on family head and spouse’s combined income at age 13	17.9	9.8

Table 7 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 7: Weighted summary statistics (complete set of categorical circumstances)

Circumstance	Obs	Percent
Birthcohort		
1978	2284551	16%
1979	2368763	17%
1980	2258074	16%
1981	2737928	19%
1982	2068559	15%
1983	2356565	17%
Sex of the individual		
Male	6979487	50%
Female	7094953	50%
Race of the individual		
White	11034274	79%
Black	2200028	16%
Other race	159772	1%
Hispanic	644211	5%
Birthweight		
High (greater than 88 lbs)	11128923	94%
Low (less than 88 lbs)	695107	6%
Marital status of mother at child’s birth		
Married	11103094	82%
Never Married	1859564	14%
Widowed	43436	0%
Divorced	324729	2%

Table 7: Weighted summary statistics (complete set of categorical circumstances) *(continued)*

Circumstance	Obs	Percent
Separated	147356	1%
Marital status of the family head at age 1		
Married	10817249	86%
Seperated	275786	2%
Divorced	274677	2%
Widowed	176720	1%
Never Married	963508	8%
Marital status of the family head at age 2		
Married	10565684	84%
Seperated	354556	3%
Divorced	381819	3%
Widowed	237857	2%
Never Married	1004880	8%
Marital status of the family head at age 3		
Married	10469423	83%
Seperated	279002	2%
Divorced	595951	5%
Widowed	163885	1%
Never Married	1071596	8%
Marital status of the family head at age 4		
Married	10505102	84%
Seperated	320012	3%
Divorced	608199	5%
Widowed	159033	1%
Never Married	954377	8%
Marital status of the family head at age 5		
Married	10393447	84%
Seperated	356244	3%
Divorced	564593	4%
Widowed	102806	1%
Never Married	1016317	8%
Marital status of the family head at age 6		
Married	10218763	83%
Seperated	499510	4%
Divorced	595761	5%
Widowed	154521	1%
Never Married	888645	7%
Marital status of the family head at age 7		
Married	10182153	83%
Seperated	379134	3%
Divorced	659656	5%
Widowed	170642	1%
Never Married	806541	7%
Marital status of the family head at age 8		
Married	9840800	82%
Seperated	405209	3%
Divorced	873556	7%
Widowed	211348	2%
Never Married	736412	6%
Marital status of the family head at age 9		
Married	9698659	81%
Seperated	558426	5%
Divorced	926427	8%
Widowed	251311	2%
Never Married	601502	5%

Table 7: Weighted summary statistics (complete set of categorical circumstances) *(continued)*

Circumstance	Obs	Percent
Marital status of the family head at age 10		
Married	9711625	81%
Seperated	460370	4%
Divorced	1107589	9%
Widowed	235367	2%
Never Married	539168	4%
Marital status of the family head at age 11		
Married	9905087	80%
Seperated	544310	4%
Divorced	1219539	10%
Widowed	233082	2%
Never Married	505555	4%
Marital status of the family head at age 12		
Married	9918686	78%
Seperated	560861	4%
Divorced	1438425	11%
Widowed	275821	2%
Never Married	581711	5%
Marital status of the family head at age 13		
Married	10060515	76%
Seperated	593519	4%
Divorced	1634293	12%
Widowed	261648	2%
Never Married	610752	5%
Marital status of the family head at age 14		
Married	9904239	74%
Seperated	696637	5%
Divorced	1680192	13%
Widowed	329867	2%
Never Married	698364	5%
Marital status of the family head at age 15		
Married	8212230	73%
Seperated	408585	4%
Divorced	1783592	16%
Widowed	321828	3%
Never Married	578562	5%
Marital status of the family head at age 16		
Married	8426088	72%
Seperated	549241	5%
Divorced	1728829	15%
Widowed	300639	3%
Never Married	663185	6%
Marital status of the family head at age 17		
Married	6459884	73%
Seperated	378472	4%
Divorced	1298937	15%
Widowed	207869	2%
Never Married	446891	5%
Marital status of the family head at age 18		
Married	7005142	74%
Seperated	476951	5%
Divorced	1221864	13%
Widowed	177755	2%
Never Married	601104	6%

Table 7: Weighted summary statistics (complete set of categorical circumstances) *(continued)*

Circumstance	Obs	Percent
Region of residence of the family at age of residence of the family at age 1		
Northeast	2619019	21%
Midwest	3564718	28%
South	4084706	33%
West	1998167	16%
Alaska or Hawaii	63752	0%
Country outside the United States	177578	1%
Region of residence of the family at age of residence of the family at age 2		
Northeast	2529264	20%
Midwest	3613495	29%
South	4173675	33%
West	1970471	16%
Alaska or Hawaii	108615	1%
Country outside the United States	183856	2%
Region of residence of the family at age 3		
Northeast	2547492	20%
Midwest	3491562	28%
South	4259632	34%
West	2062488	16%
Alaska or Hawaii	66386	0%
Country outside the United States	152297	1%
Region of residence of the family at age 4		
Northeast	2506342	20%
Midwest	3406870	27%
South	4372195	35%
West	2125394	17%
Alaska or Hawaii	39629	0%
Country outside the United States	96293	1%
Region of residence of the family at age 5		
Northeast	2569983	21%
Midwest	3311872	27%
South	4361272	35%
West	2066372	17%
Alaska or Hawaii	37995	0%
Country outside the United States	85913	1%
Region of residence of the family at age 6		
Northeast	2607491	21%
Midwest	3361601	27%
South	4204997	34%
West	2096324	17%
Alaska or Hawaii	14429	0%
Country outside the United States	72033	1%
Region of residence of the family at age 7		
Northeast	2510434	21%
Midwest	3298994	27%
South	4264621	35%
West	2030841	17%
Alaska or Hawaii	14429	0%
Country outside the United States	78807	1%
Region of residence of the family at age 8		
Northeast	2514482	21%
Midwest	3216951	27%
South	4199437	35%

Table 7: Weighted summary statistics (complete set of categorical circumstances) *(continued)*

Circumstance	Obs	Percent
West	2038734	17%
Alaska or Hawaii	16010	0%
Country outside the United States	81711	1%
Region of residence of the family at age 9		
Northeast	2441201	20%
Midwest	3212760	27%
South	4249626	35%
West	2048106	17%
Alaska or Hawaii	28258	0%
Country outside the United States	56374	0%
Region of residence of the family at age 10		
Northeast	2415201	20%
Midwest	3200752	27%
South	4333781	36%
West	2021527	17%
Alaska or Hawaii	19153	0%
Country outside the United States	63705	0%
Region of residence of the family at age 11		
Northeast	2437945	20%
Midwest	3215990	26%
South	4466505	36%
West	2111549	17%
Alaska or Hawaii	27565	0%
Country outside the United States	148019	1%
Region of residence of the family at age 12		
Northeast	2629431	21%
Midwest	3313353	26%
South	4557917	36%
West	2097444	16%
Alaska or Hawaii	27565	0%
Country outside the United States	149794	1%
Region of residence of the family at age 13		
Northeast	2685389	20%
Midwest	3456505	26%
South	4651679	35%
West	2169443	16%
Alaska or Hawaii	27565	0%
Country outside the United States	170146	1%
Region of residence of the family at age 14		
Northeast	2653042	20%
Midwest	3444188	26%
South	4753743	36%
West	2255703	17%
Alaska or Hawaii	27565	0%
Country outside the United States	175058	1%
Region of residence of the family at age 15		
Northeast	2369319	21%
Midwest	2839431	25%
South	4006965	35%
West	1928517	17%
Alaska or Hawaii	27183	0%
Country outside the United States	133382	1%
Region of residence of the family at age 16		
Northeast	2266489	19%
Midwest	3062004	26%

Table 7: Weighted summary statistics (complete set of categorical circumstances) *(continued)*

Circumstance	Obs	Percent
South	4049336	35%
West	2109530	18%
Alaska or Hawaii	19585	0%
Country outside the United States	182482	2%
Region of residence of the family at age 17		
Northeast	1862707	21%
Midwest	2280195	26%
South	2997626	34%
West	1531187	17%
Alaska or Hawaii	25602	0%
Country outside the United States	94736	1%
Region of residence of the family at age 18		
Northeast	1806690	19%
Midwest	2410147	25%
South	3348266	35%
West	1790407	19%
Alaska or Hawaii	16442	0%
Country outside the United States	118997	1%
Race of the family head at age 1		
Black	1652748	13%
Hispanic	350564	3%
Other	44295	0%
White	10460333	84%
Race of the family head at age 2		
Black	1813059	14%
Hispanic	372625	3%
Other	45611	0%
White	10348081	82%
Race of the family head at age 3		
Black	1857498	15%
Hispanic	377629	3%
Other	44295	0%
White	10300435	82%
Race of the family head at age 4		
Black	1864769	15%
Hispanic	381640	3%
Other	55997	0%
White	10244317	82%
Race of the family head at age 5		
Black	1883462	15%
Hispanic	320620	3%
Other	57408	0%
White	10171917	82%
Race of the family head at age 6		
Black	1875043	15%
Hispanic	319181	3%
Other	25767	0%
White	10137209	82%
Race of the family head at age 7		
Black	1811360	15%
Hispanic	321460	3%
Other	27699	0%
White	10037607	82%
Race of the family head at age 8		
Black	1809036	15%

Table 7: Weighted summary statistics (complete set of categorical circumstances) *(continued)*

Circumstance	Obs	Percent
Hispanic	321460	3%
Other	27794	0%
White	9909035	82%
Race of the family head at age 9		
Black	1724160	14%
Hispanic	353393	3%
Other	27699	0%
White	9931073	83%
Race of the family head at age 10		
Black	1825515	15%
Hispanic	336321	3%
Other	26383	0%
White	9865326	82%
Race of the family head at age 11		
Black	1901162	15%
Hispanic	300242	2%
Other	18363	0%
White	10187806	82%
Race of the family head at age 12		
Black	1981427	16%
Hispanic	291850	2%
Other	39521	0%
White	10462706	82%
Race of the family head at age 13		
Black	2045903	16%
Hispanic	284438	2%
Other	39426	0%
White	10790960	82%
Race of the family head at age 14		
Black	2064379	16%
Hispanic	313002	2%
Other	61278	0%
White	10870640	82%
Race of the family head at age 15		
Black	1706318	15%
Hispanic	294689	3%
Other	52401	0%
White	9247068	82%
Race of the family head at age 16		
Black	1890417	16%
Hispanic	358056	3%
Other	47770	0%
White	9393183	80%
Race of the family head at age 17		
Black	1312466	15%
Hispanic	254067	3%
Other	47222	0%
White	7178298	82%
Race of the family head at age 18		
Black	1551504	16%
Hispanic	354861	4%
Other	42658	0%
White	7541926	79%
Race of the spouse at age 1		
Black	777237	7%

Table 7: Weighted summary statistics (complete set of categorical circumstances) *(continued)*

Circumstance	Obs	Percent
Hispanic	197786	2%
Other	20892	0%
White	9797622	91%
Race of the spouse at age 2		
Black	835228	8%
Hispanic	242652	2%
Other	20287	0%
White	9578569	90%
Race of the spouse at age 3		
Black	853638	8%
Hispanic	203848	2%
Other	21603	0%
White	9536469	90%
Race of the spouse at age 4		
Black	844520	8%
Hispanic	288452	3%
Other	20276	0%
White	9486424	89%
Race of the spouse at age 5		
Black	874800	8%
Hispanic	271739	3%
Other	45419	0%
White	9320145	89%
Race of the spouse at age 6		
Black	897525	9%
Hispanic	235204	2%
Other	47446	0%
White	9118789	89%
Race of the spouse at age 7		
Black	904604	9%
Hispanic	250060	2%
Other	45514	0%
White	9074798	88%
Race of the spouse at age 8		
Black	830029	8%
Hispanic	222371	2%
Other	45419	0%
White	8806941	89%
Race of the spouse at age 9		
Black	648791	7%
Hispanic	235204	2%
Other	44803	0%
White	8810218	90%
Race of the spouse at age 10		
Black	762383	8%
Hispanic	178882	2%
Other	46119	0%
White	8745506	90%
Race of the spouse at age 11		
Black	792826	8%
Hispanic	168211	2%
Other	44803	0%
White	8906527	90%
Race of the spouse at age 12		
Black	785725	8%

Table 7: Weighted summary statistics (complete set of categorical circumstances) *(continued)*

Circumstance	Obs	Percent
Hispanic	192887	2%
Other	45419	0%
White	8990490	90%
Race of the spouse at age 13		
Black	865129	8%
Hispanic	193828	2%
Other	57092	1%
White	9004967	89%
Race of the spouse at age 14		
Black	830788	8%
Hispanic	194017	2%
Other	56381	1%
White	9035281	89%
Race of the spouse at age 15		
Black	701166	8%
Hispanic	199377	2%
Other	56381	1%
White	7499198	89%
Race of the spouse at age 16		
Black	635153	7%
Hispanic	207955	2%
Other	40671	0%
White	7828083	90%
Race of the spouse at age 17		
Black	531084	8%
Hispanic	199178	3%
Other	56926	1%
White	5854593	88%
Race of the spouse at age 18		
Black	608195	8%
Hispanic	173724	2%
Other	40671	1%
White	6431965	89%
Sex of the family head at age 1		
Male	10912681	87%
Female	1595259	13%
Sex of the family head at age 2		
Male	10795885	86%
Female	1783491	14%
Sex of the family head at age 3		
Male	10689583	85%
Female	1890274	15%
Sex of the family head at age 4		
Male	10699999	85%
Female	1846724	15%
Sex of the family head at age 5		
Male	10588219	85%
Female	1845188	15%
Sex of the family head at age 6		
Male	10369926	84%
Female	1987274	16%
Sex of the family head at age 7		
Male	10326054	85%
Female	1872072	15%
Sex of the family head at age 8		

Table 7: Weighted summary statistics (complete set of categorical circumstances) *(continued)*

Circumstance	Obs	Percent
Male	9996028	83%
Female	2071297	17%
Sex of the family head at age 9		
Male	9888708	82%
Female	2147617	18%
Sex of the family head at age 10		
Male	9885728	82%
Female	2168391	18%
Sex of the family head at age 11		
Male	10093063	81%
Female	2314510	19%
Sex of the family head at age 12		
Male	10268684	80%
Female	2506820	20%
Sex of the family head at age 13		
Male	10407861	79%
Female	2752866	21%
Sex of the family head at age 14		
Male	10383057	78%
Female	2926242	22%
Sex of the family head at age 15		
Male	8755624	77%
Female	2549173	23%
Sex of the family head at age 16		
Male	8952591	77%
Female	2736835	23%
Sex of the family head at age 17		
Male	6849798	78%
Female	1942255	22%
Sex of the family head at age 18		
Male	7622120	80%
Female	1868829	20%
Occupation of the family head at age 1		
Clerical	686833	7%
Craftsman	1892670	20%
Farm laborers	71890	1%
Farmers	154003	2%
Laborers	391645	4%
Managers	1401770	15%
Operatives	932993	10%
Private	1216	
Professional	2289181	24%
Sales	337724	4%
Service	770008	8%
Transport	556350	6%
Occupation of the family head at age 2		
Clerical	637879	6%
Craftsman	2074696	20%
Farm laborers	70044	1%
Farmers	179084	2%
Laborers	365143	4%
Managers	1521168	15%
Operatives	1035836	10%
Private	2184	
Professional	2383510	23%

Table 7: Weighted summary statistics (complete set of categorical circumstances) *(continued)*

Circumstance	Obs	Percent
Sales	734250	7%
Service	811841	8%
Transport	635519	6%
Occupation of the family head at age 3		
Clerical	705927	6%
Craftsman	2165139	20%
Farm laborers	92166	1%
Farmers	175485	2%
Laborers	414170	4%
Managers	1685026	15%
Operatives	1125263	10%
Private	15836	0%
Professional	2432579	22%
Sales	725911	7%
Service	765688	7%
Transport	625588	6%
Occupation of the family head at age 4		
Clerical	422698	4%
Craftsman	2149211	20%
Farm laborers	91358	1%
Farmers	155299	1%
Laborers	500796	5%
Managers	1926692	18%
Operatives	1065714	10%
Private	20979	0%
Professional	2474207	23%
Sales	602934	6%
Service	760345	7%
Transport	669874	6%
Occupation of the family head at age 5		
Clerical	719157	7%
Craftsman	2312736	21%
Farm laborers	113809	1%
Farmers	168882	2%
Laborers	420680	4%
Managers	1960841	18%
Operatives	902151	8%
Private	16991	0%
Professional	2549959	23%
Sales	519028	5%
Service	645594	6%
Transport	581421	5%
Occupation of the family head at age 6		
Clerical	766510	7%
Craftsman	2025110	19%
Farm laborers	73787	1%
Farmers	145804	1%
Laborers	405289	4%
Managers	2079592	19%
Operatives	801683	7%
Private	44219	0%
Professional	2423643	22%
Sales	622279	6%
Service	783945	7%
Transport	714866	7%

Table 7: Weighted summary statistics (complete set of categorical circumstances) *(continued)*

Circumstance	Obs	Percent
Occupation of the family head at age 7		
Clerical	642798	6%
Craftsman	1882176	17%
Farm laborers	155761	1%
Farmers	132161	1%
Laborers	448002	4%
Managers	2122397	20%
Operatives	810798	8%
Private	0	
Professional	2732840	25%
Sales	528283	5%
Service	762318	7%
Transport	650538	6%
Occupation of the family head at age 8		
Clerical	870497	8%
Craftsman	1878891	17%
Farm laborers	103908	1%
Farmers	145058	1%
Laborers	283860	3%
Managers	2004430	19%
Operatives	826970	8%
Private	5371	
Professional	2651749	25%
Sales	602231	6%
Service	833399	8%
Transport	559185	5%
Occupation of the family head at age 9		
Clerical	868089	8%
Craftsman	1910536	18%
Farm laborers	113446	1%
Farmers	136954	1%
Laborers	278462	3%
Managers	1923899	18%
Operatives	734935	7%
Private	19716	0%
Professional	2746647	25%
Sales	550818	5%
Service	828367	8%
Transport	694361	6%
Occupation of the family head at age 10		
Clerical	829679	8%
Craftsman	1913058	18%
Farm laborers	115295	1%
Farmers	123321	1%
Laborers	254085	2%
Managers	1970982	18%
Operatives	665138	6%
Private	49105	0%
Professional	2822369	26%
Sales	630955	6%
Service	776150	7%
Transport	549482	5%
Occupation of the family head at age 11		
Clerical	786037	7%
Craftsman	1892755	17%

Table 7: Weighted summary statistics (complete set of categorical circumstances) *(continued)*

Circumstance	Obs	Percent
Farm laborers	121043	1%
Farmers	137736	1%
Laborers	391940	4%
Managers	2215934	20%
Operatives	696154	6%
Private	60779	0%
Professional	2713342	24%
Sales	553424	5%
Service	890033	8%
Transport	636631	6%
Occupation of the family head at age 12		
Clerical	1024877	9%
Craftsman	1922431	17%
Farm laborers	41157	0%
Farmers	126694	1%
Laborers	255378	2%
Managers	2090108	18%
Operatives	828122	7%
Private	76098	1%
Professional	2999446	26%
Sales	590242	5%
Service	855693	8%
Transport	614000	5%
Occupation of the family head at age 13		
Clerical	937407	8%
Craftsman	1931831	17%
Farm laborers	68969	1%
Farmers	165258	1%
Laborers	383237	3%
Managers	2356761	20%
Operatives	778488	7%
Private	68572	1%
Professional	2885841	25%
Sales	637586	6%
Service	865800	7%
Transport	566394	5%
Occupation of the family head at age 14		
Clerical	1038869	9%
Craftsman	1994977	17%
Farm laborers	81356	1%
Farmers	153314	1%
Laborers	369467	3%
Managers	2217616	18%
Operatives	726084	6%
Private	69788	1%
Professional	3171691	26%
Sales	701342	6%
Service	912080	8%
Transport	610693	5%
Occupation of the family head at age 15		
Clerical	981899	10%
Craftsman	1755059	17%
Farm laborers	57661	1%
Farmers	118810	1%
Laborers	231416	2%

Table 7: Weighted summary statistics (complete set of categorical circumstances) *(continued)*

Circumstance	Obs	Percent
Managers	2201198	22%
Operatives	408004	4%
Private	79380	1%
Professional	2471310	24%
Sales	582000	6%
Service	831749	8%
Transport	409865	4%
Occupation of the family head at age 16		
Clerical	869616	8%
Craftsman	1599065	15%
Farm laborers	77264	1%
Farmers	138175	1%
Laborers	319330	3%
Managers	2348922	23%
Operatives	592834	6%
Private	56794	0%
Professional	2456578	24%
Sales	513673	5%
Service	924813	9%
Transport	516654	5%
Occupation of the family head at age 17		
Clerical	663356	8%
Craftsman	1314613	16%
Farm laborers	40153	0%
Farmers	115768	2%
Laborers	249076	3%
Managers	1814269	23%
Operatives	365018	5%
Private	33629	0%
Professional	1763516	22%
Sales	525584	7%
Service	657571	8%
Transport	436346	6%
Occupation of the family head at age 18		
Clerical	622230	7%
Craftsman	1506406	18%
Farm laborers	45512	0%
Farmers	98740	1%
Laborers	315126	4%
Managers	1769217	21%
Operatives	442960	5%
Private	46368	1%
Professional	1916816	23%
Sales	537337	6%
Service	718544	9%
Transport	380827	4%
Occupation of the spouse at age 1		
Clerical	1167397	27%
Craftsman	67706	2%
Farm laborers	8318	0%
Farmers	5759	0%
Laborers	76021	2%
Managers	223389	5%
Operatives	295478	7%
Private	78227	2%

Table 7: Weighted summary statistics (complete set of categorical circumstances) *(continued)*

Circumstance	Obs	Percent
Professional	1520934	36%
Sales	205367	5%
Service	603226	14%
Transport	2840	
Occupation of the spouse at age 2		
Clerical	1180826	26%
Craftsman	68260	2%
Farm laborers	0	
Farmers	5759	0%
Laborers	53382	1%
Managers	239242	5%
Operatives	367715	8%
Private	10036	0%
Professional	1529551	33%
Sales	231292	5%
Service	875830	19%
Transport	17287	0%
Occupation of the spouse at age 3		
Clerical	1325348	26%
Craftsman	99483	2%
Farm laborers	6806	0%
Farmers	0	
Laborers	29088	1%
Managers	349700	7%
Operatives	328474	7%
Private	12547	0%
Professional	1719202	34%
Sales	292507	6%
Service	833124	17%
Transport	16115	0%
Occupation of the spouse at age 4		
Clerical	1397057	26%
Craftsman	103492	2%
Farm laborers	0	
Farmers	0	
Laborers	47796	1%
Managers	382611	7%
Operatives	395015	7%
Private	68659	1%
Professional	1825713	34%
Sales	147567	3%
Service	991604	18%
Transport	32655	1%
Occupation of the spouse at age 5		
Clerical	1442733	25%
Craftsman	121609	2%
Farm laborers	0	
Farmers	9621	0%
Laborers	28331	0%
Managers	446652	8%
Operatives	419256	7%
Private	51249	1%
Professional	1876201	32%
Sales	296915	5%
Service	1128822	19%

Table 7: Weighted summary statistics (complete set of categorical circumstances) *(continued)*

Circumstance	Obs	Percent
Transport	40275	1%
Occupation of the spouse at age 6		
Clerical	1549199	26%
Craftsman	110343	2%
Farm laborers	18105	0%
Farmers	7882	0%
Laborers	35550	1%
Managers	350926	6%
Operatives	464111	8%
Private	82270	1%
Professional	1950308	32%
Sales	216390	4%
Service	1215899	20%
Transport	0	
Occupation of the spouse at age 7		
Clerical	1741594	26%
Craftsman	148659	2%
Farm laborers	20214	0%
Farmers	5759	
Laborers	66423	1%
Managers	421934	6%
Operatives	500073	8%
Private	88807	1%
Professional	2166388	33%
Sales	242804	4%
Service	1235821	19%
Transport	15368	0%
Occupation of the spouse at age 8		
Clerical	1867866	28%
Craftsman	130958	2%
Farm laborers	0	
Farmers	9403	0%
Laborers	41731	1%
Managers	416261	6%
Operatives	440721	7%
Private	134846	2%
Professional	2110705	32%
Sales	293045	4%
Service	1122425	17%
Transport	1565	
Occupation of the spouse at age 9		
Clerical	1655641	24%
Craftsman	104832	2%
Farm laborers	17479	0%
Farmers	0	
Laborers	30207	0%
Managers	546550	8%
Operatives	439070	6%
Private	71508	1%
Professional	2203360	32%
Sales	334064	5%
Service	1321907	19%
Transport	55555	1%
Occupation of the spouse at age 10		
Clerical	2047354	30%

Table 7: Weighted summary statistics (complete set of categorical circumstances) *(continued)*

Circumstance	Obs	Percent
Craftsman	62292	1%
Farm laborers	17479	0%
Farmers	0	
Laborers	49052	1%
Managers	576407	9%
Operatives	392874	6%
Private	66086	1%
Professional	2066576	31%
Sales	274580	4%
Service	1174291	17%
Transport	11825	0%
Occupation of the spouse at age 11		
Clerical	1952909	26%
Craftsman	123211	2%
Farm laborers	17101	0%
Farmers	14626	0%
Laborers	59379	1%
Managers	787974	11%
Operatives	422367	6%
Private	44069	1%
Professional	2384743	32%
Sales	317105	4%
Service	1312483	18%
Transport	19517	0%
Occupation of the spouse at age 12		
Clerical	2098354	28%
Craftsman	176927	2%
Farm laborers	31752	0%
Farmers	6126	
Laborers	65422	1%
Managers	781004	10%
Operatives	406901	5%
Private	88550	1%
Professional	2439682	32%
Sales	266826	4%
Service	1174199	15%
Transport	52669	1%
Occupation of the spouse at age 13		
Clerical	2303721	29%
Craftsman	170282	2%
Farm laborers	8677	0%
Farmers	17101	0%
Laborers	31961	0%
Managers	839484	11%
Operatives	369701	5%
Private	61572	1%
Professional	2570951	33%
Sales	190981	2%
Service	1206226	15%
Transport	79567	1%
Occupation of the spouse at age 14		
Clerical	2400083	30%
Craftsman	187566	2%
Farm laborers	25596	0%
Farmers	0	

Table 7: Weighted summary statistics (complete set of categorical circumstances) *(continued)*

Circumstance	Obs	Percent
Laborers	69474	1%
Managers	998446	12%
Operatives	326926	4%
Private	70817	1%
Professional	2611933	32%
Sales	195299	2%
Service	1078018	13%
Transport	79895	1%
Occupation of the spouse at age 15		
Clerical	2110903	31%
Craftsman	157848	2%
Farm laborers	30936	0%
Farmers	32209	0%
Laborers	22053	0%
Managers	826294	12%
Operatives	281981	4%
Private	45338	1%
Professional	2039064	30%
Sales	171228	2%
Service	952266	14%
Transport	111464	2%
Occupation of the spouse at age 16		
Clerical	1894795	27%
Craftsman	132582	2%
Farm laborers	26601	0%
Farmers	32209	0%
Laborers	99309	1%
Managers	947425	13%
Operatives	302985	4%
Private	46898	1%
Professional	2394071	34%
Sales	123981	2%
Service	967423	14%
Transport	72884	1%
Occupation of the spouse at age 17		
Clerical	1579955	29%
Craftsman	192279	4%
Farm laborers	22259	0%
Farmers	0	
Laborers	90134	2%
Managers	687286	13%
Operatives	179497	3%
Private	27187	0%
Professional	1690718	31%
Sales	136456	2%
Service	838505	15%
Transport	40020	1%
Occupation of the spouse at age 18		
Clerical	1442398	26%
Craftsman	118462	2%
Farm laborers	9621	0%
Farmers	0	
Laborers	81757	2%
Managers	844928	15%
Operatives	233753	4%

Table 7: Weighted summary statistics (complete set of categorical circumstances) *(continued)*

Circumstance	Obs	Percent
Private	81460	2%
Professional	1735845	31%
Sales	269000	5%
Service	706434	13%
Transport	57513	1%
Percent of families who used food stamps at age 1	1601023	15%
Percent of families who used food stamps at age 2	2001115	16%
Percent of families who used food stamps at age 3	2040886	16%
Percent of families who used food stamps at age 4	1963349	16%
Percent of families who used food stamps at age 5	1861363	15%
Percent of families who used food stamps at age 6	1877890	15%
Percent of families who used food stamps at age 7	1531805	13%
Percent of families who used food stamps at age 8	1509492	13%
Percent of families who used food stamps at age 9	1431159	12%
Percent of families who used food stamps at age 10	1430840	12%
Percent of families who used food stamps at age 11	1498700	12%
Percent of families who used food stamps at age 12	1480870	12%
Percent of families who used food stamps at age 13	1478533	11%
Percent of families who used food stamps at age 14	1306784	10%
Percent of families who used food stamps at age 15	1033403	9%
Percent of families who used food stamps at age 16	1203477	10%
Percent of families who used food stamps at age 17	636697	7%
Percent of families who used food stamps at age 18	755243	8%
Home ownership at age 1		
Owns Home ownership at age	7415872	59%
Pays rent	4332132	35%
Neither owns Home ownership at age nor pays rent	759936	6%
Home ownership at age 2		
Owns Home ownership at age	7686040	61%
Pays rent	4106822	33%
Neither owns Home ownership at age nor pays rent	786514	6%
Home ownership at age 3		
Owns Home ownership at age	8084627	64%
Pays rent	3936869	31%
Neither owns Home ownership at age nor pays rent	558361	4%
Home ownership at age 4		
Owns Home ownership at age	8150888	65%
Pays rent	3771127	30%
Neither owns Home ownership at age nor pays rent	624708	5%
Home ownership at age 5		
Owns Home ownership at age	8129931	65%
Pays rent	3715807	30%
Neither owns Home ownership at age nor pays rent	587669	5%
Home ownership at age 6		
Owns Home ownership at age	8088594	65%
Pays rent	3651948	30%
Neither owns Home ownership at age nor pays rent	616658	5%
Home ownership at age 7		
Owns Home ownership at age	8383303	69%

Table 7: Weighted summary statistics (complete set of categorical circumstances) *(continued)*

Circumstance	Obs	Percent
Pays rent	3267338	27%
Neither owns Home ownership at age nor pays rent	547485	4%
Home ownership at age 8		
Owns Home ownership at age	8299832	69%
Pays rent	3097225	26%
Neither owns Home ownership at age nor pays rent	670268	6%
Home ownership at age 9		
Owns Home ownership at age	8403938	70%
Pays rent	3139740	26%
Neither owns Home ownership at age nor pays rent	492647	4%
Home ownership at age 10		
Owns Home ownership at age	8840206	73%
Pays rent	2701642	22%
Neither owns Home ownership at age nor pays rent	512271	4%
Home ownership at age 11		
Owns Home ownership at age	9016298	73%
Pays rent	2962200	24%
Neither owns Home ownership at age nor pays rent	429075	4%
Home ownership at age 12		
Owns Home ownership at age	9377941	73%
Pays rent	2992754	23%
Neither owns Home ownership at age nor pays rent	404809	3%
Home ownership at age 13		
Owns Home ownership at age	9811303	75%
Pays rent	3035222	23%
Neither owns Home ownership at age nor pays rent	314202	2%
Home ownership at age 14		
Owns Home ownership at age	10016462	75%
Pays rent	3008839	23%
Neither owns Home ownership at age nor pays rent	283998	2%
Home ownership at age 15		
Owns Home ownership at age	8464831	75%
Pays rent	2549165	23%
Neither owns Home ownership at age nor pays rent	290801	3%
Home ownership at age 16		
Owns Home ownership at age	8786947	75%
Pays rent	2560188	22%
Neither owns Home ownership at age nor pays rent	342291	3%
Home ownership at age 17		
Owns Home ownership at age	6758612	77%
Pays rent	1769939	20%
Neither owns Home ownership at age nor pays rent	263502	3%
Home ownership at age 18		
Owns Home ownership at age	7338331	77%

Table 7: Weighted summary statistics (complete set of categorical circumstances) *(continued)*

Circumstance	Obs	Percent
Pays rent	1888845	20%
Neither owns Home ownership at age nor pays rent	263773	3%
Car ownership in the family at age ownership in the family at age 1		
Yes	11379770	91%
No	1128170	9%
Car ownership in the family at age 2		
Yes	11337174	90%
No	1242202	10%
Car ownership in the family at age 3		
Yes	11309831	90%
No	1270026	10%
Car ownership in the family at age 4		
Yes	9533473	91%
No	995655	10%
Car ownership in the family at age 5		
Yes	7948979	93%
No	641537	8%
Car ownership in the family at age 6		
Yes	5632808	92%
No	514386	8%
Car ownership in the family at age 7		
Yes	3726618	92%
No	325576	8%
Car ownership in the family at age 8		
Yes	1858914	94%
No	125267	6%
Car ownership in the family at age 16		
Yes	2107516	92%
No	176059	8%
Car ownership in the family at age 17		
Yes	1991758	98%
No	41043	2%
Car ownership in the family at age 18		
Yes	4639668	94%
No	322295	6%

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, 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.

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.

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.

Fabian T. Pfeffer, Ludwig Maximilian University (LMU) of Munich, University of Michigan Survey Research Center Davis Daumler, and University of Michigan Survey Research Center Esther Friedman. 2025. "PSID-SHELF, 1968–2021: The PSID's Social, Health, and Economic

Longitudinal File (PSID-SHELF), Beta Release." Inter-university Consortium for Political; Social Research (ICPSR). https://doi.org/10.3886/E194322V2.

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.

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.