Erasmus University Rotterdam

Erasmus School of Economics

Bachelor Thesis - BSc2 Econometrics and Economics

July 3, 2022

The Delusion of the American Dream:
a Temporal Analysis of Inequality of Opportunity in Income in the USA Between 2002 and 2021.

Author:
Dora Bijvoet

Supervisor:
Prof. Dr. PHBF Franses

Student Number:
493166

Second Assessor:
Prof. Dr. Ir. JC van Ours

Abstract

The United States of America (USA) have long been considered the “land of opportunities”. However, in recent decades, many have felt the USA fell short of this reputation. This paper examines the extent of its failure by evaluating Inequality of Opportunity (IOP) in income. Results are especially relevant for policy-makers to identify the sources of inequalities in the USA, and better target these. Data is collected from the General Social Survey in the USA between 2002 and 2021. I decompose IOP into the relative contribution of 10 different circumstances of birth, as well as the interaction of some using the Shapley decomposition. Using parametric and non-parametric models, I provide a lower- and upper-bound estimate of IOP. Moreover, different model specifications and inequality measures allow me to argue for the robustnesss of my findings. The parametric models estimating IOP based on the Gini coefficient is found to be the preferred approach. I find that IOP in income represents, at the least, between 40% and 55% of total income inequality and at the most between 95% and 96%. Sex and age are, on average, the two most important contributors. Covid-19 exacerbated the contribution share of sex and parental socioeconomic status as they caused 33.08% and 15.43% of IOP in 2021, making them the two most important drivers. Overall, since 2016, parental socioeconomic status, family income, and parental education significantly increased in importance. During the same time, the contribution of race decreased by 5 percentage points, making it the third to last contributor in 2021. However, this estimate may have been understated. Finally, I found that in 2014 black individuals having lived in the South at age 16 faced more inequalities than if they had lived elsewhere.

Acknowledgements

Before anything else, I would like to mention a few people without whom I could not have written this thesis, and without whom these past four years would not have been possible.

First, I would like to thank my supervisor Prof. Dr. Philips Hans Franses for guiding me throughout this process and taking the time to answer my questions in our weekly meetings.

Next, I would like to mention Greta, with who it has been a pleasure to study, nearly every day, during the past few months.

I would also like to express my deepest gratitude to Hidde, Mariska, and my friends, Sacha, Rosie, Alexandrine, and Max for always standing by me and helping me through the stressful times. Without you, these past four years would not have been the same. A special thanks to Hidde for all the love and laughs, to Mariska and Sacha for taking care of me when I got home, and to Rosie for being my best friend.

Finally, I would like to thank my parents, my Opa and Oma, and the rest of my family for supporting me throughout my bachelor's and always believing in me. Without their confidence, I could not have even hoped to graduate with a double bachelor's in Econometrics and Economics from the Erasmus University Rotterdam.

Contents

1 Introduction

The American Dream is a concept deeply rooted in the history of the United States of America (USA). It was first coined by Adams (1931) in *The Epic of America* and quickly became a national ethos. It encapsulates the ideals of freedom, prosperity, and equality, claiming that anyone and everyone can achieve success with hard work, regardless of social class or circumstances of birth such as race and gender. In fact, it relies entirely on the ideal of Equality of Opportunities (EO), affirming that America is "the land of opportunities". Initially, it was put forward by the Americans themselves, in an attempt to foster the immigration of wealth and talent to the US. However, quickly, both Americans and non-Americans started realizing that it was but an illusion. Since, harsh criticism has followed, both in academic literature as well as in the mainstream media through film and books. To this day it remains a heavily debated subject, which is increasingly relevant. Indeed, according to the estimates of Saez and Zucman (2016), in 2016 the top 1% held over 40% of the overall wealth in the United States (US), while it held "only" 25% in 1980. Even more worrying, the share held by the top 0.1% has doubled from 10% to 20% over the same time frame. Although Winship (2015) claims that there is no evidence that high or rising inequality would "chip away at the American Dream", I still wonder to what extent it is still intact. And if it is not, what are the circumstances of birth driving IOP in America?

It is scientifically relevant to investigate these questions as it will increase our understanding of IOP in income in the US by decomposing it into its exact sources, allowing us to identify which are the most important. Additionally, it is also of interest to understand how these levels have evolved over time, especially in recent decades. In turn, this is socially relevant as it will help policymakers to combat inequalities. Indeed, knowing the precise contributions of each circumstance to IOP may offer indications about how to best target inherit economic disadvantages. What is more, Bradbury and Triest (2016) found that IOP reduces economic growth. Therefore, it is essential for policymakers to develop opportunity equalizing policies: policies that compensate inequalities arising from factors beyond an individual's control while letting them be responsible for inequalities arising from factors within their control. This will only be possible once these factors can be identified.

Thus, I will use income as an indicator of success to estimate IOP. Whether or not this is a suitable measure is an entirely different, but equally valid, debate that I will be leaving to the realm of philosophy for now. In any case, this paper will seek to answer the following research question:

> *To what extent is inequality in income in the US attributable to differences in circumstances of birth, and what is the relative importance of each of these circumstances?*

To answer this research question, I will use the Shapley decomposition method to estimate the ex-ante measure of IOP in income. This method is based on Shapley values and is particularly suitable in distinguishing the relative contribution of circumstances, as it provides consistent estimates in the presence of multicollinearity. I will use parametric and non-parametric approaches to the Shapley decomposition to give an upper and lower bound of IOP.

The main findings are as follows. First, the lower bound of IOP relative to total inequality in income in the US varies between 40% and 55% between 2002 and 2021, while its upper bound varies between 95% and 96%. Over this period, IOP behaves cyclically with total inequality in income. Although IOP levels remain relatively constant, several interesting trends in the contribution share of circumstances appear. Sex and age are on average the two most important contributors to IOP in this period. After Covid-19, the contribution share of sex increased to explain 33.08% of IOP, while that of age decreased by more than 10 percentage points, to 13.14%. The importance of parental characteristics such as family income at age 16, parental education, and parental socioeconomic index especially, have been increasing between 2016 and 2021, placing them in the top 5 most important circumstances. On the contrary, the importance of race has decreased by 5 percentage points since 2016, making it the third to last contributor and representing only 3.50% of IOP. Similarly, the region an individual lived in at 16 was the third most important contributor in 2002, while it dropped to the sixth position in 2021, causing then 6.89% of IOP. Finally, in 2014, black individuals having lived in the South at age 16 faced more inequalities than if they had lived elsewhere, meaning that the interaction of these circumstances disproportionately drives IOP.

This paper contributes to the existing literature by extending the analysis of IOP and its driving sources to more recent years. Indeed, no work has studied IOP during the Covid-19 pandemic. This may be of great interest to assess the impacts of the sanitary crisis on prevailing and growing inequalities in the USA. Indeed, the most recent study evaluates IOP in income between 2005 and 2017 (Chantreuil et al., 2020). However, it focuses on just a few circumstances of birth: race, age, and education. To the best of my knowledge, this is the case for all existing literature on IOP in income in the US. Thus, with 10 circumstances and some interactions, I provide a much more relevant and accurate analysis of the situation in the US. Furthermore, my analysis also distinguishes itself from that of Wu et al. (2021) in Australia, the only paper I found which does include a large set of circumstances without grouping variables. Indeed, the sampling method I use for the Shapley decomposition has a stopping criterion based on the standard deviation of estimates and not the number of iterations, giving far more precise results. Finally, thanks to the parametric and non-parametric models, I provide an interval estimate of IOP, albeit very large.

In the subsequent section, I provide an overview of the existing literature focusing on the ideal of EO and on IOP in income specifically. Next, in section 3, I will introduce the data set used in my analysis and provide some descriptive statistics. Section 4 will outline my methodology, explaining the (sampling-based) Shapley decomposition, and the parametric and non-parametric models for estimating IOP. Finally, I will present and discuss my results in section 5, and conclude in section 6.

2 Literature review

2.1 The Ideal of Equality of Opportunity

The ideal of equality of opportunities (EO) was first formulated in the political philosophy literature (Dworkin, 1981a, 1981b). There exist many different definitions of EO and how to reach it. Nevertheless, the common consensus remains that EO refers to a situation in which individuals have the freedom to attain any desired outcome without any restriction due to factors or characteristics beyond their control (Dworkin, 1981b; Roemer, 1998). More recently, Roemer (1998) translated this idea into formal economic models based on the assumption that individuals' advantageous outcomes, such as income, are determined by (1) variables within their control and (2) variables beyond their control. These variables are coined as effort and circumstances respectively. Inequality caused by differences in circumstances is considered unethical, while inequality caused by differences in levels of effort exerted is fair and to be expected. Thus, ideally, inequality in outcome should arise exclusively from variations in individual effort. The danger of the alternative - inequality of opportunities - is that certain segments of the population systematically be disadvantaged. This refers to the two fundamental principles of justice: the *principle of reward* and the *principle of compensation*. The first states that individuals exerting different levels of effort should be rewarded accordingly in terms of outcome, while the second states that individuals disadvantaged by their circumstances of birth should be compensated in terms of outcome. The latter relates to inequalities of opportunities.

Thus, to reach this ideal of EO, the challenge lies first in the identification of the current level of inequality of opportunity, then that of its sources, and finally, the development of opportunity equalizing policies. Several measures have been developed to identify IOP. However, in the empirical literature, two measures stand out: the *ex-ante* and the *ex-post* methods (Checchi & Peragine, 2010), which we will define in more detail in the Methodology section. As aforementioned, it is in the interest of policymakers to identify the sources of IOP in income as well as their relative contributions. Therefore, several inequality index decomposition methods have been established. In this paper, I will focus on the Shapley decomposition developed by Shorrocks (2013). The Shapley decomposition is based on the Shapley value from cooperative game theory which attempts to allocate gains from cooperation to agents in a fair way (Sastre & Trannoy, 2002). Similarly, the Shapley decomposition identifies the marginal contribution of each factor to overall inequality. Thus, we can decompose inequality of opportunity into the relative importance of each circumstance.

2.2 Empirical Evidence

A number of studies have found evidence for inequality of opportunity in income[1] outside of the US, both in developed as well as developing countries.

Checchi and Peragine (2010) focus on IOP caused by differences in family backgrounds in

[1] Meaning inequalities in income caused by differences in circumstances of birth.

Italy between 1993 and 2000. Findings estimate IOP to represent around 19.5% of total inequality in this period. Although IOP is not decomposed into individual circumstance contribution, different IOP levels across regions and genders can be observed. Namely, the regions in the South suffer from a higher degree of IOP than those in the North. This is especially the case for women.

A similar study of IOP is performed in Latin America and the Caribbeans using data from 1995 to 2005. There, IOP is found to account for 20% to 34% of overall earnings inequality, with the lowest estimate found in Colombia and the highest in Brazil (Peragine, 2011). Additionally, when using household income and consumption per capita instead of personal income, IOP in Guatemala rises to a whopping 50%. This large difference may be due to a lack of personal income data. Overall, the circumstances contributing most to IOP were family background variables, in particular mother's education and father's occupation.

Further research has shown significantly smaller estimates of IOP in Australia and Sweden. Recently, (Wu et al., 2021) had a novel approach and determined both a lower and upper bound of IOP in Australia, estimating it to be between 5% and 20%. Gender was found to have the largest influence, contributing to 5% of IOP. Other significant circumstances included, in order of importance, parental occupation, age, and the number of siblings. Notice that, unlike the large majority of the empirical literature, (Wu et al., 2021) treats age as a circumstance and not a demographic variable[2], thus estimates of IOP may be biased upward. Regardless, Sweden is by far the "fairest" country in terms of inequality, as IOP has been consistently decreasing over the sample period and represents at most a third of total income inequality (Björklund et al., 2012). Thanks to the impressive amount of data available and collected, Björklund et al. (2012) have perhaps performed the most extensive and accurate estimation of IOP and its sources. Among these, Intelligence Quotient (IQ) and parental income are the main sources. Interestingly, Björklund et al. (2012) deliver an unusual interpretation of this last finding, arguing that, to a certain extent, the contribution of parental income can be seen as meritocratic. Indeed, parents "work hard" to attain higher levels of income and to be able to ensure the success of their children.

A few studies have also been performed within the US.

Pistolesi (2009) studies IOP between 1968 and 2001. He finds that the share of earnings inequality that is inherited varies between 18.6% and 43% across this period. This is surprisingly high, given that in a similar time frame, IOP was between 20% and 34% in Latin America and the Caribbean, countries which are considered far less developed. Nevertheless, we do observe a decreasing trend in the level of IOP in this period: in 1968, 33% of earnings inequality is inherited, while in 2001, this number dropped to 18.6%. Unfortunately, Pistolesi (2009) did not decompose IOP into the relative contribution of each circumstance variable. However, other studies do give some insight into the circumstances contributing to IOP in the US at that time.

[2] So does Chantreuil et al. (2020).

Marrero and Rodríguez (2011) find that race was the main contributor in the 1970s and 1980s, accounting for more than 50% of direct IOP. Direct IOP refers to inequality directly following differences in circumstances. On the other hand, indirect IOP refers to inequality following from differences in effort which in turn follow from differences in circumstances. Thus, the contribution of race to total IOP[3] is unclear. Later, in the 1990s and 2000-2007, parental education became the main contributor and the contribution of race decreased to between 5% and 10%.

Chantreuil et al. (2020) focus on the circumstances driving overall inequality in income from 2005 to 2017. This is equivalent to identifying the circumstances driving total IOP. In the overlapping periods of analysis[4], their findings seem to be in line with those of Marrero and Rodríguez (2011). Overall, according to Chantreuil et al. (2020), between 2005 and 2017, the share of overall inequality attributable to racial affiliation is at most 4%, while up to 13% of inequality is attributable to gender. Nonetheless, the contribution of race to income inequality has tended to increase from 2010 onward, while that of gender has decreased from 2005.

All in all, considering existing literature, it seems that the most important circumstances contributing to IOP in the US and the rest of the world are gender, parental characteristics (educational attainment, income, occupation, and socioeconomic status), number of siblings, and race. The latter is especially relevant in the USA. Understandably so, considering its long history with racism. Thus, I will also include these circumstances in my analysis.

Additionally, I consider the possibility of the existence of interactions between circumstances. Indeed, several findings in the literature hint toward the latter.

Marrero and Rodríguez (2011) are, to the best of my knowledge, the only authors who include cross effects in their analysis and study their relative contribution to IOP using decomposition methods. Although they found that the inclusion of quadratic and cubic terms left the IOP measure essentially unchanged, interacting parental education with race did modify the estimate towards the end of the sample period. In fact, this cross effect accounted for 35% of IOP in 2007. Additionally, Wu et al. (2021) find a gap between the parametric and non-parametric estimates of parents' occupation, suggesting interaction with other factors.

I am not only interested in investigating the existence of such cross effects, but also in investigating the possibility that the interaction of certain circumstances may contribute more to IOP than these circumstances would individually. Thus, I formulate the following sub-question:

Does the interaction of certain circumstances disproportionately contribute to IOP?

I expect to observe several disproportionately important cross effects in my analysis.

First, as Checchi and Peragine (2010) point out, IOP and intergenerational immobility seem to be more important at the extremes of the earnings distribution. In other words, the very poor

[3]Total IOP is the sum of direct and indirect IOP.

[4]The years 2005 to 2007.

stay very poor and the very rich stay very rich, because the advantages and disadvantages of the family background have more influence on their children's destinies. I thus hypothesize that parental characteristics such as parental education or parental socioeconomic status interact with parental income.

Second, similarly to Marrero and Rodríguez (2011), I expect there to be cross effects between race and parental education. Specifically, when the individual is black, I expect that parental education and other parental characteristics contribute more to IOP than average.

Finally, as Huffman and Cohen (2004) and decades of sociological research show, inequality between blacks and whites is greater in areas where the black-to-white ratio is high. Therefore, I expect to find a higher contribution of race to IOP when individuals grew up in one of these areas. Thus, there may exist cross effects between these variables.

Note that, following empirical evidence, it seems that the relative importance of these circumstances and cross effects, as well as overall levels of IOP, have changed quite a lot throughout the years (Björklund et al., 2012; Chantreuil et al., 2020; Marrero & Rodríguez, 2011; Pistolesi, 2009). Indeed, overall, IOP levels around the world have been consistently decreasing. Still, it is hard to pinpoint a definitive chronological trend for the relative contribution of individual circumstances. To investigate this trend further in the US specifically, we pose the following sub-question:

How have IOP and its driving circumstances evolved between 2002 and 2021?

Given the literature, I expect that IOP will continue to decrease over time and that the most important circumstances will be race, gender, and parental characteristics. Still, we consider that the Covid-19 pandemic might have impacted IOP in unusual ways. Indeed, Chantreuil et al. (2020) state that the increase in the contribution of race to overall inequality may originate from the financial and economic crisis of 2008 in which inequality between blacks and whites increased. Thereafter, during the recovery, it seems that this trend was not reversed, but rather emphasized. It is possible that the Covid-19 health and economic crisis may have had similar consequences. In fact, Kochhar (2020) states that blacks, Hispanics, and Asians reveal significantly higher unemployment rates than whites. Additionally, Dang and Viet Nguyen (2021) find that Covid-19 has already led to an increase in gender income inequality. Thus, we might observe an increase in IOP and the relative contribution of gender.

3 Data

3.1 Data Source

To examine the relative contributions of circumstances on inequality in income in the USA, I will use publicly available data from the General Social Survey (GSS) (Smith et al., 2021). As some variables are unavailable in other years, I will be focusing on data between 2002 and 2021.

The GSS is a national cross-sectional longitudinal survey of the independent research organization NORC at the University of Chicago. Conducted since 1972, it is a personal interview survey collecting information on a wide range of topics, including demographic characteristics, opinions, attitudes, and behaviors. It claims to be a nationally representative survey of adults in the USA and the best source to study sociological and attitudinal trends.

3.2 Variables

I have selected yearly personal income adjusted for inflation as an outcome variable. Income is a common measure considered by the empirical literature surrounding the topic of inequality of opportunity (Björklund et al., 2012; Chantreuil & Lebon, 2015; Checchi & Peragine, 2010).

Circumstances are defined as characteristics beyond an individual's control that may lead to unequal opportunities. Ideally, as many circumstances as possible should be included to avoid biased measures. Indeed, partial observability of circumstances leads to a lower bound of inequality of opportunity (Brunori, 2016). Thus, I have selected 10 circumstance variables describing individuals' socioeconomic background as well as some key characteristics which, according to the literature, may affect inequality of opportunity.

An overview of the circumstance variables and their description can be found in the table below.

Table 3.1. *Description of Circumstance Variables Considered in the Analysis*

Variable Name	Description	Classification	Possible Values
age	Respondent's age	Continuous	[18, 89]
sex	Respondent's sex	Dummy	0 = Male, 1 = Female
$race$	Respondent's race	Categorical	White, Black, Other
$born$	Was the respondent born in the USA?	Dummy	0 = Yes, 1 = No
$sibs$	Number of siblings (biological or not) during childhood	Discrete	$[0, \infty)$
$reg16$	Respondent's region of residence at 16 years old	Categorical	New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, Pacific, Foreign
$incom16$	Respondent's family income at 16 years old	Categorical	Far below average, Below average, Average, Above average, Far above average
$family16$	Did the respondent live with both parents at 16 years old?	Dummy	0 = Yes, 1 = No
$pareduc$	Highest year of school completed by either parents	Discrete	[0, 20]
$parsei10$	Highest socioeconomic index attained by either parents	Continuous	[0, 100]

For all variables listed in table 3.1, I eliminate from the sample the respondents which have not answered, have answered "Inapplicable" or "Do not know/Cannot choose", or for which the answer is unavailable. Additionally, some variable transformations have been made in order to better fit the analysis.

First, $family16$ was initially ordered into nine categories, indicating the family living situation. I will transform this variable into a dummy variable designating whether or not the respondent lived with both parents at age 16. As this variable will act as a proxy for the level of stability in the respondent's household, it will take on the value 1 if the respondent answered "mother & father", "mother & stepfather" or "father & stepmother", and 0 otherwise.

Next, I have decided to transform the initially discrete variable indicating the number of siblings into a categorical variable with three levels: "no siblings", "one or two siblings", and "more than three siblings". I did so because, each year, only a few respondents have a very large amount of siblings, but most have 2 or fewer. Thus, in order to avoid biased estimates because of these few outlier values, I have turned the number of siblings into a categorical variable.

The variable *pareduc* refers to the highest number of school years completed by either of the two parents. "Completed" means from which the parent has graduated. Elementary school is completed after 5 years, middle school after 8 years, high school after 12 years, a bachelor's degree is obtained after 16 years, a master's degree after 18 years, and a doctorate for any additional years. I prefer including this variable instead of both initial variables representing the father's and mother's education separately because it allows me to retain more observations in my sample. Besides, if I would not, the sample would have been biased, as all observations of individuals with only one parent would have been removed.

The same was done for the variable $parsei10$, which reflects the highest socioeconomic index of the two parents. This variable is based on the index established by Duncan (1961) which combines information based on income, educational data, and occupational prestige.

Finally, initially, I also wanted to include the dummy variable *immigrant_parents* indicating whether or not the respondent is from an immigrant family, as I think immigration is a sensitive topic in the US that might affect inequalities in income, especially during the Trump presidency[5]. However, no data is available for this variable in 2021. Thus, I would have had to sacrifice this year from my analysis. Given that the Covid-19 pandemic started in 2020, eliminating 2021 would have meant not being able to observe the effect of the pandemic on IOP in income, which is, in my opinion, very important. Additionally, I noticed that performing the analysis with or without the variable *immigrant_parents* almost did not affect the estimates of IOP. Therefore, I have decided to exclude this variable from my analysis.

3.3 Descriptive Statistics

Table 3.2 presents some summary statistics of the obtained data set.

[5] Trump's presidency lasted from 2017 to 2021.

Table 3.2. *Descriptive Statistics of Income and Circumstance Variables*

		2002	2004	2006	2008	2010	2012	2014	2016	2018	2021
Income		37970	38490	35574	42859	145100	38478	35564	35707	37924	43056
		(45456)	(35596)	(33600)	(69537)	(0)	(58081)	(34165)	(36911)	(37891)	(39947)
Age		41.24	42.04	41.70	42.89	52.20	43.12	44.10	44.25	44.44	48.16
		[18; 86]	[18; 86]	[18; 85]	[18; 89]	[34; 70]	[18; 86]	[19; 87]	[18; 89]	[18; 89]	[19; 89]
Born in the USA*		91.5	89.6	86.6	87.1	70.0	87.2	87.4	87.9	87.7	89.2
Stable family at 16*		83.1	84.9	84.0	83.2	95.0	82.9	79.8	79.4	77.7	84.4
*incom*16*	Extreme income	9.1	10.5	11.2	10.8	15	9.3	10.5	10.9	12.1	11.9
	Moderate income	90.9	89.4	88.7	89.2	85	90.8	89.5	89.2	87.9	88.1
pareduc		12.73	13.04	12.69	12.70	14.20	12.97	13.15	13.10	13.16	13.70
		[0; 20]	[0; 20]	[0; 20]	[0; 20]	[8; 20]	[0; 20]	[0; 20]	[0; 20]	[0; 20]	[0; 20]
*parsei*10		48.25	49.91	48.76	49.15	66.21	50.59	50.47	51.17	51.46	52.99
		[11.8; 92.8]	[11.8; 92.8]	[9.0; 92.8]	[12.6; 92.8]	[23.8; 92.8]	[12.6; 92.8]	[12.6; 92.8]	[12.4; 92.8]	[9.0; 92.8]	[10.6; 92.8]
Race*	White	79.8	79.4	73.1	76.4	85.0	76.1	75.1	74.5	72.5	80.4
	Black	13.8	13.3	13.1	13.8	0.0	14.2	15.2	16.5	16.3	10.6
*reg*16*	Northeast	20.3	17.7	16.1	16.6	10	16.9	19.7	18.7	17.3	18.3
	South	31.1	29.9	30.5	30.8	15	29.3	26.8	28	31.5	27.4
	Midwest	27.3	25.7	25.2	25.4	25	26.4	25.7	27.3	24.2	29.2
	West	15.1	19.3	18	18.4	30	17.5	17.9	17.7	18.1	25.1
Female*		51.5	49.7	51.3	47.9	25.0	51.0	51.5	50.7	53.0	52.4
Siblings		3.37	3.31	3.44	3.42	2.35	3.36	3.37	3.39	3.21	2.83
		[0; 26]	[0; 20]	[0; 32]	[0; 37]	[1; 5]	[0; 20]	[0; 21]	[0; 18]	[0; 24]	[0; 35]
Observations		1611	1522	1595	1073	20	1030	1380	1478	1249	2102

Notes. Standard deviation of income is given in round brackets. Minimum and maximum are given in square brackets as follows: [*minimum*; *maximum*]. Income standard deviation and mean are rounded to the nearest integer. All other values are rounded to the first decimal. Circumstances marked with a * are categorical variables, and their means are expressed in percentages. Not all levels are explicitly displayed but can be concluded to be the remainder share.

First, notice that the sample from 2010 is an outlier in all senses. It contains only 20 observations, while all the other year's samples contain between 1030 and 2102[6] observations. As a result, all variables have outlying means. The average income, for instance, is usually around 40000$, while in 2010 it was 145100$. Therefore, I will exclude the year 2010 from the rest of my analysis.

Besides 2010, the yearly samples are balanced. The mean of each variable remains relatively identical throughout the years and so does its range. The samples also seem to be representative of the US population in terms of all circumstances except for race. Indeed, the proportion of white individuals may be overstated in this sample. According to the US Census Bureau, in 2020, whites accounted for 61.6% of the US population (Jones et al., 2021), while they account for between 72.5% and 80.4% in this sample. Consider that this may lead to inaccurate estimates of the relative importance of race.

Finally, it should be noted that in all years, income shows very high standard errors (between 33600 and 69537). This shows that income values are spread largely around the mean, and thus points toward large inequalities in income.

[6] In 2021 surveys were increasingly conducted online and over the phone. As a result, more observations were collected, which is why the 2021 sample is larger than other years.

4 Methodology

I will now define the methodology used in my research. The following analyses should be carried out for each year available in the sample, to study the evolution of IOP in the USA.

4.1 Defining Inequality of Opportunity

By defining individual outcome, in this case, personal income, to be determined exclusively by circumstances and effort, Roemer (1998) can divide the population into a set of types. He does so by first dividing the population into groups of individuals sharing the exact same set of circumstances, and then further dividing these groups into subgroups of individuals exerting the same level of effort. These subgroups are what he calls “types” and thus contain individuals which share the same set of circumstances and exert the same level of effort. According to Roemer (1998), equality of opportunity is reached when all individuals can achieve the exact same outcome by exerting the same amount of effort, regardless of their set of circumstances. Ferreira and Gignoux (2011) have coined this definition as the strong definition of equality of opportunity as it is very demanding, requiring the distribution of outcome conditional on effort to be identical across types. One common measure of this definition of inequality of opportunity in the existing literature is the “ex-post” measure established by Checchi and Peragine (2010).

Unfortunately, this measure requires the availability of an effort variable. In our case with income, this is nearly impossible to find as all observable effort variables may very well be dependent on circumstances. Take for instance the number of hours spent studying during high school, a student whose parents have higher income and can afford private tutoring sessions may spend more time studying than classmates with lower-income parents. Therefore, like most of the empirical work on the subject (Ahmed et al., 2020; Björklund et al., 2012; Chantreuil et al., 2020; Wu et al., 2021), I will use a weaker definition which allows inequality between individuals exerting the same amount of effort, but who do not necessarily share the same set of circumstances. This definition makes effort irrelevant; thus, in my analysis types are only characterized by individuals sharing the same set of circumstances. Therefore, individuals belonging to the same type may exert different levels of effort. Equality of opportunity is achieved when the type-average income level[7] is equal across all types. One common measure of this definition of inequality of opportunity is the “ex-ante” measure (Checchi & Peragine, 2010).

4.2 Shapley Decomposition of Inequality of Opportunity

I now wish to measure and decompose ex-ante IOP using the Shapley decomposition. By decompose I mean isolate the contributions from individual circumstances to the overall level of IOP.

[7] The average income level of individuals sharing the same set of circumstances.

There exist several decomposition methods, however, there are certain properties that make the Shapley decomposition particularly attractive (Shorrocks, 2013). First, it is exact and additive. This means that we can accurately interpret the contribution of a factor as the proportion of IOP attributable to that factor. Additionally, this means that by summing up all circumstance contributions we obtain the total level of IOP. Second, it is order-independent. This is a significant asset of the Shapley decomposition. Indeed, unlike most decomposition techniques, the Shapley decomposition calculates the average contribution of a factor, over every possible subset of the factors. Thus, the Shapley values can be interpreted as the "expected marginal impact of each factor when the expectation is taken over all the possible elimination paths" (Shorrocks, 2013). In this way, the Shapley values give a more accurate estimate of circumstance contributions than other decomposition techniques. Last, but certainly not least, it provides reliable estimates even when circumstances are correlated.

As mentioned previously, the Shapley decomposition method is order-independent. Thus, we calculate the marginal contribution of the circumstance of interest k on income inequality for all possible combinations of circumstances.

Consider the entire set of circumstances K. We generate all unique permutations of K and define this set of permutations as Ω. The set Ω contains $|K|!$ elements. This amounts to $10! = 3628800$ permutations in my analysis. For each order of the circumstances r in Ω, we take the subset $S_r = \{C_i | i < k\}$, in other words all circumstances preceding circumstance k. We then compute the inequality measure $I(S_r)$, in which only the circumstances in subset S_r are "allowed to contribute" to inequality in income. Next, take the subset $S_r + k = \{S_r, k\}$ and compute the inequality measure $I(S_r + k)$. Then the Shapley contribution of circumstance k is:

$$SV_k = \frac{1}{|K|!} \sum_{r \in \Omega} (I(S_r + k) - I(S_r)^8. \tag{4.1}$$

4.3 Measuring Inequality in the Counterfactuals

Computing the inequality measures $I(S_r)$ and $I(S_{r+k})$ means we measure inequality in the counterfactual distribution of income when we allow circumstances S_r and S_{r+k} to contribute respectively. The counterfactuals refer to the distribution of inequalities once we have removed the part of inequalities due to effort variations, and only unfair inequalities remain. In other words, it is the distribution of inequalities of opportunities.

In this section, I will first describe how to obtain the counterfactual distribution of income, using the parametric and the non-parametric models. Specifically, I will explain how to handle circumstances that are "allowed" and "not allowed" to contribute. Next, I will compare the drawbacks and benefits of using the parametric and non-parametric models. Finally, I will

[8] Note that when using the variance as inequality index this is equivalent to computing the Shapley values of each variable in our model.

present the inequality measures used to measure inequality in income in the counterfactuals.

Note that, in the conventional framework, IOP is simply obtained by measuring inequality in income in the counterfactual distribution when all circumstances are allowed to contribute.

4.3.1 Obtaining the Counterfactual Distribution

Using the ex-ante definition of IOP, there exist two models to obtain the counterfactual distribution of income: the parametric introduced by Bourguignon et al. (2007) or the non-parametric introduced by Checchi and Peragine (2010).

4.3.1.1 The Parametric Model

In the parametric model we will estimate the following functional form:

$$y = g(C, E, u), \tag{4.2}$$

where C is the vector of circumstances, E is the vector of effort variables, and u is a random error term, or in other words, the variation due to unobserved factors. Note that, as aforementioned, effort may be dependent on circumstances, thus we have:

$$y = g(C, E(C, v), u), \tag{4.3}$$

where v is the random error term and can be considered as "true effort", while $E(C)$ is the indirect effect of circumstances on income. However, since effort is unobservable and I am not interested in the point estimates of effort variables, but simply in obtaining the counterfactual distribution, consider that "true effort" is comprised in the error term ϵ of equation 4.4.

Now, Sastre and Trannoy (2002) propose and compare two ways in which a circumstance can "not be allowed to contribute" to inequality in income in the counterfactual. There is the zero income distribution and the equalized income distribution. The former removes the part of income caused by the circumstances not allowed to contribute by setting them to 0, while the latter simply removes variation in the excluded circumstances by replacing them with their average value. Sastre and Trannoy (2002) find that the equalized income distribution method should be favored and so will I.

Thus, I estimate the following regression:

$$y = \bar{C}^{na\prime}\beta^{na} + C^{a\prime}\beta^{a} + \epsilon, \tag{4.4}$$

The term $\bar{C}^{na}$ is the vector of circumstances "not allowed" to contribute replacing each individual observation with the average value of each circumstance. The term C^{a} is the vector of circumstances "allowed" to contribute. Note that categorical circumstance variables are represented by a dummy for each level except the base level[9] in equation 4.4.

[9]Which level is the base level is irrelevant here as I am not interested in interpreting the estimated coefficients.

I obtain the vector of estimated coefficients $\hat{\beta} = (\hat{\beta}^{na}, \hat{\beta}^{a})$. Bear in mind that $\hat{\beta}$ cannot be interpreted as the causal effect of circumstances on income, since the variables may very well show signs of multicollinearity between each other or with unobserved circumstances in the error term (Ferreira et al., 2014) making the OLS regression results inconsistent. Therefore, the Shapley decomposition is especially suitable since, as aforementioned, Shapley values give reliable estimates, even in the presence of multicollinearity.

The counterfactual distribution of income is based on the fitted values of the regression in equation 4.4. This is because these fitted values represent the estimated part of income which is explained by circumstances. It can be seen as the minimum income that an individual with the set of circumstances $C_i = (\bar{C}_i^{na}, C_i^{a})$ can be expected to achieve. Therefore, inequality in $\hat{y}$ across the sample reflects inequality of opportunity.

Formally, for each individual i, we obtain the following counterfactual income:

$$\hat{y}_i = \bar{C}_i^{na\prime} \hat{\beta}^{na} + C_i^{a\prime} \hat{\beta}^{a}, \tag{4.5}$$

where the terms $\hat{\beta}^{na}$ and $\hat{\beta}^{a}$ represent the vectors of estimated coefficients of the circumstances "not allowed" and "allowed" to contribute respectively, and $(\bar{C}_i^{na}, C_i^{a})$ are defined as above.

Note that a weakness of this parametric approach is that the measure of inequality of opportunity will be a lower-bound estimate. This is mainly because there will inevitably be unobserved circumstances that may be wrongly attributed to true effort and randomness, i.e. the random error term (Ramos & Van De Gaer, 2012). Unfortunately, there is no way to gauge the magnitude of this bias.

4.3.1.2 The Non-parametric Model

In the non-parametric model, we make use of the "types" mentioned previously. As I use the weak definition of equality of opportunity, I divide the population into groups of people sharing the exact same set of circumstances, using only the circumstances "allowed to contribute".

In order to be able to do so, I first transform all continuous circumstance variables into categorical variables. As such, *age* now takes on seven levels: "24 or younger", "25 to 34", "35 to 44", "45 to 54", "55 to 64", "65 to 74", and "75 or older". The variable *pareduc* now takes on six levels: "no education completed" for less than 5 years of education, "elementary school completed" for between 5 and 7 years of education, "middle school completed" for between 8 and 11 years of education, "high school completed" for between 12 and 15 years of education, "bachelor completed" for between 16 and 17 years of education and "master or doctorate completed" for 18 years or more of education. Finally, the variable $parsei10$ now takes on four levels: "lower class" for an index of 25 or below, "middle class" for an index between 26 and 50, "upper middle class" for an index between 51 and 75, and "upper class" for an index of 76 and above.

Thus, the sample is divided into a set of types: $T = \{t_1, t_2, ..., t_K\}$. Assume that each type t_k for $k = 1, 2, ..., K$, contains n_k individuals. To obtain the counterfactual distribution, we set each individual's income to its type-average income $\bar{y}_{t_k}$. This gives the following counterfactual distribution of income:

$$\hat{y} = \{\mathbb{1}_{n_1}\bar{y}_{t_1}, \mathbb{1}_{n_2}\bar{y}_{t_2}, ..., \mathbb{1}_{n_K}\bar{y}_{t_K}\}, \tag{4.6}$$

in which $\mathbb{1}_{n_k}$ is a vector of ones of size n_k.

By eliminating the within types variations in income, only the between-type inequalities remain. These correspond to the inequalities due to differences in circumstances, or in other words: inequalities of opportunities.

The non-parametric approach assumes that circumstances and effort are independent of each other (Checchi & Peragine, 2010). As mentioned previously, this does not always hold. As such, this assumption is a weakness of the non-parametric approach.

4.3.1.3 Comparing the Parametric and Non-parametric Models

Wu et al. (2021) states that the non-parametric approach suffers from some issues. First, it suffers from the curse of dimensionality, as the number of types grow to be extremely large when the number of circumstances increases. In my case, there are $7 * 6 * 4 * 2 * 2 * 2 * 3 * 3 * 9 * 5 = 544320$[10] types when using the entire set of circumstances, while samples vary between 1030 and 2102 observations across years. Thus, a much larger data set than available is necessary when including all circumstances, or estimates will be biased. Contrarily, the parametric approach is much more parsimonious as it is based on a linear regression of circumstances on income.

Second, types containing few observations make it hard to capture variations, and so, disentangle the impact of circumstances and effort. Once again, in my case, it is likely that many types will be either empty or very sparse. I choose to simply remove empty types, as I cannot assign them an arbitrary average income without biasing results. This means that it is likely that even when using only a subset of circumstances, estimates will be biased.

On the other hand, contrarily to the non-parametric approach, the parametric approach does not take into account non-linearity and may thus over-smooth the data. Consequently, the non-parametric estimate can be considered an upper bound of IOP. As the parametric estimate represents the value of IOP if circumstances had a linear effect on income, it can be considered a lower bound of IOP.

Note that, summing up the contribution of each circumstance from the Shapley decomposition using the parametric model gives the exact same estimate of IOP as that obtained in the conventional framework. However, this is not the case for the Shapley decomposition with the non-parametric model. Indeed, we obtain a less biased and overstated estimate of IOP (Wu

[10] I obtain this number by multiplying the number of levels of each circumstance with each other.

et al., 2021). This is because, in the Shapley decomposition, there will be only one combination for which we use all circumstances to divide the population into types. In this case, many types will be empty or very sparse, giving an extremely unequal distribution of incomes. For all other combinations, we use only a subset. This means that only the inequality estimate of this single combination will be as biased and overestimated as the conventional IOP measure. Of course, the combination which is the empty set of circumstances will give an equally biased and the most underestimated measure of IOP. Nevertheless, by taking the average of these IOP measures over all the possible combinations, we obtain a more reasonable and less overestimated measure of IOP.

Still, even when using the Shapley decomposition, given the incredibly high number of types relative to the sample sizes, I think the non-parametric estimates of IOP will still be very biased and overstated. Thus, although I will compute the upper bound of IOP through the non-parametric approach, I think the latter will be very high and therefore useless. As such, I prefer using the parametric estimates as a reliable measure of IOP.

4.3.2 Measuring Inequality in the Counterfactual

Finally, we measure the level of inequality in the counterfactual distributions obtained:

$$IOP_{S_r} = I(\hat{y}), \tag{4.7}$$

where $I(.)$ is an inequality index and $\hat{y}$ is the counterfactual distribution obtained in equations 4.5 or 4.6 when the circumstances S_r are allowed to contribute. Thus, IOP_{S_r} is the measure of IOP when only the circumstances S_r contribute to inequality. It is equivalent to $I(S_r)$ in equation 4.1.

There exist a number of different inequality indices. I propose to use the Gini coefficient and the Mean Logarithmic Deviation (MLD) index since they both fulfill the principles of (1) symmetry (the inequality level does not change if two individuals switch incomes), (2) population invariance (the inequality level does not change if the population is "cloned"), (3) scale invariance (the inequality level does not change if all incomes are scaled by a common factor), and (4) the Pigou-Dalton transfer axiom (the inequality level increases if income is transferred from a poorer individual to a richer individual) (United Nations Development Programme, 2019). Additionally, they are universally known inequality indices.

Following the methodology proposed by Checchi and Peragine (2010), most of the empirical literature on IOP uses the MLD index. The MLD is equivalent to the generalized entropy index when its parameter is set to 0. It is defined as follows:

$$MLD(y) = \frac{1}{N}\sum_{i=1}^{N} ln\frac{\mu}{y_i}, \tag{4.8}$$

where N is the sample size, y_i is the level of income of individual i, and μ is the average level of income.

The Gini coefficient is used less frequently in papers evaluating IOP. It is defined as:

$$Gini(y) = \frac{1}{N}(N + 1 - 2\frac{\sum_{i=1}^{N}(N + 1 - i)y_i}{\sum_{i=1}^{N} y_i}). \tag{4.9}$$

Nevertheless, the MLD index has two main limitations. It is unbounded and less intuitive than other measures such as the Gini coefficient. Indeed, the Gini coefficient is always between 0 and 1 and represents the extent[11] to which the distribution of incomes deviates from a perfectly equal income distribution.

4.4 Extending the Parametric Model with Interacting Circumstances

According to the literature review, the conjugation of certain circumstances may provide particularly unfair opportunities. Thus, we subsequently include interaction terms in our initial model defined in equation 4.4.

Particularly, I first would like to study the implications of being black and growing up in a region with a high black-to-white ratio on having equal opportunities. The US census bureau states that the South is the region with the highest ratio (McKinnon, 2001). Thus, I am interested in black individuals who were living in the South at age 16. I identify individuals with this conjugation of circumstances by interacting the dummy variable corresponding to being black (i.e. $black$) to the dummy variable corresponding to growing up in the South (i.e. $South16$). I expect that these individuals will face more inequalities in income than the average black individual or the average individual who grew up in the South. If this is the case, I expect the contribution of this interaction term to be larger than the contribution of $black$ or $South16$ respectively.

Second, I would like to evaluate whether differences in parental characteristics, such as parental education or parental socioeconomic index, lead to more differences in income for black individuals than for the average individual regardless of their race. In other words, whether parental characteristics contribute more than average in determining income among black individuals. Thus, I interact the dummy $black$ with the variables $pareduc$ and $parsei10$ respectively. If parents' socioeconomic class or parental education is more determinant than average for a black individual's income, then I expect the contribution of these interaction terms to be higher than the contribution of the variables $parsei10$ and $pareduc$ respectively.

Finally, I would like to analyse the level of intergenerational immobility at the extremes of the earnings distribution. Intergenerational immobility can be defined as the extent to which parental characteristics are determinant for an individual's social class. I use income as a proxy for social class. Define the dummy variable $extremeIncom16$ as the variable which takes on value 1 when $incom16$ is either "far above average" or "far below average" and 0 otherwise. This variable corresponds to individuals whose parents, at age 16, were at the extremes of

[11] In percentages, if the Gini coefficient is multiplied by 100.

the income distribution. I interact this variable with the variables $pareduc$ and $parsei10$. If intergenerational immobility is indeed stronger at the extremes of the earnings distribution, then I expect the contributions of these interaction terms to be higher than the contributions of $pareduc$ and $parsei10$ respectively.

Thus, we evaluate these expectations formally by performing the parametric Shapley decomposition of IOP in income using the set of circumstances C_{inter} defined as follows:

$$\begin{aligned} C_{inter} = (C,\ &black * South16,\ black * pareduc,\ black * parsei10, \\ &extremeInccom16 * pareduc,\ extremeIncom16 * parsei10), \end{aligned} \tag{4.10}$$

where C is the original set of circumstances used.

4.5 The Sampling-based Shapley Decomposition

One significant drawback of the Shapley decomposition is its computational burden. To avoid this burden, Björklund et al. (2012) group selected circumstance variables and decompose IOP according to the contribution of these groups. Obviously, this sacrifices some level of detail.

Thus, I apply a sampling-based approach to the Shapley decomposition (Castro et al., 2009) using and adapting the publicly available R package written by Elbers (2021). In this approach, I randomly select m subsets of circumstances S_r. For each iteration i over the m subsets[12], I compute the marginal contribution of the circumstance of interest k:

$$MC_i = I(S_{r+k}) - I(s_r), \tag{4.11}$$

and the mean of all previously sampled marginal contributions:

$$\mu_i = \frac{1}{i}\sum_{j=1}^{i} MC_j. \tag{4.12}$$

Thus, at the l^{th}[13] iteration, I obtain a vector of means $\mu = (\mu_1, ..., \mu_l)$.

The total number of subsets m to sample is defined by a stopping criterion based on the standard deviation of μ. Indeed, once it is smaller than 0.0001, the algorithm stops sampling and the Shapley value is given by μ_m. This stopping criterion makes our sampling approach more efficient than that of Wu et al. (2021), who somewhat arbitrarily set $m = 10000$ based on simulation results using 8 circumstances.

I use this sampling-based approach only for the parametric models. Indeed, by definition, the non-parametric approach will always have high standard deviations in my analysis, as it measures inequality over relatively sparse types. Thus, the marginal contribution of circumstance k will be relatively different given the subset of circumstances considered, and μ will inevitably have high standard deviations. For the non-parametric Shapley decomposition, I adapt and use the exact Shapley decomposition package by Elbers (2021).

[12] Where $i = 1, 2, ..., m$.
[13] Where $1 < l \leq m$.

5 Results and Discussion

5.1 Inequality in Income in the USA between 2002 and 2021

Figure 5.1 shows the evolution of income inequality in the USA between 2002 and 2021 using the Gini coefficient and the MLD index.

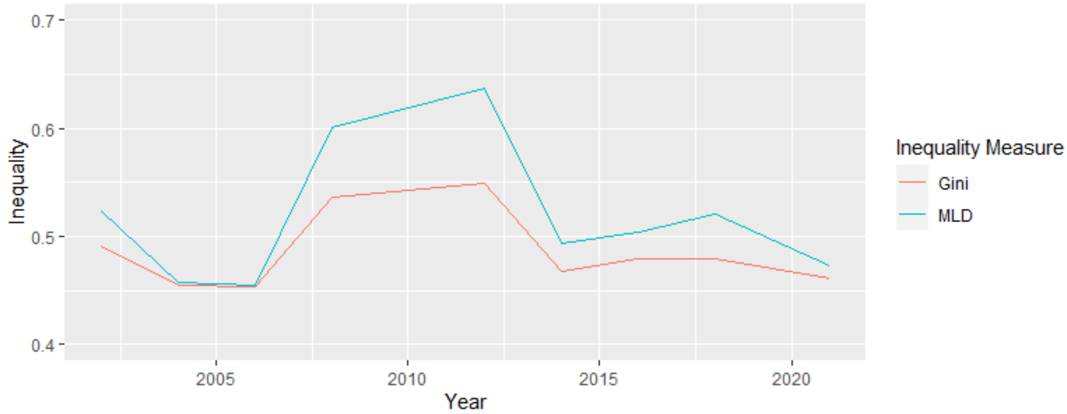

Figure 5.1. Inequality in Income in the USA Using Different Inequality Measures (2002-2021)

Looking at figure 5.1, we observe similar trends in inequality using both the Gini coefficient and the MLD index. The exact values of these inequality measures can be found in table A.1. Nevertheless, as mentioned in the Methodology, the Gini coefficient is a more intuitive measure of inequality, thus we will focus primarily on this index.

Figure 5.1 shows that inequality in income in the US decreased from 2002 to 2004 from 49.06% to 45.42%. It remained constant until 2006 but was followed by a significant increase up to 2012, reaching 54.86% of inequality. This concurs with the period of the 2008 financial crisis and may be a consequence of it. Unfortunately, due to many missing values, I have had to exclude 2010 from the analysis, thus, as frequently done in the literature, this observation has been replaced in the figure by the midpoint between the neighboring values[14]. As such, it is unclear how inequality evolved between 2008 and 2012. In 2014, inequality reverted back to 46.84% and remained relatively constant until 2018. Thereafter, there has been a small decrease, reaching levels of inequality close to those of the 2004-2006 period.

According to the World Bank (2019), the Gini index varied between 40% and 41.5% during this period. Thus, it seems that the sample used in my analysis overstates income inequality between 2002 and 2021. Indeed, as mentioned in the Data section, the standard deviation of income in most years of the sample used in my analysis is relatively high, pointing towards large inequalities in income. Still, in both cases, inequality in the US is significantly higher than in other western European countries and, since 2016, comparable to levels found in Bulgaria, the country with the most income inequalities in Europe (Eurostat, 2021).

[14]This is the case for all figures from here onward, but will not be mentioned again.

Additionally, general trends also seem to be significantly differing: according to the World Bank (2019), inequality increased between 2004 and 2006, decreased during the 2008 financial crisis, and has mostly been increasing after that. Moreover, literature shows that recessions do not usually lead to a significant increase in income inequality (Camacho & Palmieri, 2019). Thus, one must beware when interpreting the following results as they may not be representative of the entire US population.

5.2 Inequality of Opportunity in Income in the USA between 2002 and 2021

The estimates of the sampling-based Shapley approach to IOP based on the Gini and MLD indices are plotted in figures 5.2 and 5.3 respectively. The exact values can be found in tables A.2 and A.3 of the appendix. The absolute estimates can be found in tables A.4 and A.5 of the appendix.

As mentioned in the Methodology, I am able to measure IOP using the entire set of circumstances thanks to the sampling-based approach. The parametric model provides a lower bound of IOP, while the non-parametric model provides an upper bound. This can clearly be seen in the two figures below. Once again, regardless of the index chosen, relative IOP trends are nearly identical. Notice also that these trends are similar to those found in figure 5.1. Thus, IOP seems to increase and decrease along with total inequality. This points towards the efficiency of inequality-reducing policies in the US, as they seem to reduce inequalities by also targeting inequalities of opportunity and not only ethical inequalities due to effort variations. This was not always the case as Pistolesi (2009) found a negative correlation of -0.6 between IOP and inequality in income in the US between 1968 and 2001.

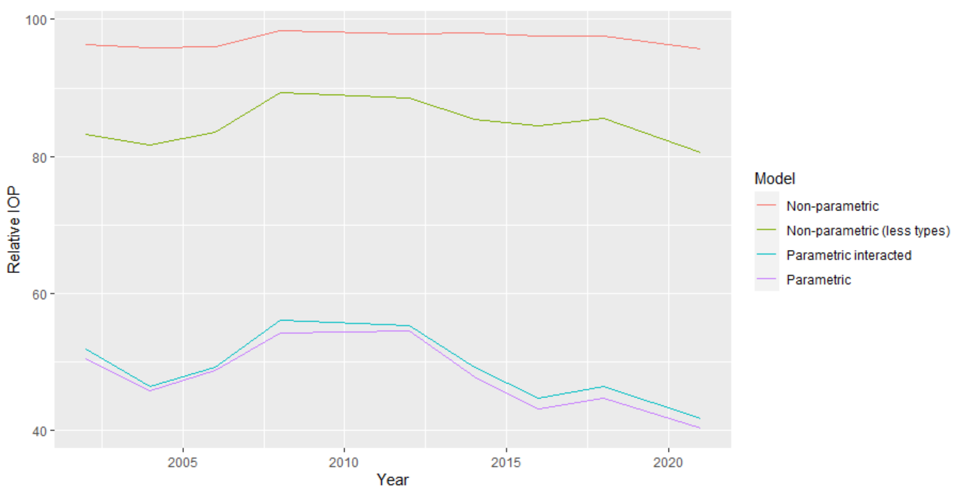

Figure 5.2. Estimates of IOP Relative to Total Inequality in Income in the USA Using the Gini Coefficient (2002-2021)

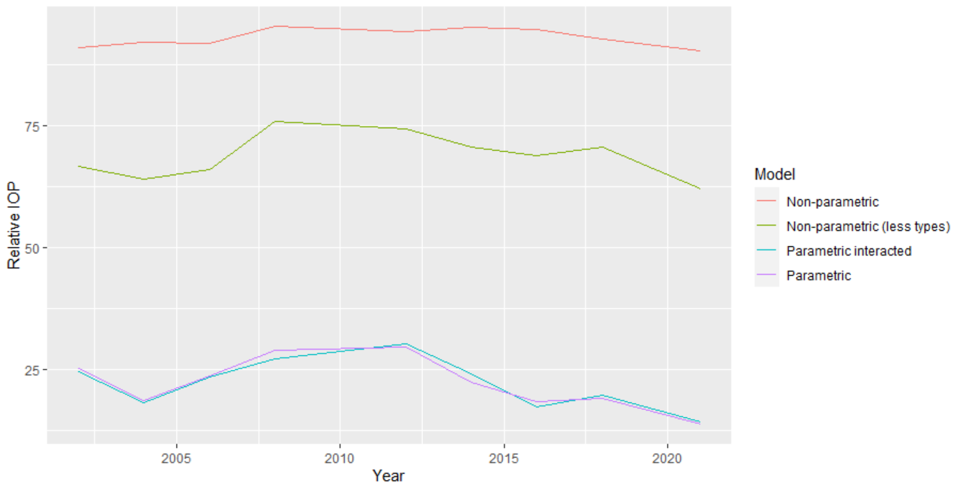

Figure 5.3. Estimates of IOP Relative to Total Inequality in Income in the USA Using the MLD Index (2002-2021)

I focus now specifically on the parametric results. Throughout the period of 2002 to 2021, the lower bound of IOP using the Gini coefficient is at most around 55% of total income inequality (between 2008 and 2012), and at the least around 40% (in 2021). Using the MLD index, estimates are much lower. The lower bound of IOP is at most around 29% of total income inequality (between 2008 and 2012), and at the least around 13% (in 2021).

Overall, levels of IOP remained relatively constant. Nonetheless, they have been decreasing since 2012 and are now lower than ever. This is an encouraging sign for the US as a whole. Yet they are still far behind in terms of social development, as they show much higher levels of IOP than similar economically developed Western countries. Indeed, comparing the MLD-based IOP estimates: Wu et al. (2021) found IOP in income to be at least 9% in 2017 in Australia, while I estimate IOP to be at least around 19% in the same period. This gap is not recent, as, in 2009, Bourguignon et al. (2007) estimated IOP to be 20% in Colombia, while I estimate it to be at least around 29% in the same period.

Next, I focus on the non-parametric results. As mentioned in the Methodology, the non-parametric Shapley estimation should result in a less overestimated and biased upper bound measure of relative IOP than the conventional non-parametric estimation (Wu et al., 2021). However, as expected, due to the incredibly large amount of types, IOP using both the Gini and MLD indices remains substantially higher than the parametric estimations. Indeed, using the Gini coefficient, the upper bound of relative IOP is estimated to be between 95% and 96% over the entire period, and between 90% and 95% using the MLD index. In both cases, there is a subtly increasing trend before 2008, followed by a plateau until 2014, and after that, a subtly decreasing trend. Overall, this coincides with the general trend found in the parametric estimation of IOP, pointing toward the robustness of my previous findings.

The theory mentioned above, explaining the overestimation of the non-parametric results, is confirmed by the non-parametric estimation of IOP using fewer types. In this model we reduce the number of levels for the variables age, $pareduc$, $parsei10$, $reg16$ and $incom16$, such that we obtain $3 * 3 * 3 * 2 * 2 * 2 * 3 * 3 * 5 * 3 = 29160$ types. In the figures above, it can be seen that for both the Gini and MLD indices, IOP estimates drop to a much lower level, albeit still very high: between 80% and 89% for the Gini coefficient and between 64% and 75% for the MLD index. Additionally, the trends in IOP over time resemble even more that of the IOP results from the parametric models. This once again points to the robustness and accuracy of my previous findings. All in all, this shows that although my non-parametric estimates are overestimated, they are consistent.

Nevertheless, they remain a useless upper bound. Additionally, given the large number of circumstances used, I consider the bias in the parametric model due to unobserved circumstances to be relatively small. Thus, from here onward I will focus primarily on the parametric analyses and consider the parametric estimates a reasonable measure of IOP.

Notice that, even when considering only the parametric estimates, my results are significantly higher than those found by Pistolesi (2009) and Marrero and Rodríguez (2011). In 2001, Pistolesi (2009) estimated IOP to be around 24% using the Gini index and around 16% using the MLD index. In 2002, I found IOP to have a lower bound of 50.4% using the Gini index and 25.2% using the MLD index. Furthermore, based on the MLD index, Marrero and Rodríguez (2011) estimated IOP to be between 4% and 7% in the period of 2002 to 2007, while I found a lower bound between 25.2% and 29%. In both cases, the difference between these lower numbers and my parametric results can be explained by the fact that the authors consider significantly fewer circumstances than I do. Indeed, Marrero and Rodríguez (2011) consider only parental education and race, while Pistolesi (2009) considers only parental education, parental occupation, race, place of birth, and age.

Finally, observe that, regardless of the model used, relative IOP estimates based on the MLD index are much lower than IOP estimates based on the Gini coefficient. Brunori (2016) suggests that this is a direct consequence of the higher sensitivity of the MLD index to outliers. When constructing the counterfactual distributions of income, by definition one removes the extreme values from total inequality. Thus, this smoothing means that an index such as the MLD will "detect" less inequality. In light of this property, from here onward I prefer relying on the Gini-based results of IOP.

Thus, comparing the relative Gini-based IOP estimates of the parametric interacted model to the parametric model in figure 5.2, it can be seen that the IOP estimates of the interacted model are consistently around 1% higher than the IOP estimates of the non-interacted model. This indicates that at least some of the interactions mentioned in section 4.4 do significantly contribute to IOP. Unfortunately, this finding is sensitive to the inequality index used, as it does not always hold when using the MLD index.

5.3 Decomposing Inequality of Opportunity in Income in the USA between 2002 and 2021

The results of the parametric and non-parametric sampling-based Shapley decompositions using Gini coefficients are presented in figures 5.4 and 5.5 respectively.

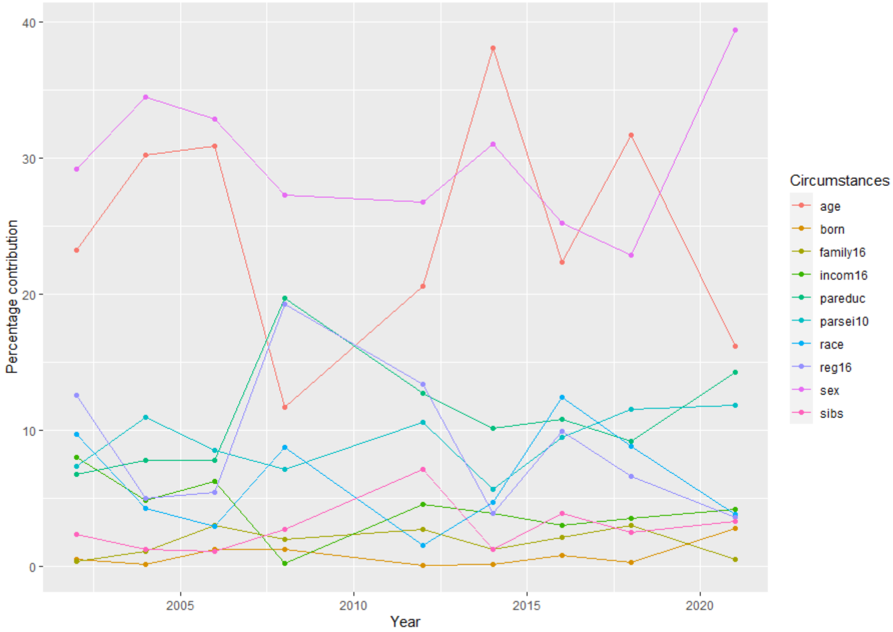

Figure 5.4. Parametric Estimates of the Relative Contribution of Circumstance Variables to IOP in the USA Using the Gini Coefficient (2002-2021)

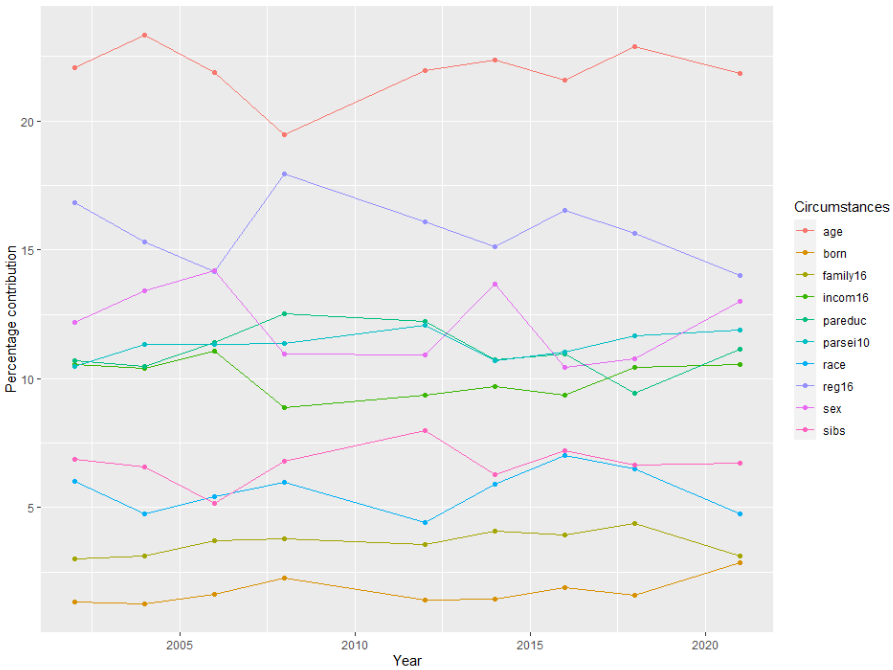

Figure 5.5. Non-parametric Estimates of the Relative Contribution of Circumstance Variables to IOP in the USA Using the Gini Coefficient (2002-2021)

The parametric and non-parametric sampling-based Shapley decompositions using the MLD index are presented in figures A.1 and A.2 of the appendix respectively. The exact values of both Gini- and MLD-based parametric models are presented in tables A.6 and A.7 of the appendix respectively. Table 5.1 below gives the relative contribution of circumstances to total inequality in income, allowing me to compare my results to those of other authors regardless of the number of circumstances used.

Table 5.1. *Parametric Estimates of the Relative Contribution of Circumstance Variables to Total Inequality in Income Using the Gini Coefficient (2002-2021)*

	2002	2004	2006	2008	2012	2014	2016	2018	2021
age	10.13	11.34	12.73	6.53	10.17	15.03	8.27	11.14	5.29
born	0.30	0.09	0.45	1.11	0.21	0.19	0.55	0.31	1.26
family16	0.63	0.96	1.56	1.40	1.77	1.28	1.34	1.85	0.44
incom16	5.64	3.68	4.52	2.10	4.72	3.11	1.88	3.39	3.16
pareduc	3.94	3.85	4.03	9.03	6.72	4.27	4.85	4.02	4.85
parsei10	4.49	6.16	5.54	4.81	6.17	4.11	5.36	5.93	6.21
race	3.63	1.95	1.63	3.79	1.08	2.27	3.81	3.53	1.41
reg16	7.02	3.54	3.28	10.99	8.20	2.87	5.31	3.50	2.78
sex	12.44	13.08	14.68	12.18	11.53	13.49	9.57	9.25	13.32
sibs	2.24	1.17	0.36	2.34	3.98	1.23	2.26	1.81	1.56

Notes. Values are rounded to the second decimal.

Looking at the parametric estimates of IOP in figure 5.4, I find that on average, between 2002 and 2021, the two largest contributors are sex and age. This was also the case in 2017 Australia (Wu et al., 2021). They stand out from the rest of the circumstances as, in all years except for 2008, 2012, and 2021, they contribute, at the least, 5% more to IOP than any other circumstance. Second, looking at table 5.1, it seems I also found comparable results to Chantreuil et al. (2020) for both circumstances. Indeed, they found that the relative contribution of sex to total inequality varied between 9% to 13% between 2005 and 2017, and between 6% to 9% for age. I found values between 9.25% and 14.68% for sex, and between 6.53% and 15.03% for age.

It should be noted that in all years except for 2014 and 2018, sex contributes consistently more to IOP than age, and as such is the overall most important driver of IOP in the US. Generally, between 2002 and 2021 it fluctuated between 20% and 30%. More recently, it seems to have increased in importance as it reached its maximum of 33.08% in 2021. This may point toward a worrying increase in the gender gap in income, which should pressingly be addressed by US policymakers. Specifically, between 2018 and 2021, the importance of sex increased by more than 12 percentage points. Thus, this increase may be a direct cause of Covid-19. This is in line with the findings of Dang and Viet Nguyen (2021) claiming that the gender gap has increased as a result of the pandemic.

On the other hand, age was the most important contributor in 2014, when it represented 31.42% of IOP. More recently, it has decreased by more than 10 percentage points, representing 13.14% of IOP in 2021, a contribution close to its minimum of 12.03% reached in 2008, making it the third-largest contributor. Overall, it seems that in important socioeconomic crises, such as the 2008 crisis or the recent Covid-19 pandemic, the contribution of age decreases drastically. From these results, I cannot conclude with certainty whether older or younger Americans are favored during these periods. However, experience accumulated throughout one's career may be a determinant factor in income, and more often than not, experience is directly correlated with age. Thus, I can speculate that older Americans may especially be favored during socioeconomic crises, in which it is clearly harder for a young graduate to enter the job market than for an established employee to maintain his or her position.

Next, parents' socioeconomic status increasingly represents an important determinant of IOP. In 2002 it was the source of 8.90% of unfair inequalities, while in 2021 it has grown to be responsible for 15.43% of IOP, making it the second most important contributor. Since 2014 especially, the importance of parents' socioeconomic status in creating unequal opportunities in income has nearly doubled in the US.

Subsequently, an individual's region at age 16 seems to have been especially important in 2002 and 2008, as it was, respectively, the third and second most important contributor to IOP, more recently however it has decreased to be the source of only 6.89% of IOP, dropping to the sixth rank in terms of importance. This may be partially due to the increasing mobility and decentralization in the US, thanks to the development of new technologies such as the internet. Indeed, one can speculate that more and more people in the US do not grow up living in the same area as they spent their childhood. However, I could hypothesize that in the earlier years of 2002-2008, this variable may have incidentally acted as a proxy for an individual's current living area. Therefore, the fact that certain regions of the US were disproportionately affected by the 2008 financial crisis (Connaughton & Madsen, 2012) may explain the importance of the variable $reg16$ in that year.

Then, at 12.04%, parental education was the fourth most important contributor in 2021. Its importance has increased from 7.81% in 2002 to 16.63% in 2008 and then decreased again, fluctuating between approximately 9% and 12% until 2021. The relative importance of parental education has behaved similarly to that of the region an individual lived in at 16 years old. Moreover, they seem to both follow the same trend as IOP estimates, reaching their peaks in 2008.

Thereafter, an individual's family's income at age 16 has also been a significant contributor. It was responsible for 11.19% of IOP in 2002, then decreased to 3.86% in 2008, making it the ante-penultimate contributor. More recently, in 2021, it was the source of 7.84% of IOP and the fifth largest contributor. Covid-19 seems to have exacerbated the importance of family income, as its importance increased by nearly 4 percentage points between 2018 and 2021. Once again, this is a very worrying trend, as it means that inequality in income is increasingly inherited amongst Americans. This is understandable given the staggering costs of obtaining a higher education in the US coupled with recent incidents such as the 2019 college admissions bribery scandals. This shows how essential it is for the US government to quickly provide more equal opportunities in education.

After, comes race. My findings for its contribution share are similar to those of Chantreuil et al. (2020) between 2005 and 2016. Indeed, he found that the relative importance of race to total inequality in income increased between 2005 and 2007, decreased between 2007 and 2010, and increased again between 2010 and 2016. Overall, he also found that the latter fluctuated between 1% and 4%. Looking at table 5.1, this is also the case for my results.

Since 2016, the contribution of race decreased by 5 percentage points, representing 8.82% of IOP in 2016 and only 3.50% in 2021. This makes it one of the least important circumstances in driving IOP. More than 4 percentage points of this decrease have happened between 2018 and 2021: a period during which Covid-19 emerged. This is surprising considering the fact that Kochhar (2020) finds that blacks, Hispanics, and Asians suffer from significantly higher unemployment rates than whites since the start of the pandemic. Considering this, and the fact that the share of whites is often overstated in my sample, I cannot conclude with certainty that racism in income has truly decreased since 2016 as it is possible that the importance of race was underestimated.

Another even smaller contributor is the number of siblings, which increased from 2002 to 2012, representing 7.29% of IOP, and subsequently decreased again to 3.88% in 2021. Similarly, the contribution of the family situation at age 16 - in other words growing up with both parents or not - is constantly low throughout the period, fluctuating between 1.09% and 4.13%. Finally, the variable indicating whether an individual was born in the US was, on average, the smallest contributor to IOP between 2002 and 2021. Its contribution remains relatively constant across the entire period: between 0.18% and 3.12%.

Overall, we find similar results for the MLD-based Shapley decompositions. It can be seen in figure A.1 that the order of importance of circumstances' contribution to IOP and the general trends are almost identical. Furthermore, comparing the parametric and non-parametric Gini-based Shapley decompositions plotted in figures 5.4 and 5.5 respectively, we notice that although the levels of the circumstances' relative contribution shares are different, there are some similarities in their trends over time. Indeed, they seem to be identical in direction over the years, but with lesser magnitude. Moreover, except for the two largest contributors - sex and age -, region at 16, race, and number of siblings, the order of importance of circumstances seems to be more or less the same throughout the years. Thus, our previous findings are weakly robust to the model and index chosen and seem relatively reliable.

5.4 Interacting Circumstances

Figure 5.6 presents the evolution of the relative contribution of all circumstances and interaction terms over the period of 2002 to 2021 using the Gini coefficient. Figure 5.7 does the same, but including only the interaction terms and variables which are interacted.

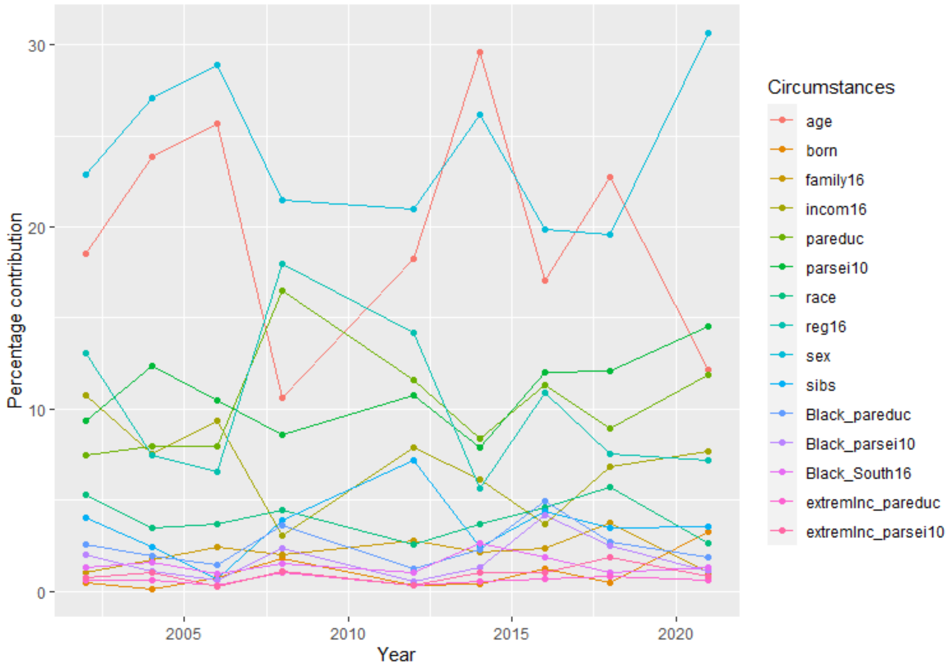

Figure 5.6. Estimates of the Relative Contribution of Circumstance Variables to IOP in the Parametric Model with Interactions Using the Gini Coefficient (2002-2021)

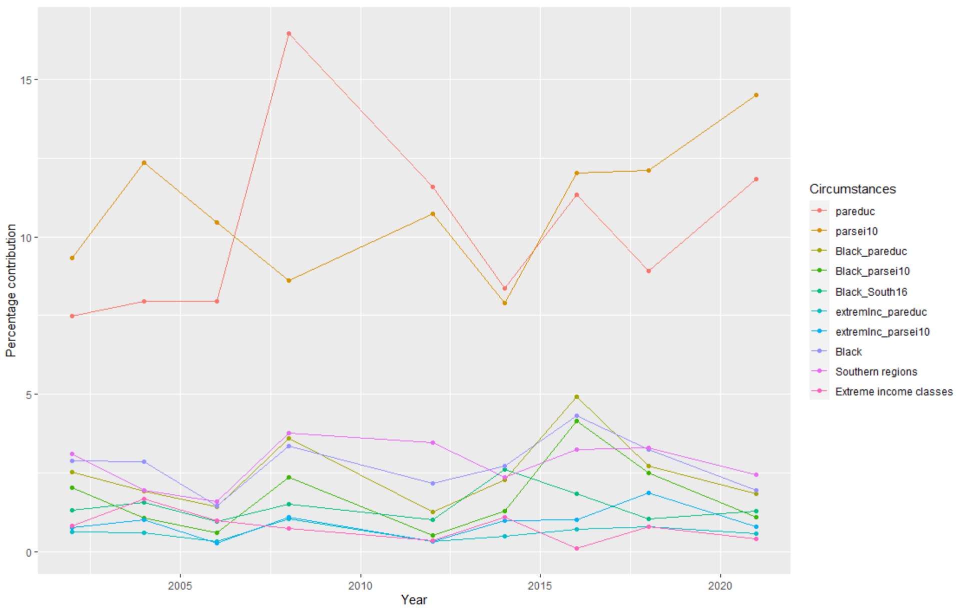

Figure 5.7. Parametric Estimates of the Relative Contribution of Interacted Circumstance Variables to IOP Using the Gini Coefficient (2002-2021)

The same results using the MLD index are presented in figures A.3 and A.4 of the appendix respectively. The exact values of these figures are displayed in tables A.10, A.12, A.11, and A.13 respectively.

First and foremost, it should be noted that, looking at figure 5.6, the contribution shares of the included interaction terms are significantly smaller than most circumstances, this is also the case using the MLD index. Notice also, that, comparing this figure to figure 5.4, the parametric interacted Shapley decomposition gives nearly identical results to those of the parametric Shapley decomposition. This attests to the robustness of our previous findings to the model chosen.

Second, it can be seen from figure 5.7 above that, in all years and for both inequality measures, the contribution share of all interaction terms involving parental education or parental socioeconomic index is significantly smaller (at least 5 percentage points smaller, if not more) than the contribution share of parental education or parental socioeconomic index itself. This indicates that differences in parental education or in parental socioeconomic index of black individuals do not lead to more inequality in income than for an individual when race is not taken into account. This stands in contrast with the findings of Marrero and Rodríguez (2011), which indicate that cross effects between race and parental education accounted for almost 35% of IOP in 2007 in the US. Nevertheless, this gap may once again be explained by the smaller number of circumstances included in Marrero's analysis.

Moreover, this also holds for individuals whose family income was at the extremes of the income distribution. For those, differences in parental education and parental socioeconomic index do not disproportionately lead to inequalities in income compared to individuals with a moderate family income at age 16.

Focus now on the term interacting Black individuals to individuals living in the South at age 16. It can be seen from figure 5.7 that in 2014 this interaction term contributed more to the Gini-based IOP estimate than did living in the South at 16 years old, and nearly as much as did being black. This also holds for the MLD-based IOP estimates (plotted in figure A.3) except for the fact that the interaction term is two times larger than both individual circumstance contributions and that it was slightly larger than the individual contribution share of living in the South at 16 in 2004 and 2008. Considering this and the fact that the Gini coefficient provides more reliable IOP estimates, it is safe to conclude that being black and having lived in the South at age 16 disproportionately lead to inequalities in income in 2014. In other words, the average black individual who lived in the South at 16 may have faced more inequalities in income than did the average black individual or the average individual who lived in the South at 16. This is in line with the findings of Chantreuil et al. (2020) claiming that race contributes significantly more to income inequality in the Southern US states than in other parts of the country.

The parametric model without interaction terms defined in equation 4.4 assumes a strictly linear relationship between income and circumstances while the non-parametric model considers non-linearity. Thus, if there is a large gap between the parametric estimate of a circumstance contribution to inequality and the non-parametric estimate, we may assume that there may exist interactions between this circumstance and others (Wu et al., 2021).

Obviously, since the parametric and non-parametric estimates of IOP are different we compare the estimates of the contribution share of said circumstance to total inequality in income. For all circumstances, the Gini- and MLD-based results are displayed, respectively, in tables A.8 and A.9 of the appendix. Keep in mind that, as aforementioned, due to the small yearly samples relative to the large number of types, the non-parametric estimates are likely overestimated. Thus, it is difficult to quantify a "large gap".

Nevertheless, some circumstance estimates stand out. The Gini-based results of these circumstances are displayed in table 5.2. First, it can be seen that, regardless of the inequality measure used, the difference in the contribution share estimate of sex is relatively small. More so for the Gini-based estimates, where it is approximately a 1 percentage point difference. Thus, just like Wu et al. (2021) did, I conclude that it is unlikely that sex interacts with other circumstances. On the other hand, notice that the contribution shares of age and reg16 are quite large: around 10 percentage points or more for the Gini-based estimates and around 15 percentage points for the MLD-based estimates. Thus, it seems that there exist some interactions between an individual's age or the region in which they lived at 16 years old, and other circumstances, which could contribute significantly to IOP in income.

Table 5.2. *Comparing Parametric and Non-parametric estimates of the Relative Contribution of Circumstances to Total Inequality in Income Using the Gini Coefficient (2002-2021)*

		2002	2004	2006	2012	2014	2016	2018	2021
age	Parametric	10.13	11.34	12.73	10.17	15.03	8.27	11.14	5.29
	Non-parametric	22.08	23.35	21.90	21.97	22.35	21.59	22.89	21.86
reg16	Parametric	7.02	3.54	3.28	8.20	2.87	5.31	3.50	2.78
	Non-parametric	16.81	15.31	14.14	16.10	15.13	16.55	15.64	14.03
sex	Parametric	12.44	13.08	14.68	11.53	13.49	9.57	9.25	13.32
	Non-parametric	12.17	13.40	14.21	10.94	13.67	10.44	10.79	13.02

Notes. Values are percentages and rounded to the second decimal.

6 Conclusion

This paper examines the evolution of inequality of opportunity in income in the US between 2002 and 2021. Therefore, I studied the relative contribution of several circumstances, and the interactions of some, to IOP in income.

First, I found that IOP in income seems to behave cyclically with total inequality, reaching its peak between 2008 and 2012. This points towards the efficiency of inequality reducing policies in targeting not only inequalities due to effort variations, considered ethical, but also inequalities of opportunity, considered unethical. Regarding inequalities of opportunity, I concluded that IOP in income represented, at the least, between 40% and 55% of total income inequality and at the most between 95% and 96% in the period of 2002 to 2021. Nevertheless, I consider the lower bound a sufficient and reasonable estimate of IOP. Although it is an encouraging sign that levels of IOP have been decreasing since 2012 and were lower than ever in 2021, the US is still far behind similar economically developed countries in terms of social development as the latter show much lower levels of IOP.

While IOP remained relatively constant over the years, I found that the importance of its driving circumstances varied significantly.

Overall, I found that sex and age were the two largest contributors to IOP in income between 2002 and 2021. Moreover, the recent increase in importance of sex points toward a worrying development of the gender gap in income. Second, it seems that Covid-19 may have accelerated this increase. Thus, I recommend US policymakers pressingly address this issue, rather than focusing on restricting women's reproductive rights.

Similarly, the importance of parental characteristics such as socioeconomic status, educational attainment, and family income has increased since 2016, making them, respectively, the second, fourth, and fifth most important contributors in 2021. The contribution share of parents' socioeconomic index especially has nearly doubled between 2014 and 2021. Once again Covid-19 seems to have exacerbated the importance of family income, meaning that inequality in income is increasingly inherited amongst Americans. Thus, US policymakers should focus on opportunity harmonizing policies in income, such as making education more accessible and facilitating inter-generational mobility.

On the other hand, since 2016, the importance of race decreased, making it the 8^{th} circumstance in driving IOP. Surprisingly, most of this decrease happened during Covid-19. However, due to the overstatement of white individuals in the sample, the importance of race may have been underestimated.

Furthermore, presumably due to the increasing mobility and decentralization in the US, the region one lived in at 16, dropped from being the 3^{rd} largest contributor of IOP in income in 2002, to being the 6^{th} in 2021. Nevertheless, it should be noted that this circumstance does not say anything about local inequalities which may arise from living in a disadvantaged neighborhood and which may still be very high in 2021.

Finally, I found that the combination of being black and having lived in the South at 16 disproportionately contributed to IOP in income in 2014. This highlights the need for Southern governors to focus on anti-racism policies and catch up to the rest of the country. On the other hand, I did not find parental characteristics to be a more deciding factor in income for black individuals or for individuals at the extremes of the parental income distribution.

All in all, my analysis shows that in the period 2002 to 2021, the American Dream represented nothing more than that: a dream. In this period, equality of opportunity was certainly not attained. Thus, I have to agree with Adams's critics and recognize the American Dream to be a great delusion.

It should be noted that one significant limitation of my analysis is that my sample may not be representative of the level of inequality in the US between 2002 and 2021. Therefore, it is difficult to evaluate the reliability of my findings with certainty. Additionally, the small sample size is especially problematic for the non-parametric analysis. Indeed, combined with a large number of circumstances, I find extremely overstated levels of IOP, which give a useless upper bound. Besides, white individuals are over-represented in nearly all yearly samples. Thus, this may lead to inaccurate estimations of the relative importance of race. Finally, I consider age to fit the definition of a circumstance given that it is not something one can exert any control over. However, it is likely that age also acts as a proxy for (work) experience. In this case, one may argue that experience is not an unfair consideration for income, but a result of effort. Therefore, it is questionable whether it should be included as a circumstance.

Thus, in the future, it may be interesting to perform the same analysis, treating age as a demographic variable and not a circumstance, and compare the results to ours. Many empirical papers have done so already by dividing each yearly sample into cohorts of age (Checchi & Peragine, 2010). However, the disadvantage of this method is that we will obtain results valid only for that age cohort in that year. Another interesting and arguably more telling solution would be to use age-adjusted income as an outcome variable. We have not done so in our analysis because we would have obtained an outcome variable with very large standard errors due to the small size of our sample.

Therefore, further research should place particular importance on gathering a large, representative sample. This is especially important for evaluating the relative importance of race. On top of that, a larger sample would give less overestimated non-parametric results and a more reasonable upper bound. Alternatively, it would be useful to focus on finding an appropriate way to handle empty types without eliminating them from the analysis and thus hopefully obtain less overestimated non-parametric results.

Additionally, to refine the estimate of IOP, further research may focus on improving the parametric model to obtain a higher lower bound, and thus a tighter interval of IOP. For this, they should focus on investigating the potential interactions between $reg16$ (or age if it is included as a circumstance variable) and other circumstances, as that is the variable for which I found the largest difference between the parametric and non-parametric estimates.

Finally, it may be of particular interest to policymakers to compare IOP estimates using constant income before tax as an outcome variable with IOP estimates using constant disposable income (as is done in this analysis). Thereby, policymakers may be able to evaluate the efficiency of fiscal policies and develop new policies targeting certain circumstances or IOP as a whole.

7 References

Adams, J. T. (1931). *The Epic of America*. Boston: Little, Brown, Company.

Ahmed, H. M. M., Osman, E. A. M., & Mahran, H. A. (2020). Inequality of Opportunity in The Labor Market: Evidence from Sudan. *Journal of Social Science Studies*, *7*(2), 38–57. https://doi.org/10.5296/jsss.v7i2.16833

Björklund, A., Jäntti, M., & Roemer, J. E. (2012). Equality of opportunity and the distribution of long-run income in Sweden. *Social Choice and Welfare*, *39*(2-3), 675–696. https://doi.org/10.1007/s00355-011-0609-3

Bourguignon, F., Ferreira, F. H. G., & Menéndez, M. (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

Bradbury, K., & Triest, R. K. (2016). Inequality of Opportunity and Aggregate Economic Performance. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *2*(2), 178–201. https://doi.org/10.7758/RSF.2016.2.2.08

Brunori, P. (2016). *How to Measure Inequality of Opportunity: A Hands-On Guide*, Dipartimento di Scienze Economiche e Metodi Matematici, University of Bari. www.lifecoursecentre.org.au

Camacho, M., & Palmieri, G. (2019). Do economic recessions cause inequality to rise? *Journal of Applied Economics*, *22*(1), 304–320. https://doi.org/10.1080/15140326.2019.1620982

Castro, J., Gómez, D., & Tejada, J. (2009). Polynomial calculation of the Shapley value based on sampling. *Computers and Operations Research*, *36*(5), 1726–1730. https://doi.org/10.1016/J.COR.2008.04.004

Chantreuil, F., Fourrey, K., Lebon, I., & Rebière, T. (2020). *Decomposing US Income Inequality à la Shapley: Race Matters, but Gender Too* (tech. rep.). IZA Institute of Labor Economics. www.iza.org

Chantreuil, F., & Lebon, I. (2015). Gender contribution to income inequality. *Economics Letters*, *133*, 27–30. https://doi.org/10.1016/j.econlet.2015.05.009

Checchi, D., & Peragine, V. (2010). Inequality of opportunity in Italy. *Journal of Economic Inequality*, *8*(4), 429–450. https://doi.org/10.1007/s10888-009-9118-3

Connaughton, J. E., & Madsen, R. A. (2012). U.S. State and Regional Economic Impact of the 2008/2009 Recession. *Journal of Regional Analysis and Policy*, *42*(3), 177–187. https://doi.org/10.22004/AG.ECON.143779

Dang, H. A. H., & Viet Nguyen, C. (2021). Gender inequality during the COVID-19 pandemic: Income, expenditure, savings, and job loss. *World Development*, *140*, 105–296. https://doi.org/10.1016/J.WORLDDEV.2020.105296

Duncan, O. D. (1961). A socioeconomic index for all occupations. *Class: Critical Concepts*, *1*, 388–426.

Dworkin, R. (1981a). What is Equality? Part 1: Equality of Welfare. *Philosophy & Public Affairs*, *10*(3), 185–246. http://www.jstor.org/stable/2264894

Dworkin, R. (1981b). What is Equality? Part 2: Equality of Resources. *Philosophy & Public Affairs*, *10*(4), 283–345. http://www.jstor.org/stable/2265047

Elbers, B. (2021). Compute Shapley-Shorrocks value decompositions. https://github.com/elbersb/shapley#readme

Eurostat. (2021). Living conditions in Europe - income distribution and income inequality.

Ferreira, F. H. G., Lakner, C., Lugo, M. A., & Özler, B. (2014). *Inequality of Opportunity and Economic Growth : A Cross-Country Analysis*, World Bank. https://doi.org/10.1596/1813-9450-6915

Ferreira, F. H., & Gignoux, J. (2011). The measurement of inequality of opportunity: Theory and an application to Latin America. *Review of Income and Wealth*, *57*(4), 622–657. https://doi.org/10.1111/j.1475-4991.2011.00467.x

Huffman, M., & Cohen, P. (2004). Racial Wage Inequality: Job Segregation and Devaluation across U.S. Labor Markets. *American Journal of Sociology*, *109*(4), 902–936. https://doi.org/10.1086/378928

Jones, N., Marks, R., Ramirez, R., & Ríos-vargas, M. (2021). 2020 Census Illuminates Racial and Ethnic Composition of the Country. https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html#:~:text=The%20White%20alone%20adult%20population,2010%20to%208.8%25%20in%202020.

Kochhar, R. (2020). Unemployment rose higher in three months of COVID-19 than it did in two years of the Great Recession. https://www.pewresearch.org/fact-tank/2020/06/11/unemployment-rose-higher-in-three-months-of-covid-19-than-it-did-in-two-years-of-the-great-recession/

Marrero, G. A., & Rodríguez, J. G. (2011). Chapter 9 Inequality of Opportunity in the United States: Trends and Decomposition. In J. G. Rodríguez (Ed.), *Inequality of opportunity: Theory and measurement* (pp. 217–246). Emerald Group Publishing Limited. https://doi.org/10.1108/S1049-2585(2011)0000019012

McKinnon, J. (2001). Majority of African Americans Live in 10 States; New York City and Chicago Are Cities With Largest Black Populations. https://www.census.gov/newsroom/releases/archives/census_2000/cb01cn176.html#:~:text=The%20region%20with%20the%20highest,the%20West%20(6%20percent).

Peragine, V. (2011). Review of "Measuring inequality of opportunities in Latin America and the Caribbean" by Ricardo Paes de Barros, Francisco H. G. Ferreira, José R. Molinas Vega, and Jaime Saavedra Chanduvi, World Bank and Palgrave Macmillan, 2009. *Journal of Economic Inequality*, *9*(1), 137–143. https://doi.org/10.1007/s10888-010-9151-2

Pistolesi, N. (2009). Inequality of opportunity in the land of opportunities, 1968-2001. *Journal of Economic Inequality*, *7*(4), 411–433. https://doi.org/10.1007/s10888-008-9099-7

Ramos, X., & Van De Gaer, D. (2012). Empirical Approaches to Inequality of Opportunity: Principles, Measures and Evidence. *Journal of Economic Surveys*, *30*(5), 855–883. www.ecineq.org

Roemer, J. E. (1998). *Equality of Opportunity*. Harvard University Press. https://doi.org/10.4159/9780674042872

Saez, E., & Zucman, G. (2016). Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data. *Quarterly Journal of Economics*, *2*, 519–578.

Sastre, M., & Trannoy, A. (2002). Shapley Inequality Decomposition by Factor Components: Some Methodological Issues. *Journal of Economics*, *9*(1), 51–89.

Shorrocks, A. F. (2013). Decomposition procedures for distributional analysis: A unified framework based on the Shapley value. *Journal of Economic Inequality*, *11*(1), 99–126. https://doi.org/10.1007/s10888-011-9214-z

Smith, T. W., Davern, M., Freese, J., & Morgan, S. L. (2021). *General Social Surveys, 1972-2021 [machine-readable data file]* (tech. rep.). NORC. Chicago, NORC.

United Nations Development Programme. (2019). *Human development report 2019 : beyond income, beyond averages, beyond today: inequalities in human development in the 21st century* (tech. rep.). United Nations. https://www.un-ilibrary.org/content/books/9789210044967

Winship, S. (2015). Has Rising Income Inequality worsened Inequality of Opportunity in the United States? *Social Philosophy and Policy*, *31*(2), 28–47. https://doi.org/10.1017/S0265052514000247

World Bank. (2019). Gini index - United States. https://data.worldbank.org/indicator/SI.POV.GINI?end=2019&locations=US&start=2002&view=chart

Wu, D., Rao, P., Rohde, N., Trivedi, P., & Tang, K. K. (2021). Measuring Inequality of Opportunity Using Sampling-based Shapley Decomposition: Application to Australian Data 2001-2017. *36th IARIW Virtual General Conference*. https://iariw.org/wp-content/uploads/2021/08/wu_et_al_paper.pdf

A Appendix

A.1 Abbreviation list

USA = United States of America
IOP = Inequality of opportunity
EA = Ex-ante
EP = Ex-post
IQ = Intelligence quotient

A.2 Inequality in Income

Table A.1. *Inequality in Income in the USA between 2002 and 2021 Using Different Inequality Measures*

	2002	2004	2006	2008	2012	2014	2016	2018	2021
Gini	0.491	0.454	0.453	0.536	0.549	0.468	0.479	0.479	0.461
MLD	0.524	0.458	0.454	0.601	0.638	0.494	0.504	0.520	0.473

Notes. Values are rounded to the third decimal.

A.3 Inequality of Opportunity

Table A.2. *Parametric and Non-parametric Estimates of IOP Relative to Total Inequality in Income Using the Gini Coefficient (2002-2021)*

Year	Non-parametric	Non-parametric (less types)	Parametric interacted	Parametric
2002	0.963	0.832	0.518	0.504
2004	0.959	0.816	0.464	0.458
2006	0.959	0.835	0.493	0.488
2008	0.983	0.894	0.561	0.543
2012	0.979	0.885	0.553	0.545
2014	0.980	0.854	0.492	0.478
2016	0.976	0.844	0.447	0.432
2018	0.976	0.856	0.463	0.447
2021	0.957	0.806	0.417	0.403

Notes. Values are rounded to the third decimal.

Table A.3. *Parametric and Non-parametric Estimates of IOP Relative to Total Inequality in Income Using the MLD Index (2002-2021)*

Year	Non-parametric	Non-parametric (less types)	Parametric interacted	Parametric
2002	0.910	0.666	0.244	0.252
2004	0.921	0.641	0.182	0.185
2006	0.920	0.661	0.234	0.236
2008	0.953	0.759	0.271	0.290
2012	0.944	0.743	0.302	0.295
2014	0.953	0.706	0.241	0.223
2016	0.947	0.688	0.173	0.182
2018	0.927	0.706	0.195	0.190
2021	0.904	0.621	0.141	0.136

Notes. Values are rounded to the third decimal.

Table A.4. *Parametric and Non-parametric Estimates of Absolute IOP Using the Gini Coefficient (2002-2021)*

Year	Non-parametric	Non-parametric (less types)	Parametric interacted	Parametric
2002	0.473	0.408	0.254	0.247
2004	0.435	0.371	0.211	0.208
2006	0.434	0.378	0.223	0.221
2008	0.527	0.479	0.301	0.291
2012	0.537	0.485	0.303	0.299
2014	0.459	0.400	0.230	0.224
2016	0.468	0.404	0.214	0.207
2018	0.468	0.410	0.222	0.214
2021	0.441	0.372	0.192	0.186

Notes. Values are rounded to the third decimal.

Table A.5. *Parametric and Non-parametric Estimates of Absolute IOP Using the MLD Index (2002-2021)*

Year	Non-parametric	Non-parametric (less types)	Parametric interacted	Parametric
2002	0.477	0.349	0.128	0.132
2004	0.421	0.293	0.083	0.085
2006	0.418	0.300	0.106	0.107
2008	0.572	0.456	0.163	0.174
2012	0.602	0.474	0.193	0.188
2014	0.470	0.348	0.119	0.110
2016	0.477	0.346	0.087	0.092
2018	0.482	0.367	0.102	0.099
2021	0.428	0.294	0.067	0.065

Notes. Values are rounded to the third decimal.

A.4 Decomposing Inequality of Opportunity

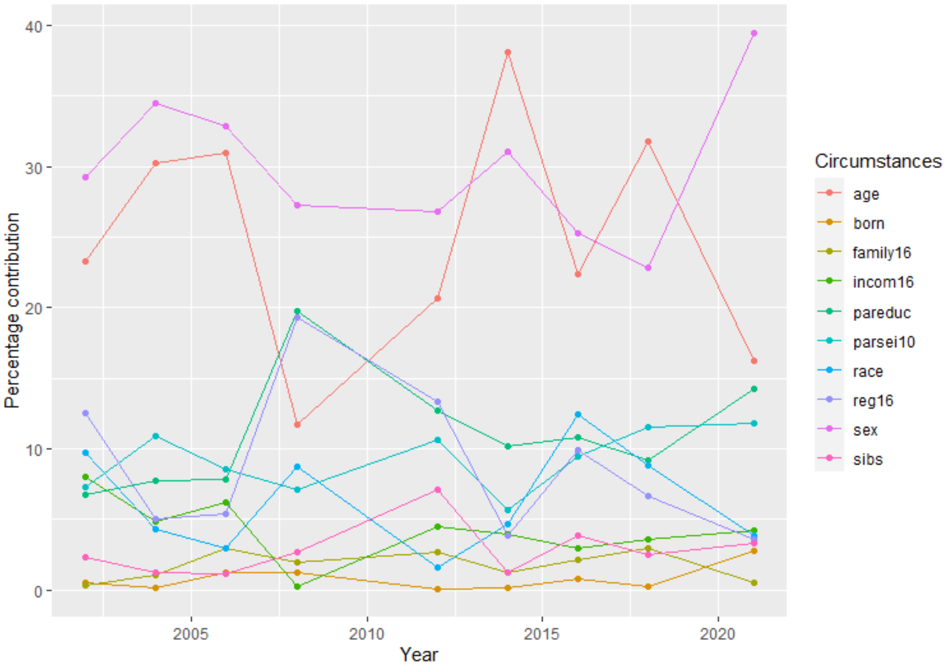

Figure A.1. Parametric Estimates of the Relative Contribution of Circumstance Variables to IOP in the USA Using the MLD Index (2002-2021)

Table A.6. *Parametric Estimates of the Relative Contribution of Circumstance Variables to IOP Using the Gini Coefficient (2002-2021)*

	2002	2004	2006	2008	2012	2014	2016	2018	2021
age	20.08	24.75	26.10	12.03	18.64	31.42	19.16	24.90	13.14
born	0.59	0.19	0.93	2.04	0.38	0.39	1.27	0.68	3.13
family16	1.24	2.10	3.20	2.59	3.25	2.67	3.09	4.13	1.09
incom16	11.19	8.02	9.26	3.86	8.66	6.50	4.35	7.58	7.84
pareduc	7.81	8.40	8.25	16.63	12.33	8.92	11.22	9.00	12.04
parsei10	8.90	13.44	11.36	8.86	11.31	8.59	12.41	13.25	15.43
race	7.19	4.26	3.35	6.99	1.98	4.75	8.82	7.88	3.50
reg16	13.92	7.72	6.73	20.25	15.03	6.00	12.30	7.83	6.89
sex	24.65	28.55	30.09	22.45	21.13	28.18	22.16	20.68	33.08
sibs	4.43	2.56	0.73	4.31	7.29	2.57	5.22	4.06	3.88

Notes. Values are percentages and rounded to the second decimal.

Table A.7. *Parametric Estimates of the Relative Contribution of Circumstance Variables to IOP Using the MLD Index (2002-2021)*

	2002	2004	2006	2008	2012	2014	2016	2018	2021
age	23.25	30.22	30.94	11.72	20.63	38.10	22.35	31.72	16.22
born	0.49	0.12	1.20	1.20	0.04	0.15	0.79	0.28	2.76
family16	0.34	1.09	2.99	1.96	2.70	1.20	2.14	2.98	0.50
incom16	8.04	4.83	6.24	0.23	4.53	3.90	2.98	3.55	4.21
pareduc	6.77	7.76	7.82	19.74	12.69	10.15	10.83	9.21	14.28
parsei10	7.32	10.94	8.52	7.10	10.59	5.62	9.49	11.51	11.84
race	9.69	4.28	2.92	8.77	1.55	4.69	12.40	8.79	3.84
reg16	12.55	4.99	5.40	19.29	13.35	3.87	9.94	6.62	3.60
sex	29.21	34.50	32.86	27.27	26.82	31.07	25.23	22.87	39.48
sibs	2.35	1.26	1.11	2.71	7.11	1.25	3.85	2.48	3.27

Notes. Values are percentages and rounded to the second decimal.

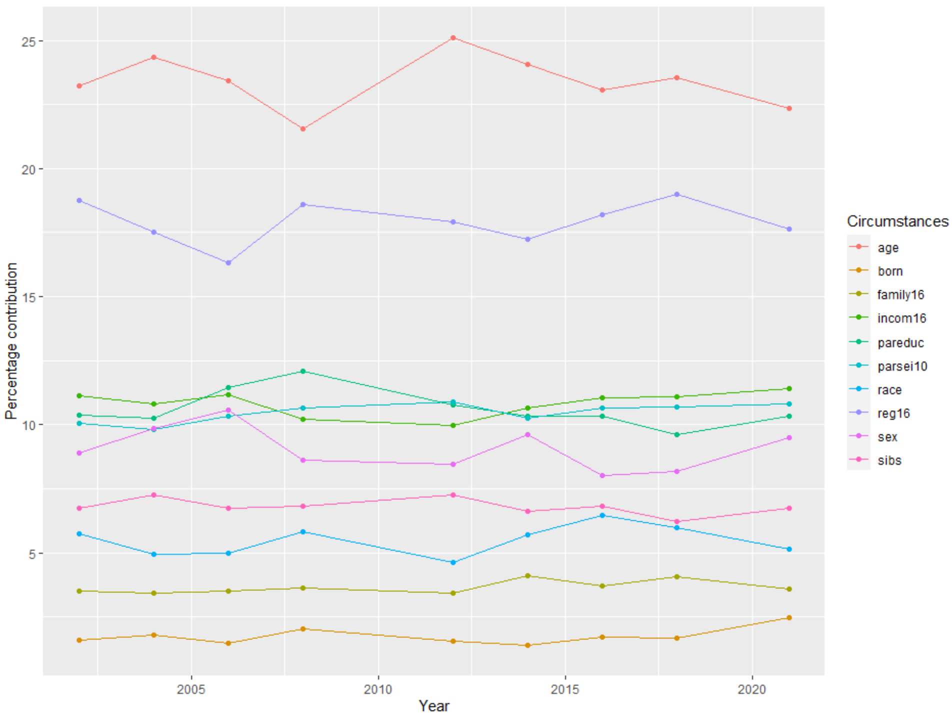

Figure A.2. Non-parametric Estimates of the Relative Contribution of Circumstance Variables to IOP in the USA Using the MLD Index (2002-2021)

A.5 Interacting Circumstances

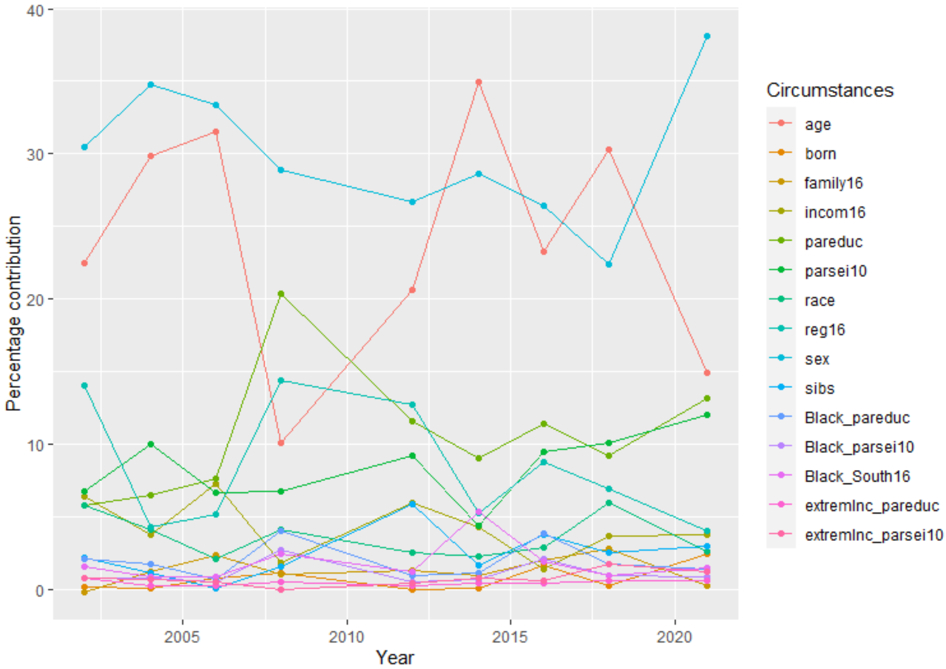

Figure A.3. Parametric Estimates in the Interacted Model of the Relative Contribution of Circumstance Variables to IOP in the USA Using the MLD Index (2002-2021)

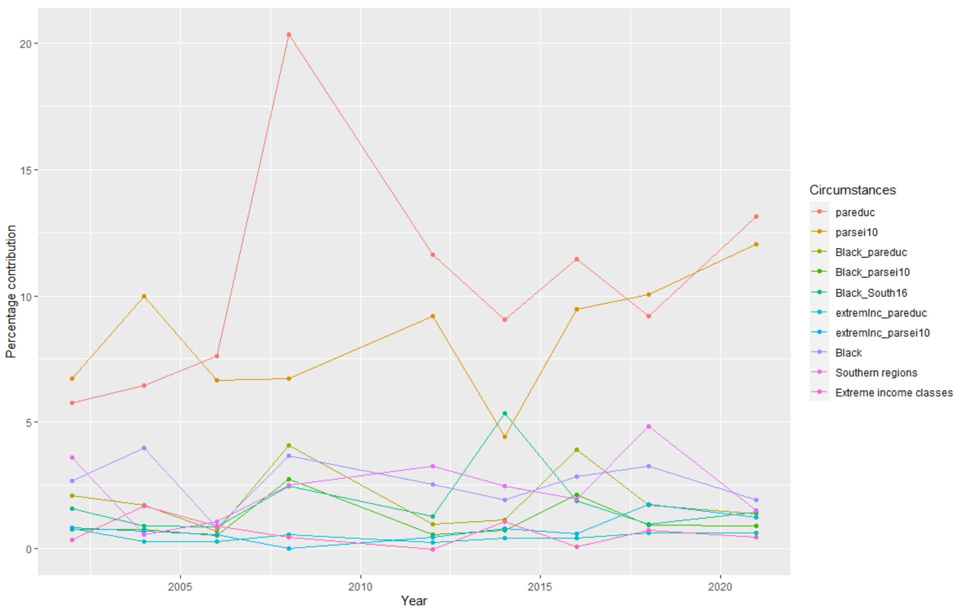

Figure A.4. Parametric Estimates of the Relative Contribution of Interacted Circumstance Variables to IOP Using the MLD Index (2002-2021)

Table A.8. *Comparing Parametric and Non-parametric estimates of the Relative Contribution of Circumstances to Total Inequality in Income Using the Gini Coefficient (2002-2021)*

		2002	2004	2006	2008	2012	2014	2016	2018	2021
age	Parametric	10.13	11.34	12.73	6.53	10.17	15.03	8.27	11.14	5.29
	Non-parametric	22.08	23.35	21.90	19.46	21.97	22.35	21.59	22.89	21.86
born	Parametric	0.30	0.09	0.45	1.11	0.21	0.19	0.55	0.31	1.26
	Non-parametric	1.36	1.26	1.64	2.26	1.40	1.45	1.89	1.61	2.85
family16	Parametric	0.63	0.96	1.56	1.40	1.77	1.28	1.34	1.85	0.44
	Non-parametric	3.00	3.12	3.70	3.77	3.55	4.07	3.92	4.39	3.13
incom16	Parametric	5.64	3.68	4.52	2.10	4.72	3.11	1.88	3.39	3.16
	Non-parametric	10.55	10.39	11.06	8.90	9.36	9.70	9.35	10.43	10.57
pareduc	Parametric	3.94	3.85	4.03	9.03	6.72	4.27	4.85	4.02	4.85
	Non-parametric	10.69	10.46	11.42	12.54	12.21	10.73	10.95	9.46	11.14
parsei10	Parametric	4.49	6.16	5.54	4.81	6.17	4.11	5.36	5.93	6.21
	Non-parametric	10.47	11.34	11.34	11.38	12.06	10.71	11.05	11.68	11.90
race	Parametric	3.63	1.95	1.63	3.79	1.08	2.27	3.81	3.53	1.41
	Non-parametric	6.01	4.77	5.42	5.97	4.43	5.90	7.04	6.49	4.77
reg16	Parametric	7.02	3.54	3.28	10.99	8.20	2.87	5.31	3.50	2.78
	Non-parametric	16.81	15.31	14.14	17.95	16.10	15.13	16.55	15.64	14.03
sex	Parametric	12.44	13.08	14.68	12.18	11.53	13.49	9.57	9.25	13.32
	Non-parametric	12.17	13.40	14.21	10.96	10.94	13.67	10.44	10.79	13.02
sibs	Parametric	2.24	1.17	0.36	2.34	3.98	1.23	2.26	1.81	1.56
	Non-parametric	6.87	6.59	5.17	6.82	7.98	6.27	7.21	6.63	6.74

Notes. Values are percentages and rounded to the second decimal.

Table A.9. *Comparing Parametric and Non-parametric estimates of the Relative Contribution of Circumstances to Total Inequality in Income Using the MLD Index (2002-2021)*

		2002	2004	2006	2008	2012	2014	2016	2018	2021
age	Parametric	5.86	5.60	7.31	3.40	6.09	8.49	4.07	6.02	2.21
	Non-parametric	21.12	22.39	21.54	20.53	23.71	22.93	21.83	21.81	20.20
born	Parametric	0.12	0.02	0.28	0.35	0.01	0.03	0.14	0.05	0.38
	Non-parametric	1.46	1.66	1.37	1.93	1.46	1.32	1.62	1.55	2.23
family16	Parametric	0.09	0.20	0.71	0.57	0.80	0.27	0.39	0.57	0.07
	Non-parametric	3.19	3.17	3.22	3.47	3.23	3.90	3.50	3.75	3.24
incom16	Parametric	2.03	0.89	1.47	0.07	1.34	0.87	0.54	0.67	0.57
	Non-parametric	10.12	9.94	10.26	9.75	9.41	10.14	10.45	10.29	10.33
pareduc	Parametric	1.71	1.44	1.85	5.72	3.75	2.26	1.97	1.75	1.95
	Non-parametric	9.45	9.44	10.52	11.52	10.17	9.85	9.78	8.92	9.33
parsei10	Parametric	1.85	2.02	2.01	2.06	3.13	1.25	1.73	2.18	1.62
	Non-parametric	9.16	9.04	9.52	10.16	10.27	9.79	10.10	9.91	9.79
race	Parametric	2.44	0.79	0.69	2.54	0.46	1.05	2.26	1.67	0.52
	Non-parametric	5.22	4.55	4.60	5.55	4.36	5.42	6.11	5.54	4.64
reg16	Parametric	3.17	0.92	1.28	5.59	3.95	0.86	1.81	1.26	0.49
	Non-parametric	17.07	16.11	15.01	17.71	16.92	16.44	17.23	17.60	15.93
sex	Parametric	7.37	6.39	7.77	7.90	7.92	6.92	4.60	4.34	5.39
	Non-parametric	8.09	9.05	9.74	8.20	7.99	9.16	7.59	7.57	8.59
sibs	Parametric	0.59	0.23	0.26	0.79	2.10	0.28	0.70	0.47	0.45
	Non-parametric	6.12	6.70	6.19	6.50	6.86	6.31	6.48	5.76	6.11

Notes. Values are percentages and rounded to the second decimal.

Table A.10. *Estimates of the Relative Contribution of Circumstance Variables to IOP in the Parametric Model with Interactions Using the Gini Coefficient (2002-2021)*

	2002	2004	2006	2008	2012	2014	2016	2018	2021
age	18.51	23.83	25.64	10.60	18.26	29.60	17.09	22.72	12.14
born	0.46	0.10	0.72	1.82	0.34	0.37	1.20	0.45	3.24
family16	1.01	1.69	2.44	2.03	2.80	2.11	2.37	3.72	1.01
incom16	10.73	7.50	9.36	3.08	7.91	6.13	3.65	6.87	7.65
pareduc	7.49	7.95	7.96	16.47	11.58	8.37	11.33	8.91	11.84
parsei10	9.33	12.35	10.47	8.61	10.75	7.88	12.02	12.10	14.51
race	5.31	3.47	3.69	4.45	2.57	3.69	4.57	5.69	2.62
reg16	13.05	7.43	6.53	17.94	14.16	5.64	10.87	7.56	7.17
sex	22.84	27.09	28.91	21.49	21.01	26.15	19.84	19.59	30.67
sibs	4.00	2.41	0.70	3.88	7.16	2.40	4.40	3.46	3.56
Black_pareduc	2.53	1.92	1.42	3.61	1.25	2.28	4.93	2.72	1.85
Black_parsei10	2.04	1.06	0.59	2.37	0.52	1.28	4.16	2.50	1.09
Black_South16	1.33	1.58	0.96	1.50	1.03	2.62	1.84	1.05	1.30
extremInc_pareduc	0.62	0.60	0.34	1.04	0.33	0.50	0.70	0.80	0.58
extremInc_parsei10	0.76	1.01	0.26	1.11	0.34	0.99	1.03	1.86	0.78

Notes. Values are percentages and rounded to the second decimal.

Table A.11. *Estimates of the Relative Contribution of Circumstance Variables to IOP in the Parametric Model with Interactions Using the MLD Index (2002-2021)*

	2002	2004	2006	2008	2012	2014	2016	2018	2021
age	22.45	29.84	31.57	10.12	20.64	34.97	23.30	30.28	14.92
born	0.19	0.10	0.82	1.15	0.00	0.10	1.66	0.29	2.46
family16	-0.19	1.20	2.41	1.09	1.35	0.97	2.00	2.82	0.30
incom16	6.44	3.81	7.27	1.82	5.96	4.32	1.42	3.65	3.77
pareduc	5.78	6.46	7.63	20.36	11.63	9.07	11.46	9.19	13.15
parsei10	6.74	9.98	6.65	6.74	9.18	4.41	9.48	10.06	12.05
race	5.80	4.10	2.10	4.13	2.50	2.29	2.87	5.93	2.59
reg16	14.07	4.29	5.19	14.36	12.72	5.25	8.76	6.90	4.06
sex	30.51	34.74	33.41	28.87	26.69	28.64	26.41	22.42	38.15
sibs	2.20	1.14	0.07	1.54	5.86	1.66	3.75	2.50	3.02
Black_pareduc	2.09	1.71	0.68	4.07	0.97	1.12	3.90	1.72	1.38
Black_parsei10	0.75	0.76	0.50	2.74	0.56	0.70	2.13	0.94	0.88
Black_South16	1.58	0.88	0.87	2.46	1.26	5.34	1.90	0.95	1.45
extremInc_pareduc	0.79	0.28	0.29	0.54	0.24	0.40	0.39	0.62	0.63
extremInc_parsei10	0.82	0.70	0.55	0.00	0.45	0.78	0.58	1.74	1.22

Notes. Values are percentages and rounded to the second decimal.

Table A.12. *Parametric Estimates of the Relative Contribution of Interacted Circumstance Variables to IOP Using the Gini Coefficient (2002-2021)*

Year	incom16		pareduc	parsei10	race		reg16		Black _pareduc	Black _parsei10	Black _South16	extremInc _pareduc	extremInc _parsei10
	Other income classes	Extreme income classes			Non-black	Black	Other regions	Southern regions					
2002	0.099	0.008	0.075	0.093	0.024	0.029	0.099	0.031	0.025	0.020	0.013	0.006	0.008
2004	0.058	0.017	0.080	0.123	0.006	0.029	0.055	0.020	0.019	0.011	0.016	0.006	0.010
2006	0.084	0.010	0.080	0.105	0.022	0.015	0.050	0.016	0.014	0.006	0.010	0.003	0.003
2008	0.023	0.007	0.165	0.086	0.011	0.033	0.142	0.038	0.036	0.024	0.015	0.010	0.011
2012	0.075	0.004	0.116	0.107	0.004	0.022	0.107	0.035	0.013	0.005	0.010	0.003	0.003
2014	0.050	0.011	0.084	0.079	0.010	0.027	0.033	0.024	0.023	0.013	0.026	0.005	0.010
2016	0.035	0.001	0.113	0.120	0.003	0.043	0.076	0.032	0.049	0.042	0.018	0.007	0.010
2018	0.061	0.008	0.089	0.121	0.024	0.033	0.043	0.033	0.027	0.025	0.010	0.008	0.019
2021	0.072	0.004	0.118	0.145	0.007	0.020	0.047	0.024	0.019	0.011	0.013	0.006	0.008

Notes. Values are percentages and rounded to the second decimal.

Table A.13. *Parametric Estimates of the Relative Contribution of Interacted Circumstance Variables to IOP Using the MLD Index (2002-2021)*

Year	incom16		pareduc	parsei10	race		reg16		Black _pareduc	Black _parsei10	Black _South16	extremInc _pareduc	extremInc _parsei10
	Other income classes	Extreme income classes			Non-black	Black	Other regions	Southern regions					
2002	0.061	0.003	0.058	0.067	0.031	0.027	0.105	0.036	0.021	0.008	0.016	0.008	0.008
2004	0.021	0.017	0.065	0.100	0.001	0.040	0.037	0.006	0.017	0.008	0.009	0.003	0.007
2006	0.064	0.009	0.076	0.066	0.013	0.008	0.041	0.011	0.007	0.005	0.009	0.003	0.005
2008	0.014	0.005	0.204	0.067	0.004	0.037	0.119	0.025	0.041	0.027	0.025	0.005	0.000
2012	0.060	0.000	0.116	0.092	0.000	0.025	0.095	0.033	0.010	0.006	0.013	0.002	0.005
2014	0.033	0.010	0.091	0.044	0.004	0.019	0.028	0.025	0.011	0.007	0.053	0.004	0.008
2016	0.014	0.000	0.115	0.095	0.000	0.028	0.068	0.019	0.039	0.021	0.019	0.004	0.006
2018	0.029	0.007	0.092	0.101	0.027	0.033	0.021	0.048	0.017	0.009	0.010	0.006	0.017
2021	0.033	0.004	0.131	0.120	0.007	0.019	0.025	0.015	0.014	0.009	0.014	0.006	0.012

Notes. Values are percentages and rounded to the second decimal.