UCO Cahiers de recherche en économie-gestion
N° 2025-eco-18

UCO Working Paper Series in Economics and Business Administration
N° 2025-eco-18

Novembre 2025

Long Run Evolution of Inequality of Opportunity across Age and Cohorts: Evidence from the French Permanent Demographic Sample

Geoffrey Teyssier
UCO Angers

Laurent Toulemon
INED

UCO UNIVERSITÉ CATHOLIQUE DE L'OUEST

EGEI
Éthique et Gouvernance de l'Entreprise et des Institutions

Université catholique de l'Ouest
Faculté de Droit, Économie, Gestion et Science politique
3 place André Leroy, 49008 Angers, France
Responsable : Karl-F. ISRAEL, kisrael@uco.fr
https://recherche.uco.fr/equipe/egei

Long Run Evolution of Inequality of Opportunity across Age and Cohorts: Evidence from the French Permanent Demographic Sample

2024

by Geoffrey Teyssier (INED, ENSAI) and Laurent Toulemon (INED)

This paper offers a unique age-cohort analysis of French wage inequalities and inequality of opportunity using the French Demographic Sample. We observe the childhood circumstances of individuals born between 1958 and 1980 in census data from 1975 and 1990 and track their yearly wages throughout their salaried life in administrative payroll data from 1984 to 2018. We find that inequalities increase substantially throughout life and decrease slowly across cohorts. Additionally, we show that father's education is the main driver of inequality of opportunity and that one's own education not only transmits but amplifies the effect of childhood circumstances on wages.

1. Introduction

A growing body of the literature about economic inequalities pursues a more ambitious agenda than solely characterizing the overall level of inequality in income or in other monetary outcomes such as earnings or consumption. It seeks, in addition, to quantify the level of inequality of opportunity (IO), defined as the part of total inequality explained by "circumstances" (Roemer 1993) which are determined at birth or during childhood (e.g., social background or sex). Estimates of IO have been burgeoning in the last two decades and now exist in diverse countries; see in this regard the reviews by Brunori, Ferreira and Peragine (2013), Ferreira and Peragine (2015), Roemer and Trannoy (2015), Ramos and van de Gaer (2016) or Palmisano and Peragine (2022). These estimates have normative appeal. They represent the fraction of inequality that can be considered unambiguously unfair, to the extent that people cannot be held responsible for their unchosen circumstances. They may also help to better understand people's demand for redistribution, as preferences appear in practice to match the normative view according to which IO should be eliminated whereas total inequality should not necessarily be so (among others, see Konow 2000 for a laboratory experiment, Fong 2001 for a social attitude survey, or Cappelen et al. 2014 for a neuroeconomic study). Research in the field has also been motivated by the important detrimental impact that IO may have on economic development, especially in the form of foregone investments in human capital. This line of thought, which suggests there is room to improve both growth and equity through appropriate policies, has been chiefly defended by the World Bank (2005) and has since then received empirical support in the US (Marrero and Rodríguez 2013) or Brazil (Teyssier 2017a).

The evolution over time of IO is however still poorly known. Its estimates have in most countries been restricted to one or a handful of years, due to lacking microdata on circumstances. We fill this gap regarding France thanks to a unique dataset, the Permanent Demographic Sample, also known as "Échantillon Démographique Permanent" (EDP). The EDP tracks individuals born between October 1st and 4th (about one percent of the population) who live in France across various data sources dating back to the late 1960s, including census and administrative payroll data (Jugnot 2014). We observe various circumstances in the censuses, where individuals can be detected in their parental home when they were still kids or teenagers, and we observe the earnings of the same individuals all along their working salaried life in the payroll data. We can then study the evolution of IO over a long period (1984-2018), for the first time in France or elsewhere by cohort and on a yearly basis by age.

Tables A7.1 and A7.2 in the (Web) appendix put our data is in perspective with the rest of the

empirical literature about the evolution of IO in income or earnings.[1] We are aware of only two studies attempting to disentangle its evolution across cohort from that over the lifecycle, as we do: Andreoli et al. (2021) and Bussolo, Checchi and Peragine (2023). The cohorts in the two studies are observed in France and other European countries. They are however not observed at comparable points of their lifecycle. For instance, the earliest cohort (1945-55) in Andreoli et al. (2021) is only observed after 50 years old (in 2005 or later) whereas the latest one (1975-85) is only observed before 44 years old (in 2019 or earlier). [2] Simple comparisons across age and cohorts are confusing the two temporal dimensions in such situations. Bussolo et al. (2023) rely on Deaton and Paxson's (1994b) decomposition to fix this issue. This decomposition yields the age profile of an average cohort and the cohort profile of an average age group, but it assumes away any interaction effects between age and cohorts.

The other studies disregard the evolution of IO over the lifecycle. They make comparisons across cohorts only, which they observe at a single but comparable age range across the successive waves of a survey.[3] Marrero and Rodríguez (2016) consider the longest time period (1970-2009) and the largest number of waves (33 of them), thanks to the yearly frequency of the Panel Study of Income Dynamics (PSID) in the US. Other countries where numerous temporal observations (about 15 or more) have been obtained include Australia (Wu et al. 2021), Germany (Brunori and Neidhöfer 2021), South Korea (Moramarco, Palmisano and Peragine 2020) or the UK (Bussolo et al. 2023).

Regarding France, estimates of IO based on traditional measurement methods have been obtained in only a few cohorts, observed five years apart or more. Lefranc, Pistolesi and Trannoy (2004, 2009) document IO in France in the five waves of the Family Budget (*Budget de Famille*, BdF) survey between 1979 and 2000, and Bussolo et al. (2023) do it additionally in the sixth 2005 wave. Lefranc, Pistolesi and Trannoy (2011) use the 1977 and 1993 waves of the Training and Vocational Skills (*Formation et Qualification Professionnelle*, FQP) survey. Dhérbecourt and Flamand (2023) compares the years 2011 and 2019 and is one of the few

[1] Our review of the literature is more comprehensive in Tables A7.1 and A7.2. It is as exhaustive as possible regarding the studies considering France in at least two points in time. We only review the studies with more than five temporal observations per country otherwise. Studies considering a single point in time are not reviewed.

[2] To be fair though, this is an extreme example and the other cohorts in the two studies are observed over age intervals that do intersect in varying degrees.

[3] Another branch of the literature makes cross-age or cross-cohort comparisons over a single period (Bourguignon, Ferreira and Menéndez 2007; Golley, Zhou and Wang 2019; Dai and Li 2021). These comparisons are interesting in themselves but are not meant to reflect one temporal dimension in particular, since cohorts are completely separated age groups in a given period. Andreoli et al. (2021), cited earlier, can be linked to these studies since the bulk of its cross-cohort comparisons is made separately by survey year.

studies relying on the EDP, as we will comment later. The other studies, quite numerous, compare France and other European countries across the 2005 and 2011 waves of the EU Statistics on Income and Living Conditions (EU-SILC) survey, and more rarely across the 2019 wave as well (e.g., Andreoli et al. 2021). This annual survey contains retrospective questions about parental background only in these three waves. Still, two studies manage to consider additional waves by relying on non-traditional methods to measure IO: del Valle-Inclán (2019) from 2003 to 2016 and Carranza (2022) from 2005 to 2019. [4]

In comparison, thanks to the EDP, we can study the yearly evolution in France of IO in earnings from 1984 to 2018 in four cohorts born between 1958 and 1980. The EDP presents several cumulative advantages over typical survey data for the purpose of disentangling this evolution across the age and cohort dimensions. First, its sample size is large enough to estimate IO separately by cohort-age groups. This first point avoids imposing the functional form assumption of no interaction effect between age and cohort: cross-cohort comparisons can be made at each given age, and cross-age ones can be made within each cohort.

Second, the sample size of the EDP is also large enough to define narrow cohorts. We consider cohorts of four-year interval, which consist of three consecutive even birth years: 1958, 1960 and 1962 for instance for the earliest cohort. In comparison, most studies in the field define cohorts of twenty years or more. Besides increasing the number of cohorts observed, this second point is important to properly distinguish the age and cohort dimensions. Consider for illustrative purposes a situation where two cohorts are observed over the same but large age range, say 30 to 50 years old, each in a distinct survey year. The two cohorts are such that: (i) inequality is the same across the two cohorts at each and every year of age; (ii) it increases with yearly age. In this simple example, there is no cohort effect (and no interaction effect between age and cohort for that matter) owing to (i). Yet, owing to (ii), a simple comparison of the two cohorts would typically indicate that inequality has increased if there is a demographic change such that the average age is larger in the latest cohort (e.g., 41 versus 39). This comparison is valid as such, but it is easy to attribute it to the cohort effect whereas it in fact represents the age effect. The confusion disappears if the cohorts are sufficiently narrow to neglect this intra-

[4] Del Valle-Inclán (2019) builds a proxy for parental education or occupation out of own capital income, which he argues has a large "dynastic" component. Carranza (2022) uses the panel rotating structure of the EU-SILC to perform a fixed-effects regression on each wave. The income variation explained by the fixed-effects constitutes an upped-bound estimate of IO, as it captures the effect of virtually all circumstances but also of additional time-invariant factors (Niehues and Peichl 2014). Combined with traditional estimates, which are lower-bounds of the true IO resulting from all circumstances, including unobserved ones, these upper-bounds provide valuable intervals. Their variation is less informative though, as it tends to mimic mechanically that of total inequality.

cohort age effect.

Third, the long period coverage of the EDP combined with its yearly frequency means that we can follow precisely each cohort over a large part of its working life. The trends we uncover for each cohort can be considered reliable, as they are based on numerous data points and are marked by a high continuity. Temporal comparisons can otherwise be misleading, as they can for instance reflect a break in the data series or a temporary change in the economic conditions of the country under study. Besides, the period coverage of the EDP is also lengthy enough to capture meaningful long-term changes between cohorts: almost a generation separates our earliest and latest cohorts, which were born respectively around 1960 and 1978.

The rest of the paper is organized as follows. Section 2 presents the EDP data (subsection 2.1) and how we treat it (subsections 2.2 to 2.5), before ending with some descriptive statistics about wages (subsection 2.6). Section 3 describes how we compute inequalities (subsection 3.1) and presents our main results about their evolution by age and cohorts (subsection 3.2). Section 4 discusses these results. It puts them in perspective with the empirical literature about inequality and intergenerational mobility (subsections 4.1 and 4.2). It also rationalizes them theoretically (subsection 4.3), based on a model inspired from Haider and Solon (2006) about mobility. Additionally, it explores the implication of one of our results – namely, a steep increase of inequalities over the life course – on the comparability of inequality estimates across studies with differing age ranges (subsection 4.4). Section 5 provides a methodological clarification on how to perform a Shapley decomposition of inequality by circumstances (subsection 5.1) and presents the ensuing results about their relative importance (subsection 5.2). Section 6 explores the mediating role of education between childhood circumstances and adult wages. It proposes an analytical framework to simultaneously decompose the variances explained by education and/or circumstances (subsection 6.1), and it presents the ensuing results (subsection 6.2). Section 7 concludes.

2. Data

2.1.Presentation of the EDP

The EDP is an exceptionally rich socioeconomic panel constructed by the French Institute of Statistics (*Institut de la statistique et des études économiques*, INSEE). It combines census and various administrative data about the French population since the late 1960s. In more detail, it follows individuals born between October 1st and 4th across the following data sources, as well as over time within each of them: payroll data since 1967, civil register data since 1967, census data since 1968, electoral register data since 1990, and tax return data since 2011. Twelve other days of birth have been added to the EDP since 2002 but not retroactively, so individuals born these days cannot be observed before 2002 even if they were born earlier. The different data sources are made available in separate tables, to which is added a synthetic table called *Table Individu* that summarizes a few key demographic information from the above sources. These tables can be merged together thanks to an anonymous identifier specific to each individual. INSEE is in charge of constructing this identifier. It has access to this end to highly identifying information such as the name and national identification number, which are then not included in the EDP itself for confidentiality reasons.

The EDP has nevertheless been relatively underused in the literature, for three main reasons: its access is restricted due to privacy regulations; few of its documentation is in English; [5] and its heterogeneous sources require a heavy harmonization treatment, as they were not originally meant to be used together and are mainly made available in a raw format. We are aware of only two studies that have relied on the EDP to study IO so far: Dhérbecourt (2018) and Dherbécourt and Flamand (2023). [6] They compute IO in one or two points in time (the pooled period 2011-2014 for the former and the years 2011 and 2019 for the latter) out of the tax data of the EDP, whereas we compute IO across several cohorts in many years out of the payroll data of the EDP. Also, they compute IO respectively in terms of household equivalent disposable income and of "active" income, which gathers wages, self-employment income, unemployment benefits and all other sources related to one's activity. These outcome variables are more comprehensive than wages alone, on which we focus on. Household income in particular depicts better one's standard of living or purchasing power. However, they are only available since 2011 in the EDP

[5] See Jugnot (2014) and the user guide (INSEE 2022) for a detailed documentation in French. See Robert-Bobée and Gualbert (2021) and its English translation, which can be found on INSEE's website, for an introduction. See also the useful website https://utiledp.site.ined.fr/en/, in both languages.

[6] But see, among others, Kenedi and Siruge (2023) and Sicsic (2023) on the closely related topic of intergenerational income mobility, Galland and Rouault (1998) on intergenerational social mobility, and McAvay (2018) on ethnic segregation.

and they aggregate several income sources, which may each follow a different dynamic in terms of inequality. Furthermore, household income neglects intra-household inequalities and tends in consequence to understate the lack of opportunity of women.

2.2.Outline of our empirical strategy

We use the EDP in three interrelated steps, which we summarize briefly below before going into more details in the next subsections. First, we use the date and place of birth in the *Table Individu* to identify individuals born in metropolitan France between the 1st and 4th of October of an even year between 1958 and 1980, a subsample with most complete and comparable information. Odd years of birth are excluded because they start being observed in the payroll data only in 2002. Individuals born outside metropolitan France are also excluded because, as explained in subsection 2.4, the cohorts would otherwise not be comparable with respect to immigration. These pieces of demographic information can be found as well in the censuses but are of better quality in this synthetic table (INSEE 2022), in which we also obtain the circumstance sex. Second, we obverse these individuals when they were children or teenagers in the 1975 and 1990 censuses, where we obtain some circumstances about their fathers and place of residence. Third, we observe these same individuals once adults in the payroll data between 1984 and 2018, where we retrieve their salary information and where we compute inequality arising from circumstances every year by cohort. For a matter of consistency, we restrict our observations to salaried people from the nonagricultural private sector. Our cohorts comprise three consecutive even birth years; the ages at which they are observed are shown in Table 1. The other data sources of the EDP are not used for the following reasons: the civil and electoral registers contain no further information about income or circumstances, and the fiscal data used by Dhérbecourt (2018) or Dherbécourt and Flamand (2023) are too recent.

2.3.Birth cohorts observed as children or teenagers and as adults

We focus on cohorts born the first four days of October between 1958 and 1980 because they are the only ones which satisfy the both following criteria. First, they can be observed at least until their late thirties in the last 2018 wave of the payroll data, which is a minimal requirement given that the evolution of IO with age drastically flattens around 30 to 35 years old (see Figure 2). Second, they can also be observed before 17 in the 1975 or 1990 censuses (see Table 1).

Table 1: Age by year and year of birth in the census and payroll data

	Year of birth (cohort)											
	1958 (C60)	1960 (C60)	1962 (C60)	1964 (C66)	1966 (C66)	1968 (C66)	1970 (C72)	1972 (C72)	1974 (C72)	1976 (C78)	1978 (C78)	1980 (C78)
Census years (before adulthood)												
1975	17	15	13	11	9	7	5	3	1			
1990										14	12	10
Payroll years (during adulthood)												
1984	26	24	22									
1985	27	25	23									
1986	28	26	24									
1987	29	27	25									
1988	30	28	26									
1989	31	29	27	25	23	21						
1990												
1991	33	31	29	27	25	23						
1992	34	32	30	28	26	24						
1993	35	33	31	29	27	25						
1994	36	34	32	30	28	26						
1995	37	35	33	31	29	27	25	23	21			
1996	38	36	34	32	30	28	26	24	22			
1997	39	37	35	33	31	29	27	25	23			
1998	40	38	36	34	32	30	28	26	24			
1999	41	39	37	35	33	31	29	27	25			
2000	42	40	38	36	34	32	30	28	26			
2001	43	41	39	37	35	33	31	29	27	25	23	21
2002	44	42	40	38	36	34	32	30	28	26	24	22
2003	45	43	41	39	37	35	33	31	29	27	25	23
2004	46	44	42	40	38	36	34	32	30	28	26	24
2005	47	45	43	41	39	37	35	33	31	29	27	25
2006	48	46	44	42	40	38	36	34	32	30	28	26
2007	49	47	45	43	41	39	37	35	33	31	29	27
2008	50	48	46	44	42	40	38	36	34	32	30	28
2009	51	49	47	45	43	41	39	37	35	33	31	29
2010	52	50	48	46	44	42	40	38	36	34	32	30
2011	53	51	49	47	45	43	41	39	37	35	33	31
2012	54	52	50	48	46	44	42	40	38	36	34	32
2013	55	53	51	49	47	45	43	41	39	37	35	33
2014	56	54	52	50	48	46	44	42	40	38	36	34
2015	57	55	53	51	49	47	45	43	41	39	37	35
2016	58	56	54	52	50	48	46	44	42	40	38	36
2017	59	57	55	53	51	49	47	45	43	41	39	37
2018	60	58	56	54	52	50	48	46	44	42	40	38

Note: The table shows the ages at which we consider individuals before adulthood in the censuses and during adulthood in the payroll data. We study them by cohorts of three consecutive even birth years (e.g., C60 for 1958, 1960 and 1962) throughout the paper. The payroll data was unfortunately not collected in 1983 and 1990 because the INSEE had an overload of work with the 1982 and 1990 censuses.

This 17 years old age limit is chosen because we have noticed a sudden rise in the proportion of people who do not live with their father at 18, which marks the end of high school and the beginning of majority in France. Indeed, we need to observe people when they still live with their father in the censuses, as paternal circumstances are obtained therein via a link variable in combination with other variables about the head of the family in the household. Note that this link variable does not distinguish between fathers and step fathers. We set it to missing when the induced age difference with the father is strictly below 15 or strictly above 50, which concerns around one percent of cases (see Table 2).

We consider three consecutive even birth years per cohort to have enough observations in each. This makes four cohorts: born in 1958, 1960 or 1962 ("C60"); in 1964, 1966 or 1968 ("C66"); in 1970, 1972 or 1974 ("C72"); and in 1976, 1978 or 1980 ("C78"). We start studying them in the payroll data when their youngest members reach 21 years old. This means for instance that the earliest cohort (C60) should start being studied in 1983 but, because the payroll data was exceptionally not collected this year, it starts in fact to be so in 1984.

2.4.Circumstances from census data

Researchers about IO typically try to consider as many circumstances as possible, provided their sample size allows it, in order to obtain a comprehensive picture of this inequality. IO estimates can otherwise be severely biased downwards (see Ferreira and Gignoux 2011b for a proof), at the risk of sending an overly optimistic message to policy-makers (Kanbur and Wagstaff 2016).

We consider sex (man/ woman) in the *Table Individu* and the following five other circumstances that we recode comparably across the 1975 and 1990 censuses: father's education (low/ medium/ high)[7]; father's occupation (low/ high/ medium and others)[8]; father's citizenship (French/ foreign) [9]; age difference with the father (23 or less/ 24 to 30/ 31 or more); and area of

[7] "Low" education corresponds to completed primary education ("niveau CEP", reached normally at age 11 with age defined as the one attained during the year) or less. "Medium" corresponds to secondary education attended but not completed ("niveau BEPC" and "niveau CAP-BEP"). "High" corresponds to completed secondary education ("niveau BAC général ou technique", reached normally at age 18) or more.

[8] "Low" occupation gathers fathers in low-skilled professions. "High" gathers those in high-skilled professions as well as those who are employers or self-employed. "Medium and others" is a residual category that mainly gathers fathers in intermediate professions. The variable is constructed out of the profession and employment statuses, whose combination is given in more detail in Table A6.1 in the appendix.

[9] We count naturalized fathers in the foreign category to capture the migratory origin and to enhance the comparability of this variable over time.

residence during youth (rural unit/ small urban unit / medium urban unit/ large urban unit).[10] This makes six circumstances and 2.3.3.3.2.4=432 "types" or combinations of circumstances, which is a rich set of information compared to other studies (see Table A7.2). The number of observations per type remains acceptable, given that the final sample on which our benchmark measures of inequality are computed (in the payroll data, after retrieving circumstances in the censuses and applying casewise deletion on them) has 4,962 observations for the smallest year-cohort. This number is of 846,744/105≈8,064 for the average year-cohort, where 846,744 is total number of observation over which we compute inequality across our 105 year-cohorts.

Each of the above six circumstances is rather standard and deserves no further comment at this point, except perhaps the age difference with the father (i.e., his age at childbirth). This variable is motivated by the sociodemographic literature about young mothers, who tend to be underprivileged in various ways (e.g., Robert-Bobée et al. 2006; Tomkinson, Breton and Mazuy 2017). We do not consider maternal circumstances because household heads are by default men in the 1975 census and because no variables about the partner of the head have been included in the EDP from this census. Now, in presence of homogamy, we can expect children of young fathers to be disadvantaged as well, which the data confirms (see Figure 1).

We use one census per cohort to retrieve circumstances: the 1975 one for the birth years 1958-1974 and the 1990 one for the birth years 1976-1980 (see Table 1). The EDP contains all French censuses since 1968 (exhaustive until 1999, and in the form of a rotating annual survey since 2004) but we do not use them all for the following reasons. No variables about the head nor a fortiori the parents have been included in the EDP from the 1968 census. The same applies to parental education regarding the 1982 census. Other parental information is available in 1982 but with a high rate of missingness because the INSEE coded family links only for a random quarter of households in the original census (Boudoul 1983, Jugnot 2014). The 1999 and later censuses are too recent, as the additional cohorts they would bring would not be observed sufficiently late during adulthood in the payroll data.

The variables coming from the 1975 and 1990 censuses contain very few missing values, with two exceptions (see Table 2). First, variables about the heads were not integrated in the EDP for 11.3 percent of individuals in 1975.[11] These individuals with missing information are

[10] A urban unit is a densely populated area that is made of one or several contiguous municipalities. Small urban units gather less than 100,000 inhabitants. Medium ones gather between 100,000 and 2 million inhabitants. There is only one large unit of more than 2 million inhabitants; it corresponds to the region Île-de-France.

[11] The reason why this information is missing in the EDP is not entirely clear. We understand from Jugnot (2014) or INSEE (2022) that these constructed variables have been directly extracted from the census files to the EDP

Table 2: Percentage of missingness in the censuses, by cause and year of birth

	Year of birth (age in 1975 if born before 1975 and age in 1990 otherwise)											
	1958 (17)	1960 (15)	1962 (13)	1964 (11)	1966 (9)	1968 (7)	1970 (5)	1972 (3)	1974 (1)	1976 (14)	1978 (12)	1980 (10)
Does not live with the father	13.7	10.9	8.6	7.5	7.5	6.3	5.9	5.2	4.6	11.7	11.0	10.6
Implausible age diff. with the father (<15 or >50)	1.4	1.3	1.2	1.4	1.3	1.2	1.0	1.5	1.5	1.0	0.8	0.8
Other (unanswered)	0.5	0.5	0.3	0.8	0.8	0.7	0.9	0.6	0.9	0.0	0.0	0.0
Total	15.5	12.6	10.0	9.7	9.6	8.2	7.7	7.3	7.1	12.7	11.8	11.4

Notes: The table shows the percentages of missingness on circumstances that we face in the censuses. They are computed among the cohorts of interest described in subsection 2.3, in the 1975 census for the pre-1975 birth years and in the 1990 census for the post-1975 birth years. The cohorts are not restricted to individuals who appear as well in the payroll data but this affects in any case little the percentages shown. We face in the end similar rates of missingness by birth year when computing inequality, as there is no missing value on earnings in the payroll data.

The percentages are shown by birth year to emphasize how the co-residency with the father, a necessary condition to observe paternal circumstances, decreases with age. The co-residency is lower at comparable ages for the 1976-1980 than for the 1960-1966 birth years because there are more single mothers in 1990 than in 1975. Two other causes of missingness are considered: having an implausible age difference with one's (potential) father, in which case we purposely set paternal circumstances to missing; and a residual category made of "don't know/ don't want answer" among those who live with their plausible father. The three sources do not always exactly add up to the totals shown, due to rounding.

There is also a fourth source of missingness in 1975, which affects a random fraction of individuals in some municipalities, as explained in the main text. This fourth source of missingness is not counted as such in the table. Instead, we use sampling weights among the sampled individuals to obtain representative statistics here or elsewhere in the paper.

exclusively located in 130 municipalities (out of approximately 36,000 in total in France), of which they represent a random 80 percent fraction (INSEE 2022). The sample sizes of the cohorts observed in 1975 are still satisfyingly large, as already mentioned, and we use a weight variable to obtain representative statistics about them. The weight is of five for the remaining 20 percent of individuals from the 130 above-mentioned municipalities, and it is otherwise of one (it is of one for all municipalities regarding the C78 cohort, observed in the 1990 census).

Second, we cannot retrieve the paternal circumstances of people who did not live their father at the time of the census, in 1975 or 1990. Those with observed paternal circumstances, who lived with their father during their youth, remain nevertheless representative of a target population worthy of interest. They are in any case representative of the bulk of their underlying cohort, as the absence of co-residency with the father is moderate among young people aged 17 or less. Besides, this co-residency is unlikely to have a large impact on our inequality estimates because it is largely explained by the specific age before 17 at which we observe individuals (see Table

(i.e., they have not been reconstructed for the purpose of the EDP) and that they have not been coded exhaustively in the first place in the original 1975 census data. Yet, we have had access to the latter piece of data as well and it does not seem to be concerned by the missingness. A possible explanation is that these variables have been recoded exhaustively in the census files more recently, after the construction of the EDP files.

2), which has no direct connection with our variables of interest. Now, there may be an indirect connection through the year of birth. The co-residency, which decreases rapidly as children get older, is for instance at its lowest among the 1958 birth year, observed at 17 in 1975 and among which the induced missingness reaches 13.7 percent. Our concern here is this may alter our cross-cohort comparisons at a nominally comparable age in the payroll data: the 1958 birth year is underrepresented among the C60 cohort, and so this cohort gives less weight to its oldest members than the other cohorts. We address this concern in the appendix (Figure A2.1), where we consider reweighting birth years within cohorts, with virtually no impact on our results.

The fact that our cohorts are not observed at the same age during their youth raises two other issues. First, there is a problem of differential selection into immigration: it risks being artificially high in the cohorts observed at the end of adolescence, since immigrants can only appear in a census if this one takes place after they have arrived in France. We solve this problem of differential selection by restricting our cohorts to individuals born in metropolitan France. Individuals born in French overseas are excluded as well because these territories were not part of the 1975 census data included in the EDP. Second, our circumstances variables besides sex are not time-invariant during childhood. For instance, a father may move, change jobs or receive post-school training.[12] Circumstances may thus not capture exactly the same reality depending on the age at which they are observed. We show in the appendix (Figure A2.2) that this is not a serious source of concern thanks to people born in 1974, who can alternatively be observed at age 1 in 1975 and at age 16 in 1990.

2.5.Earnings from payroll data

The payroll data comes from the "All Employee Panel", which follows individuals born in October (of which the first four days are included in the EDP) over their working salaried life. This yearly panel has been constructed by the INSEE out of administrative data which employers have the legal duty to fill. Started in 1967, it was originally restricted to individuals born an even year who are employed in what we call for short the "nonagricultural private sector" which, more precisely, excludes domestic workers and overseas territories but includes state-owned companies. Starting from 1988, its scope has gradually but not retroactively been enlarged, and it now covers all employees born in October in the country. As already explained,

[12] The age difference with the father and his citizenship of origin may also vary between the two censuses, either due to measurement errors or to a change in the identity of the father, who can actually be the stepfather.

we focus on cohorts made of even birth years and we start studying them during adulthood in 1984. This is before the enlargement of 1988, so we restrict the payroll data of the EDP to individual-year observations that correspond to the nonagricultural private sector. [13]

The earnings variable on which we compute inequality corresponds to yearly net wage (after payment of social contributions but not of income tax). Extreme values at the top or the bottom can have a disproportionate influence on inequality estimates in general (Cowell and Victoria-Feser 1996; Cowell and Flachaire 2007). The top 0.1 percent of the payroll data happens to be quite volatile, but the top 1 percent is more stable (see Figure A6.1 in the appendix). We therefore winsorize the top 1 percent to smooth our inequality series. We do so separately by years and birth years to treat all cohort-age groups comparably (as opposed to a winsoriztion by year only) and to avoid inadvertently introducing an intra-cohort bias related to age (as opposed to a winsorization by year-cohort). We proceed somehow differently regarding the bottom of the distribution because it differs from a substantive point of view from the rest of our sample, which is more firmly entrenched in the labor market. Namely, we drop the observations that correspond to strictly less than one hundred paid days in the year, whose wages are in general low in yearly terms, as well as those whose full-year equivalent wage is strictly below the legal minimum, which can happen for people who work on a part-time basis during their paid period. [14] Alternative treatments of the two tails of the distribution are considered in the appendix, where we also compare the results obtained from bottom-sensitive versus top-sensitive inequality indices, as further discussed in section 3.

Lastly, we rescale earnings in our estimation sample (i.e., after performing casewise deletion on circumstances) so that age effects are on average eliminated within our cohorts of three birth years. We proceed in each wave of the payroll data by dividing individuals earnings by their

[13] We use five variables to select this sector: *nes5* and *cs2* to exclude the agricultural sector, *domempl* and *cat* to exclude the public sector and domestic workers, and *dept* to exclude overseas territories (see the user guide, INSEE 2022, for the variable labels). Some of them are entirely missing before the 1988 enlargement (or before a later enlargement, depending on the variable) but the meaning of these missing values is implicit: they refer to the nonagricultural private sector. We impute these missing values accordingly, meaning that the corresponding observations are kept. The other missing values (less than 0.1 percent of the sample) are imputed with the closest available value in time for each individual. There remain less than 20 observations whose belonging to the nonagricultural private sector is unclear in the end. We drop them. No other observation is dropped due to a missing value on one of the payroll data's variables.

[14] Yearly data on the legal minimum net wage is obtained from the following series of the INSEE: "sm01" before 2005, "sm02" between 2006 and 2012, and "000879878" afterwards. In 2000, the standard legal number of working hours in full-time employment was reduced from 39 to 35 hours per week. The minimum wage, defined in hourly terms by law, was not increased substantially but employers were forbidden from cutting monthly wages owing to this reduction of working hours. Hence, the 39-hours basis was in effect maintained in monthly or yearly terms. We thus avoid a break in the series by converting the post-2000 figures, originally given on a 35-hours basis in the "sm02" and "000879878" tables, to a 39-hours basis.

average in the corresponding birth year. Denoting the non-rescaled earnings of individual *i* in year *t* by w_{it}, our rescaled earnings variable is therefore

$$y_{it} = \frac{w_{it}}{\frac{1}{N_{bt}} \sum_{i \in b} w_{it}}, \quad (1)$$

where N_{bt} is the number of individuals in the combination of year *t* and birth year *b*.

2.6.Descriptive statistics about wages and circumstances

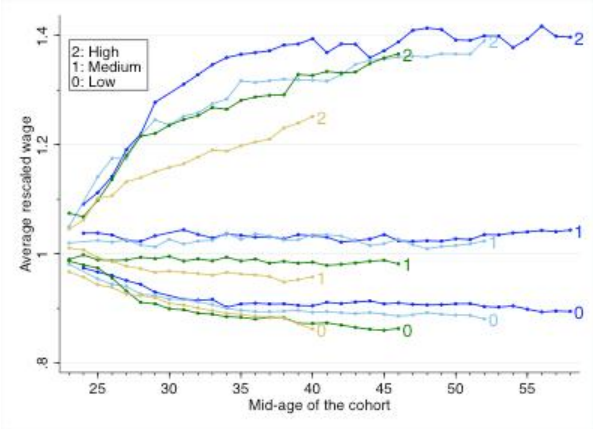
We conclude this section with Figure 1, which shows the average wage of each category of circumstances over the life course of our cohorts. The corresponding population shares are relegated to the appendix (Figure A1.1). Each panel is dedicated to one circumstance and all panels are put on the same scale to ease comparisons across circumstances. The x-axis corresponds to the "mid-age" of a cohort (e.g., 30 years old in 1990 for the C60 cohort, born between 1958 and 1962). We will often refer to this mid-age simply as age when commenting this or other figures throughout the paper. The y-axis is computed over rescaled wages y_{it}. Note that, by eq. (1), the average of y_{it} is one in each year-year of birth. It is a fortiori one in each year-cohort. Hence, the y-axis gives the average wage of a circumstance category in a given cohort-age group relative to the grand average, of one, of this cohort-age group. It can also be interpreted as the ratio of the wage share of this category to its population share. Several points are worth noticing about the figure.

First, these is no surprise as to which categories are advantageous or not. The disadvantaged ones gather women; individuals who grew up in a rural area; and those whose father is young (aged 23 or less at the birth of the child), of foreign origin, or has a low educational or occupational status. These individuals are on average poorer than the other members of their cohort at the same age in virtually all of the 105 combinations of years and cohorts considered. Some intermediate categories about which we had no strong prior expectations are also below the average (i.e., one), but only slightly so. They gather individuals who grew up in small urban areas; those with "old" fathers (aged 31 or more at birth); and, for the two most recent cohorts only, those with middle-educated fathers. We note in passing that the relationship between wages and the age of the father at birth is not monotonic.

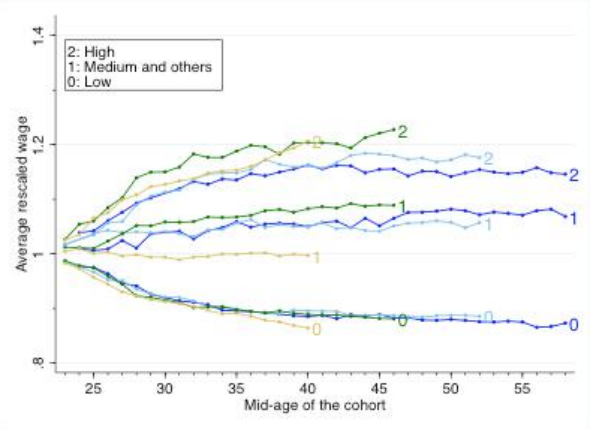
Second, the disparity between circumstances mainly comes from the advantaged categories. For instance, individuals with high paternal education are around 40 percent above the average

Figure 1: Average rescaled wages over the life course, by cohorts and circumstances

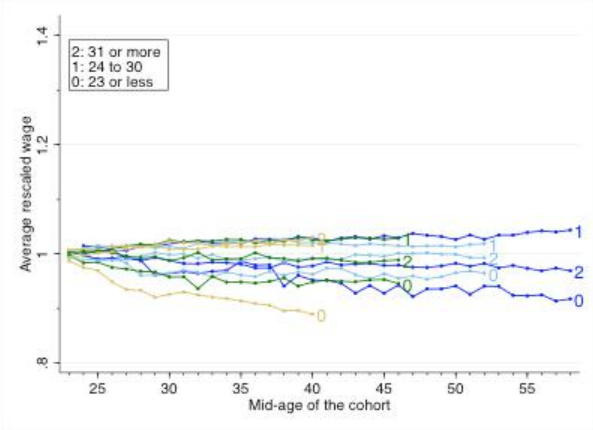
By father's education



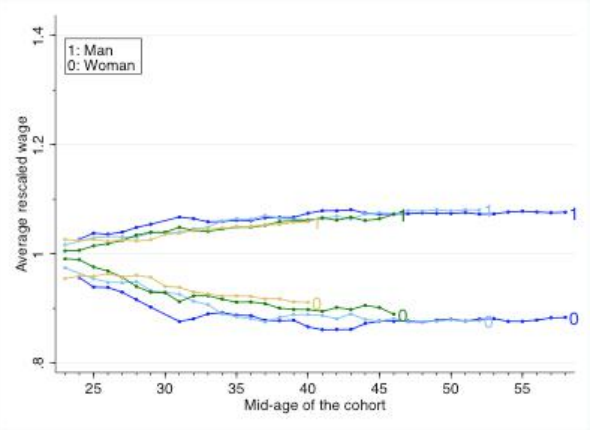
By father's occupation



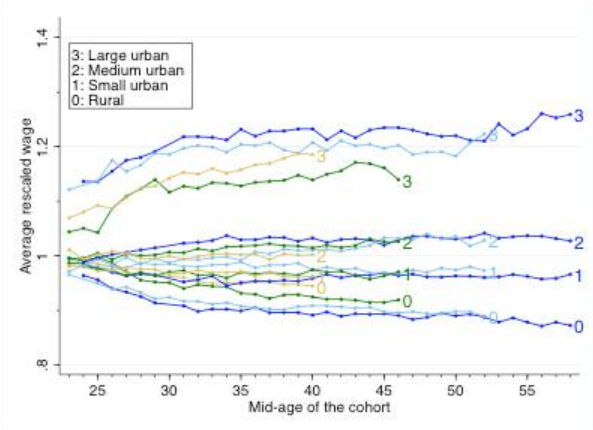
By age difference with the father



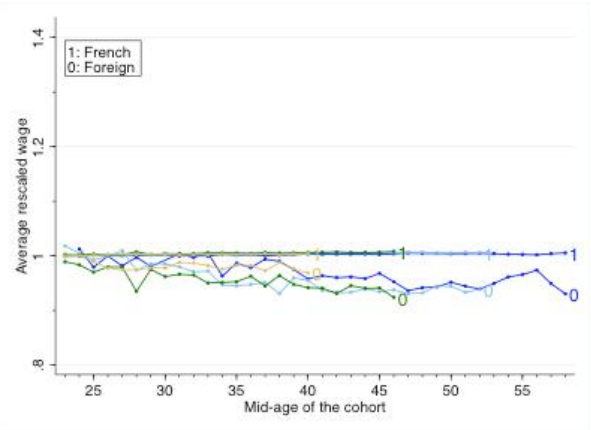
By sex

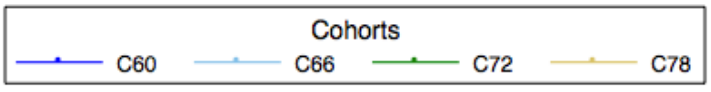


By area of residence during youth



By father's citizenship





Note: The figure shows the average wage of each circumstance category by year-cohort or, equivalently, by (mid) age-cohort. Wages are rescaled by year and year of birth, and so the overall average of each year-cohort is set to one. The estimation sample is the same in all panels (i.e., after casewise deletion on all circumstances).

in the earliest C60 cohort after 40 years old whereas those of French paternal origin are barely (indistinguishably so graphically) above it in almost all cohort-age groups. The disadvantaged categories fall in contrast within a rather tight wage range, never dropping below 0.86 (i.e., 14 percent less than the average). We attribute this result to two notorious features of the French labor market: the legal minimum wage is high by international standards, and there are tax deductions on low wages. These deductions aim at fighting unemployment among the low-skilled by cutting labor costs, but they give firms a strong incentive to resist wage increases at the bottom and thereby pull many wages (and the average with them) towards the minimum.

Third, circumstances can be ranked as follows with respect to their capacity to explain wage differentials. Father's education comes first by far. On average across year-cohorts between mid-ages 31 and 40, when our estimates are more or less stabilized and when all cohorts are observed every year, the difference between its High and Low categories represent 39 percent of the year-cohort average. The corresponding number is 27 percent for father's occupation (High versus Low), 26 percent for the area of residence during youth (Large urban versus Rural), 15 percent for sex (Man versus Woman), 7 percent for the age difference with the father (24 to 30 versus 23 or less), and 3 percent for father's citizenship (French versus Foreign). In Figure A1.2 in the appendix, we show the simple regression coefficients associated with these wage differentials together with their 95% confidence intervals. They are significantly different from zero in the vast majority of the combinations of years and cohorts considered, except for father's citizenship. The latter has a large standard error and is not significant regarding the earliest C60 cohort, in which individuals of foreign paternal origin are fewer (see Figure A1.1 about population shares).

Fourth, Dhérbecourt and Flamand (2023) also compare circumstances based on the earnings differences they generate in France. Their ranking resembles ours in terms of order but not in terms of magnitude. They find greater earnings differentials than we do, particularly by migratory origin: 19 percent of their sample average separate people of African origin and those whose two parents were born in France. We have considered breaking our Foreign category more finely to compare our results with theirs but this is not alone what drives the discrepancy: the 3 percent difference mentioned earlier between our French and Foreign categories only rises to 9 percent between the French and African ones.[15] The discrepancy also stems from our

[15] We do not rely on this finer categorization of father's citizenship in our main analysis because people of African origin represent only 2.5 percent of our estimation sample on average across year-cohorts (vs. 9.0 percent for all persons of foreign origin). This would result into very sparsely populated combinations of circumstances, which would introduce noise in our IO estimates and bias them upwards. Besides, this would bring little signal to our IO

exclusion of very low yearly wage earners, many of whom do not work on a full-year, full-time basis: the 3 percent difference between our French and Foreign categories rises to 8 percent when all wage earners are considered; and the 9 percent difference between the French and African categories rises to 20 percent, a number close to that found by Dhérbecourt and Flamand (2023).These results indicate that second generation immigrants, particularly those of African origin, suffer from a greater disadvantage at the hiring stage (finding a full-year, full-time job) than at the remuneration stage (in hourly terms, once hired), which aligns with the findings of Aeberhardt et al. (2010).

Fifth, these wage differentials tend to increase with age and to decrease, but more slowly and less robustly, as time goes by across cohorts. An eyeball comparison of confidence intervals in Figure A1.2 shows that the increase is often significant after a few years of age at the beginning of adulthood. Note that it would be even more pronounced had we not rescaled earnings in the first place since earnings (and their differences) tend to increase in absolute terms over the life course. Regarding the decrease across cohorts, the confidence intervals associated with our extreme cohorts (C60 and C78) do not overlap, or barely do so, for father's education (High and Medium vs. Low), sex and place of residence during youth (Large urban and Medium urban vs. Rural). The cross-cohort evolution is unclear for father's citizenship. It goes in the other (upward) direction for his age: as is visible in Figure 1, having a young father (aged 23 or less at birth) is more disadvantageous in the C78 cohort than in the three earlier cohorts. Becoming a parent early in life has likely become more challenging owing in part to the expansion of tertiary education. For father's occupation, we can also detect a clear widening gap across cohorts between the High and the intermediate "Medium and others" categories. [16]

estimates anyway: the R-squared from the simple regression of earnings on father's citizenship is consistently lower than one percent in each of our 105 year-cohorts when it is coded into French, African, European and Other.

[16] The gap between the Low and intermediate categories has tightened but rescaled wages in the Low category have remained rather stable across cohorts and so the relevant comparison here is between the High and intermediated categories.

3. Benchmark results about the evolution of inequality

3.1.Foreword on inequality measurement

This section shows our benchmark results about the evolution of inequality across age and cohorts. Three measures of inequality are considered: IO in relative terms ("*IOR*"); IO in absolute terms ("*IOA*"); and total earnings inequality ("*Itot*"), which is computed as well after performing casewise deletion on circumstances. *IOR* is the share of total inequality resulting from circumstances. It is the ratio of *IOA* to *Itot*: *IOR=IOA/Itot*, where *IOA* is the inequality that exists between the types formed by the six circumstances considered in Figure 1. It is comprised between 0 and 100 percent since *Itot* is the sum of *IOA* and a residual within-type term, which are both positive. This decomposition within and between types is called "ex-ante" in the literature. Following Ferreira and Gignoux (2011b), we achieve it parametrically, meaning that we compute *IOA* by applying an inequality index over a counterfactual distribution made of the predicted values of earnings regressed on circumstances. This approach is parametric in that it assumes away the interaction effects that circumstances may have between each other. It is more robust against sampling error than the alternative non-parametric approach, which uses the average earnings value of each type instead of its predicted value.

We use the *GE2* inequality index, also known as the half squared coefficient of variation. Accordingly, we compute *Itot* in each year-cohort as

$$GE2_{ct}(\{y_{it}\}) = \frac{1}{2N_{ct}} \sum_{i \in c} \left[\left(\frac{y_{it}}{\frac{1}{N_{ct}} \sum_{i \in c} y_{it}} \right)^2 - 1 \right], \tag{2}$$

where $\{y_{it}\} = (y_{1t}, \dots, y_{it}, \dots, y_{N_{ct}})$ is the vector of earnings in the year-cohort *ct*, to which individual i belongs. The corresponding formula for our parametric measure of *IOA* is $GE2_{ct}(\{\hat{y}_{it}\})$. It is simply obtained by replacing in eq. (2) the actually observed earnings values y_{it} with their OLS predicted value $\hat{y}_{it} = y_{it} - u_{it}$ (we abstract from the hat notation on the residual and its underlying coefficients) obtained from the year-cohort regression

$$y_{it} = \beta_{0ct} + \boldsymbol{X_i \beta_{ct}} + u_{it}, \tag{3}$$

where $\boldsymbol{X_i}$ is our row vector of categorical circumstance variables, which are time-invariant once adult, and where $\boldsymbol{\beta_{ct}}$ is its associated column vector of coefficients. Ferreira and Gignoux

(2011b) use in contrast the *GE0* inequality index, also known as the mean log deviation. [17] Both indices are members of generalized entropy family, which gathers all scale invariant inequality indices that are additively decomposable (Bourguignon 1979). The *GE0* index is more often used in IO the literature because it is the only one to be additionally path-independent (Foster and Shneyerov 2000). [18] This practice is justifiable in a non-parametric setting. However, when applied over a counterfactual income distribution estimated parametrically, only the *GE2* index satisfies the more essential between-type transfer principle, according to which IO must increase when some money is transferred from one type to a richer one (Teyssier 2017b).

Note that our measure of *IOR* corresponds in fact to the R-squared from eq. (3). We have $GE2_{ct}(\{\hat{y}_{it}\})/GE2_{ct}(\{y_{it}\}) = V_{ct}(\{\hat{y}_{it}\})/V_{ct}(\{y_{it}\})$ because the variance equals the *GE2* index times twice the squared average of the distribution under consideration and because $\{y_{it}\}$ and $\{\hat{y}_{it}\}$ share by construction the same average. Note also that our measures of *Itot* and *IOA* could equivalently be computed with the variance since the variance collapses to twice the *GE2* index when computed over a distribution whose average is one. This is the case of $\{y_{it}\}$ (and thus of $\{\hat{y}_{it}\}$) since, as already mentioned, we have set its average to unity in each year-cohort in the process of doing so in each year-year of birth (hence, we have $\frac{1}{N_{ct}}\sum_{i\in c} y_{it} = 1\ \forall c,t$ inside eq. (2)). Now, we prefer to refer to the *GE2* index because the variance is not scale invariant in general, when the distribution is not rescaled. This property, which always holds for the R-squared, is in our view essential for studying the evolution of inequality with age. Earnings tend indeed to grow with age in a multiplicative rather than additive fashion, as suggested by Mincer (1974)'s log-linear specification, which is known to depict well the evolution of earnings over the lifecycle (Polachek 2007). A non-scale invariant index such as the variance (without rescaling) would in in this framework mechanically increase as earnings grow with age, which would not be very informative.

[17] They compute *Itot* as $GE0_{ct}(\{y_{it}\}) = \frac{1}{N_{ct}}\sum_{i\in c} \ln\left(\left(\frac{1}{N_{ct}}\sum_{i\in c} y_{it}\right)/y_{it}\right)$. However, they do not compute *IOA* as $GE0_{ct}(\{\hat{y}_{it}\})$ but as $GE0_{ct}(\{\tilde{y}_{it}\})$, where the typical element in $\{\tilde{y}_{it}\}$ is $\tilde{y}_{it} = \exp\left(\widehat{\ln(y_{it})}\right)$. They rely on exponentiated predicted logs to build their counterfactual income distribution because predicted incomes can be negative, in which case the *GE0* index is undefined. Teyssier (2017b) shows that no inequality index applied over $\{\tilde{y}_{ct}\}$ can satisfy the between-type transfer principle, defined later in the paragraph.

[18] Scale invariant indices do not change when all incomes vary in the same proportion. Additively decomposable ones are such that the residual term (*Itot-IOA*) is more than just a residual and is expressible as a weighted average of the inequality within each group. Path-independent ones are such that both *IOA* and (*Itot-IOA*) can be computed directly over a counterfactual distribution from which the other part of inequality has been removed.

3.2.Results

Figure 2 plots the evolution of inequality with age separately by cohorts. The main takeaway is that, on the overall, *IOR*, *IOA* and *Itot* increase with age and decrease, but less rapidly, as time goes by across cohorts. As one would expect, the first two (*IOR* and *IOA*) go in the same direction as most wage differentials explained by circumstances in Figure 1.

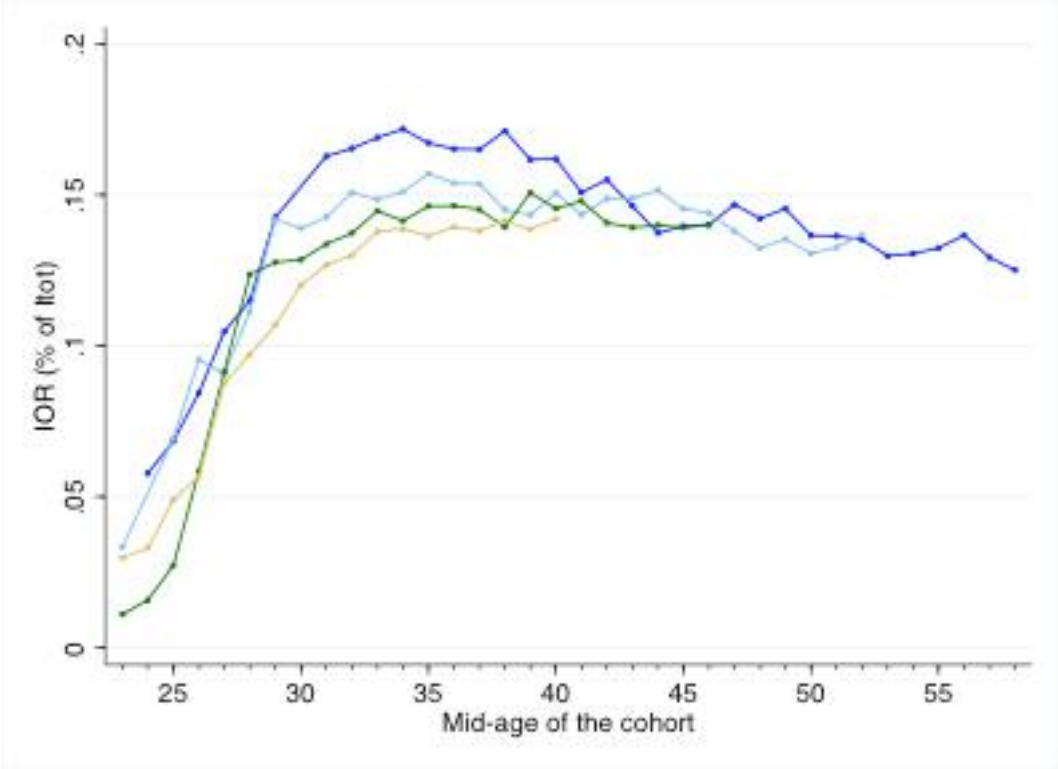
Panel (a) shows that *IOR* follows a clear concave relationship with age: it increases very fast at the beginning of adulthood, starts to decelerate around (depending on the cohort) 28 years old, reaches a maximum or a near-maximum around 35, and stagnates or slowly decreases afterwards. Its overall increase over the life course is substantial. At its minimum, before age 24, *IOR* represents from 1.1 to 5.8 percent of total inequality depending on the cohort. At its maximum, between ages 34 and 40, it represents from 14.2 to 17.2 percent of total inequality.

The decrease of *IOR* across cohorts is not constant over the life course, which shows that some interaction effects are at play between the age and the cohort dimensions. We can distinguish three phases. The cross-cohort decrease is the clearest during the "middle phase", from 28 to 42 years old, in which it applies to virtually all pairwise combinations of years and cohorts. During this phase, the gap between cohorts first widens in absolute value until around 31, then stagnates until around 35, and finally tightens back. There seems to be no systematic differences left across cohorts in the "final phase", after age 42, but this result is not highly robust (see Figure A2.3 in the appendix, where the cross-cohort decrease is maintained and is only less marked than at younger ages). Note that cross-cohort comparisons are less instructive in this final phase because we lose the C78 cohort after 40 and the C72 one after 46. The remaining cohorts are not only fewer but are also separated by fewer birth years. In the "initial phase", before age 28, the decrease from one cohort to the next tightens with age if we omit the three particularly low starting values (below 3 percent) of the C72 cohort, in green. These three values, and the rest of our *IOR* estimates before 25, should be interpreted cautiously. They are subject to higher sampling variability because they concern young adults who are not yet firmly entrenched in the labor marker. Moreover, *IOR* tend to overreact to small variations of *Itot* or *IOA* at these young ages, in which *Itot* and *IOA* are themselves very low.

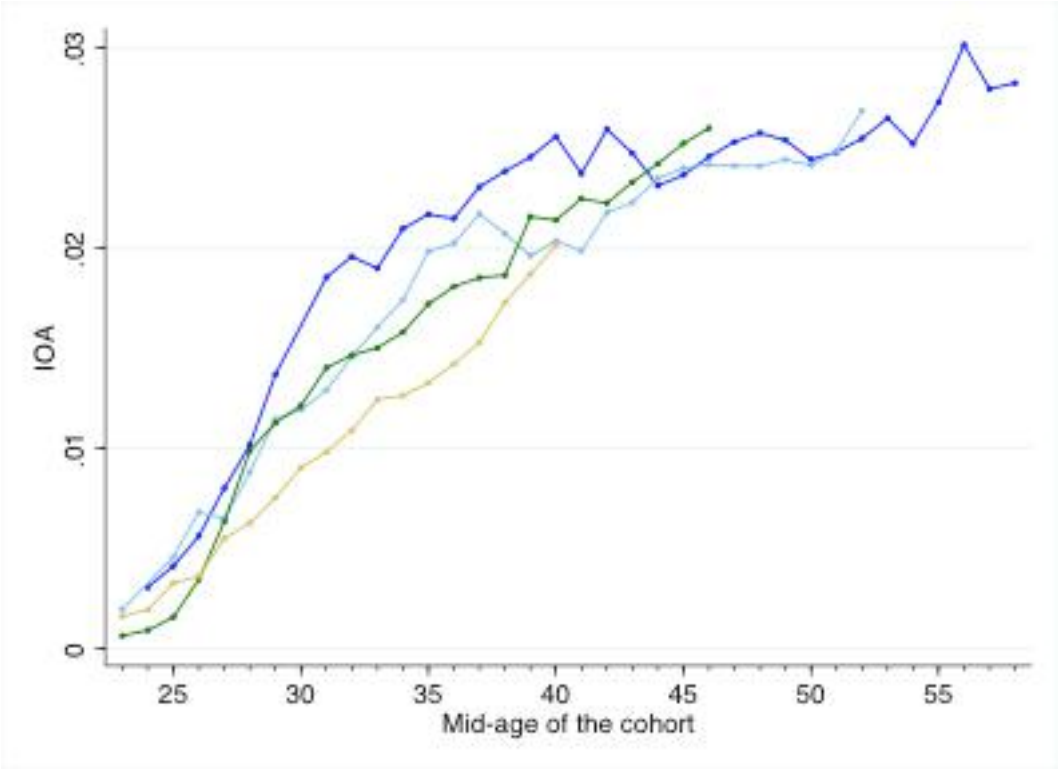
Overall, the decrease across cohorts is less important, or slower, than its increase with age. *IOR* decreases *at most* (at 31 years old) by only 3.6 percentage points of total inequality between the earliest and latest cohorts, which are separated by 18 years of birth. In comparison, *IOR* gains for instance *at least* (in the latest C78 cohort) 8.7 percentage points between 25 and 35 years

Figure 2: Inequality over the life course, by cohort

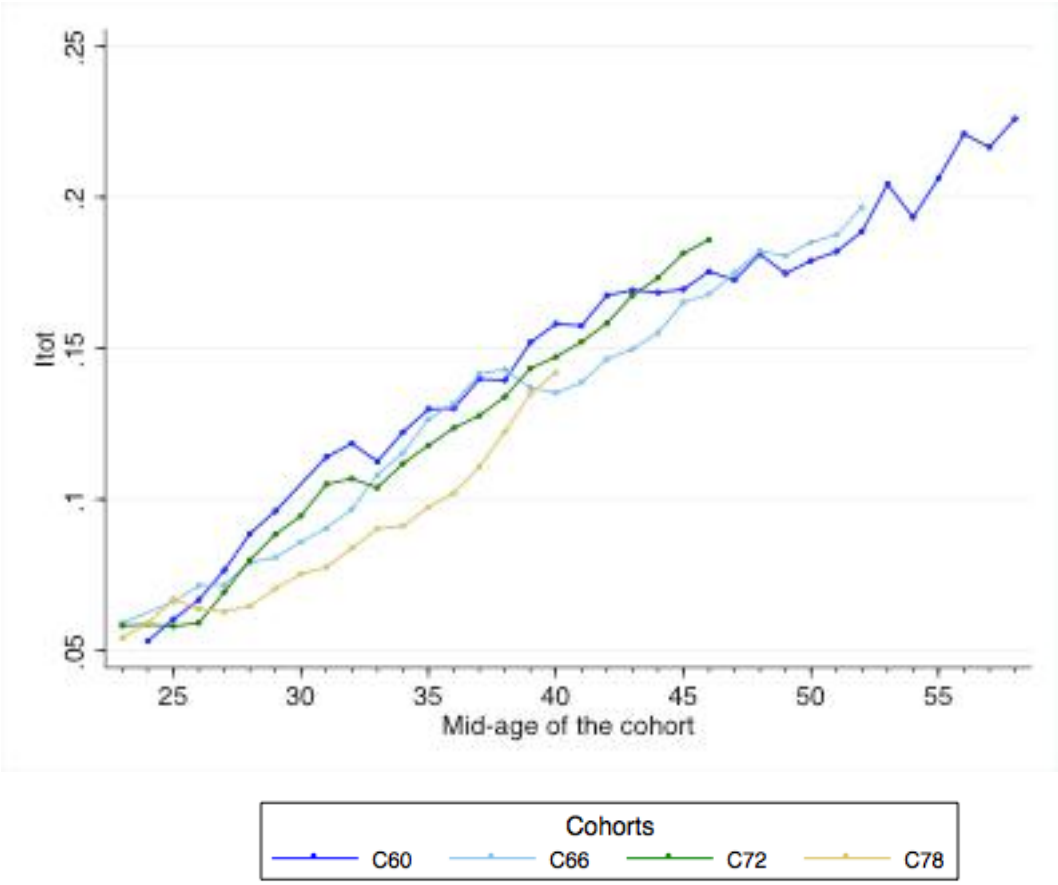
(a) Inequality of opportunity in relative terms (*IOR*)



(b) Inequality of opportunity in absolute terms (*IOA*)



(c) Total earnings inequality (*Itot*)



Note: *IOR* is the ratio of *IOA* to *Itot* and *IOA* is the part of earnings inequality explained by father's education, father's occupation, father's citizenship, age difference with the father, area of residence during youth and sex.

old. Finally, the concavity of the curves slightly decreases across cohorts: *IOR* starts to decelerate at a slightly younger age (around 28 years old) and its maximum occurs at a slightly older age (around 35) in the latest cohorts. This indicates that the evolution of *IOR* with age is becoming more gradual across cohorts.

Panels (b) and (c) show respectively the evolution of *IOA* and *Itot*. Two main points are worth noticing. First, both tend to increase linearly with age, almost perfectly so in the case of *Itot*. We put more emphasis on *IOR* throughout the paper for this reason: it brings newer information about IO given prior knowledge about *Itot*. Second, they both evolve on the overall in the same direction as *IOR*, be it in the age or cohort dimensions. This means that the numerator *IOA* varies proportionally faster than the denominator *Itot*. [19] For instance, *IOA* decreases by 39 percent whereas *Itot* decreases by 25 percent between the earliest and the latest cohorts at age

[19] But as we have seen, *IOR* stop varying much in the age dimension after 35 years old and in the cohort dimension after 42. We conclude that *IOA* and *Itot* evolve approximately at the same speed after these ages.

35. The speed difference is even more pronounced in the age dimension: *IOA* increases for instance by 305 percent (a more than fourfold increase) and *Itot* by "only" 45 percent from 25 to 35 years old in the latest C78 cohort.[20]

We perform some robustness checks in Figures A2.1 to A2.7 of the appendix and comment below the main changes we uncover compared to Figure 2. In short, the shapes of the curves and their underlying cross-age and cross-cohort evolutions are robust, particularly those of *IOR*. The values of *IOR* are also quite robust. In more detail, we drop the 99th wage percentile in Figure A2.3 instead of winsorizing it. This reduces inequality between the top and the rest of the distribution, particularly at late ages, which translates into smaller slopes for the *IOA* and *Itot* curves. The reason is that the 99th percentile increases faster than the rest of the distribution over the life course, as shown in Figure A6.1. We include all wage earners in Figure A2.5 instead of excluding "poor" individuals, whose full-year equivalent wage is below the legal minimum, as well as those with less than one hundred paid days during the year. This increases inequality between the bottom and the rest of the distribution, particularly for *Itot* at young ages. Its curves now decrease fast until age 27 before rising again. A comparison with Figure 2A.4, where poor individuals are retained but not those who work less than one hundred days, shows that it is mainly the disparity in the length of work that generates this additional inequality at the bottom. Lastly, we compute inequality with the *GE0* index in Figure 2A.7, which is more sensitive to the bottom of the distribution than the *GE2* one. We obtain similar results, except that *IOR* stabilizes at a higher level after age 30, around 17.5 percent of *Itot* instead of 15.

[20] The corresponding figures for *IOR* are a 18 percent (not percentage points) decrease in the cohort dimension and a 178 percent increase in the age dimension.

4. Discussion

We now discuss how our benchmark results from Figure 2 relate to the literature. Concerning the decrease of inequality across cohorts, we focus specifically on the literature about France (subsection 4.1). Concerning its increase with age, we broaden the perspective to other countries (subsection 4.2) and to theoretical arguments (subsection 4.3), as we believe this result to be less country specific. Lastly, we investigate what our results imply regarding the comparability of the inequality estimates one can find in the literature (subsection 4.4).

4.1.Cross-cohort decrease of French inequalities

The literature about French inequalities is abundant, but it is more limited when it comes to decomposing their evolution across age and cohorts. D'Albis and Badji (2022) is one exception regarding *Itot*. Based on Deaton and Paxson (1994b)'s parametric decomposition of the two time dimensions, they document a stagnation if not a slight increase across cohorts born in the 1960s and 1970s. Their results are about *Itot* in income and are not directly comparable to ours though. Now, Verdugo (2014) documents a fall of *Itot* in earnings since the late sixties, so as Bozio, Breda and Guillot (2023) when earnings are defined as net wages (as is the case here and as opposed to net wages plus payroll taxes). Birth years are being replaced every year in their aggregate measures, computed over the working age population, so their result is coherent with the cross-cohort decrease that we show in panel (c). They attribute this fall, which is not observed in the US and other Western countries, to three main changes in the French economy: the expansion of secondary and tertiary education and the ensuing decline of the skill premium (Verdugo 2014); the increasing progressivity of payroll taxation (Bozio et al. 2023); and the rise of the real minimum wage (both papers).

By the same token, the fall of *IOA* and *IOR* in earnings found by Lefranc et al. (2011) between 1977 and 1993 as well as by del Valle-Inclán (2019) between 2003 and 2016 are in line with the cross-cohort decreases that we show in panels (a) and (b). We note that their values of *IOR* are similar to ours after 30 years old: they vary somewhere around 15 percent. Andreoli and Fusco (2017) also consider *IOA* in earnings and find in contrast that it increases slightly from 2005 to 2011. Contrarily to others though, they consider gross earnings (before payment of social contributions), which explains the discrepancy if we extrapolate from the results of Bozio et al. (2023) about *Itot*. Finally, Bussolo et al. (2023) find as we do that *IOR* decreases across cohorts born between 1960 and 1980, but they consider *IOR* in income rather than earnings.

4.2.Cross-age increase of inequalities: empirics

While numerous studies have examined the evolution of *Itot* in earnings in France, to our knowledge (and our surprise), none had explored its progression along the lifecycle holding the cohort dimension fixed until now. [21] Now, a strong positive relationship between *Itot* in earnings and age has already been well-documented in other countries, especially in the US, albeit with varying shapes and with less regularity regarding the young ages at which the labor participation is not yet stabilized. [22] Panel (c) confirms this pattern in France for the first time.

We are unaware of studies about the evolution of *IOA* over the lifecycle, in France or elsewhere. Now, its increase shown in panel (b) find echoes in the mobility literature, where estimates of intergenerational earnings elasticity have repeatedly been found to increase with age (see Solon 1999 or Black and Devereux 2011 for a review, and Kenedi and Siruge 2023 for a study about France). These estimates correspond to the slope coefficient from a regression akin to eq. (3) but where earnings are in logs and where these is only one explanatory circumstance variable, the log of parental earnings. The inequality explained by such a regression can be seen as a measure of *IOA* close to ours. It is likely to rise with age even faster than the intergenerational elasticity since it equals this elasticity times the inequality in (log) parental earnings, which can be expected to rise as well. Indeed, although the observed values on parental earnings are measured during the childhood of the individual and do not vary during his adulthood, the variable itself does vary owing to sample selection into the labor market. Its variance is likely to be smaller at the early stages of adulthood, when employment is less widespread.

Turning to *IOR*, Bussolo et al. (2023) estimate its age profile in France, Germany, Great-Britain and Italy. Their profile is concave, as ours in panel (a), and it peaks around age 40, which is close to our peak age, around 35. However, they document a strong decline of *IOR* with age after the peak. We find instead a plateau at comparable ages or, for our earliest C60 cohort only, a moderate decline. One possible explanation for this discrepancy, besides the use of income instead of wages in their analysis, is that the age profile they estimate out of Deaton and Paxson (1994b)'s decomposition represents that of an average cohort born between 1900 and 1989. This average cohort is centered around an earlier birth year than our cohorts, and it could be the

[21] But see d'Albis and Badji (2022), who find that *Itot* in income increases with age until at least 55-59 years old.
[22] About the US, see Deaton and Paxson (1994a), Heathcote, Storesletten and Violante (2005) or Huggett, Venture and Yaron (2006, 2011). Deaton and Paxson (1994a) consider Great Britain and Taiwan as well. See also Creedy and Hart (1979) for an early study about Great Britain, and Bönke, Corneo, and Lüthen (2015) about Germany. The latter find a negative relationship until around 35 years old (and then a positive one) but this is largely because they attribute an earnings value of zero to young adults who have not yet entered the labor market.

case that the age profile of *IOR* after 40 years old has turned across generations from declining to plateauing. Another possibility is that they treat age as a circumstance variable, whose intra-cohort effect is declining over the life course, as we further discuss in the subsection 4.4. Figure A3.3 in the appendix lends support to this second explanation (which does not contradict the first), where we compare *IOR* estimated over various age ranges when the year of birth is treated as an additional circumstance.

Our results about *IOR* can also be related to Nybom and Stuhler (2017), who study in Sweden how the correlation between the (log) incomes of fathers and sons evolves with the age of sons. This intergenerational correlation can be interpreted as a measure of relative inequality since it is the square root of the R-squared from a typical mobility regression. Its age profile has a concave increasing shape that is very similar to that of *IOR* in panel (a). Interestingly, Nybom and Stuhler (2017) also explore the slope from this regression, which, as we have already argued, is closer in spirit to a measure of absolute inequality. It has a rather linear age profile, akin to *IOA* in panel (b).

4.3.Cross-age increase of inequalities: theory

The accumulation of human capital is one leading theoretical explanation for the increasing age profile of *Itot* in panel (c). In Huggett et al. (2006) for instance, individuals with high learning abilities forego their high earnings potential early in life to increase instead their human capital, thanks to which they can enjoy from even higher earnings later. As suggested by Deaton and Paxson (1994a), screening models can also offer a convincing explanation: when abilities are only slowly revealed to firms, they are induced to compress wages among their younger employees. The same mechanisms easily extend to explain the rise of *IOA* with age in panel (b). It suffices that abilities be related to circumstances, whether through genes, child rearing, or more broadly through the social environment. Alternatively, it suffices that the effect of abilities on earnings, rather than abilities themselves, depends on circumstances. [23]

The above arguments explain why *IOA* and *Itot* increase with age but not why *IOA* increases faster than *Itot* in Figure 2, at least until 35 years old. In other words, they do not explain why *IOR* increases until this age. An econometric model à la Haider and Solon (2006), originally proposed in the context of mobility, can help us capture the missing part of the story. Regressing

[23] The two alternatives may coincide if one sees abilities as being acquired during childhood.

current earnings w_{it} instead of lifetime earnings w_i (in logs) on circumstances $\boldsymbol{X_i}$ yields

$$\ln(w_{it}) = (\lambda_{0ct} + \lambda_{ct} b_{0c}) + \lambda_{ct} \boldsymbol{X_i b_c} + (\lambda_{ct} \varepsilon_i + v_{it}) \quad (4)$$

instead of

$$\ln(w_i) = b_{0c} + \boldsymbol{X_i b_c} + \varepsilon_i, \quad (5)$$

where λ_{0ct}, λ_{ct} and v_{it} come from the linear projection of the logs of w_{it} on w_i:

$$\ln(w_{it}) = \lambda_{0ct} + \lambda_{ct} \ln(w_i) + v_{it}. \quad (6)$$

Here, a measure of *IOA* based on the variance of logs is

$$V_{ct}\left(\{\widehat{\ln(w_{it})}\}\right) = \lambda_{ct}^2 V_c(\{\boldsymbol{X_i b_c}\}), \quad (7)$$

with $V_c(\{\boldsymbol{X_i b_c}\})$ its long-run component, based on w_i. The corresponding measure of *IOR* is

$$\frac{V_{ct}\left(\{\widehat{\ln(w_{it})}\}\right)}{V_{ct}(\{\ln(w_{it})\})} = \frac{\lambda_{ct}^2 V_c(\{\boldsymbol{X_i b_c}\})}{\lambda_{ct}^2 V_c(\{\boldsymbol{X_i b_c}\}) + \lambda_{ct}^2 V_c(\{\varepsilon_i\}) + V_{ct}(\{v_{it}\})}, \quad (8)$$

where *Itot* in the denominator has been obtained using the fact that the residuals from eq. (4)-(6) are uncorrelated with their corresponding explanatory variables, by definition of OLS. We are referring to estimated sample quantities out of simplicity here but we could instead replace the left hand-sides of eq. (7)-(8) by their probability limits and refer to population quantities in the right-hand sides. [24] Also, we assume that the sample is constant over the life course of the cohorts to simplify the discussion below. This explains why we do not index the variances $V_c(\{\boldsymbol{X_i b_c}\})$ and $V_c(\{\varepsilon_i\})$ by *t*.

The coefficient λ_{ct} can safely be assumed to rise from less to more than one over the life course (i.e., with *t*). This rise of λ_{ct} reflects in eq. (6) the tendency for individuals with high lifetime earnings to also experience high earnings growth rates. In Haider and Solon (2006), this results from an extremely simplified reduced form equation where $\ln(w_{it})$ grows linearly as a function of *t*, but this would be implied by a more structural model relating abilities to human capital accumulation or to the screening behavior of firms, along the lines given above. Hence, *IOA* in eq. (7) rises over time with λ_{ct} provided that λ_{ct} is positive, which is typically the case (it can be negative at the very beginning of adulthood, owing to negative autocorrelation in earnings,

[24] This would require zero covariances of the explanatory variables with what would then be population errors in eq. (4)-(6). We see these zero covariances as true by definition since we see eq. (4)-(5), so as eq.(6), as linear projections that are not aimed at estimating causal or partial effects.

but this is not the rule empirically).

Regarding *IOR* in eq. (8), its evolution is governed as well by the variance of the transitory component of earnings $V_{ct}(\{v_{it}\})$, the missing part of our story so far, which reflects their volatility and naturally enters the residual part of inequality (*Itot-IOA)*. This variance (or close proxies of it) has been found to follow a U-shaped evolution over the life course and to attain a minimum around 40 years old (see Baker and Solon 2003 about Canada, Mazumder 2005 about the US, and Blundell, Graber and Mogstad 2015 about Norway). This U-shaped pattern aligns with the intuition that individuals are most entrenched in the labor market at the midlife of their career. Blundell (2014) puts forward that individuals are still sorting themselves into firms and occupations early in their careers and are more prone to health shocks later, resulting in higher earnings volatility at both ends. Combined to the increase of λ_{ct}, the U-shaped evolution of $V_{ct}(\{v_{it}\})$ fully accounts for the rise of *IOR* that we observe until almost 40 years old. It is also coherent with the plateau that we observe afterwards. [25]

We now extend the discussion to the *GE2* index, which is the index we actually use and is Lorenz-consistent, unlike the variance of logs. To this end, we consider what happens to eq. (7)-(8) if we replace $\ln(w_{it})$ with y_{it} in eq. (4) and (6). [26] As a reminder, y_{it} is the rescaled version of the original earnings w_{it}: the average of y_{it} has been set to unity within each year-year of birth in eq. (1).

First consider what happens if we replace $\ln(w_{it})$ with $\ln(y_{it})$. This would change practically nothing to the variances of log in eq. (7)-(8) because they are scale invariant. In fact, this would change absolutely nothing if intra-cohort age effects in eq. (1) can be neglected, such that w_{it} and y_{it} differ only by a multiplicative constant within each year-cohort.

Now consider what happens if we replace $\ln(y_{it})$ with y_{it}. The explained variance and its share in the left-hand sides of eq. (7)-(8) would now be $V_{ct}(\{\hat{y}_{it}\})$ and $V_{ct}(\{\hat{y}_{it}\})/V_{ct}(\{y_{it}\})$ and they would correspond to $GE2_{ct}(\{\hat{y}_{it}\})$ and $GE2_{ct}(\{\hat{y}_{it}\})/GE2_{ct}(\{y_{it}\})$, as we have already explained. Regarding the right-hand sides, their formulas would still apply but the coefficients

[25] We do not seek to fully account for the plateau as this result, which depends on the relative sizes and speeds of $\lambda_{ct}^2 V_c(\{\boldsymbol{X_i b_c}\}) + \lambda_{ct}^2 V_c(\{\varepsilon_i\})$ and $V_{ct}(\{v_{it}\})$, is too specific. The concave profile of *IOR* also depends on those relative sizes and speeds but it requires some benign assumptions on that matter. The same applies to the increasing profile of *Itot*, which, no matter how structural one's model is, must in the end depend on some "ultimate" residual term.

[26] An alternative approach would be to keep eq. (4) and (6) as they are and to study how the *GE2* index evolves when applied over the exponentiated predicted values from eq. (4). We do not pursue this approach because the ensuing measures of IO do not satisfy the between-type transfer principle (see footnote 17) and are not the ones we actually use.

and residuals inside would change. In particular, we would now have a perfect analogy between eq. (4) and our actual regression (3), with $\lambda_{ct}\boldsymbol{b_c} = \boldsymbol{\beta_{ct}}$ for instance. The big question, then, is whether the old and new versions of each of λ_{ct} and $V_{ct}(\{v_{it}\})$, defined by eq. (6), would *evolve* similarly over time. If so, the variance of logs and the *GE2* index would also evolve similarly. A "no" to this question is in general unlikely but not impossible. In the present case, it is most unlikely because earnings are concentrated near the average (itself close the minimum, see subsection 2.6). Since the average of y_{it} is one and since $\ln(y_{it}) \approx (y_{it} - 1)$ for $y_{it} \approx 1$, the log-log specification with $\ln(y_{it})$ in the left-hand side of eq. (6) can be seen as approximation for the level-log specification with $(y_{it} - 1)$ instead. This, in turn, would give the same slope λ_{ct} and residual variance $V_{ct}(\{v_{it}\})$ as the level-log specification with y_{it} instead.

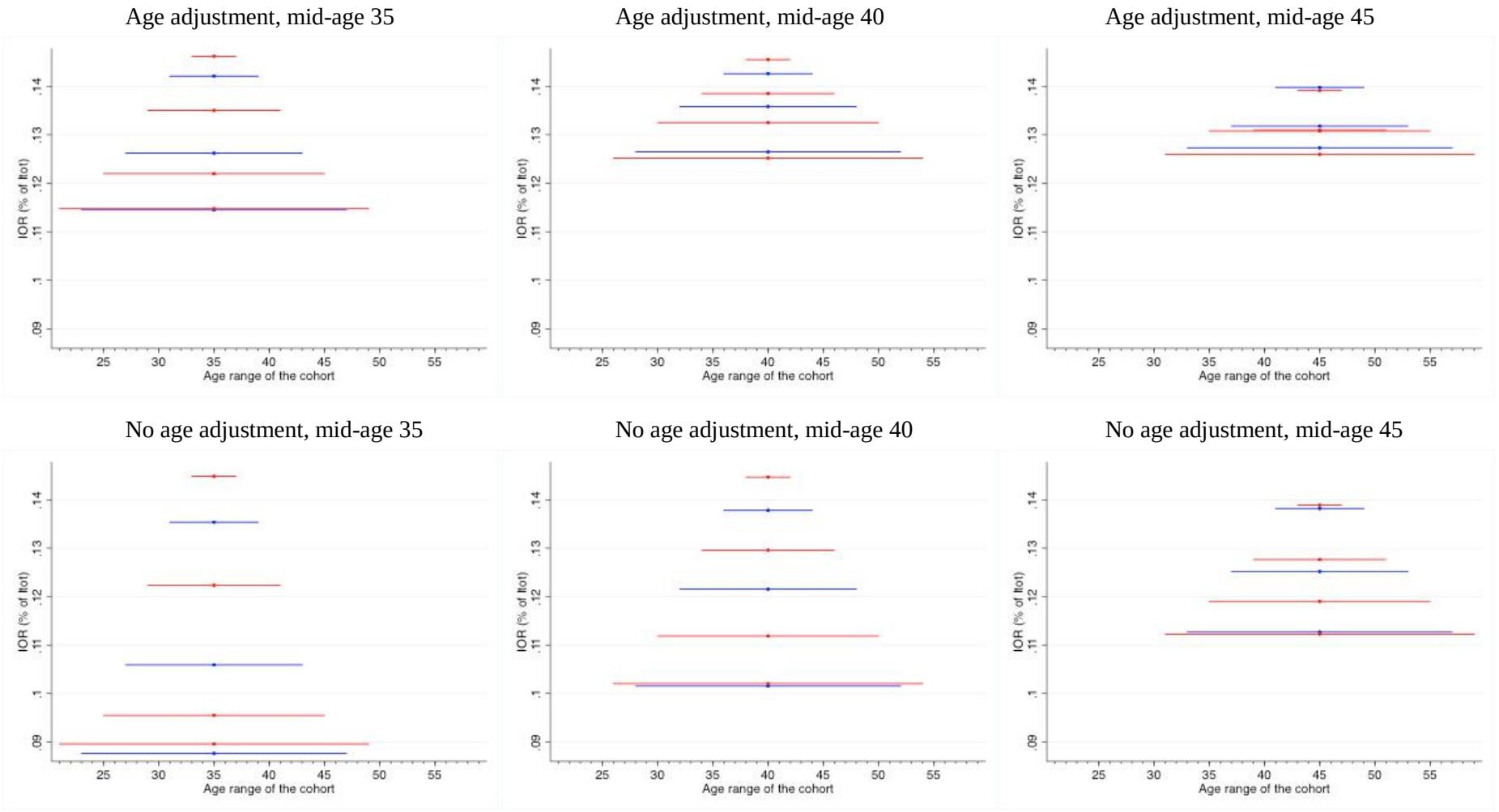
4.4.Implications for the comparability of inequality estimates in the literature

The steep age profiles we observe in Figure 2 suggest that inequality may substantially differ depending on the age range over which it is computed. Practices in this regard are not entirely homogeneous in the IO literature. Typically, aggregate estimates are computed over age ranges starting between 20 and 30 years old and ending between 50 and 60 (see Table A7.1). In this subsection, we explore the comparability of such aggregate estimates as well as their comparability with narrower cohort-level estimates.

To this end, we compare several cohorts of varying lengths but centered around the same birth year at three selected mid-ages typically found in the literature: 35, 40 and 45. We define 1972 as our central birth year because it is the one around which we can gather the largest number of birth years at those mid-ages. Hence, we compute inequality in each cohort in the years 2007, 2012 and 2017 respectively for the three mid-ages. Our narrowest cohort, already labeled as C72 in the previous sections, gathers the birth years 1970, 1972 and 1974. We then define six larger cohorts by gradually adding an even birth year to each side of C72, up until the largest one, which gathers the even birth years from 1958 to 1986.

Figure 3 shows the results from this comparative exercise for *IOR*. Let us begin with the top panels, where *IOR* is computed as in the rest of the paper. Two main sources of variation emerge: *IOR* decreases with the length of the age range, particularly at mid-age 35; and it is substantially lower at mid-age 35 than at 40 or 45 in the five largest cohorts but not in the two smallest ones. Given the concave increasing profile of the C72 cohort in Figure 2 (a), , with a plateau reached around age 35, the rationale in both cases is that the share in the estimation

Figure 3: Comparability of *IOR* over selected age ranges



Note: The figure shows how *IOR* varies with the mid-age of cohorts (from left to right across panels), represented by a point, and their age ranges (within panels), represented by a horizontal line. All cohorts are centered around the birth year 1972. *IOR* is computed in 2007 in the left panels (mid-age 35), in 2012 in the central panels (mid-age 40), and in 2017 in the right panels (mid-age 45). The red and blue colors are just aimed at improving readability and have no particular meaning. In the top panels, average earnings has been rescaled to one in each birth year before computing *IOR*, as in the rest of the paper; in the bottom panels, they have not.

sample of individuals younger than this age, particularly those in their twenties and early twenties, drives down *IOR*. Our recommendation, then, is to start computing *IOR* at an age where earnings are stabilized, unless younger individuals are central to the analysis at hand. As a rule of thumb, this means at the very least excluding people in their twenties.

Similar recommendations are commonly found in the mobility literature, where those types of age-related comparability issues have received more attention. This literature sees mobility estimates based on lifetime earnings as the benchmark and those based on current earnings as biased, particularly at young ages. We do not seek to compare our *IOR* estimates to a benchmark here.[27] Our point is simply that they are sensitive to the age range covered, enough to make many comparisons across studies doubtful. One should beware that apparently small differences in age ranges can otherwise render *IOR* comparisons misleading: in the top left panel, *IOR* at mid-age 35 is, for instance, around 10 percent higher for people aged 27-43 compared to 23-47 (12.6 vs. 11.5% of *Itot*) or for those aged 29-41 compared to 25-45 (13.5 vs. 12.2% of *Itot*). It is even higher for our narrowest cohort (14.6% of *Itot* at age 33-37).

Our second recommendation to enhance comparability is to apply the age adjustment used throughout the paper, whereby earnings averages have been equalized across yearly ages (and year of earnings data) prior to computing inequality. In the bottom panels, we do not apply this adjustment. The discrepancies across age ranges are now more pronounced: the 10 percent difference mentioned earlier regarding the age ranges 27-43 versus 23-47 is now around 20 percent (10.6 vs. 8.8% of *Itot*), and it is around 28 percent for the age ranges 29-41 versus 25-45 (12.2 vs. 9.5% of *Itot*). The reason is that within cohort inequality due to age (i.e., between yearly ages) has been removed in the top but not in the bottom panels. In the bottom panels, it inflates the denominator of *IOR* (*Itot*) much more than its numerator (*IOA*) since the age at which we observe individuals during adulthood has little to do with their childhood circumstances.[28] This reinforces the tendency of *IOR* to be low when people in their twenties are included: in the top panels, *IOR* within each yearly ages (best proxied by the C72 cohort) is low at young adult ages, which drives down the cohort-level estimates; in the bottom panels,

[27] We note, however, that even our highest narrow cohort-level estimates underestimate *IOR* based on lifetime earnings: $V_c(\{\boldsymbol{X_i b_c}\})/[V_c(\{\boldsymbol{X_i b_c}\}) + V_c(\{\varepsilon_i\})]$ is the R-squared from eq. (5), based on lifetime earnings, and it must be higher for all t than the R-squared described by eq. (8), based on current earnings.

[28] Of course, this presupposes that the birth year itself is not treated as a circumstance, as some authors do. We show in Figures A3.3 and A3.4 what happens to *IOR* and *IOA* when it is, although we find this choice debatable since the birth year captures intra-cohort age effects that are encountered by everyone over their complete life course. Note that the age adjustment is not an option in this case, as the birth year would only bring noise to the estimates.

inequality between yearly ages in the denominator further drives them down since wages evolve rapidly at the beginning of adulthood (see, e.g., Figure 1). This also reinforces the tendency of *IOR* to be low when the age range is large: in the top panels, this is because larger age ranges tend to include younger individuals as well; in the bottom ones, inequality between yearly ages tends to mechanically increase with the length of the age range and further drives *IOR* down.

The corresponding figures for *Itot* and *IOA* are shown in the appendix (Figures A3.1 and A3.2). The main differences, compared to Figure 3 about *IOR*, are the following. First, the mid-age now plays a key role for comparability, more important than the length of the age range, and the estimates keep increasing between mid-ages 40 and 45. This is not surprising given the linear profiles of *IOA* and *Itot* in Figure 2. Moreover, this shows that while two *IOR* estimates may be comparable, their underlying *IOA* estimates may not be, and inversely. Second, the age adjustment affects *IOA* little since inequality due to age mainly goes into the residual part of inequality (*Itot-IOA*). Third, *Itot* increases with the length of the age range when this age adjustment is not performed (whereas it otherwise decreases with it), as does inequality due to age, resulting in a downward pyramidal figure. Fourth, the length of the age range however plays a modest role for *Itot*, except at mid-age 35 without the adjustment because inequality due to age is high at the beginning of adulthood (relative to the low levels of *Itot* at those ages).

5. The relative importance of circumstances

5.1.Foreword on the Shapley decomposition of inequality of opportunity

We now assess the relative importance of circumstances by performing a Shapley decomposition of the R-squared from model (3), which we recall is numerically equivalent to our measure of *IOR* based on the *GE2* index. The importance of circumstance x_l **(**in the year-cohort ct) is given by its Shapley value, Sh_{lct}:

$$Sh_{lct} = \sum_{K} \frac{|K|!\,(J-1-|K|)!}{J!}.\left(R^2_{ct}(K_l) - R^2_{ct}(K)\right), \tag{9}$$

where J is the number of circumstances in the "full" model (3), which is six here, K is a subset of them that excludes x_l, $K_l = (K \cup x_l)$ is the corresponding subset that includes x_l, and $R^2_{ct}(.)$ is the R-squared associated with a set. [29] The term $\left(R^2_{ct}(K_l) - R^2_{ct}(K)\right)$ represents the marginal contribution of x_l to $R^2_{ct}(K_l)$. Sh_{lct} is a weighted average of the marginal contributions of x_l across all possible models that can be built with it and a subset of the other circumstances present in the full model (3). [30]

Shapley is the only decomposition rule that satisfies the three basic properties of "efficiency", "symmetry" and "strong monotonicity" (Young 1985). [31] These essential properties hold true if we replace $R^2_{ct}(.)$ with any real-valued function $f(.)$ defined over all sets of $S = K_l, K$ and such that $f(\emptyset) = 0$, which encompasses virtually any inequality index. They hold true, in particular, whether or not the coefficients implicit in $R^2_{ct}(.)$ are updated across sub-models. We would like to use this subsection to provide a methodological clarification on this – overlooked – matter. We argue that the coefficients should be updated. Accordingly, and writing the full model (3) in non-vector notation as

[29] Some circumstances have strictly more than two categories and x_l is in this case a vector of several dummy variables, all of which are included in K_l and none of which in K. Strictly speaking, this corresponds to an Owen (1977) decomposition, a generalization of Shapley (1953)'s.

[30] The summation is done over all the combinations K of $|K|$ circumstances among J ones, with $J - 1 \geq |K| \geq 0$. There are 2^{J-1} such combinations. The term $|K|!\,(J-1-|K|)!$ acts as a weight for each K. It corresponds to the number of permutations of $(J-1)$ circumstances such that a sequential elimination of the first element eventually yields the set K. Equivalently, one could suppress this weight (i.e., replace it by 1) in eq. (9) and perform the summation over $J!$ permutations instead of 2^{J-1} combinations. This would be more computationally expensive though, as $J!$ is in general much greater than 2^{J-1}.

[31] Efficiency implies that the Shapley values add up to what is being decomposed, *IOR* here. Symmetry implies that two circumstances must have the same Shapley values if they can be substituted without altering the R-squared of any of the sub-models that can be built from (3). Strong monotonicity implies that a Shapley value must increase if the marginal contribution of its underlying circumstance increases in all sub-models.

$$y_{it} = \beta_{0ct} + \sum_{j=1}^{J} \beta_{jct} x_{ji} + u_{it}, \qquad (3bis)$$

we define the function $R^2_{ct}(.)$ as

$$R^2_{ct}(S) \equiv R^2_{ct}\left(S; \boldsymbol{\beta^S_{ct}}\right) = \frac{V_{ct}\left(\left\{\sum_{j \in S} \beta^S_{jct} x_{ji}\right\}\right)}{V_{ct}(\{y_{it}\})} \; \forall S = K_l, K \qquad (10)$$

and not as

$$R^2_{ct}(S) \equiv R^2_{ct}(S; \boldsymbol{\beta_{ct}}) = \frac{V_{ct}\left(\left\{\sum_{j \in S} \beta_{jct} x_{ji}\right\}\right)}{V_{ct}(\{y_{it}\})} \; \forall S = K_l, K. \qquad (11)$$

Above, x_{ji} denotes the value taken by circumstance x_j (which can potentially correspond to x_l, whose Shapley value is being calculated) for individual i, "$j \in S$" is a shorthand notation for "j: $x_j \in S$", and the superscript S serves to stress that the coefficients in eq. (10) are updated, as opposed to those in eq. (11), which come from the full model (3 or 3bis) only. We treat as β_{jct} and β^S_{jct} as scalars for ease of exposition, although they are in fact vectors when x_{ji} has more than two categories.

There are three main reasons for updating the coefficients across sub-models. First, failing to update them amounts to imposing an unrealistic restriction on the data, namely that circumstances are uncorrelated (in which case $\beta^S_{jct} = \beta_{jct} \; \forall j, S$). When the full model is structural, which is not the case here, one may be tempted to stick with its coefficients anyway, on the basis that they are those of substantive interest. This choice would be debatable though. Suppose for instance that the partial effect of x_{li} is zero, such that $\beta_{lct} = 0$ in the full model. Then, we trivially have $R^2_{ct}(K_l; \boldsymbol{\beta_{ct}}) = R^2_{ct}(K; \boldsymbol{\beta_{ct}}) \; \forall K$ and the "unupdated" Shapley value based on eq. (11) must be zero. Yet, x_{li} might be a deep cause of y_{it} that affects it through the other regressors. Second, two highly correlated regressors may obtain very different Shapley values when the coefficients are not updated, which is not coherent. This difference must be greater (in absolute value) than when the coefficient are updated. Israeli (2007) proves this result and stands out in the literature about the Shapley decomposition of the R-squared by explicitly making the distinction between $\boldsymbol{\beta_{ct}}$ and $\boldsymbol{\beta^S_{ct}}$. Third, and most importantly in the context of IO, negative Shapley values can occur when the coefficients are not updated. This is plainly incoherent with the lower-bound result according to which IO can only increase as additional circumstances are considered, which is itself a direct consequence of the transfer principle.

Yet, most studies implementing a Shapley decomposition of IO either stick to the coefficients of the full model (3) without explicitly stating it, or provide too little information to determine whether they have updated them. [32] They have largely been influenced in this respect by the literature that decomposes inequality by income sources, which has first imported the Shapley method from the field of game theory into that of inequality. [33] This literature writes income as the sum of various sources (e.g., labor plus capital incomes). It is not regression-based, which explains the relative silence of subsequent IO studies on whether the coefficients should be updated. It has also induced them to treat the summands $\beta_{lct}x_{li}$ of the full model as sources, an analogy that implicitly leaves the coefficients unchanged across sub-models since the value an of an income source does not depend on which other sources are considered next to it.

The influence of the "source approach" is most visible when considering that eq. (9) combined with the "unupdated" eq. (11) reduces to the often used Shapley-Shorrocks decomposition:

$$Sh_{lct} = \frac{cov_{ct}(\{\beta_{lct}x_{li}\}, \{y_{it}\})}{V_{ct}(\{y_{it}\})}. \tag{12}$$

The right-hand side is an application to the explained part of the full model (3) of Shorrocks' (1982) "natural" decomposition of the variance by sources, which has later been proven to be equivalent to a Shapley decomposition. [34] Eq. (12) has been seen as the basis for a "reconciliation" (Cowell and Fiorio 2011) between the decompositions of inequality of by

[32] A review of Shapley decompositions of IO, which is not exhaustive but still comprehensive enough to offer a reliable overview of current practices, reveals that the following IO studies do not update the coefficients, based on clues whose rationale is given in the main text: (i) Marrero and Rodriguez (2011), Ferreira and Gignoux (2011a), Ferreira, Gignoux and Aran (2011), as they rely on the Shapley-Shorrocks decomposition; (ii) Deutsch, Pi Alperin, and Silber (2018), but with less certainty, as they appear to rely on a variant of this decomposition in a logistic setting; iii) Björklund, Jäntti et Roemer (2012) and Hederos Jäntti et Lindahl (2017), as they obtain negative Shapley values. In contrast, Cabrera et al. (2021) use a Stata module by Juárez and Soloaga (2014), which we have tested and which does update the coefficients. We however advise against the use of this module, as we have discovered that it treats categorical circumstances as continuous and that it, in fact, decomposes the variance of logs. Wu et al. (2021) are in between: they do not update the coefficients for their parametric measure of IO, as they explicitly treat the summands $\beta_{lt}x_{li}$ as sources, but they implicitly update the coefficients of the fully saturated model underlying their non-parametric measure. We were unable to determine whether the following studies update the coefficients: Suárez-Álvarez & López-Menéndez (2018, 2021), Golley et al. (2019), Shi (2019), Yang, Gustafsson and Sicular (2021), Salas-Rojo and Rodríguez (2022), and Kovacic and Orso (2022).

[33] Shapley has been the main tool of decomposition by income sources since Shorrocks (1982), as we discuss in the main text below. The IO literature, which began more recently with Roemer (1993), uses in contrast the inequality indices themselves (the generalized entropy ones in particular) as the main tool to decompose inequality by population subgroups. Shapley only serves as an auxiliary tool in this case, to further decompose the between group term. Two other, less prominent, applications of Shapley exist in the context of subgroup decomposition. It can serve to assess the contribution of a particular group (for instance men, as opposed to the difference between men and women), as in Chantreuil and Trannoy (2011; sect. V), or to bypass the problem of path-dependence raised by Foster and Shneyerov (2000), as in Shorrocks (2013, sect. 5).

[34] A proof of the equivalence can be found in Chantreuil and Trannoy (2011, 2013) or Shorrocks (2013). Chantreuil and Trannoy (2011, 2013) attribute this important result to an unpublished manuscript by Auvray and Trannoy (1992), which we were unable to find.

population subgroups (to which IO belongs) and income sources. In our view, the reconciliation can only be apparent though, as the issue of negative Shapley values shows. While these are unacceptable when the decomposition is truly by subgroups, as we have already argued, they can make perfect sense when it is truly by sources. This is for instance the case when the variance of a source is larger than that of overall income, which can happen if its sources are sufficiently negatively correlated.

5.2.Results

Figure 4 shows the evolution of our Shapley values. Each cohort is depicted in a panel where the upper contour of the overall shaded area corresponds to the *IOR* curve of this cohort in our main Figure 2. Distinct colors represent the contributions of our six circumstances, ordered increasingly from bottom to top according to their average Shapley values across all year-cohorts. In the appendix, we show the Shapley values in the form of curves rather than shaded areas (Figure A4.1) as well as in percentage of IO rather than *Itot* (Figures A4.2 and A4.3, which can equivalently be read in percentage of *IOR* or *IOA* since the denominator *Itot* in eq. (10) is the same whatever the circumstance considered).

We start commenting the cohort-averages of our Shapley values between mid-ages 31 and 40, when the Shapley values are fairly stabilized and when all cohorts are observed every year. Circumstances can be ranked in three groups of relative importance. Father's education is by far the most important driver of IO. Depending on the cohort, it explains from 5.2 to 6.8 percent of *Itot*, which translates to 38.5 to 47.7 percent of IO. Father's occupation, sex, and the area of residence during youth come next. They explain the bulk of the remaining IO. Each accounts for around 20 percent of IO in the two earliest cohorts (C60 and C66). The importance of father's occupation increases across cohorts and attains 27.8 percent of IO in the last C78 one, in which sex and the area of residence during youth each explains around 15 percent of IO. The age difference with the father and his citizenship play a marginal role, with the former explaining from 1.0 to 2.9 percent of IO and the latter having a virtually null contribution, below 0.6 percent of IO.

This ranking of circumstances into three groups of relative importance in Figure 4 confirms their eyeball comparison by average wages in Figure 1. However, Figure 1 as compared to Figure 4 understates the relative importance of the circumstance sex. The reason is that sex has the specificity of being distributed almost independently from the other circumstances and that

Figure 1 displays unconditional wage averages. An unconditional analysis overstates the importance of these other circumstances, which share part of their explanatory power in common. It double counts the earnings variation co-explained by these circumstances by attributing it entirely to each of them. This is the very reason why the R-squared from the simple linear regressions with one circumstance only (shown in Figure A4.4) add up across circumstances to more than the R-squared of the full model (3). [35] This is also the reason why unconditional wage differences by sex are sensibly smaller than those by area of residence during youth in Figure 1 whereas the two variables play a comparable role according to their Shapley values in Figure 4. [36]

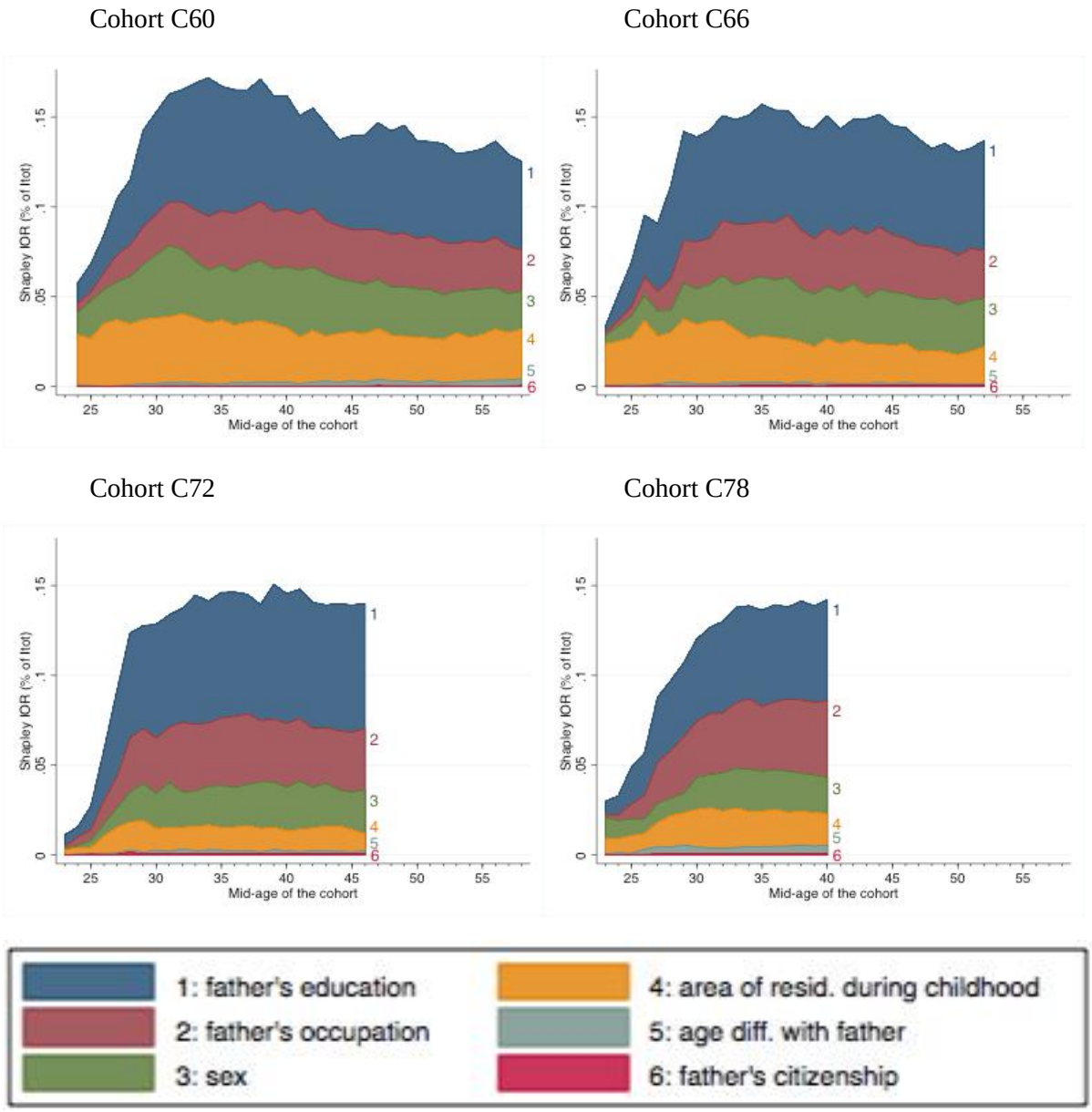
Based on this reasoning, Figure 1 should overstate the importance of father's education and occupation in particular, since these are two strongly related parental characteristics. Yet, it even more strongly overstates the importance of the following circumstances: area of residence during youth, age difference with the father, and citizenship of the father. The first lags behind father's occupation in Figure 4 but generates at least as much wage differences as father's occupation in Figure 1. The latter two play a marginal role in Figure 4 but generate wage differences that are significant in many year-cohorts in Figure 1 (both practically and statistically, see Figure A1.2 with the confidence intervals). The reason is that most of the inequality explained by each of these three circumstances comes from a single sparsely populated category: as is visible in Figure 1, their other categories are close to the sample average. These sparsely populated small categories weighs little when computing the R-squared or its Shapley values from Figure 4. They consist of people who grew up in a large urban area, who represent around 15 percent of the sample (on average across year-cohorts); of people with a relatively young father, aged 23 or less at their birth, who also represent around 15 percent of the sample; and of people with a foreign father, who represent around 9 percent of it.

We now turn to the evolution of our Shapley values. Across cohorts, most progress has been made with respect to sex and the place of the residence during childhood. On average between

[35] The marginal contributions to the R-squared of the full model ($R^2_{ct}(K_l) - R^2_{ct}(K)$ with $R^2_{ct}(K_l)$ the "full" R-squared), shown in Figure A4.5, suffer from polar opposite weaknesses: they add up to less than it since they discard all the earnings variation co-explained by several circumstances. The whole point of the Shapley decomposition is to distribute all this co-explained variation across circumstances, without discarding nor double counting it.

[36] This can be deduced by comparing Figures 1 and A1.2 about unconditional wage differences with Figure A1.3 about conditional ones. In the unconditional case, 15 percent of the sample average separate men from women on average across year-cohorts between mid-ages 31 and 40, while 26 percent separate people who grew up in Île-de-France from those who grew up in a rural area. In the conditional case, these differences are respectively of 18 and 20 percent.

Figure 4 : Shapley decomposition of *IOR* by circumstances



Note: The figure shows how the Shapley contributions to *IOR* of our six circumstances evolve. Each panel decomposes the corresponding cohort-level curve about *IOR* from our main Figure 2.

mid-ages 31 and 40, their combined effect has fallen from 6.7 percent of *Itot* in the initial C60 cohort to 4.1 percent of *Itot* in the most recent C78 one, which accounts for almost all of the corresponding cross-cohort fall in *IOR*, from to 16.6 to 13.7 percent of *Itot*. The corresponding contribution of father's education decreases as well, from 6.7 to 5.3 percent of *Itot*, but this is offset by father's occupation. As a result, their combined contribution remains relatively stable in percentage of *Itot* (a fall from 9.7 to 9.1), although it increases in percentage of IO (from 58.3 to 66.3). Regarding the evolution over the life course, circumstances start at low Shapley values

and then mostly describe a concave increasing profile (see Figure A4.1), as overall *IOR* does.

Their ranking is rather stable over the lifecycle, besides for the area of residence during youth. The latter often ranks first or second before 25 years old whereas it mostly ranks fourth later in life, which is not surprising given that young adults are likely to live in or nearby the parental home. We note however that the prominent role of this circumstance early in life has decreased across cohorts: at age 25, it accounts for around 38.3 percent of IO in the C60 cohort but only for 19.6 percent of it in the C78 cohort. One explanation that comes to mind is the massification of higher education in France, which entails moving to large cities for many young adults (and staying there once graduated).

6. The mediating role of education

6.1.Foreword on the decomposition of the variances explained by education and circumstances

We now consider the role played by (own) education in shaping inequalities. To this end, we propose a framework that simultaneously decomposes the explained variances of the following three models, by education E_i and/or circumstances $\boldsymbol{X_i}$:

$$y_{it} = \beta_{0X} + \boldsymbol{X_i}\boldsymbol{\beta_X} + u_{it}, \qquad (3ter)$$

$$y_{it} = \beta_{0E} + \beta_E E_i + r_{it}, \qquad (13)$$

$$y_{it} = \alpha_0 + \boldsymbol{X_i}\boldsymbol{\alpha_X} + \alpha_E E_i + e_{it}. \qquad (14)$$

Eq. (3ter) corresponds to our model (3) rewritten with a new notation convention: in this section, we index the coefficients according to the variables they are attached to and, to lighten notations, we keep the *ct* subscript implicit. We treat education E_i as fixed at the beginning of adulthood, and we represent β_E as a scalar for notational convenience, although E_i will sometimes be measured as a categorical variable with more than two categories.

The explained variance of the "extended" model (14) is trivially:

$$V(\{\boldsymbol{X_i\alpha_X} + \alpha_E E_i\}) = V(\{\boldsymbol{X_i\alpha_X}\}) + V(\{\alpha_E E_i\}) + 2cov(\{\boldsymbol{X_i\alpha_X}\}, \{\alpha_E E_i\}). \quad (15)$$

We could rely on the Shapley method to distribute among E_i and $\boldsymbol{X_i}$ the above covariance term but this is not our objective, as childhood circumstances causally precede education. Instead, we nest the explained variances of the "small" models (3) and (13) inside eq. (15) using "omitted variable" algebra. Before proceeding, note that our discussion about the variance does not necessarily extend to the *GE2* index in this section. In particular, $V(\{\boldsymbol{X_i\alpha_X}\})$ and $V(\{\alpha_E E_i\})$ do not correspond to the *GE2* index, as the means of $\{\boldsymbol{X_i\alpha_X}\}$ and $\{\alpha_E E_i\}$ need not equal one.

Regarding the explained variance of model (3) , plug the linear projection

$$E_i = \gamma_{0X} + \boldsymbol{X_i\gamma_X} + \vartheta_i \quad (16)$$

into eq. (14) and reorganize terms to get $\boldsymbol{\beta_X} = \boldsymbol{\alpha_X} + \alpha_E\boldsymbol{\gamma_X}$. Hence:[37, 38]

$$V(\{\boldsymbol{X_i\beta_X}\}) = V(\{\boldsymbol{X_i}(\boldsymbol{\alpha_X} + \alpha_E\boldsymbol{\gamma_X})\})$$

$$= V(\{\boldsymbol{X_i\alpha_X}\}) + V(\{\alpha_E E_i\})\frac{V(\{\boldsymbol{X_i\gamma_X}\})}{V(\{E_i\})} + 2cov(\{\boldsymbol{X_i\alpha_X}\}, \{\alpha_E E_i\}). \quad (17)$$

Symmetrically for the explained variance of model (13), $\boldsymbol{X_i}$ can be written as

$$\boldsymbol{X_i} = \boldsymbol{\gamma'_{0E}} + E_i\boldsymbol{\gamma'_E} + \boldsymbol{\eta_i} \quad (18)$$

by replacing each element in $\boldsymbol{X_i} = \left(X_{1i} \dots X_{ji} \dots X_{Ji}\right)$ with its linear projection on E_i and by rearranging the ensuing vector of sums as a sum of vectors. Above, $\boldsymbol{\gamma'_{0E}}$ and $\boldsymbol{\gamma'_E}$ are row vectors, which we write as transposes because we view coefficient vectors as columns by default. Plugging eq. (18) into (14) yields $\beta_E = \alpha_E + \boldsymbol{\gamma'_E\alpha_X}$. Hence:

$$V(\{\beta_E E_i\}) = V(\{(\alpha_E + \boldsymbol{\gamma'_E\alpha_X})E_i\})$$

$$= V(\{\alpha_E E_i\}) + V(\{\boldsymbol{X_i\alpha_X}\})\frac{V(\{E_i\boldsymbol{\gamma'_E\alpha_X}\})}{V(\{\boldsymbol{X_i\alpha_X}\})} + 2cov(\{\boldsymbol{X_i\alpha_X}\}, \{\alpha_E E_i\}). \quad (19)$$

[37] To obtain eq. (17), write $V(\{\boldsymbol{X_i}(\boldsymbol{\alpha_X} + \alpha_E\boldsymbol{\gamma_X})\}) = V(\{\boldsymbol{X_i\alpha_X}\}) + V(\{\alpha_E\boldsymbol{X_i\gamma_X}\}) + 2cov(\{\boldsymbol{X_i\alpha_X}\}, \{\alpha_E\boldsymbol{X_i\gamma_X}\})$. Next, write $V(\{\alpha_E\boldsymbol{X_i\gamma_X}\}) = \alpha_E^2 V(\{\boldsymbol{X_i\gamma_X}\})[V(\{E_i\})/V(\{E_i\})] = V(\{\alpha_E E_i\})[V(\{\boldsymbol{X_i\gamma_X}\})/V(\{E_i\})]$. Lastly, use the uncorelatedness of ϑ_i in eq. (16) with each variable in $\boldsymbol{X_i}$ to get $cov(\{\boldsymbol{X_i\alpha_X}\}, \{\alpha_E\boldsymbol{X_i\gamma_X}\}) = cov(\{\boldsymbol{X_i\alpha_X}\}, \{\alpha_E E_i\})$.

[38] Palomino, Marrero and Rodríguez (2019) propose an alternative decomposition to eq. (17). They use the *GE0* index but, otherwise, they essentially regress $\boldsymbol{X_i\beta_X}$ on $\boldsymbol{E_i}$ to split $V(\{\boldsymbol{X_i\beta_X}\})$ into its explained and residual parts.

We obtain our final simultaneous decomposition of inequalities by nesting as follows eq. (17) and (19) into (15):

$$V(\{\boldsymbol{X_i\alpha_X} + \alpha_E E_i\}) = \overbrace{\overbrace{\overbrace{A}^{V(\{\boldsymbol{X_i\alpha_X}\})} + \overbrace{B}^{V(\{\boldsymbol{X_i\beta_X}\})} + C}^{V(\{\boldsymbol{X_i\beta_X}\})} + \underbrace{\underbrace{D + E}_{V(\{\alpha_E E_i\})}}_{V(\{\beta_E E_i\})}, \quad with \tag{20}$$

$$A = V(\{\boldsymbol{X_i\alpha_X}\})\left(1 - \frac{V(\{E_i\boldsymbol{\gamma'_E\alpha_X}\})}{V(\{\boldsymbol{X_i\alpha_X}\})}\right)$$

$$B = V(\{\boldsymbol{X_i\alpha_X}\})\frac{V(\{E_i\boldsymbol{\gamma'_E\alpha_X}\})}{V(\{\boldsymbol{X_i\alpha_X}\})}$$

$$C = 2cov(\{\boldsymbol{X_i\alpha_X}\}, \{\alpha_E E_i\})$$

$$D = V(\{\alpha_E E_i\})\frac{V(\{\boldsymbol{X_i\gamma_X}\})}{V(\{E_i\})}$$

$$E = V(\{\alpha_E E_i\})\left(1 - \frac{V(\{\boldsymbol{X_i\gamma_X}\})}{V(\{E_i\})}\right).$$

Components A through D add up to our measure of *IOA*, $V(\{\boldsymbol{X_i\beta_X}\})$, which captures the total effect of circumstances on the variance of wages. Components A and B add up to $V(\{\boldsymbol{X_i\alpha_X}\})$, which only captures their direct effect, holding education (but not other unobserved factors) fixed. Symmetrically, components B through E capture the total effect of education $V(\{\beta_E E_i\})$, while D and E capture its direct effect $V(\{\alpha_E E_i\})$ holding circumstances fixed. Now, education and circumstances are not symmetrical from a causal point of view. Only component E, in yellow in Figure 5 below, is truly attributable to education, which we call it its "pure" effect. The other components, shown in different shades of blue, are better attributed to circumstances. Component D captures their indirect effect through education. It is the fraction – equal to $V(\{\boldsymbol{X_i\gamma_X}\})/V(\{E_i\})$, the R-squared from eq. (16) – of the direct effect of education $(D + E)$ that was caused by circumstances in the first place. Component C represents a "correlated effect", capturing the influence of unobserved factors that are common causes of both observed

circumstances and education, at the origin of their correlation.[39] These common causes are better seen as unobserved circumstances since, like the observed ones, they cannot be chosen by individuals. Component B is a spurious indirect effect of education through circumstances. It exists mathematically since the models from this subsection do not impose a priori the sense of causation, but we find it empirically to be negligible. Lastly, component A represents the pure effect of circumstances.

6.2.Results

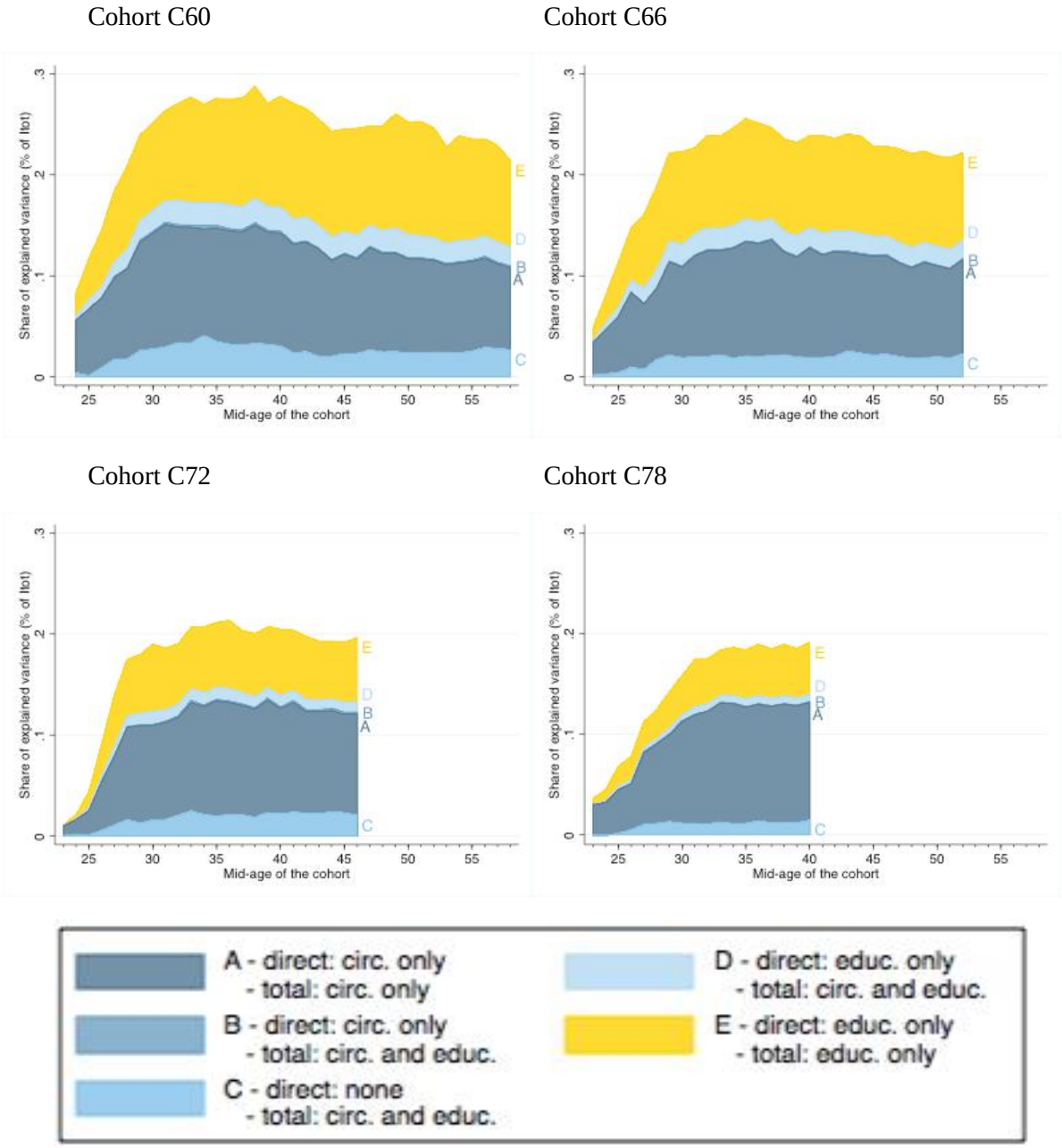
Figure 5 shows the evolution of components A through E in share of *Itot*, with education measured rather coarsely in the last exhaustive census, from 1999. It splits people who attended at least at upper secondary education ("lycée"), our high educational group, from those who did not. We are forced to adopt this coarse partitioning to preserve the comparability of our cohorts, as the youngest members of our sample, born in 1980, had not yet completed their secondary education by 1999. They should normally have completed it in the year they turned 18, in 1998, but a non-negligible share of them had fallen behind (e.g., by repeating classes) and will have completed it after the census.

Four main points are worth noticing. First, it is interesting to compare *IOR* (depicted in various shades of blue) with the pure effect of education (in yellow), given that education is major theoretical wage determinant. *IOR* reaches approximately 15 percent of *Itot* after age 30, which may seem modest at first glance. However, in comparison, the pure effect of education reaches around 10 percent of *Itot* after this age in the two earliest cohorts (C60 and C66) and only around 5 percent in the two latest ones (C72 and C78). We consider breaking our education variable more finely in Figure A5.1, at the cost of dropping the latest C78 cohort.[40] It then explains approximately as much as *IOR*, which does not overturn our point: the effect of circumstances is not small compared to that of education.

[39] But circumstances themselves are also at the origin of their correlation with education since they cause it. Accordingly, part of this correlated effect is spurious and captures in fact the indirect effect of circumstances. Note that common consequences of education and circumstances (e.g., one's occupation, which plays a mediating role between them and wages) are not part of this correlated effect: they do not explain why circumstances are correlated with education in the first place.

[40].We code own education into four categories. The two lowest categories match those of father's education: completed at most primary education ("low"), and secondary education attended but not completed ("medium"). To capture the educational expansion that took place among our cohorts, we split the educational levels corresponding to the "high" category of father's education into two: completed at most secondary education ("medium high"), and completed at least a tertiary degree ("very high"; e.g., bachelor's).

Figure 5: Shares of variance explained by circumstances and education



Note: The figure shows how components *A* trough *E* from eq. (20) evolve in share of *Itot*. The legend specifies whether these components capture the direct or total effect of circumstances or education on the variance of wages. In each panel, the area in various shades of blue decomposes almost exactly the corresponding cohort-level curve about *IOR* from our main Figure 2. The only discrepancy comes from the exclusion in this figure, as opposed to Figure 2, of people whose education is not observed in the 1999 French census. We display component *C* at the bottom rather than at the middle of each panel because it contains a few (almost indistinguishable) negative values.

Second, the pure effect of education evolves similarly to *IOR* in the age (upwards, concavely) and cohort (downwards) dimensions. The cross-cohort decrease reflects the expansion of secondary and tertiary education and the ensuing decline of the skill premium, as already documented by Verdugo (2014). This decrease is much less marked when education is finely

categorized (but is still more rapid than *IOR*'s) though. This is due to the impressive expansion that took place between the C66 and C72 cohorts: around 50 percent of the C60 and C66 cohorts attended at least at upper secondary education, while nearly 70 percent of the C72 cohort did (see Figure A1.4). As a result, educational inequalities largely occur within this high educational group among the C72 cohort, which is overlooked under the coarse categorization.

Third, most of IO is made of the pure effect of circumstances, which operates through channels beyond education (e.g., parental connections): component *A* represents 65.9 percent of IO (i.e., of components *A* through *D*) on average between mid-ages 31 and 40 in the earliest C60 cohort, and this share rises continuously across cohorts to attain in 85.1 percent in the last C78 cohort. Hence, if wages inequalities related to education (i.e., components *B* through *E*) were turned off by setting everyone at the same education level, IO (*IOA*, but also *IOR* if we assume that *Itot* remains unchanged) would fall by only 14.9 percent in the C78 cohort. This indicates that the educational expansion that took place in France has contributed to reducing IO but that there is little room left for a further expansion to considerably reduce it.[41]

Fourth, circumstances account for a larger share of inequalities related to education than of overall *Itot*. On average between mid-ages 31 and 40, components *B* through *D*, which relate to both education and circumstances, represent between 29.4 and 36.5 percent of components *B* through *E* related to education, depending on the cohort.[42] This is more than twice the corresponding figure for *IOR* in every cohort, which indicates that the French educational system not only transmits the effect of childhood circumstances onto the labor market but, worse, amplifies it. This only confirms the well-known importance of social background in shaping the schooling achievement of students in France (e.g., see Thélot and Vallet 2000).

[41] This 14.9 percent potential reduction of IO through educational expansion is understated by our coarse partitioning of education, by around 10 percentage points if we extrapolate from the C72 cohort. However, when it comes to *IOR*, this potential reduction is also overstated by implicitly assuming that educational inequalities are entirely replaced by some residual inequality. If these inequalities were eliminated altogether instead, its potential reduction would be less important since *Itot* in its denominator would fall. In practice, an educational expansion falls between these two polar scenarios, as wage rigidities mean that removing a source of inequality (education) only partially eliminates the inequality previously caused by it. In the end, the 14.9 percent potential reduction of *IOR* for the C78 cohort is probably not too far off: for C72, the corresponding number is 24.3 percent (coarse categorization, first scenario), which falls within the range given by the two scenarios with a fine categorization of education (20.4 percent in the second scenario and 34.6 in the first).

[42] These shares are slightly lower when education is finely categorized, between 27.1 and 33.1 percent among the first three cohorts. It remains to be seen whether measuring education even more precisely, particularly tertiary education, would further reduce them. Indeed, the French tertiary is dual with, schematically, elitist "Grandes Ecoles" where social reproduction is high (Benveniste 2023) on one side, and more universally accessible universities on the other.

7. Conclusion

This paper offers a unique age-cohort analysis of French wage inequalities and inequality of opportunity. Using the French Demographic Sample, we observe the circumstances of individuals born between 1958 and 1980 during their childhood, in the 1975 and 1990 censuses. We then follow the yearly wages of the same individuals, once adult, in administrative payroll data from 1984 to 2018. We focus on adults working in the nonagricultural private sector, whom we split into four narrow birth cohorts, and we compute inequalities at the year-cohort level.

We find that total wage inequality increases fast and linearly with age, nearly quadrupling over the salaried life, and that it decreases, but much more slowly, across cohorts. Absolute inequality of opportunity, the part of wage inequality explained by childhood circumstances, shows a qualitatively similar evolution across age and cohorts. Relative inequality of opportunity, in share of wage inequality, decreases as well across cohorts but follows a concave pattern over the life course. It increases very quickly at the beginning of adulthood, starting at around five percent or lower among individuals in their early twenties and reaching around 15 percent by age 35, and it stabilizes afterwards.

We explain this potentially counterintuitive result – the influence of childhood circumstances grows over the life course although they are fixed at the beginning of adulthood – thanks to an econometric model à la Haider and Solon (2006), about intergenerational mobility. We also discuss how the rapid increase in inequalities over the life course implies that apparently small differences in the age range covered can render inequality comparisons across studies misleading. The midpoint of the age range matters more than its length for total inequality and absolute inequality of opportunity, dues to their linear trends. Conversely, the length of the age range matters more for relative inequality of opportunity, due to its concave trend. However, focusing on individuals older than 30 and removing intra-cohort age effects prior to computing to inequality can greatly enhance the comparability of the relative estimates, even when the length of the age range differs.

We also explore which of our circumstances matter most in shaping inequalities. We begin with a methodological clarification on how to perform a Shapley decomposition of inequality by circumstances. Specifically, we argue that it is fundamentally different from a Shapley decomposition by income sources and that treating circumstances as incomes sources, as has often been done, contradicts the principle of transfer. We then find that father's education is by far the most important driver, accounting alone for around 40 to 50 percent of inequality of

opportunity, depending on the cohort. Father's occupation, sex and the place of residence during youth explain the bulk the remaining inequality of opportunity while father's age at childbirth and father's citizenship play a negligible role. The virtually null role of father's citizenship, which may seem surprising, is due to two main factors. First, people of foreign origin represent a small share (around 9 percent) of our sample and therefore carry little weight on the inequality estimates. Second, they suffer from a greater disadvantage in the labor market at the hiring than at the remuneration stage.

Finally, we propose an analytical framework to study the mediating role of education between childhood circumstances and adult wages, based on the simultaneous decomposition of the variances explained by education and/or circumstances. We find that the 15 percent of wage inequality explained by circumstances at prime ages is not small compared to the inequality explained by education. Additionally, we find that education not only transmits the effect of childhood circumstances on wages but, worse, amplifies it. Lastly, we find that educational expansion has contributed to reducing inequality of opportunity across cohorts by lifting the majority of the more recent cohorts towards the same higher educational level, but that there is limited scope for a further expansion to reduce it any more.

References

Aeberhardt, R., Fougère, D., Pouget, J., & Rathelot, R. (2010). Wages and employment of French workers with African origin. *Journal of Population Economics*, 23, 881–905.

Andreoli F., & Fusco A. (2017). The evolution of inequality of opportunity across Europe: Evidence from EU-SILC. In *Income and Living Conditions in Europe*, 435–448. Luxembourg: Publications Office of the European Union.

Andreoli, F., Fusco, A., Kyzyma, I., & Van Kerm, P. (2021). *New estimates of inequality of opportunity across European cohorts (and some insights on the long-term impact of educational policy).* 36th IARIW Virtual General Conference e-Oslo, Norway.

Auvray, C., & Trannoy, A. (1992): *Décomposition par source de l'inégalité des revenues à l'aide de la valeur de Shapley*. Journées de Microéconomies Appliqué, Sfax.

Baker, M., & Solon, G. (2003). Earnings dynamics and inequality among Canadian men, 1976–1992: Evidence from longitudinal income tax records. *Journal of Labor Economics*, 21, 289–321.

Benveniste, S. (2023*). Like Father, Like Child: Intergenerational Mobility in the French Grandes Écoles throughout the 20th Century*. AMSE WP 2023 Nr. 18.

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, 675–696.

Black, S. E., & Devereux, P. J. (2011). Recent developments in intergenerational mobility. In *Handbook of labor economics* (Volume 4, 1487–1541). Amsterdam: Elsevier.

Blundell, R. (2014). Income dynamics and life-cycle inequality: mechanisms and controversies. *The Economic Journal*, 124, 289–318.

Blundell, R., Graber, M., & Mogstad, M. (2015). Labor income dynamics and the insurance from taxes, transfers, and the family. *Journal of Public Economics*, 127, 58–73.

Bönke, T., Corneo, G., & Lüthen, H. (2015). Lifetime earnings inequality in Germany. *Journal of Labor Economics*, 33, 171–208.

Boudoul, J. (1983). Organisation du recensement de la population de 1982. *Journal de la société française de statistique*, 124, 3–15.

Bourguignon, F. (1979). Decomposable Income Inequality Measures. *Econometrica*, 47, 901–920.

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

Bozio, A., Breda, T., & Guillot, M. (2023). Using payroll taxes as a redistribution tool. *Journal of Public Economics*, 226, 1–11.

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

Brunori, P., & Neidhöfer, G. (2021). The evolution of inequality of opportunity in Germany: A machine learning approach. *Review of Income and Wealth*, 67, 900–927.

Brzezinski, M. (2015). *Inequality of opportunity in Europe before and after the Great Recession*. ECINEQ WP No. 353.

Bussolo, M., Checchi, D., & Peragine, V. (2019). *Long-term evolution of inequality of opportunity*. ECINEQ WP No. 485.

Bussolo, M., Checchi, D., & Peragine, V. (2023). Long-term evolution of inequality of opportunity: Educated parents still matter. *The Journal of Economic Inequality*, 21, 277–323.

Cabrera, L., Marrero, G. A., Rodríguez, J. G., & Salas-Rojo, P. (2021). Inequality of opportunity in Spain: new insights from new data. *Hacienda Publica Espanola*, 237, 153–185.

Cappelen, A. W., Eichele, T., Hugdahl, K., Specht, K., Sørensen, E. Ø., & Tungodden, B. (2014). Equity theory and fair inequality: A neuroeconomic study. *Proceedings of the National Academy of Sciences*, 111, 15368–72.

Carranza, R. (2022). Upper and Lower Bound Estimates of Inequality of Opportunity: A Cross-National Comparison for Europe. *Review of Income and Wealth*.

Chantreuil, F., & Trannoy, A. (2011). Inequality Decomposition Values. *Annals of Economics and Statistics*, 101/102, 13–36.

Chantreuil, F., & Trannoy, A. (2013). Inequality decomposition values: the trade-off between marginality and efficiency. *The Journal of Economic Inequality*, 11, 83–98.

Checchi, D., Peragine, V., & Serlenga, L. (2016). Inequality of Opportunity in Europe: Is There a Role for Institutions?. In *Inequality: Causes and consequences* (Research in Labor Economics, Volume 43, 1–47). Bingley: Emerald Group Publishing Limited.

Cowell, F. A., & Fiorio, C. V. (2011). Inequality decompositions—a reconciliation. *The Journal of Economic Inequality*, 9, 509–528.

Cowell, F. A., & Flachaire, E. (2007). Income distribution and inequality measurement: The problem of extreme values. *Journal of Econometrics*, 141, 1044–1072.

Cowell, F. A., & Victoria-Feser, M.-P. (1996). Robustness Properties of Inequality Measures. *Econometrica*, 64, 77–101.

Creedy, J., & Hart, P. E. (1979). Age and the Distribution of Earnings. *The Economic Journal*, 89, 280–293.

d'Albis, H., & Badji, I. (2022). Inequality within generation: Evidence from France. *Research in Economics*, 76, 69–83.

Dai, X., & Li, J. (2021). Inequality of opportunity in China: Evidence from pseudo panel data. *China Economic Review*, 68, 101637.

Deaton, A., & Paxson, C. (1994a). Intertemporal choice and inequality. *Journal of Political Economy*, 102, 437–467.

Deaton, A., & Paxson, C. (1994b). Saving, growth, and aging in Taiwan. In *Studies in the Economics of Aging*, 331–362. Chicago: University of Chicago Press for the National Bureau of Economic Research.

Deutsch, J., Pi Alperin, M. N., & Silber, J. (2018). Using the Shapley decomposition to disentangle the impact of circumstances and efforts on health inequality. *Social Indicators Research*, 138, 523–543.

del Valle-Inclán, H. (2019). *Estimating inequality of opportunity in many new periods: The Capital Income approach*. ECOBAS Working Papers 2019 – 01.

Dherbécourt C. (2018). *Nés sous la même étoile ? Origine sociale et niveau de vie*. La Note d'analyse de France Stratégie n° 68.

Dherbécourt C., & Flamand, J. (2023). *Inégalité des chances: ce qui compte le plus*. La Note d'analyse de France Stratégie n° 120.

Ferreira, F. H., & Gignoux, J. (2011a). *The measurement of educational inequality: Achievement and opportunity*. IZA DP No. 3181.

Ferreira, F. H. G., & Gignoux, J. (2011b). The measurement of inequality of opportunity: Theory and an application to Latin America. *Review of Income and Wealth*, 57, 622–657.

Ferreira, F. H., Gignoux, J., & Aran, M. (2011). Measuring inequality of opportunity with imperfect data: the case of Turkey. *The Journal of Economic Inequality*, 9, 651–680.

Ferreira, F. H. G., & Peragine, V. (2015). *Equality of opportunity: theory and evidence*. World Bank Policy Research Working Paper No. 7217.

Fong, C. (2001). Social preferences, self-interest, and the demand for redistribution. *Journal of Public Economics*, 82, 225–246.

Foster, J. E., & Shneyerov, A. A. (2000). Path independent inequality measures. *Journal of Economic Theory*, 91, 199–222.

Galland, O., & Rouault, D. (1998). Devenir cadre dès trente ans: une approche longitudinale de la mobilité sociale. *Économie et Statistique*, 316, 97–107.

Golley, J., Zhou, Y. and Wang, M. (2019), Inequality of Opportunity in China's Labor Earnings: The Gender Dimension. *China & World Economy*, 27, 28–50.

Haider, S., & Solon, G. (2006). Life-cycle variation in the association between current and lifetime earnings. *American Economic Review*, 96, 1308–1320.

Heathcote, J., Storesletten, K., & Violante, G. L. (2005). Two Views of Inequality over the Life Cycle. *Journal of the European Economic Association*, 3, 765–775.

Hederos, K., Jäntti, M., & Lindahl, L. (2017). Gender and inequality of opportunity in Sweden. *Social Choice and Welfare*, 49, 605–635.

Huggett, M., Ventura, G., & Yaron, A. (2006). Human capital and earnings distribution dynamics. *Journal of Monetary Economics*, 53, 265–290.

Huggett, M., Ventura, G., & Yaron, A. (2011). Sources of Lifetime Inequality. *American Economic Review*, 101, 2923–2954.

INSEE. (2022). *Base étude 2020 de l'Echantillon Démographique Permanent : Manuel de l'utilisateur*.

Israeli, O. (2007). A Shapley-based decomposition of the R-square of a linear regression. *The Journal of Economic Inequality*, 5, 199–212.

Juárez, F. W. C., & Soloaga, I. (2014). Iop: Estimating Ex-Ante Inequality of Opportunity. *The Stata Journal*, 14, 830–846.

Jugnot, S. (2014). *La constitution de l'échantillon démographique permanent de 1968 à 2012*. Insee Document de Travail N° F1406, 1–83.

Kanbur, R., & Wagstaff, A. (2016). How Useful Is Inequality of Opportunity as a Policy Construct? In *Inequality and Growth: Patterns and Policy*, 131–150. London: Palgrave Macmillan.

Kenedi, G., & Sirugue, L. (2023). Intergenerational income mobility in France: A comparative and geographic analysis. *Journal of Public Economics*, 226, 104974.

Konow, J. (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions. *American Economic Review*, 90, 1072–1091.

Kovacic, M., & Orso, C. E. (2022). Trends in inequality of opportunity in health over the life cycle: The role of early-life conditions. *Journal of Economic Behavior & Organization*, 201, 60–82.

Lefranc, A., Pistolesi, N., & Trannoy, A. (2004). Le revenu selon l'origine sociale. *Economie et Statistique*, 371, 49–82.

Lefranc, A., Pistolesi, N., & Trannoy, A. (2009). Equality of opportunity and luck: Definitions and testable conditions, with an application to income in France. *Journal of Public Economics*, 93, 1189–1207.

Lefranc, A., Pistolesi, N., & Trannoy, A. (2011). Measuring Circumstances: Francs or Ranks, does it Matter?. In *Inequality of Opportunity: Theory and Measurement* (Research on Economic Inequality, Volume 19, 131–156). Bingley: Emerald Group Publishing Limited.

Marrero, G. A., & Rodríguez, J. G. (2011). Inequality of opportunity in the United States: Trends and decomposition. In *Inequality of Opportunity: Theory and Measurement* (Research on Economic Inequality, Volume 19, 217–246). Bingley: Emerald Group Publishing Limited.

Marrero, G. A., & Rodríguez, J. G. (2013). Inequality of opportunity and growth. *Journal of Development Economics*, 104, 107–122.

Marrero, G. A., & Rodríguez, J. G. (2016). Macroeconomic determinants of inequality of opportunity in the United States: 1970–2009. In *Inequality after the 20th Century: Papers from the Sixth ECINEQ Meeting* (Research on Economic Inequality, Volume 24, 307–330). Bingley: Emerald Group Publishing Limited.

Martinez, A., Rampino, T., Western, M., Tomaszewski, W., & Roque, J. D. (2017). Estimating the Contribution of Circumstances that Reflect Inequality of Opportunities. *Economic Papers*, 36, 380–400.

Mazumder, B. (2005). Fortunate sons: New estimates of intergenerational mobility in the United States using social security earnings data. *Review of Economics and Statistics*, 87, 235–255.

McAvay, H. (2018). The ethnoracial context of residential mobility in France: Neighbourhood out-migration and relocation. *Population, Space and Place*, 24, 1–16.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. New York: Columbia University Press.

Moramarco, D., Palmisano, F., & Peragine, V. (2020). *Inter-temporal inequality of opportunity*. ECINEQ WP No. 554.

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

Nybom, M., & Stuhler, J. (2017). Biases in standard measures of intergenerational income dependence. *Journal of Human Resources*, 52, 800–825.

Owen, G. (1977). Values of games with a priori unions. In *Essays in Mathematical Economics & Game Theory*, 76–88. Berlin: Springer.

Palmisano, F., & Peragine, V. (2022). Inequality of opportunity: theoretical considerations and recent empirical evidence. In *Advances in Economic Measurement*, 349–386. Singapore: Palgrave Macmillan.

Palomino, J. C., Marrero, G. A., & Rodríguez, J. G. (2019). Channels of inequality of opportunity: The role of education and occupation in Europe. *Social Indicators Research*, 143, 1045–1074.

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

Polachek, S. W. (2007). *Earnings over the lifecycle: The mincer earnings function and its applications*. IZA DP No. 3181.

Ramos, X., & van de Gaer, D. (2016). Approaches to inequality of opportunity: Principles, measures and evidence. *Journal of Economic Surveys*, 30, 855–883.

Ramos, X., & van de Gaer, D. (2021), Is Inequality of Opportunity Robust to the Measurement Approach?. *Review of Income and Wealth*, 67, 18–36.

Robert-Bobée, I., & Gualbert, N. (2021). L'échantillon démographique permanent : en 50 ans, l'EDP a bien grandi ! *Courrier des statistiques*, 6, 47–63.

Robert-Bobée, I., Rendall, M., Couet, C., Lappegard, T., Rønsen, M., & Smallwood, S. (2006). Âge au premier enfant et niveau d'études: une analyse comparée entre la France, la Grande-Bretagne et la Norvège. In *Donnée sociales : La société française*, 69–76. Paris : INSEE.

Roemer, J. E. (1993). A pragmatic theory of responsibility for the egalitarian planner. *Philosophy & Public Affairs*, 22, 146–166.

Roemer, J. E., & Trannoy, A. (2015). Equality of opportunity. In *Handbook of income distribution* (Volume 2A, 217–300). Amsterdam: Elsevier.

Salas-Rojo, P., & Rodríguez, J. G. (2022). Inheritances and wealth inequality: a machine learning approach. *The Journal of Economic Inequality*, 20, 27–51.

Shapley, L. S. (1953). A value for n-person games. In *Contributions to the Theory of Games* (Volume II, 307–317). Princeton: Princeton University Press.

Shi, X. (2019). Inequality of opportunity in energy consumption in China. *Energy policy*, 124, 371–382.

Shorrocks, A. F. (1982). Inequality Decomposition by Factor Components. *Econometrica*, 50, 193–211.

Shorrocks, A.F. (2013). Decomposition procedures for distributional analysis: a unified framework based on the Shapley value. *The Journal of Economic Inequality*, 11, 99–126.

Sicsic, M. (2023). Who Climbs Up the Income Ladder? An Analysis of Intergenerational Income Mobility in France. *Economie et Statistique, prepublication 2023*.

Solon, G. (1999). Intergenerational Mobility in the Labor Market. In *Handbook of labor economics* (Volume 3, 1761–1800). Amsterdam: Elsevier.

Suárez-Álvarez, A., & López-Menéndez, A. J. (2018). Income Inequality and Inequality of Opportunity in Europe: Are they on the Rise? In *Inequality, Taxation and Intergenerational Transmission* (Research on Economic Inequality, Volume 26, 149–196). Bingley: Emerald Group Publishing Limited

Suárez-Álvarez, A., & López-Menéndez, A. J. (2021). Dynamics of inequality and opportunities within European countries. *Bulletin of Economic Research*, 73, 555–579.

Teyssier, G. (2017a). *Inequality of opportunity: new measurement methodology and impact on growth*. Seventh ECINEQ Meeting, New York City.

Teyssier, G. (2017b). *Inequality of Opportunity and the Parametric Decomposition of Inequality by Population Subgroups*. Eighth ECINEQ Meeting, Paris.

Thélot, C., & Vallet, L.-A. (2000). La réduction des inégalités sociales devant l'école depuis le début du siècle. *Economie et Statistique, 334, 3–32*.

Tomkinson, J., Breton, D., & Mazuy, M. (2017). Entry into motherhood during adolescence in France. *Quetelet Journal*, 5, 73–98.

Toulemon, L., & Mazuy, M. (2001). Les naissances sont retardées mais la fécondité est stable. *Population*, 56, 611–644.

Verdugo, G. (2014). The great compression of the French wage structure, 1969–2008. *Labour Economics*, 28, 131–144.

World Bank. (2005). *World development report 2006: Equity and development*. Washington D.C.: The World Bank, Oxford University Press.

Wu, D., Rao, P., Trivedi, P., Tang, K. K., & Rohde, N. (2021). *Measuring inequality of opportunity using sampling-based Shapley decomposition: Application to Australian data 2001-2017*. 36th IARIW Virtual General Conference e-Oslo, Norway.

Yang, X., Gustafsson, B., & Sicular, T. (2021). Inequality of opportunity in household income, China 2002–2018. *China Economic Review*, 69, 101684.

Young, H. P. (1985). Monotonic solutions of cooperative games. *International Journal of Game Theory*, 14, 65–72.

Zhang, Y., & Eriksson, T. (2010). Inequality of opportunity and income inequality in nine Chinese provinces, 1989–2006. *China Economic Review*, 21, 607–616.

Web Appendix to:

Long Run Evolution of Inequality of Opportunity across Age and Cohorts: Evidence from the French Permanent Demographic Sample

2024

by Geoffrey Teyssier (INED) and Laurent Toulemon (INED)

Table of contents

I. Variations on Figure 1 3
Figure A1.1 : Population shares over the life course, by cohorts and circumstances 3
Figure A1.2: Simple regression coefficients over the life course, by cohorts (with 95% C.I.) 5
Figure A1.3: Multiple regression coefficients 9
Figure A1.4: Population shares and average rescaled wages, by own education 13
II. Variations on Figure 2 14
Figure A2.1: Inequality over the life course, by cohort: birth years reweighted 14
Figure A2.2: Inequality over the life course: birth year 1974 at age 1 versus 16 16
Figure A2.3: Inequality over the life course, by cohort: top 1 percent wages excluded 17
Figure A2.4: Inequality over the life course, by cohort: poor people included 19
Figure A2.5: Inequality over the life course, by cohort: all wage earners included 20
Figure A2.6: Inequality over the life course, by cohort: full-year equivalent wage 21
Figure A2.7: Inequality over the life course, by cohort: *GE0* index 23
III. Variations on Figure 3 25
Figure A3.1: Comparability of *IOA* over selected age ranges 25
Figure A3.2: Comparability of *Itot* over selected age ranges 26
Figure A3.3: Comparability of *IOR* over selected age ranges: treating birth year as a circumstance (no age adjustment) 27
Figure A3.4: Comparability of *IOA* over selected age ranges: treating birth year as a circumstance (no age adjustment) 28
IV. Variations on Figure 4 29
Figure A.4.1: Shapley decomposition of *IOR* by circumstances: curves 29
Figure A4.2: Shapley decomposition of *IOR* by circumstances: in percentage of IO 30
Figure A.4.3: Shapley decomposition of IOR by circumstances: in perc. of IO, curves 31
Figure A4.4: Marginal contributions of circumstances: simple linear models 32
Figure A4.5: Marginal contributions of circumstances: full multiple linear model 33

V. Variations on Figure 5 34

Figure A5.1: Shares of variance explained by circumstances and education: quaternary education 34

Figure A5.2: Shares of variance explained by circumstances and education: curves 35

Figure A5.3: Shares of variance explained by circumstances and education: quaternary education, curves 36

VI. Further data description 37

Table A6.1: Coding of father's occupation 37

Figure A6.1: Average and top wages over the life course 38

VII. Literature review 39

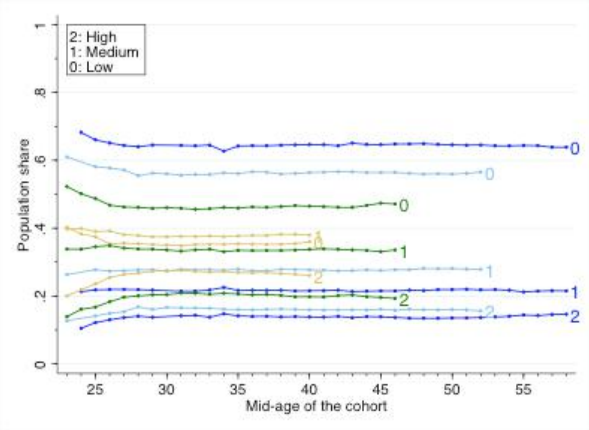
Table A7.1: Review of the literature about the evolution of IO in income or earnings : time dimension 39

Table A7.2: Review of the literature about the evolution of IO in income or earnings: circumstance and income variables 42

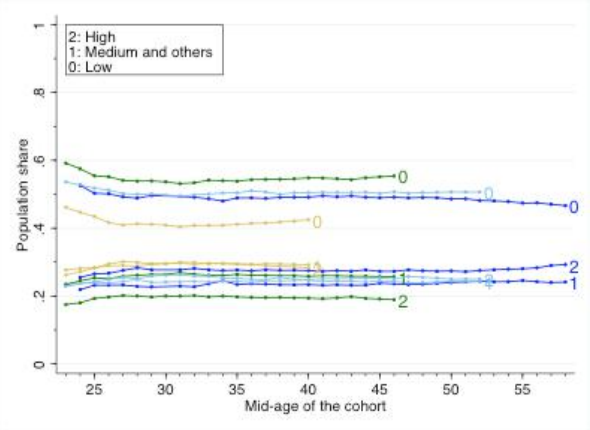
I. Variations on Figure 1

Figure A1.1 : Population shares over the life course, by cohorts and circumstances

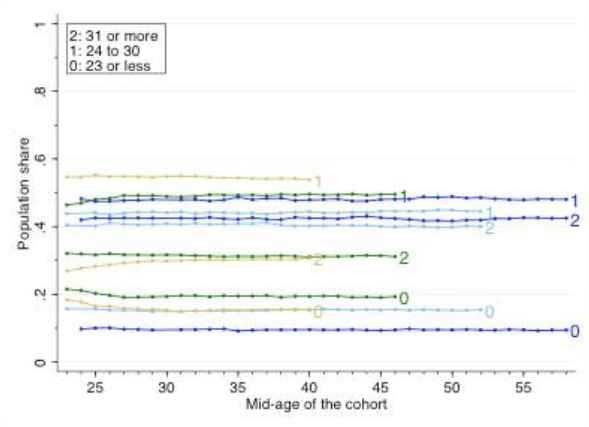
By father's education



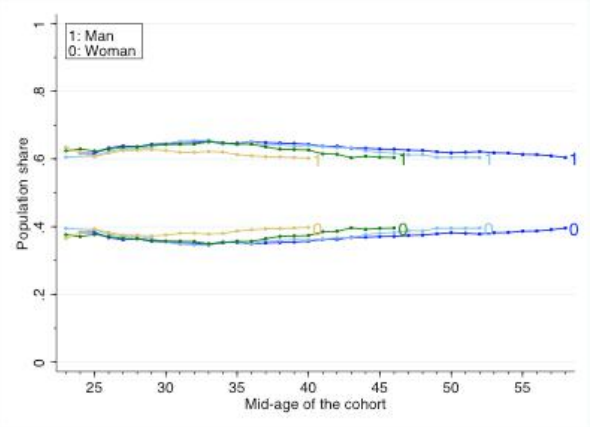
By father's occupation



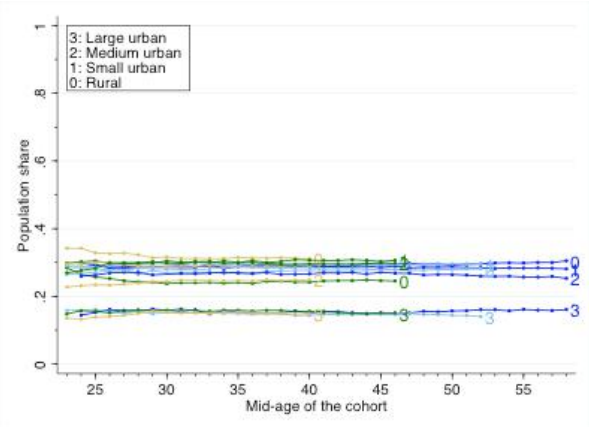
By age difference with the father



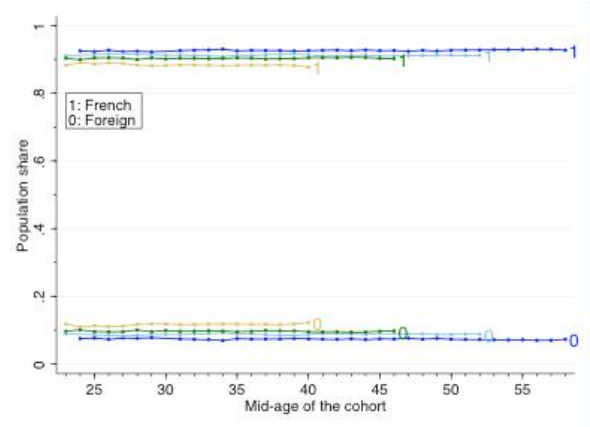
By sex

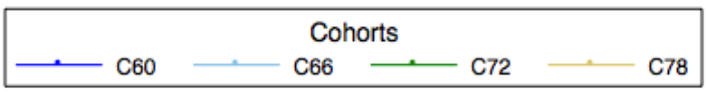


By area of residence during youth



By father's citizenship



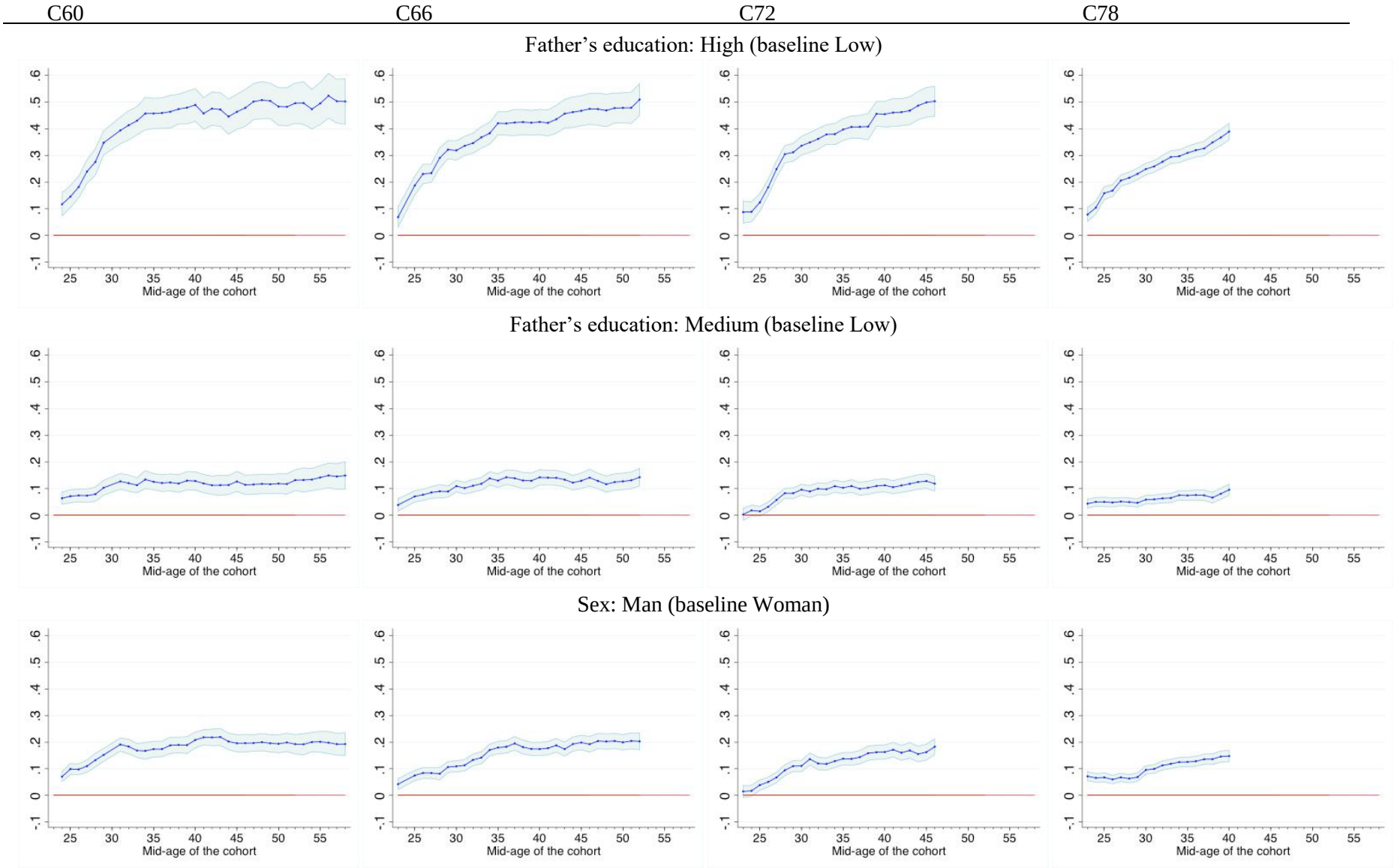


Note: The figure shows the population share of each circumstance category by year-cohort in our benchmark sample over which we compute inequality, which mainly gathers salaried people from the "nonagricultural

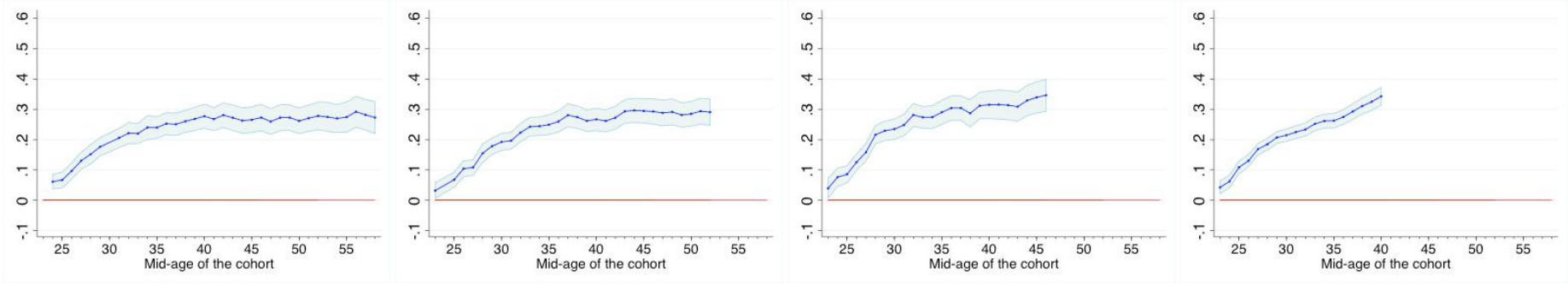
private sector", after casewise deletion on circumstances.

Comment: The shares are stable over the lifecourse, which is not surprising. However, we note a slight decrease for women around age 33, associated with their lower participation rate due to maternity. The cross-cohort evolutions are more pronounced, reflecting the expansion of secondary and tertiary education, the shift of the working population towards higher-skilled professions, and the rise of immigration in France. The rejuvenation of fathers between the C60 and C78 cohorts may seem surprising, as it does not reflect the more recent demographic trends. It does however reflect the drop in the age of mothers at childbirth that gradually occurred in France from the late 1940s to the late 1970s, as already documented by Toulemon and Mazuy (2001).

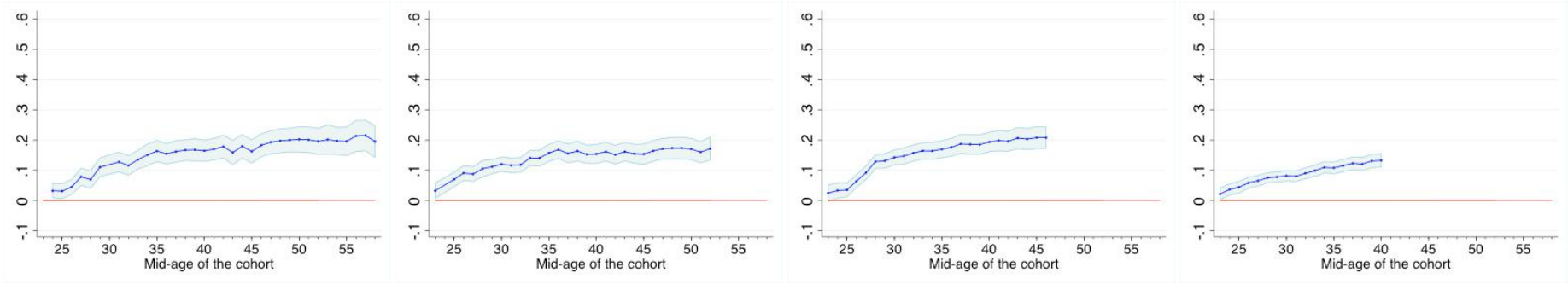
Figure A1.2: Simple regression coefficients over the life course, by cohorts (with 95% C.I.)



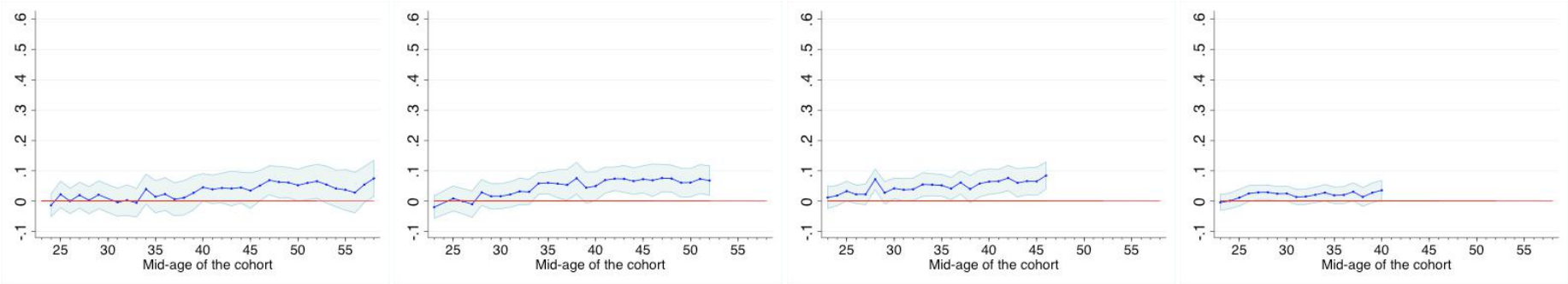
Father's occupation: High (baseline Low)



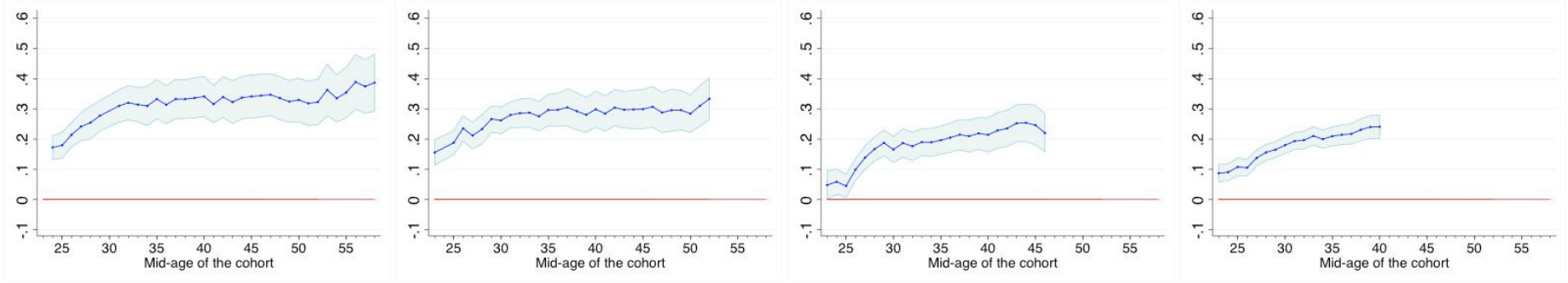
Father's occupation: Medium and others (baseline Low)



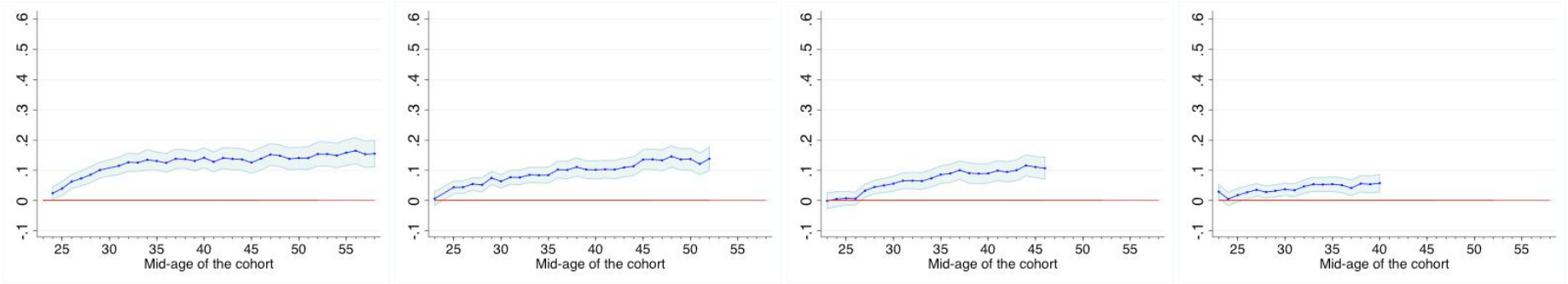
Father's citizenship: French (baseline Foreign)



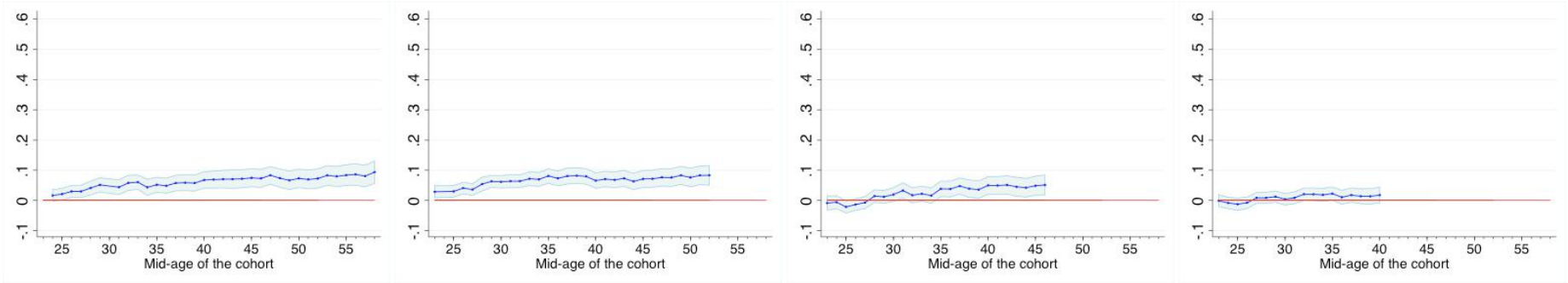
Place of residence during youth: Large urban (baseline Rural)



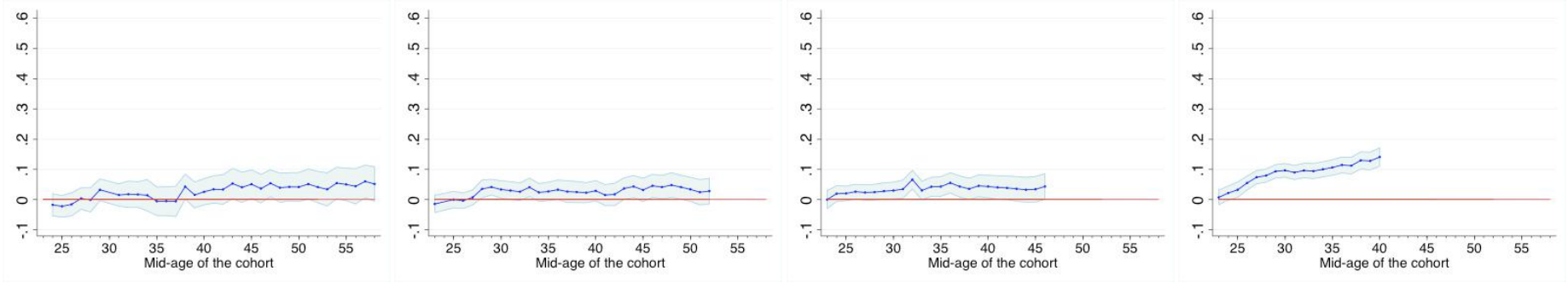
Place of residence during youth: Medium urban (baseline Rural)



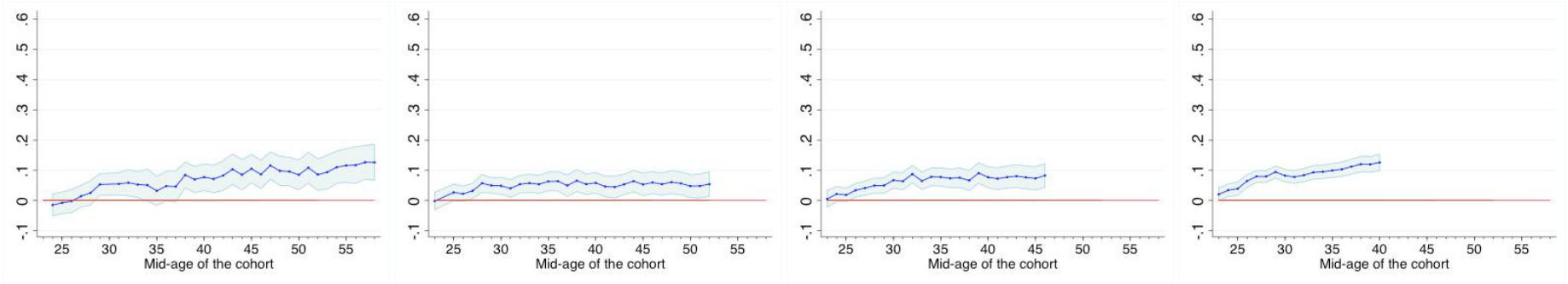
Place of residence during youth: Small urban (baseline Rural)



Age difference with the father: 31 or more (baseline 23 or less)

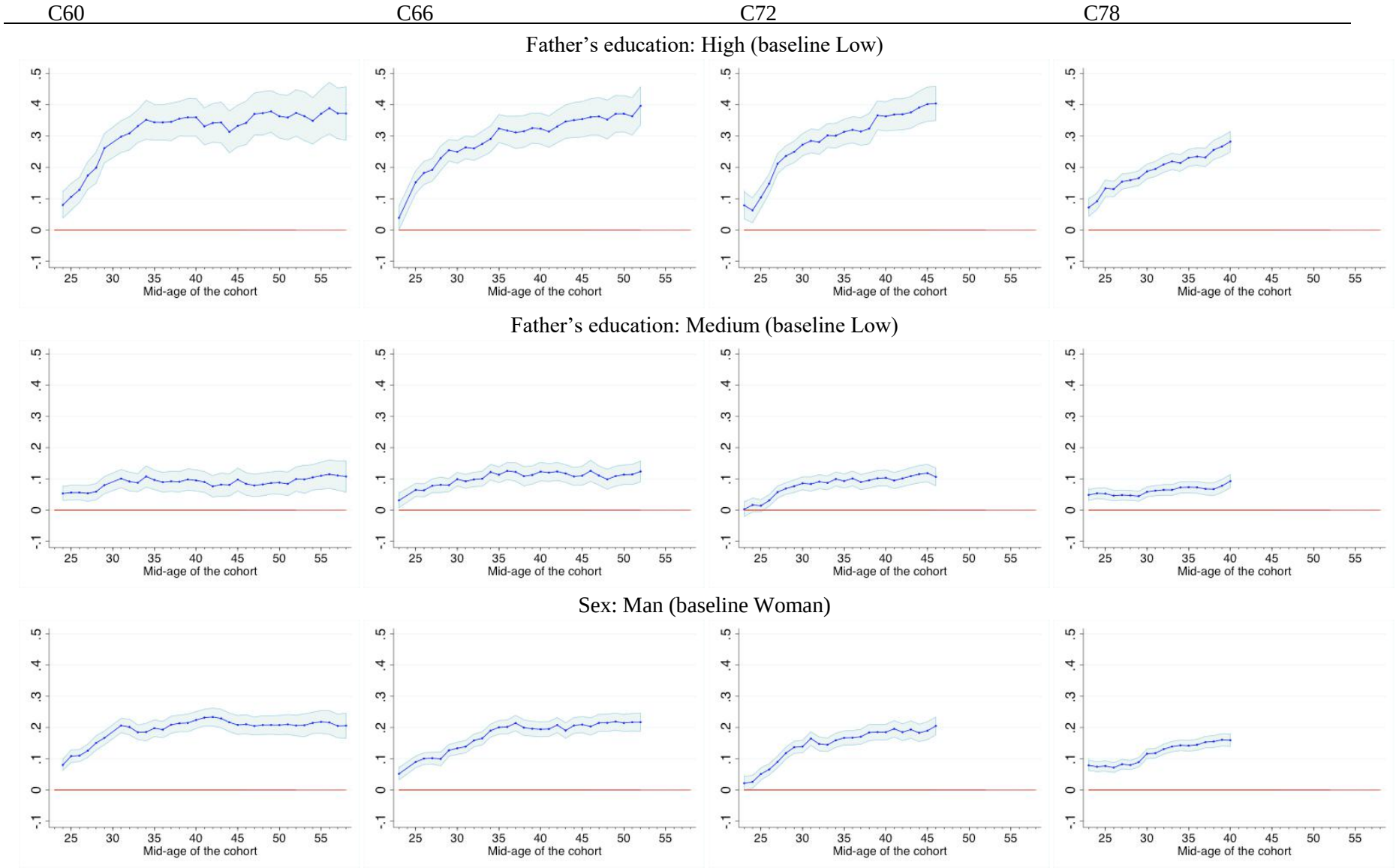


Age difference with the father: 24 to 30 (baseline 23 or less)

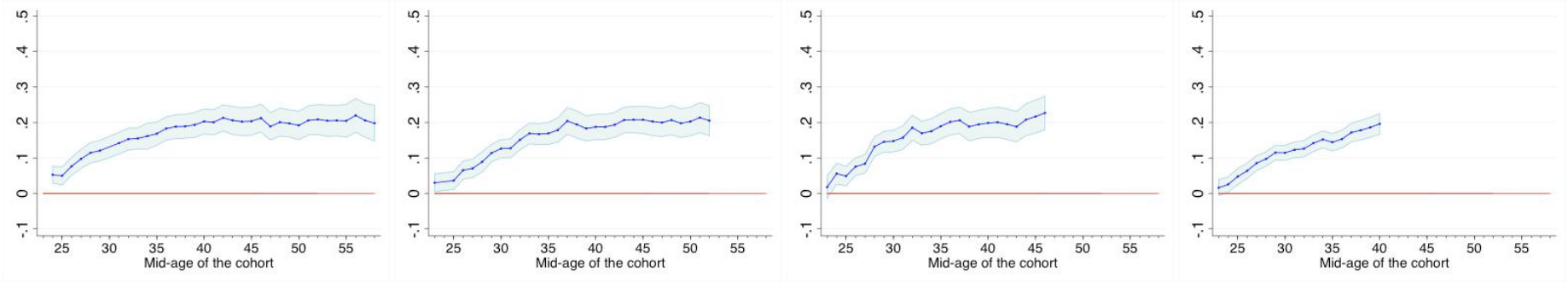


Note: The figure shows the OLS coefficients and associated 95% confidence intervals from alternative simple linear models of wages regressed on one circumstance at a time, at the year-cohort level. The regressions are performed over rescaled wages, so a value of, for instance, 0.1 means that the average wage of the underlying category equals the average of the baseline plus 10 percent of the overall wage average (of the year-cohort). All regressions are estimated on the same sample as inequality in Figure 2.

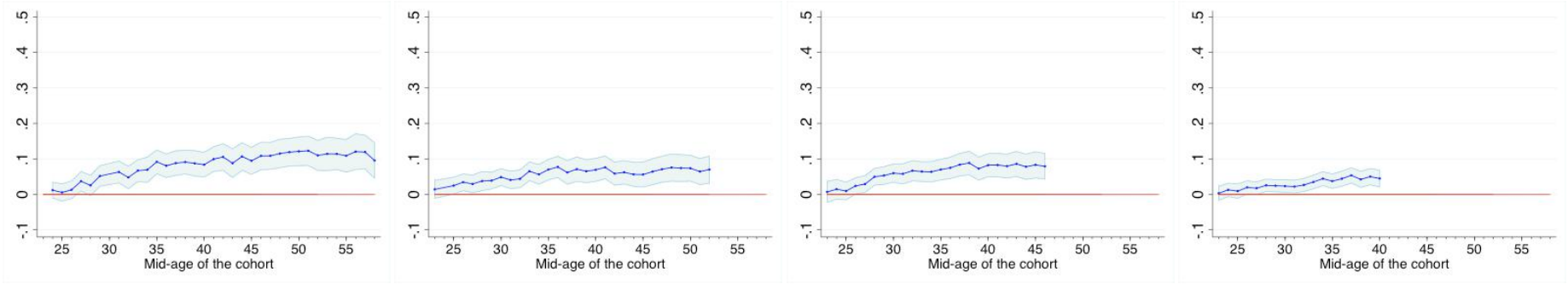
Figure A1.3: Multiple regression coefficients



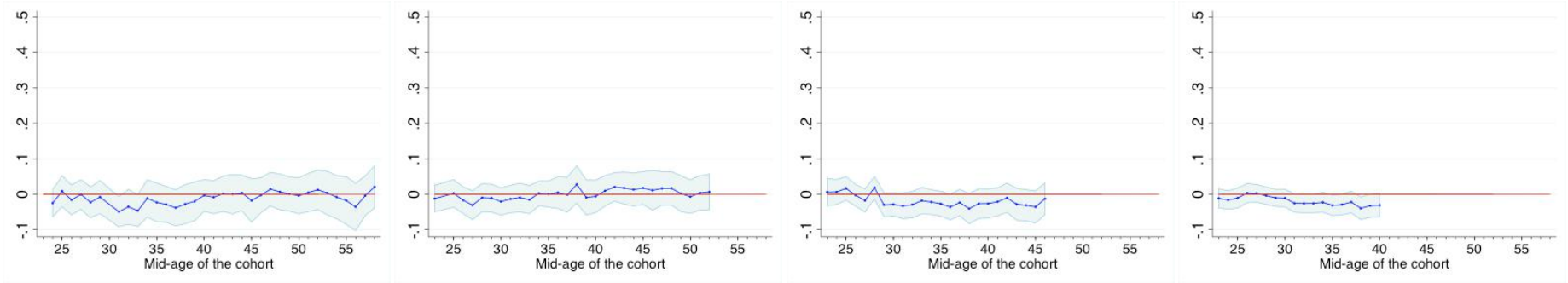
Father's occupation: High (baseline Low)



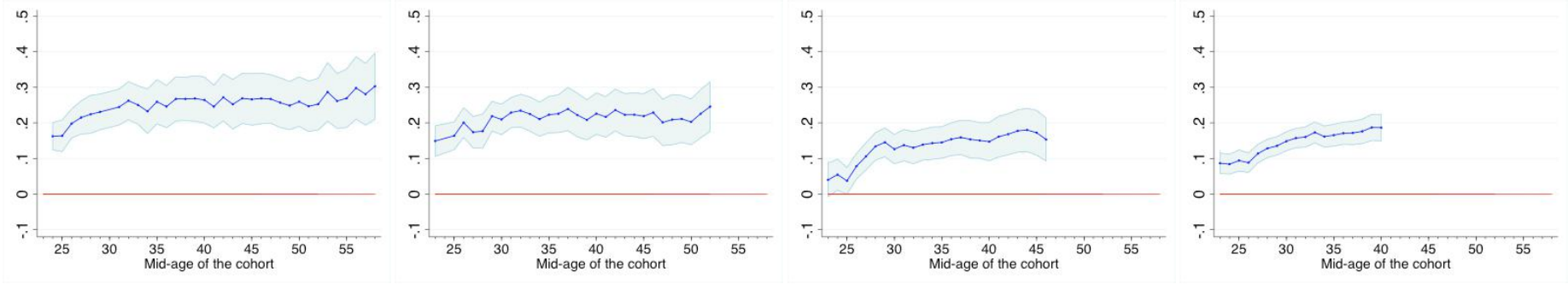
Father's occupation: Medium and others (baseline Low)



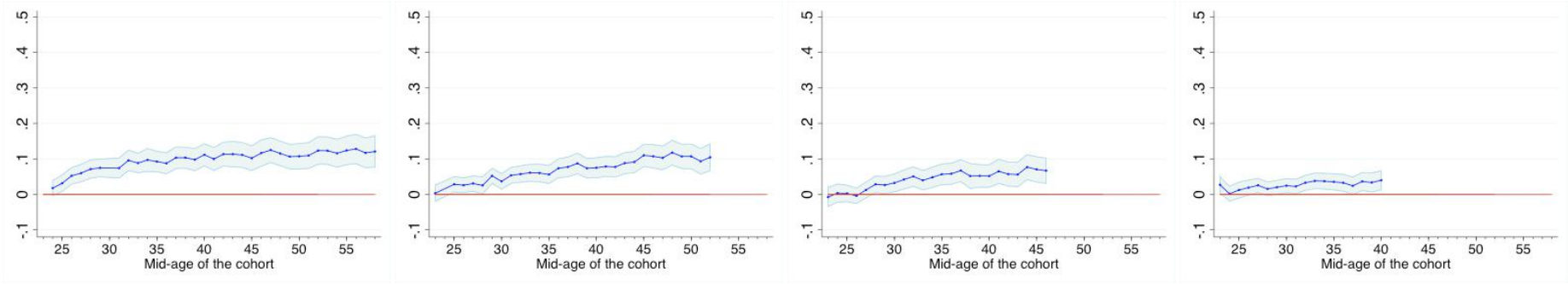
Father's citizenship: French (baseline Foreign)



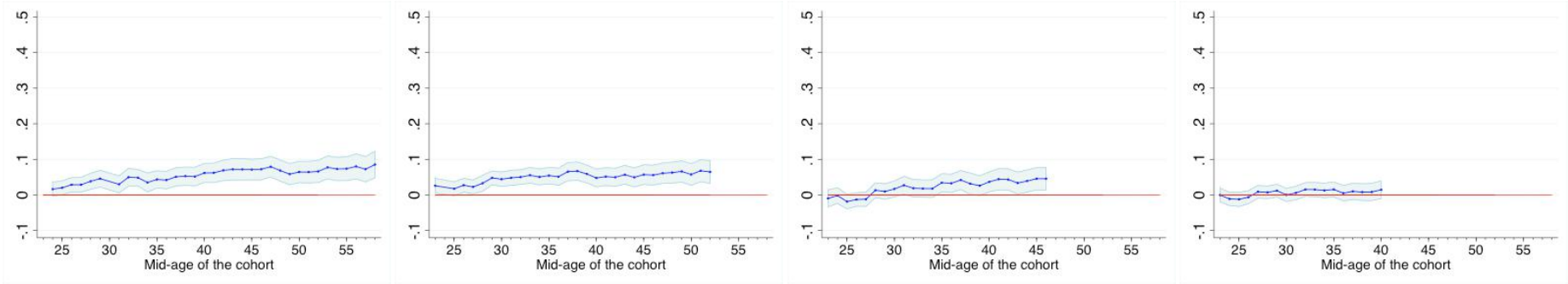
Place of residence during youth: Large urban (baseline Rural)



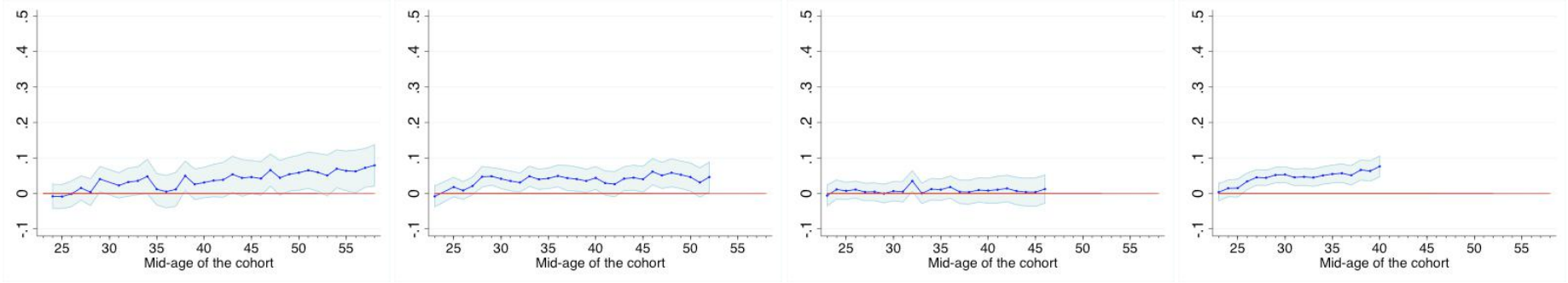
Place of residence during youth: Medium urban (baseline Rural)



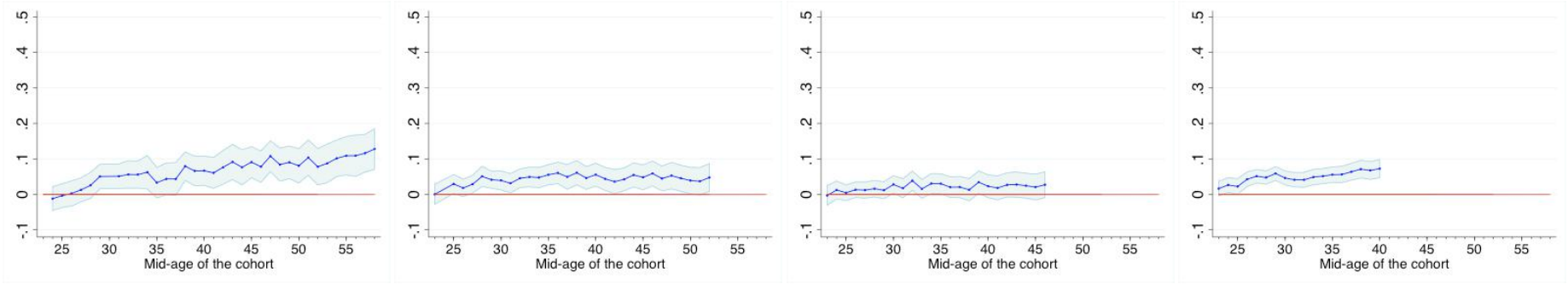
Place of residence during youth: Small urban (baseline Rural)



Age difference with the father: 31 or more (baseline 23 or less)



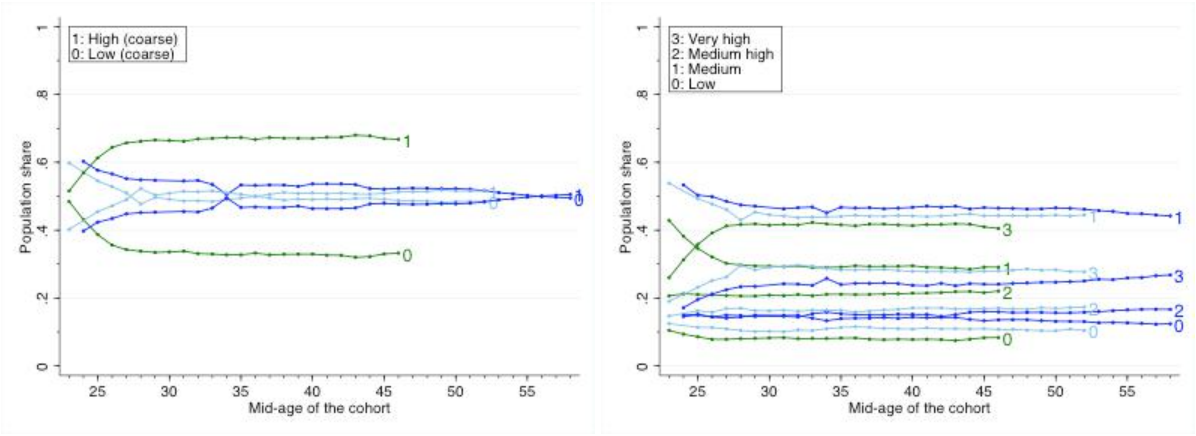
Age difference with the father: 24 to 30 (baseline 23 or less)



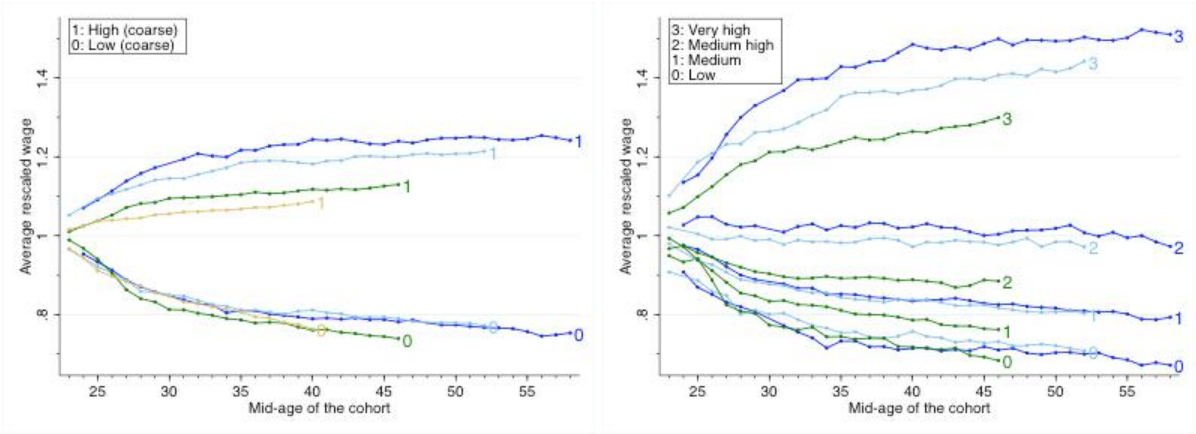
Note: The figure shows the OLS coefficients and associated 95% confidence intervals from the multiple linear model (3) of wages regressed on circumstances, at the year-cohort level. The regressions are performed over rescaled wages, so a value of, for instance, 0.1 means that the average wage of the underlying category equals the average of the baseline plus 10 percent of the overall wage average (of the year-cohort). All regressions are estimated on the same sample as inequality in Figure 2

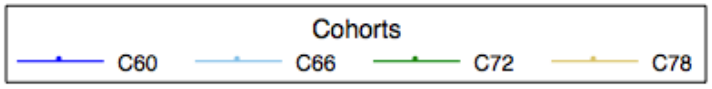
Figure A1.4: Population shares and average rescaled wages, by own education

Population shares, coarse categ. Population shares, fine categ.



Ave. rescaled wages, coarse categ. Ave. rescaled wages, fine categ.





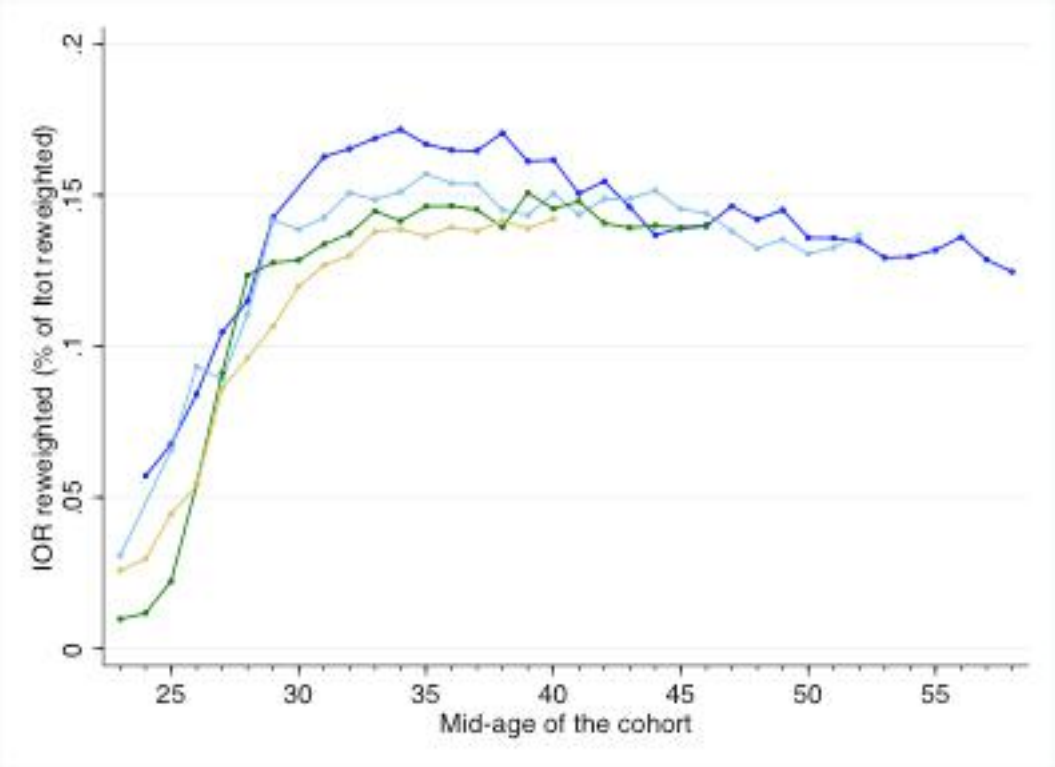
Notes : The figure shows the population shares the average rescaled wages over the life course by cohorts and educational categories, alternatively under our coarse and fine categorization of own education. The fine categorization is not available for the last C78 cohort.

The fine categorization is not nested within the coarse one. The High group under the coarse categorization (upper secondary education attended, at least) is normally reached at age 16, with age defined as the one attained during year. The Medium group under the fine categorization (lower secondary education attended but upper secondary not completed) is normally reached at age 12, the Medium high group (upper secondary education completed, at most) is normally reached at age 18.

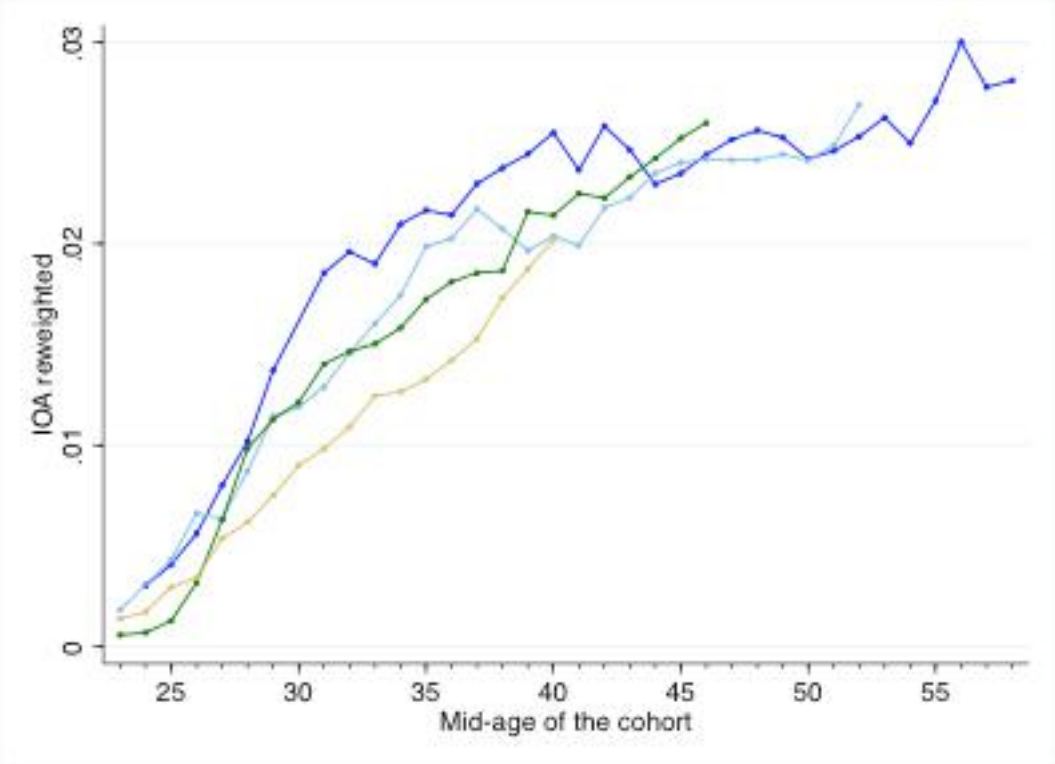
II. Variations on Figure 2

Figure A2.1: Inequality over the life course, by cohort: birth years reweighted

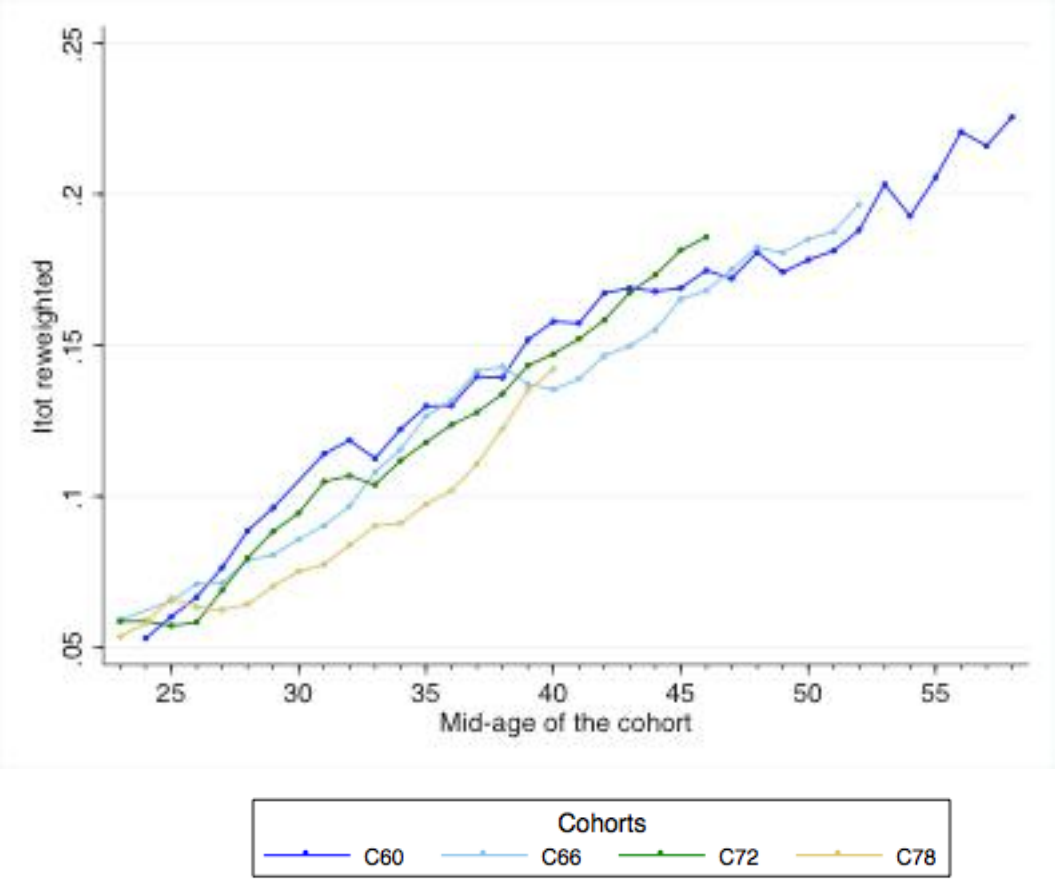
(a) Inequality of opportunity in relative terms (*IOR*)



(b) Inequality of opportunity in absolute terms (*IOA*)



(c) Total earnings inequality (*Itot*)

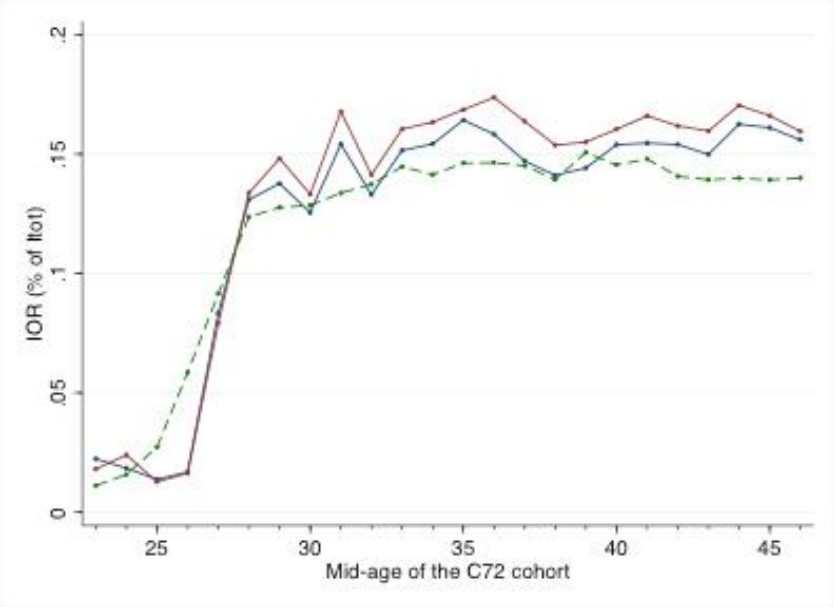


Note: The figure is the same as Figure 2 but with birth years reweighted prior to computing inequality. We reweight individuals, in each year of the payroll data, by the ratio of the population size of their birth year before casewise deletion on circumstances to that after casewise deletion. This reweighting ensures that all birth years are equally weighted before and after the deletion. It thereby prevents the deletion from inadvertently introducing an intra-cohort age effect.

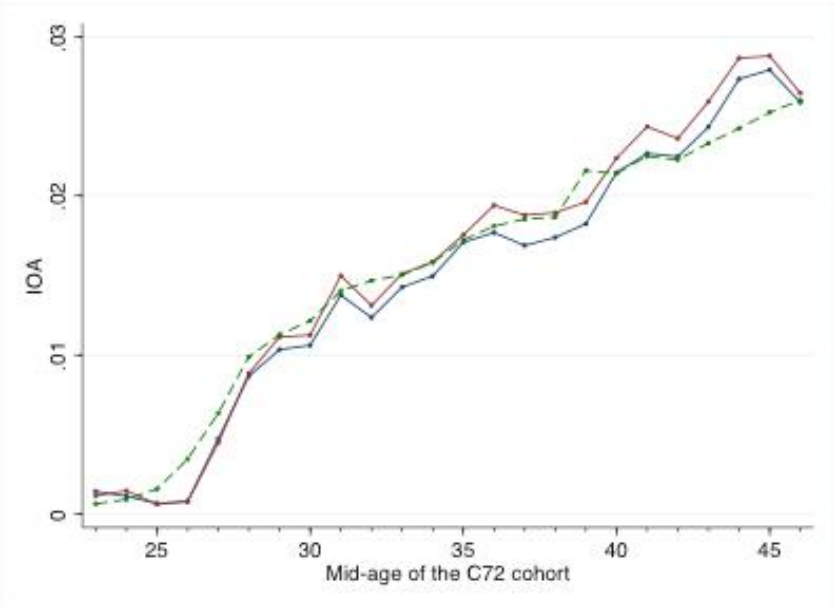
Comment: The figure is nearly indistinguishable from Figure 2, which indicates that casewise deletion on circumstances has not introduced a bias in Figure 2. Although the reweighted inequality estimates shown here are statistically different from (lower than) the non-reweighted ones from Figure 2 (with paired t-test above 5 in absolute value for each of *IOR*, *IOA* and *Itot* across our 105 year-cohort observations), they are not practically different. The correlation between the reweighted and non-reweighted estimates is above 0.99, and their average relative difference is below 1.5 percent in absolute value for each of *IOR*, *IOA* and *Itot* across our 105 year-cohort observations.

Figure A2.2: Inequality over the life course: birth year 1974 at age 1 versus 16

(a) Inequality of opportunity in relative terms (*IOR*)



(b) Inequality of opportunity in absolute terms (*IOA*)

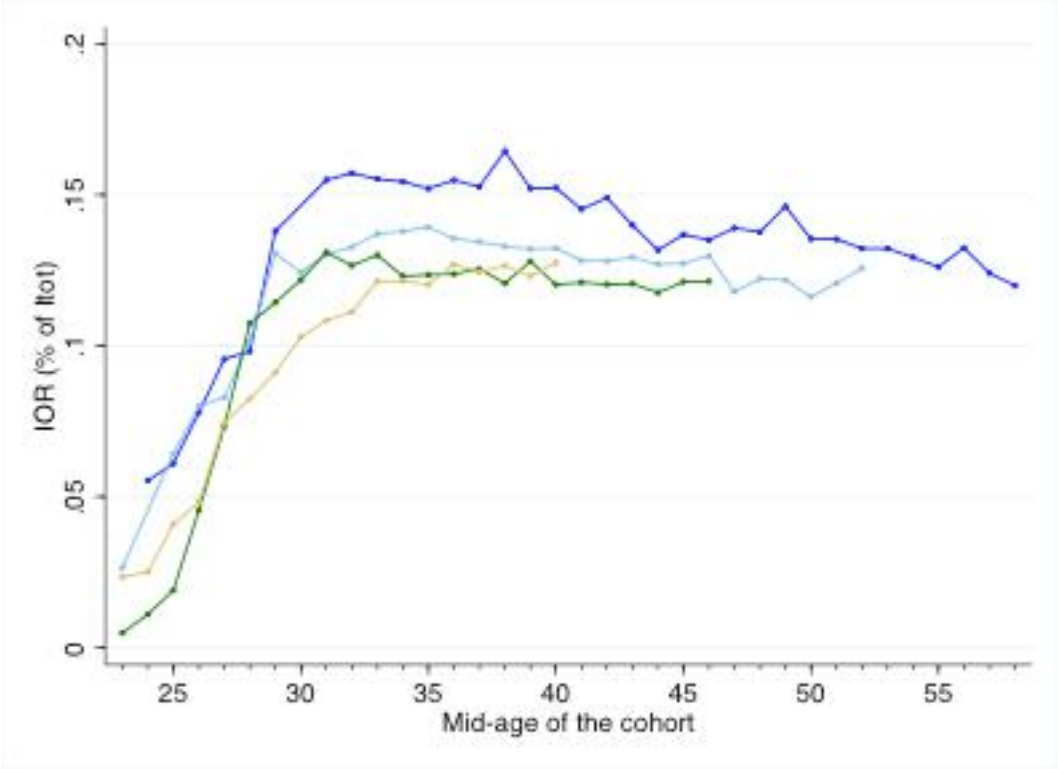




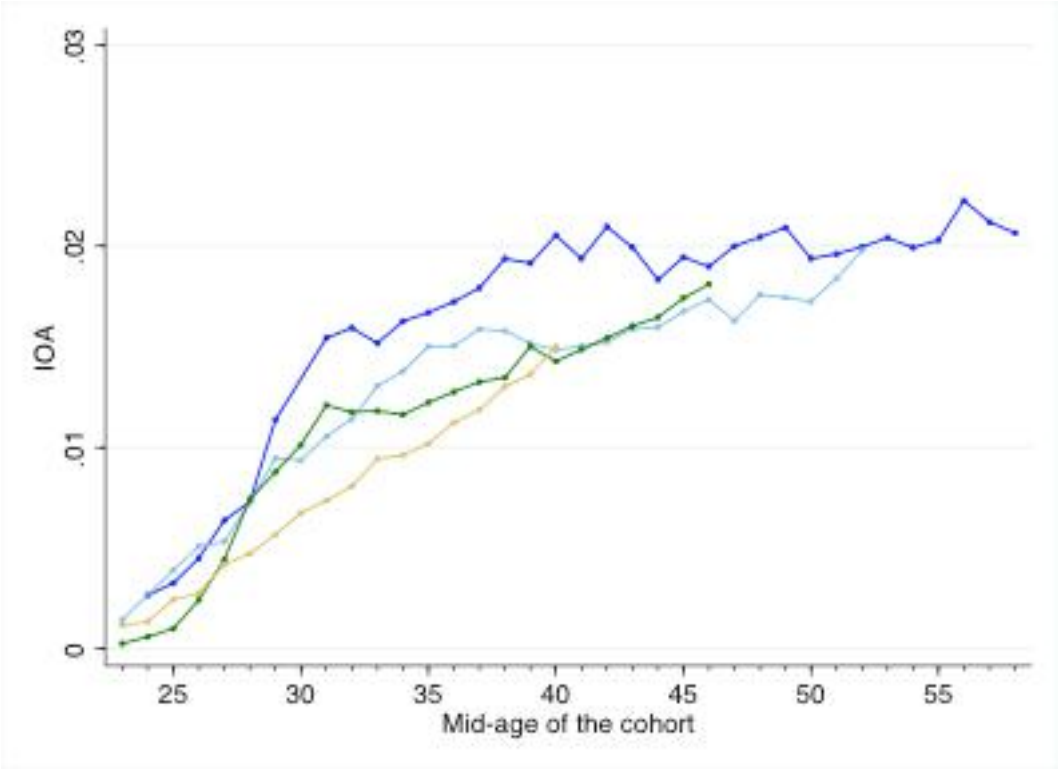
Note: The solid curves show that, among people born in 1974, *IOR* and *IOA* do not change much depending on whether their circumstances are retrieved at age 1 in the 1975 census or at age 16 in the 1990 census. For this comparison exercise, we restrict our sample to the common subsample of people born in 1974 who have all their circumstances observed in both 1975 and 1990. The dashed curves correspond to the original ones from Figure 2 for the C72 cohort, which also includes people born in 1970 and 1974, where circumstances are observed in the 1975 census but are not required to be observed in 1990 as well.

Figure A2.3: Inequality over the life course, by cohort: top 1 percent wages excluded

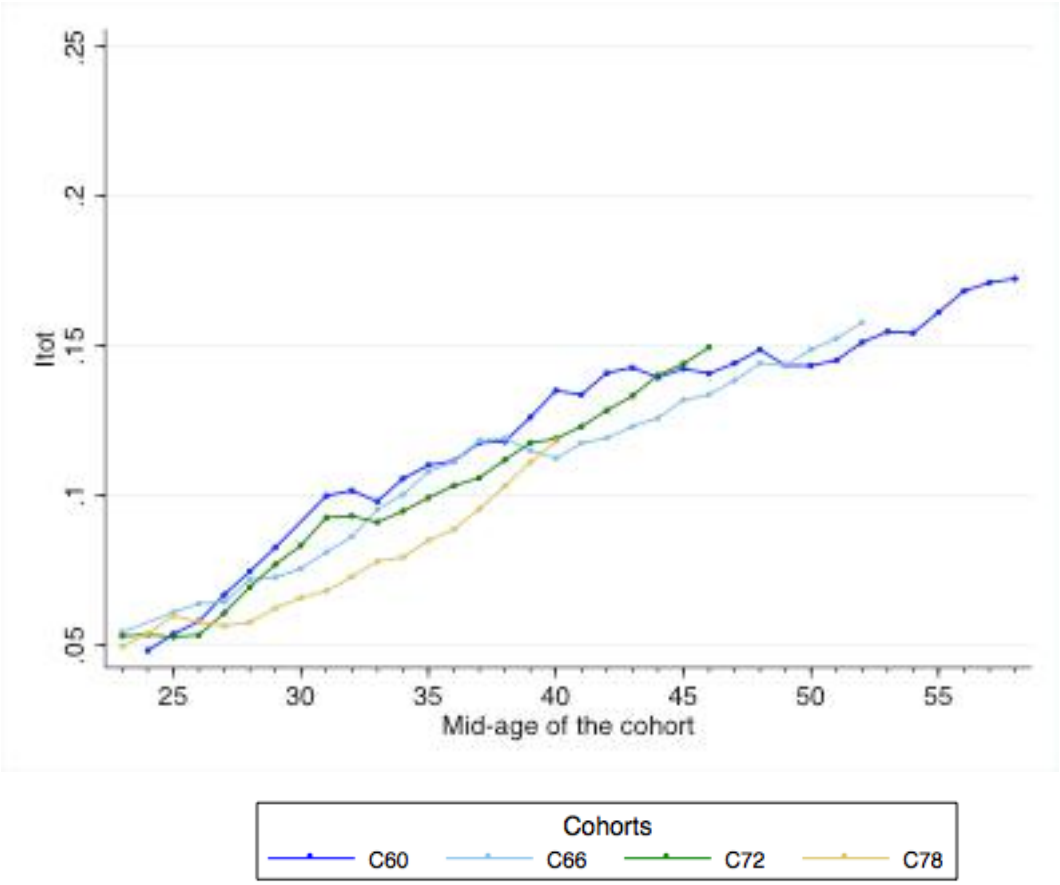
(a) Inequality of opportunity in relative terms (*IOR*)



(b) Inequality of opportunity in absolute terms (*IOA*)



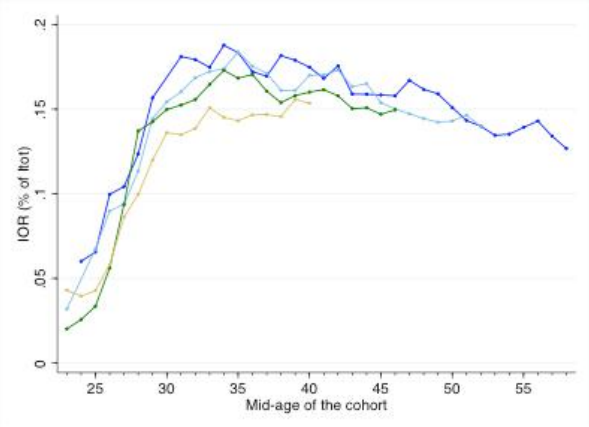
(c) Total earnings inequality (*Itot*)



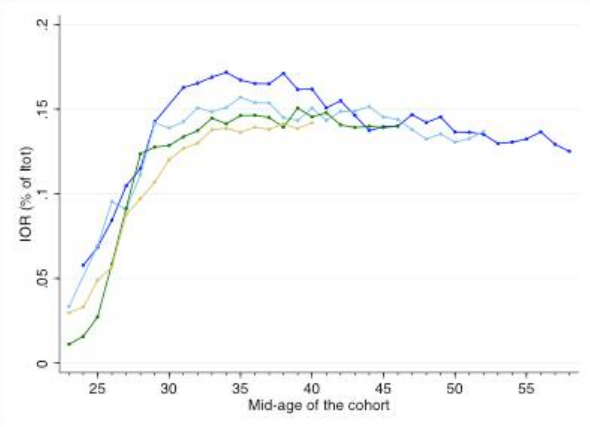
Note: The figure is the same as Figure 2 but with the 99th percentile (shown in Figure A6.1) dropped rather than winsorized before computing inequality. The scale is kept the same as in Figure 2 to ease comparisons with it.

Figure A2.4: Inequality over the life course, by cohort: poor people included

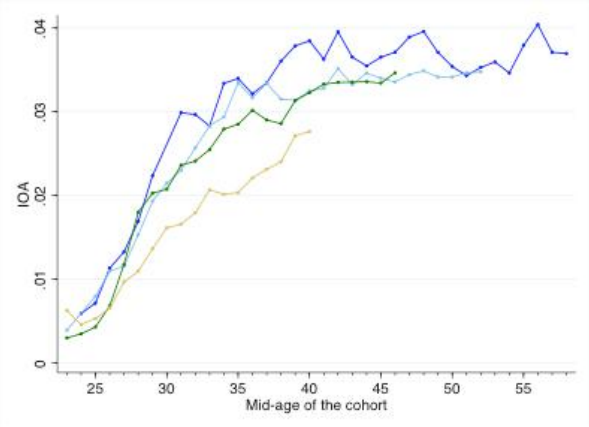
(a1) *IOR*, poor included



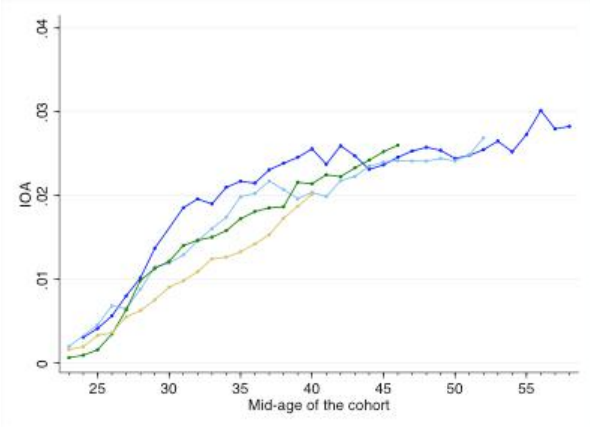
(a2) *IOR*, poor excluded (benchmark)



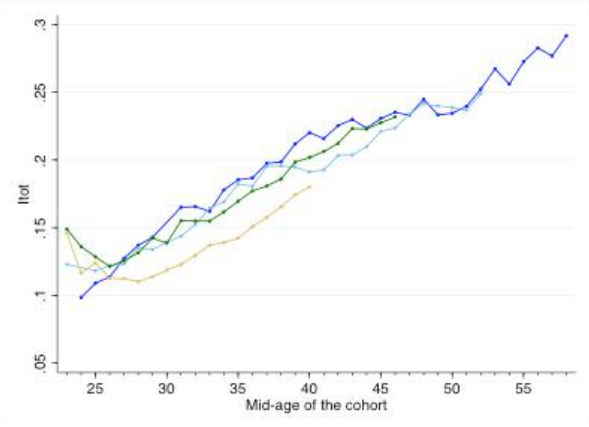
(b1) *IOA*, poor included



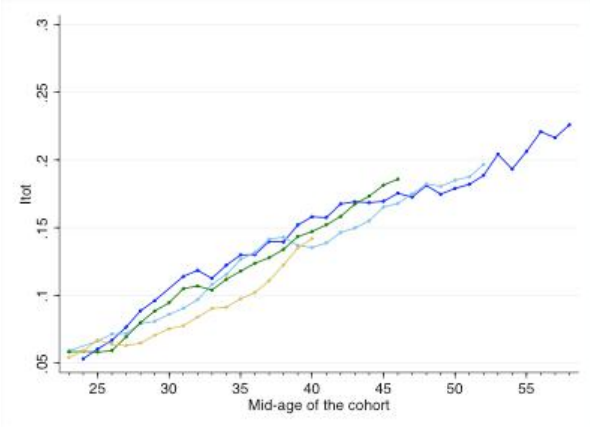
(b2) *IOA*, poor excluded (benchmark)

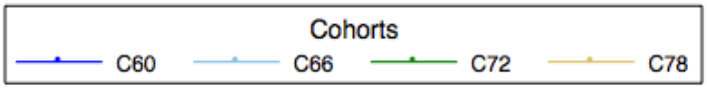


(c1) *Itot*, poor included



(c2) *Itot*, poor excluded (benchmark)

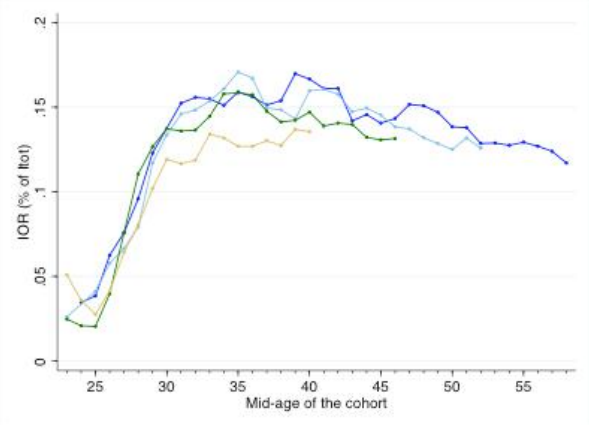




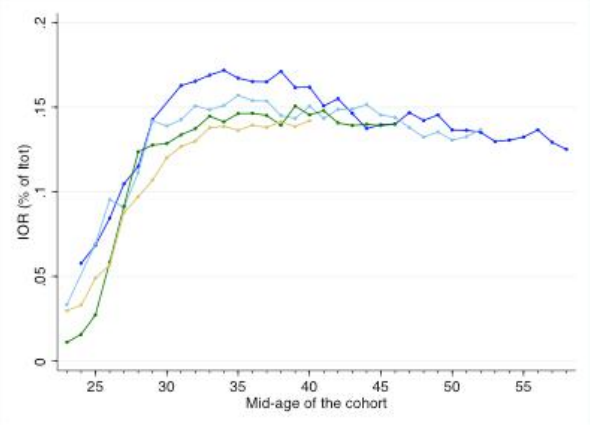
Note: The left panels reproduce Figure 2 but without excluding "poor" individuals, whose full-year equivalent wage is below the legal minimum. The right panels exactly reproduce Figure 2 but use the same y-scale as the left ones for comparison purposes. Including poor individuals increases the maxima of *IOA* and *Itot*, making it impossible to use the smaller y-scale from Figure 2 for the left panels.

Figure A2.5: Inequality over the life course, by cohort: all wage earners included

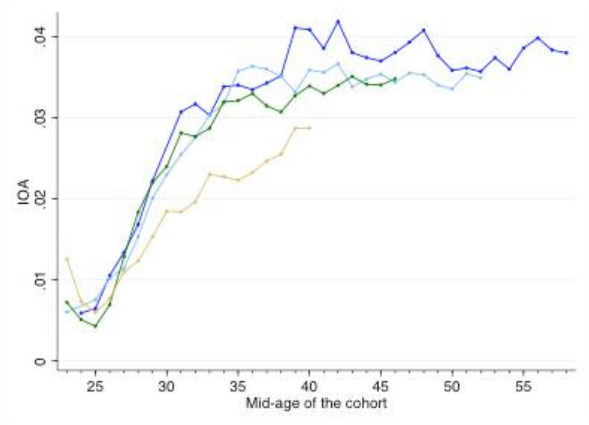
(a1) *IOR*, all



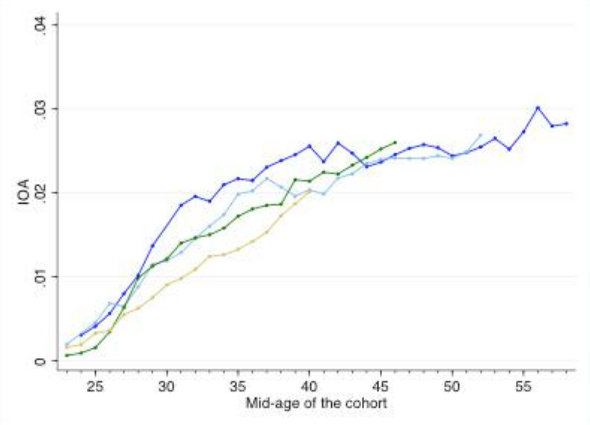
(a2) *IOR*, not all (benchmark)



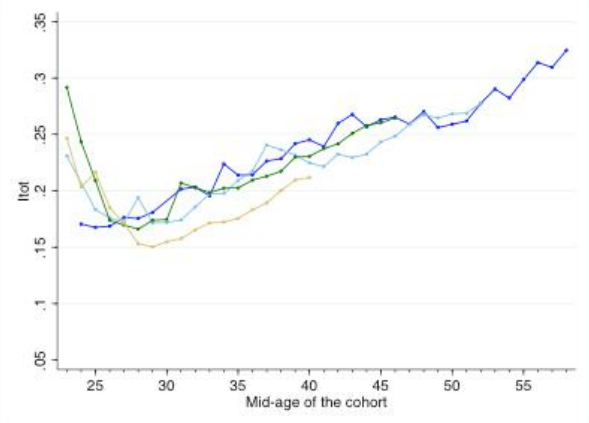
(b1) *IOA*, all



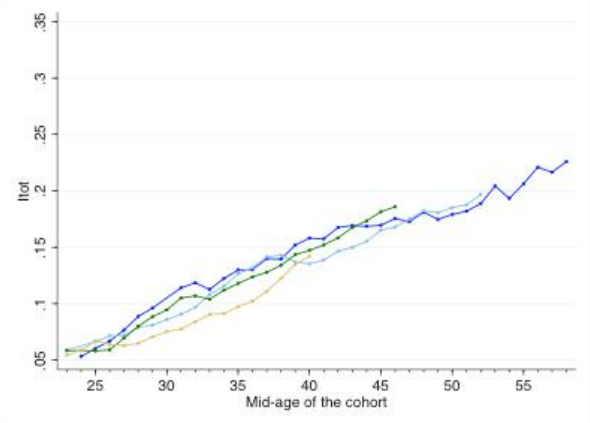
(b2) *IOA*, not all (benchmark)

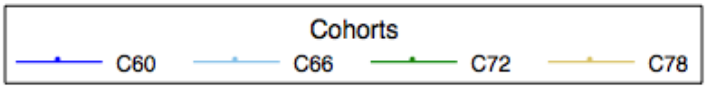


(c1) *Itot*, all



(c2) *Itot*, not all (benchmark)

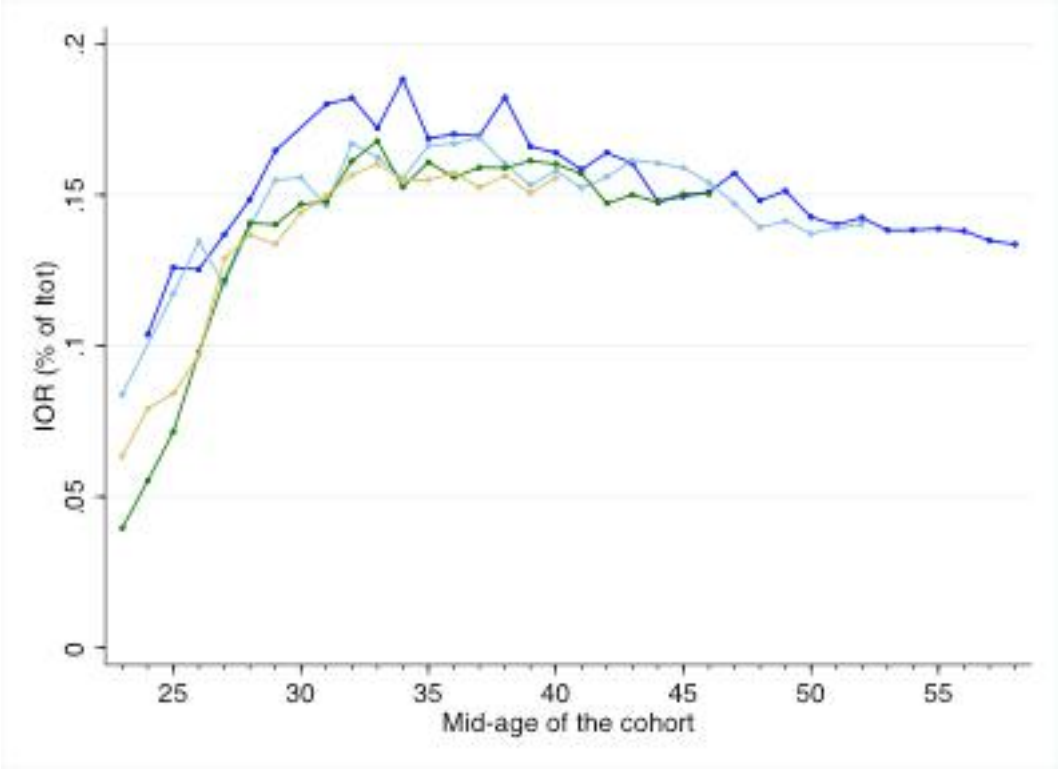




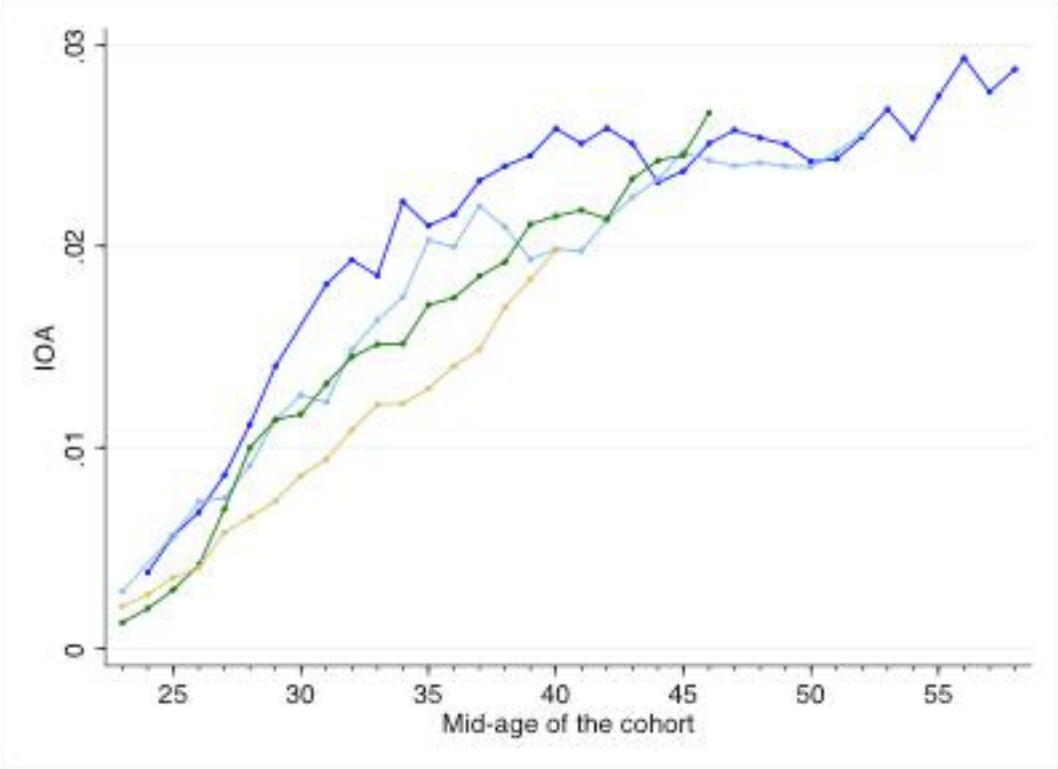
Note: The left panels reproduce Figure 2 but include all wage earners from the private nonagricultural sector, after casewise deletion on circumstances. That is, they do not exclude "poor" individuals, whose full-year equivalent wage is below the legal minimum, nor those with less than one hundred paid days during the year. The right panels exactly reproduce Figure 2 but use the same y-scale as the left ones for comparison purposes. Including all wage earners increases the maxima of *IOA* and *Itot*, making it impossible to use the smaller y-scale from Figure 2 for the left panels.

Figure A2.6: Inequality over the life course, by cohort: full-year equivalent wage

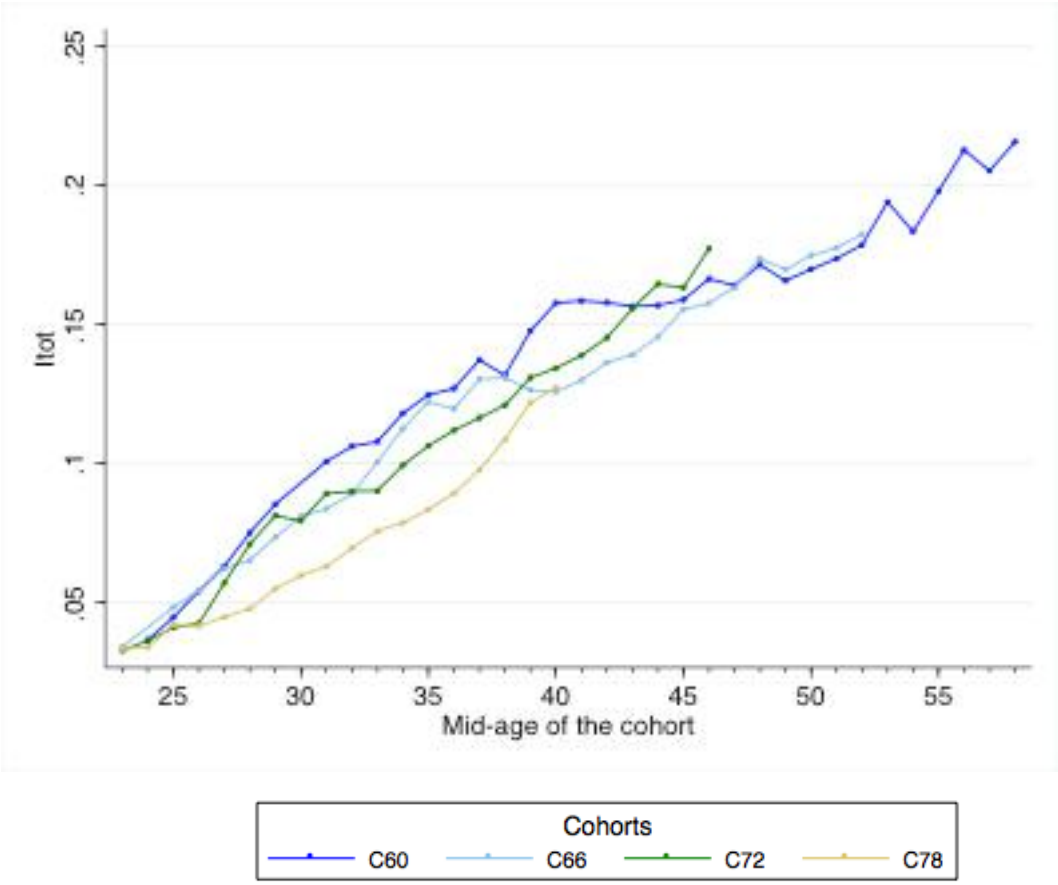
(a) Inequality of opportunity in relative terms (*IOR*)



(b) Inequality of opportunity in absolute terms (*IOA*)



(c) Total earnings inequality (*Itot*)

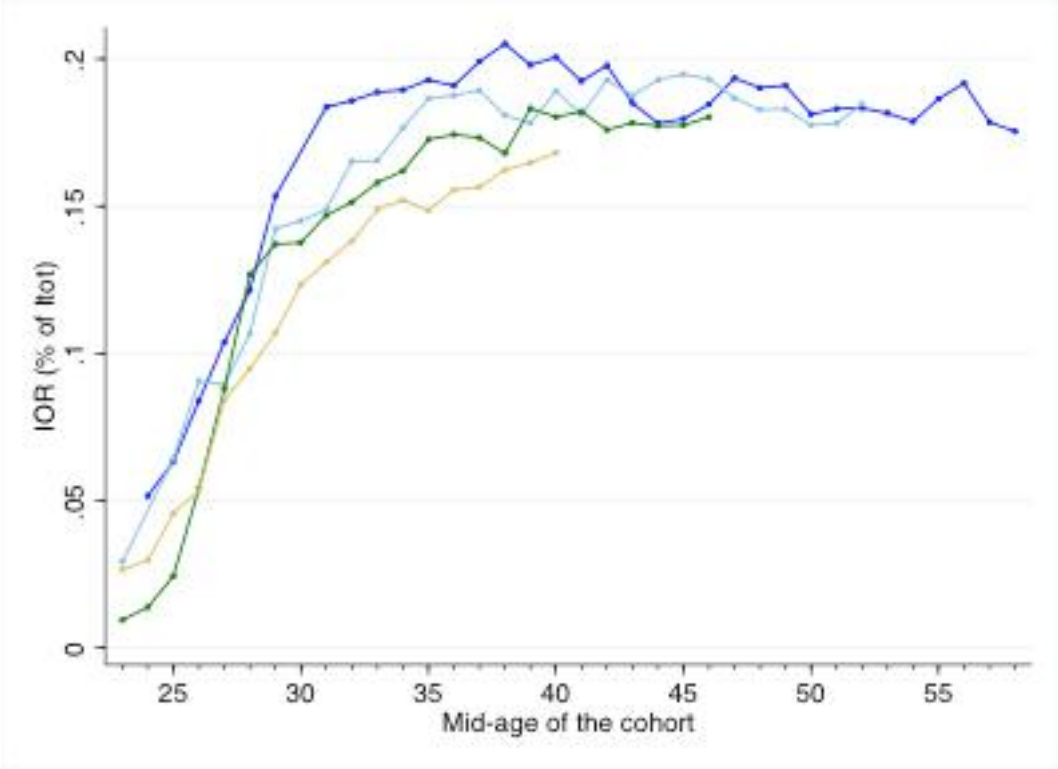


Note: The figure is the same as Figure 2 but is computed over full year equivalent wages. This means that we transform our original individual wage variable, w_{it} in eq. (1), as follows before proceeding to the computations from eq. (1) to (3): w_{it} is divided by the number of paid days during the year and multiplied by 360, which is the maximal number of paid days in the payroll data. The y-scale differs from that of Figure 2 for *IOA* and *Itot*, but the difference is marginal and should not hinder any comparisons with Figure 2.

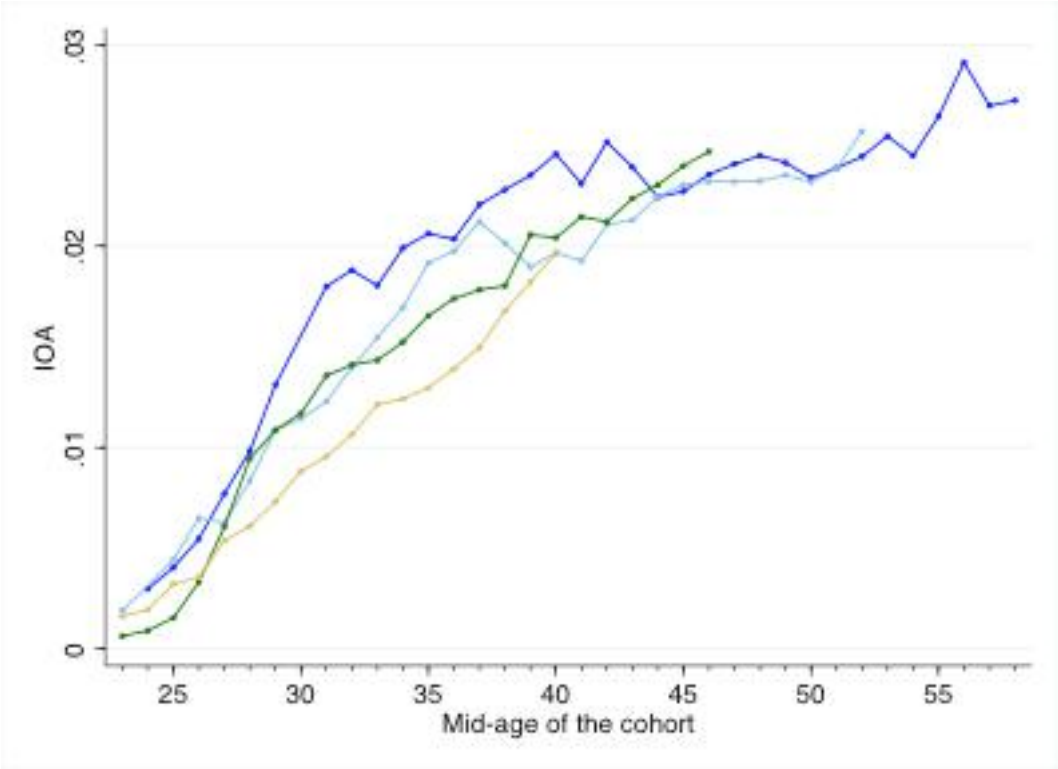
Comment: The figure is similar to Figure 2, which indicates that our sample definition in Figure 2 already neutralizes inequality due differential of length of work quite well. Unfortunately, we do not have access to hourly wages over the entire study period, which would fully neutralize the length of work. We note however that *IOR* at very young ages is higher here than in Figure 2. One explanation is that young adults with favorable circumstances pursue longer studies than others, which makes them more likely to be in a transition year from their studies to their working life.

Figure A2.7: Inequality over the life course, by cohort: *GE0* index

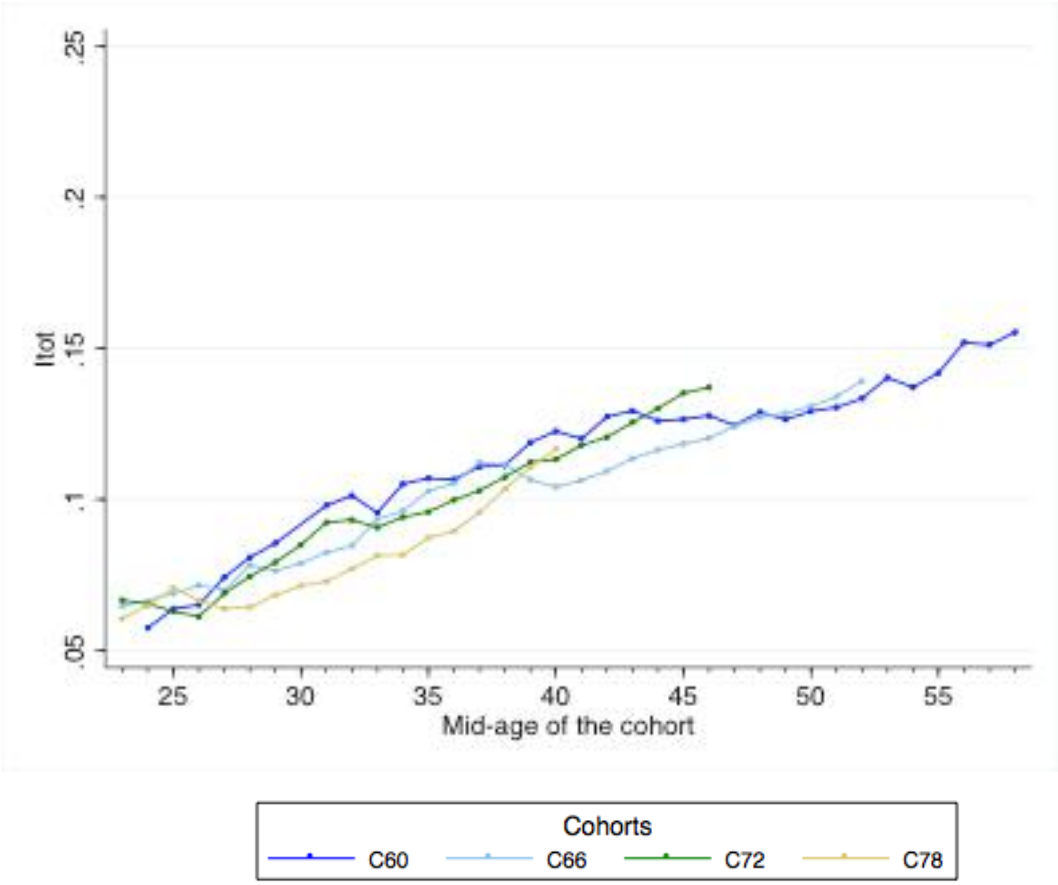
(a) Inequality of opportunity in relative terms (*IOR*)



(b) Inequality of opportunity in absolute terms (*IOA*)



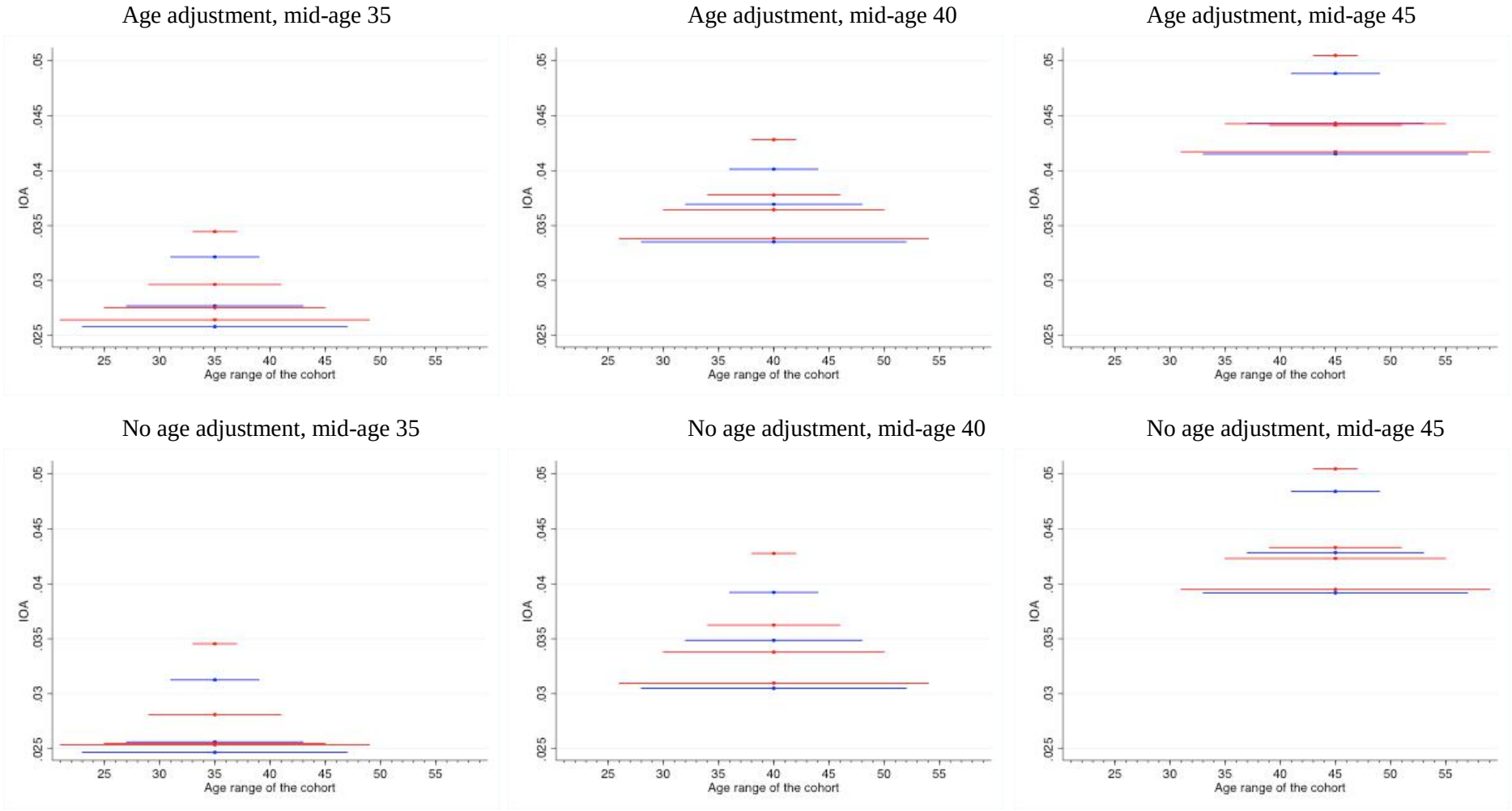
(c) Total earnings inequality (*Itot*)



Note: The figure is the same as Figure 2 but relies on the *GE0* inequality index instead of the *GE2* index. For *IOA*, the *GE0* index is applied directly over the predicted wages from eq. (3), as we face no negative predicted values. The y-scale differs from that of Figure 2 for *IOR* and *IOA*, but the difference is marginal and should not hinder any comparisons with Figure 2.

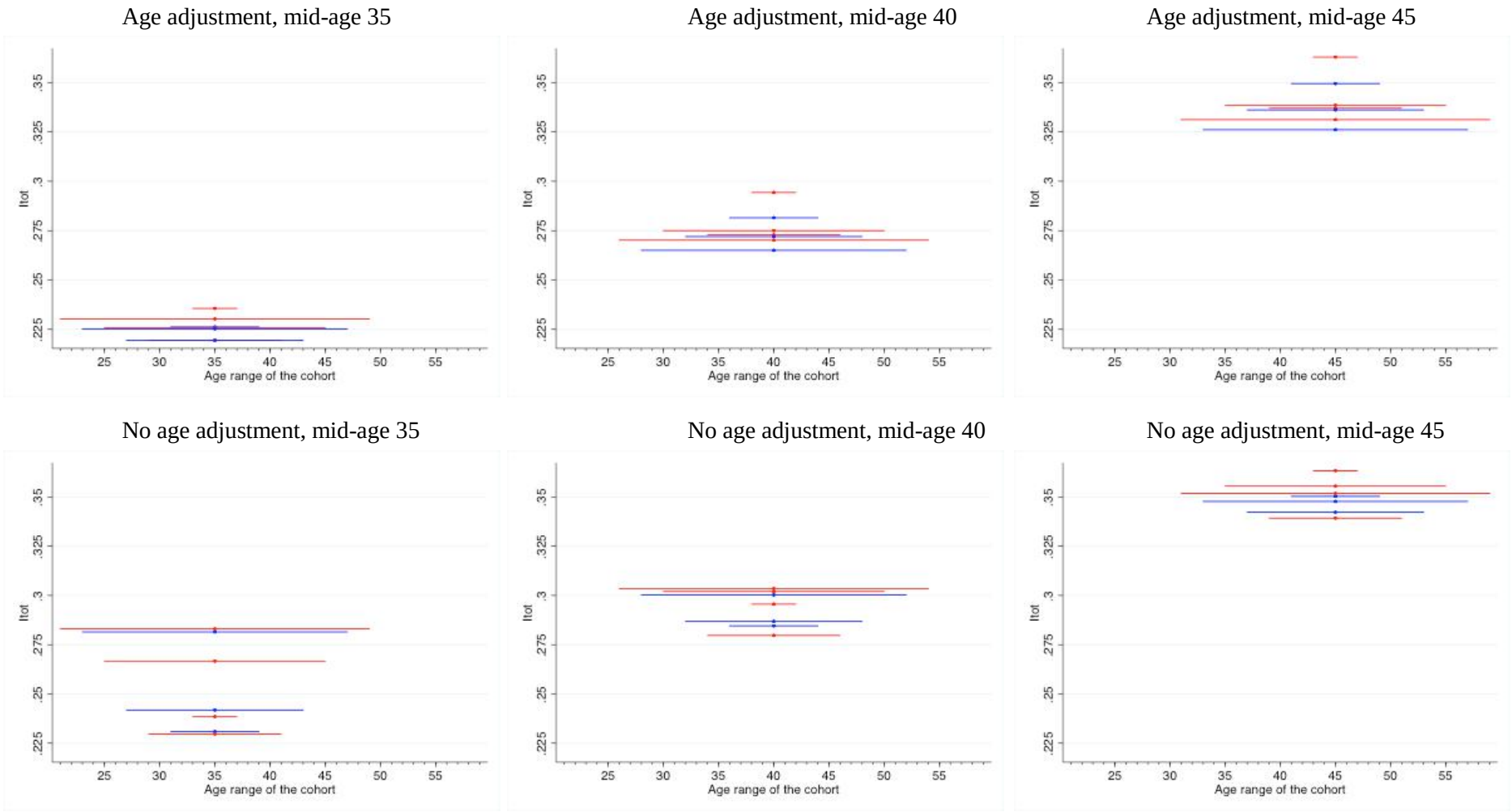
III. Variations on Figure 3

Figure A3.1: Comparability of *IOA* over selected age ranges



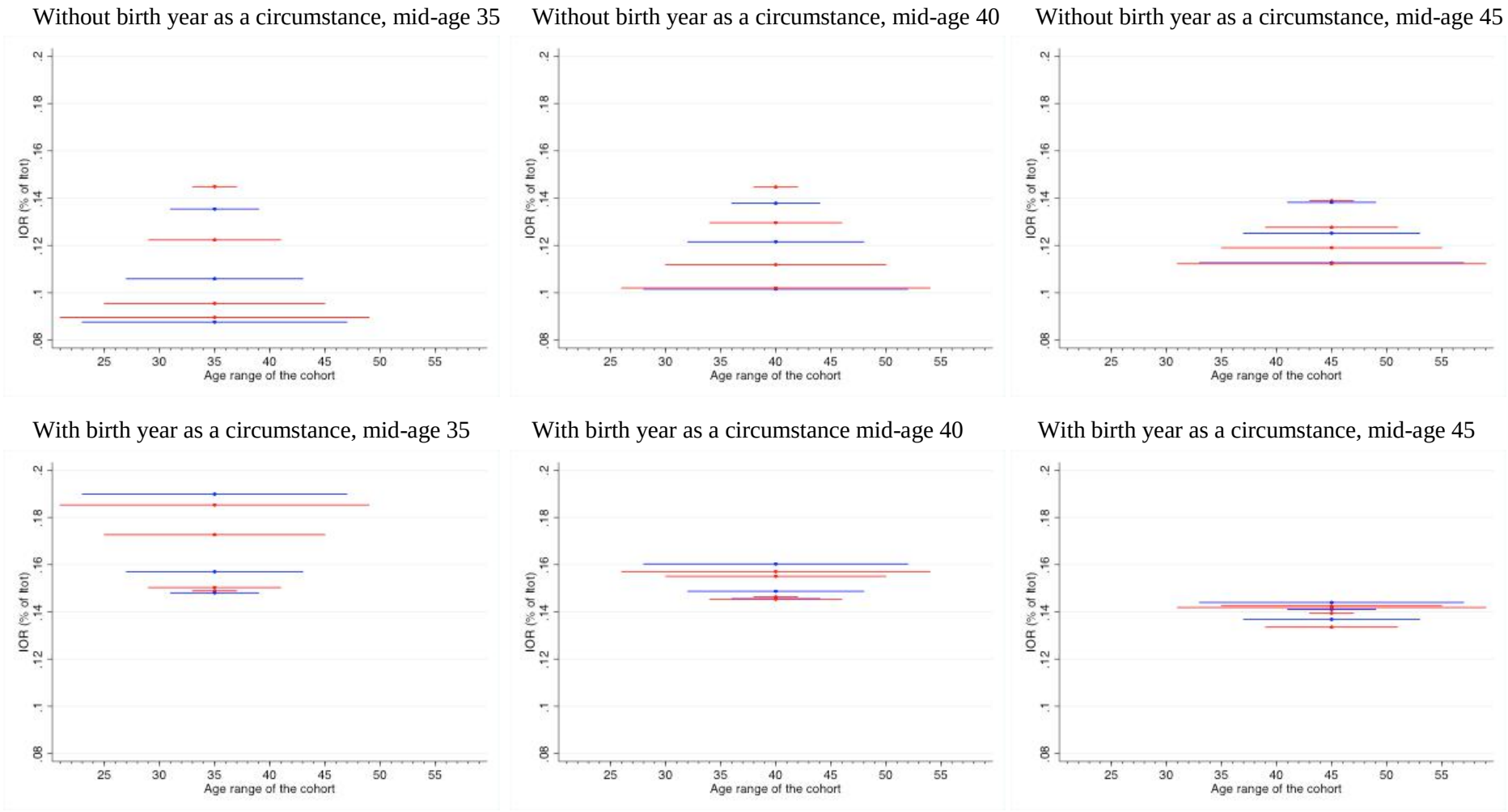
Note: The figure is the same as Figure 3 but with *IOA* instead of *IOR*.

Figure A3.2: Comparability of *Itot* over selected age ranges



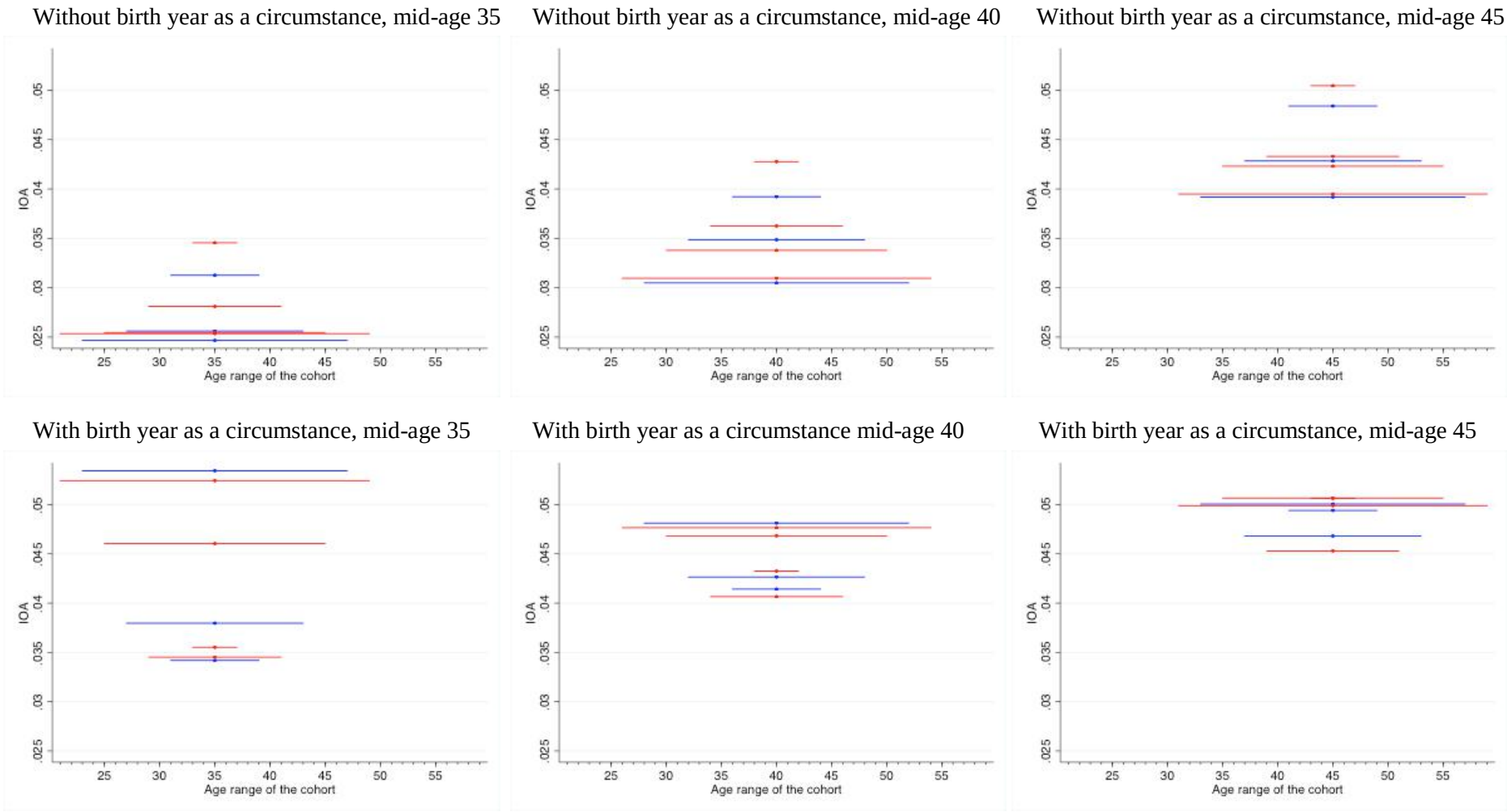
Note: The figure is the same as Figure 3 but with *Itot* (computed as well after casewise deletion on circumstances) instead of *IOR*.

Figure A3.3: Comparability of *IOR* over selected age ranges: treating birth year as a circumstance (no age adjustment)



Note: The figure compares *IOR* over selected age ranges when birth year dummies are treated as an additional circumstance (bottom panels) or not (top panels). The age adjustment used throughout the paper is not applied in any panel, as the birth year would only bring noise to the estimates from the bottom panels. The top panels without age as a circumstance exactly reproduce the bottom panels from Figure 3, but with a larger scale for the y-axis.

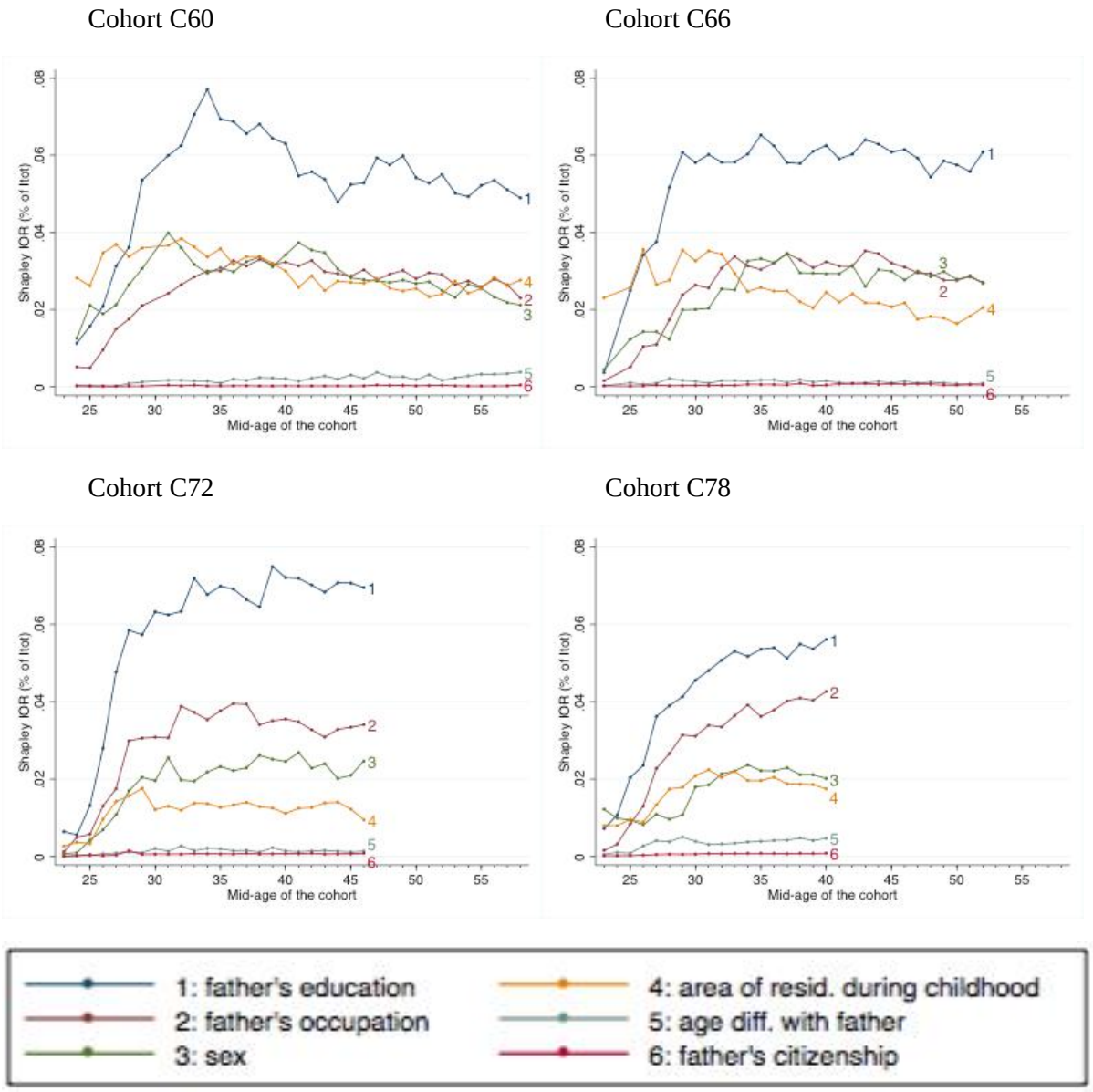
Figure A3.4: Comparability of *IOA* over selected age ranges: treating birth year as a circumstance (no age adjustment)



Note: The figure compares *IOA* over selected age ranges when birth year dummies are treated as an additional circumstance (bottom panels) or not (top panels). The age adjustment used throughout the paper is not applied in any panel, as the birth year would only bring noise to the estimates from the bottom panels. The top panels without age as a circumstance exactly reproduce the bottom panels from Figure A3.1, but with a larger scale for the y-axis.

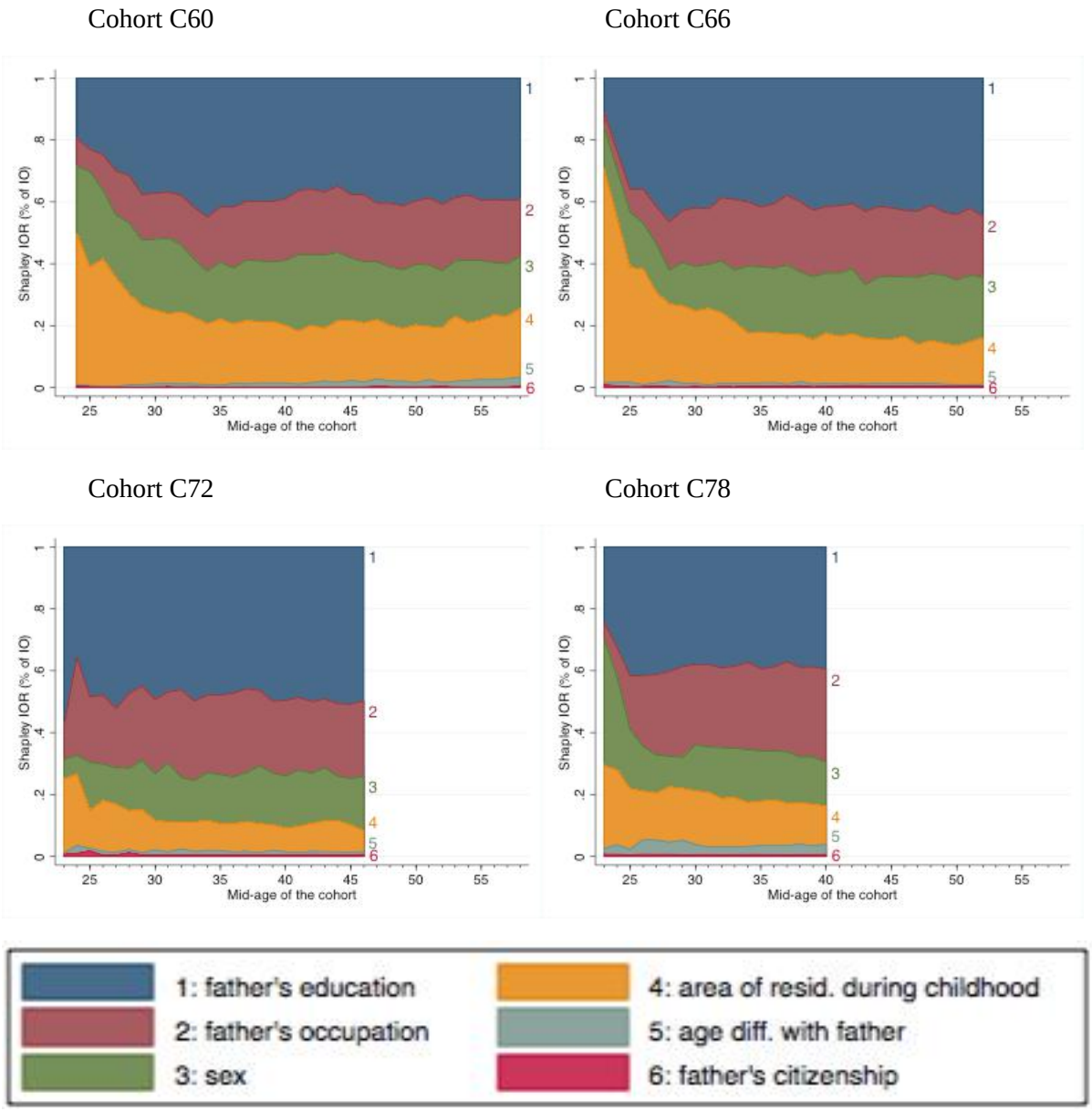
IV. Variations on Figure 4

Figure A.4.1: Shapley decomposition of *IOR* by circumstances: curves



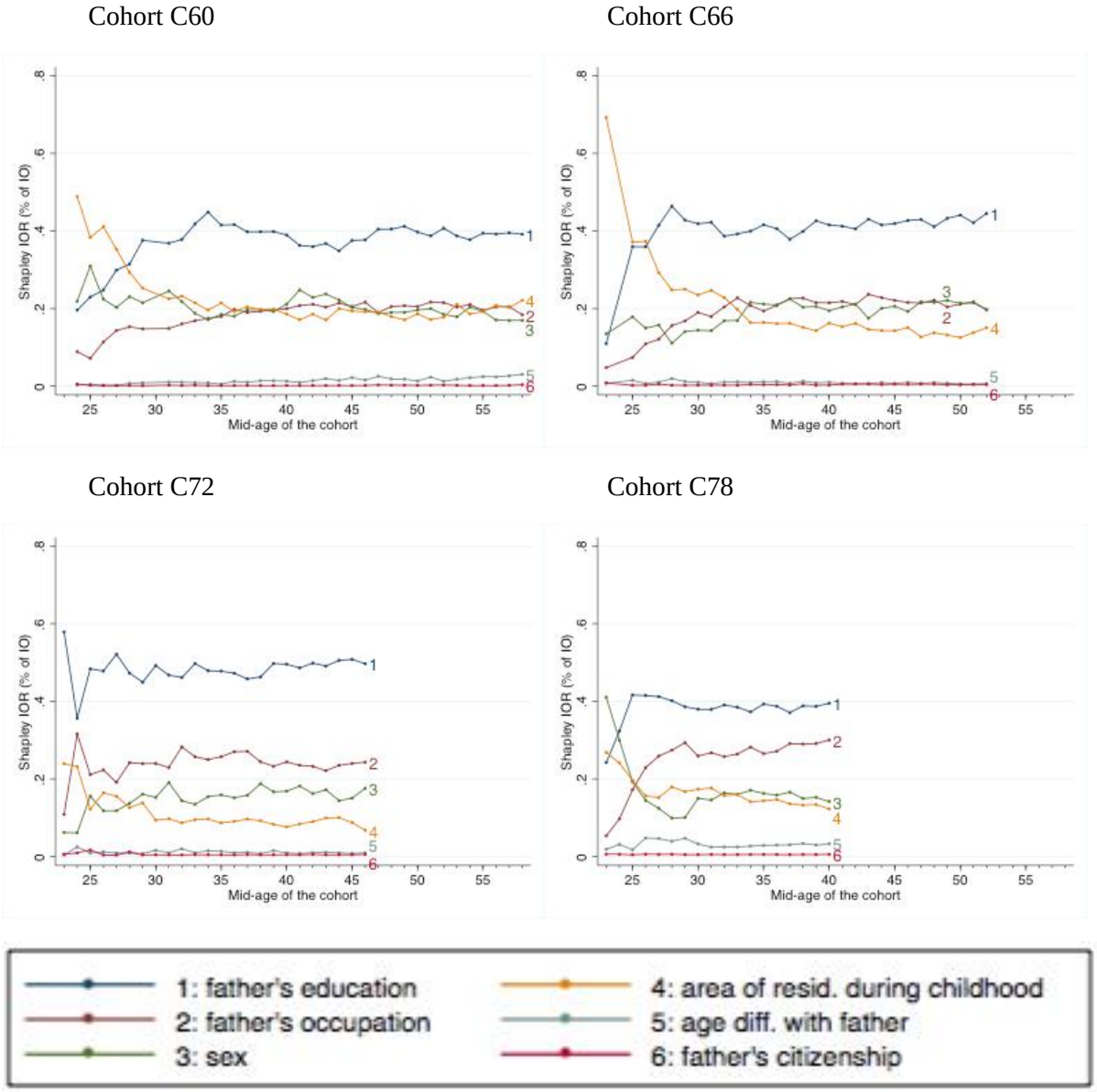
Note: The figure shows the same Shapley values as Figure 4, but in the form of curves rather than cumulative areas, making it easier to follow the evolution of each Shapley value.

Figure A4.2: Shapley decomposition of *IOR* by circumstances: in percentage of IO



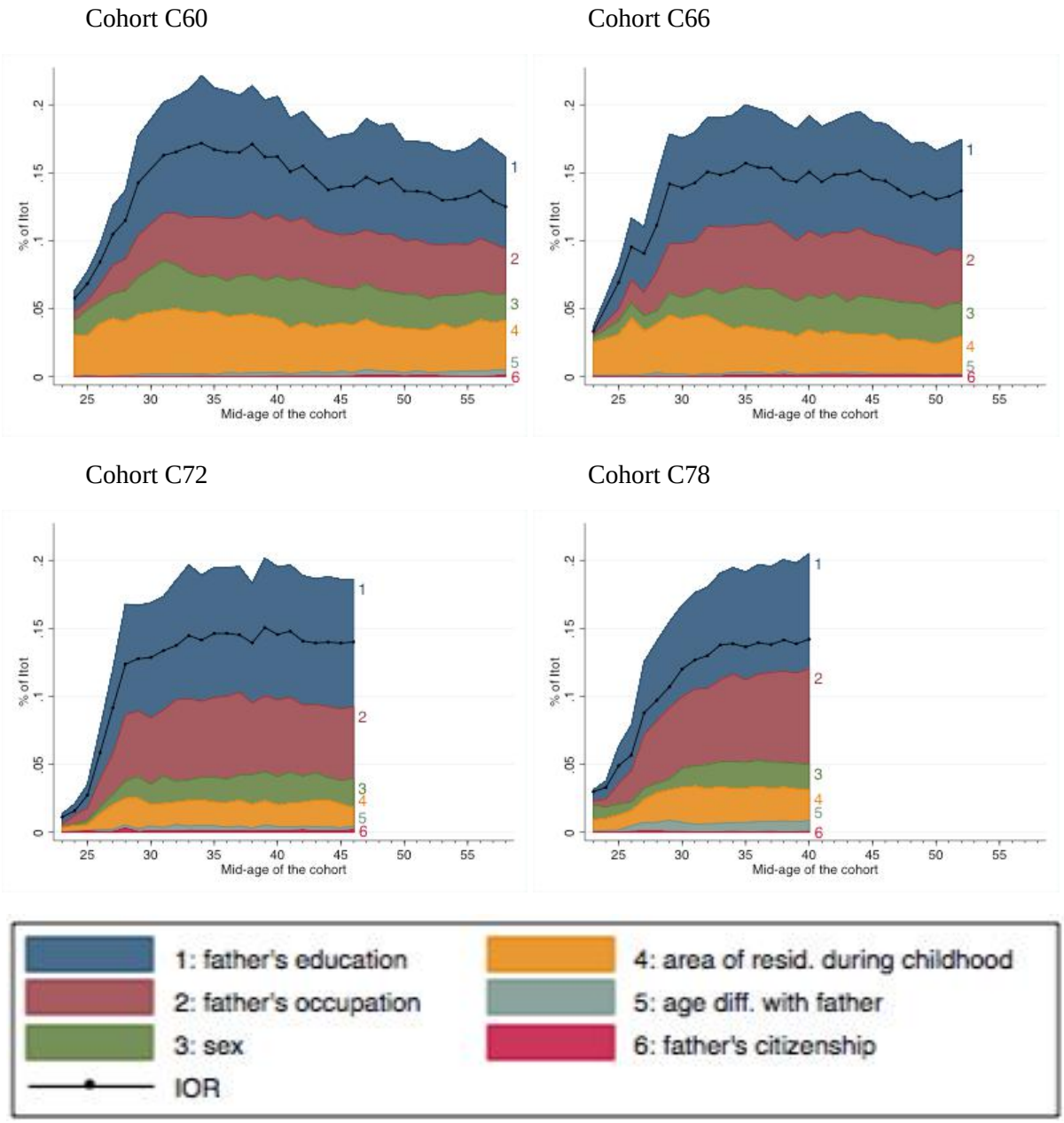
Note: The figure shows the same Shapley values as Figure 4, but in percentage of IO (be it *IOR* or *IOA*) rather than in percentage of *Itot*. We display the Shapley values in the same order from bottom to top as in Figure 4, although the place of residence during childhood has a higher average Shapley value than sex in percentage of IO (18.3 versus 18.1 percent of IO across our 105 year-cohort observations, respectively). The converse is true in percentage of *Itot* (2.2 versus 2.3 percent of *Itot*). This discrepancy arises because the place of residence during childhood has high Shapley value relative to the other circumstances at the beginning of adulthood, when IO itself is low.

Figure A.4.3: Shapley decomposition of IOR by circumstances: in perc. of IO, curves



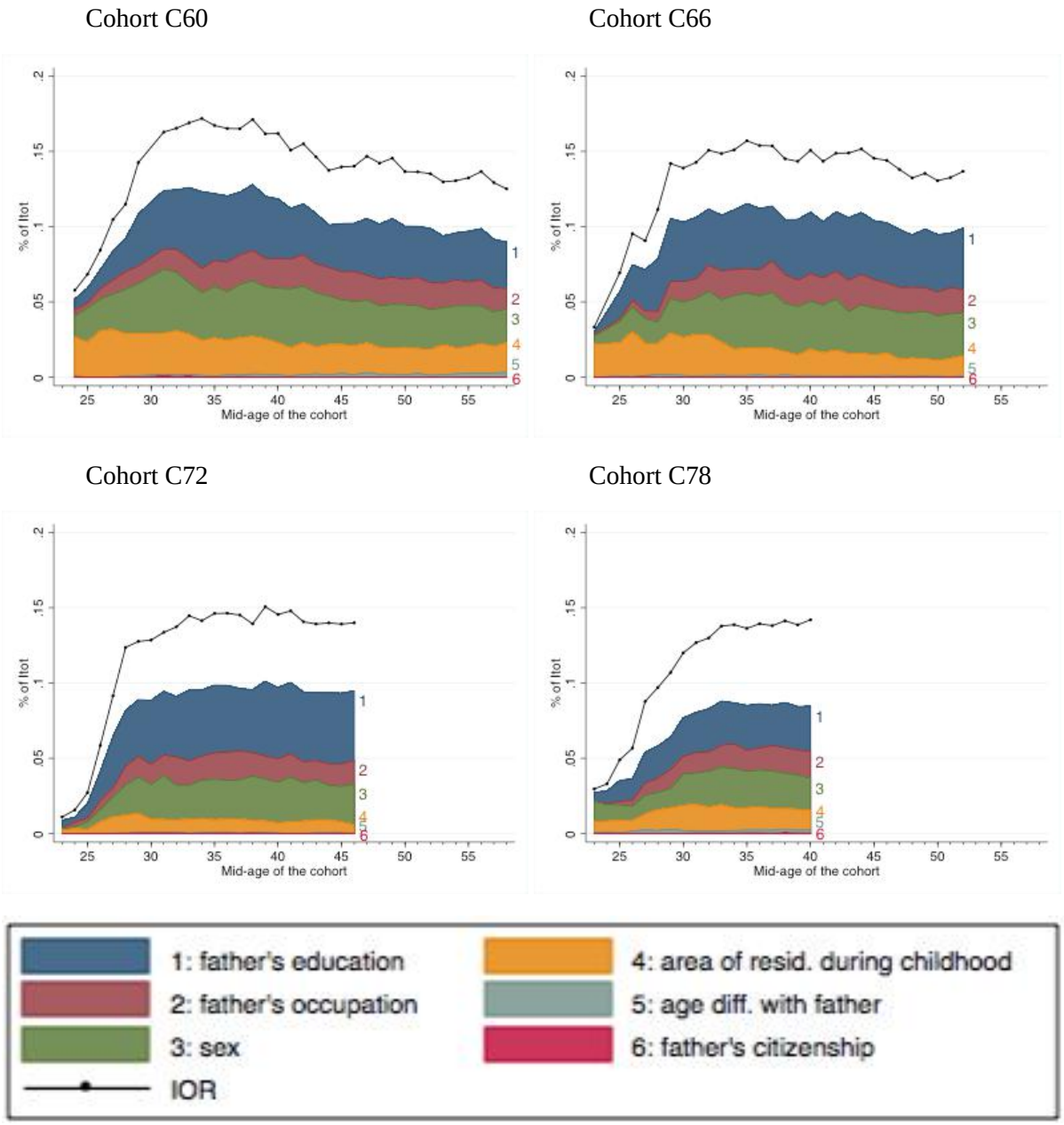
Note: The figure shows the same Shapley values as Figure A4.2 (in percentage IO), but in the form of curves rather than cumulative areas, making it easier to follow the evolution of each Shapley value.

Figure A4.4: Marginal contributions of circumstances: simple linear models



Note: The figure shows in shaded areas the R-squared values obtained from the simple linear regressions with one circumstance only (after casewise deletion on all). These R-squared are also marginal contributions since removing the circumstance from each of these simple linear regressions yields an R-squared of zero. We display them in the same order from bottom to top as in Figure 4, although the place of residence during childhood has a higher average R-squared value than sex (2.8 versus 2.0 percent of *Itot* across our 105 year-cohort observations, respectively).The converse is true in Figure 4 about Shapley values. This discrepancy arises because an unconditional analysis, as the one presented here, understates the relative importance of sex. We explain why in subsection 5.2, as well as why the simple linear R-squared values add up to more than the R-squared from our full model (3). This can be seen from the black curves, which lie below the upper contour of the overall shaded areas and which reproduce the *IOR* curves from our benchmark Figure 2.

Figure A4.5: Marginal contributions of circumstances: full multiple linear model

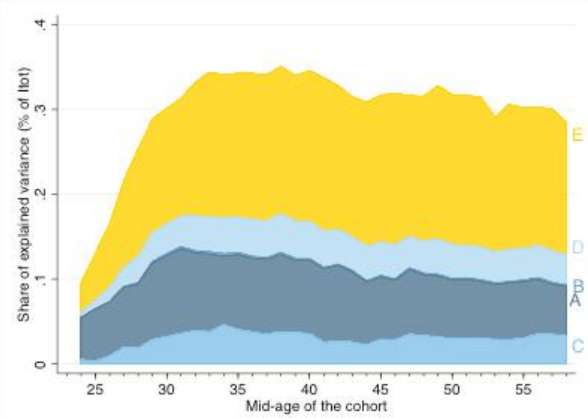


Note: The figure shows in shaded areas the marginal contributions of circumstances to the R-squared our full model (3), which includes all circumstances. These contributions are obtained as the R-squared from the full model minus that from the same model but which excludes the circumstances under scrutiny. We display them in the same order from bottom to top as in Figure 4, although their average values (across our 105 year-cohort observations) rank differently than that of the Shapley values from Figure 4: sex now ranks second instead of third, area of residence during childhood now ranks third instead of fourth, and father's occupation now ranks fourth instead of second. As we explain in subsection 5.2, the reason for this discrepancy is that a fully partial analysis, as the one presented here, overstates the relative importance of sex, which correlates little with the other circumstances, and understates that of father's occupation, which correlates a lot with them (with father's education in particular). We also explain in subsection 5.2 why these marginal contributions add up to less than the R-squared of the full model (3). This can be seen from the black curves, which lie above the upper contour of the overall shaded areas and which reproduce the *IOR* curves from our benchmark Figure 2.

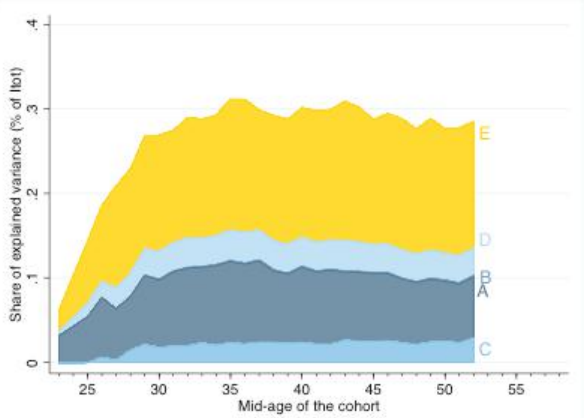
V. Variations on Figure 5

Figure A5.1: Shares of variance explained by circumstances and education: quaternary education

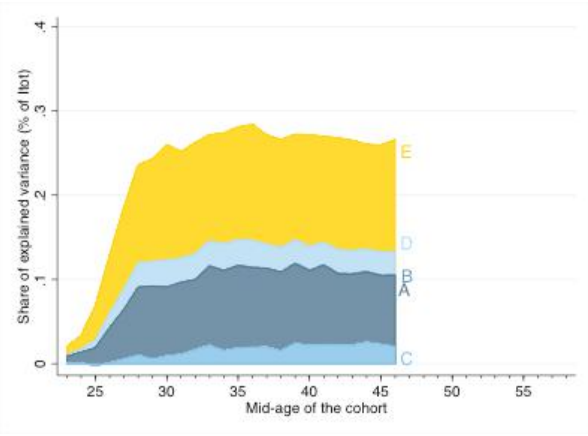
Cohort C60

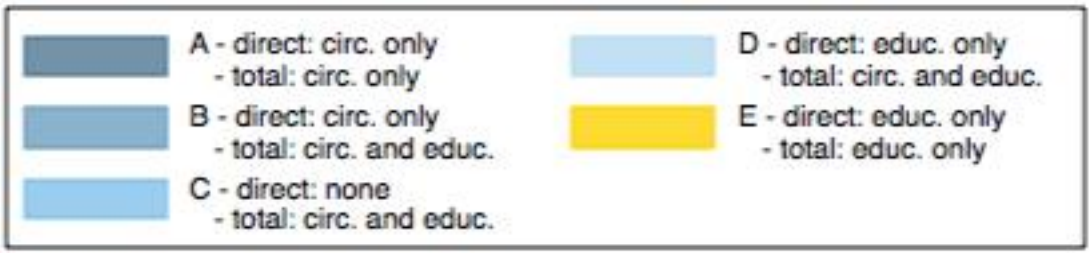


Cohort C66



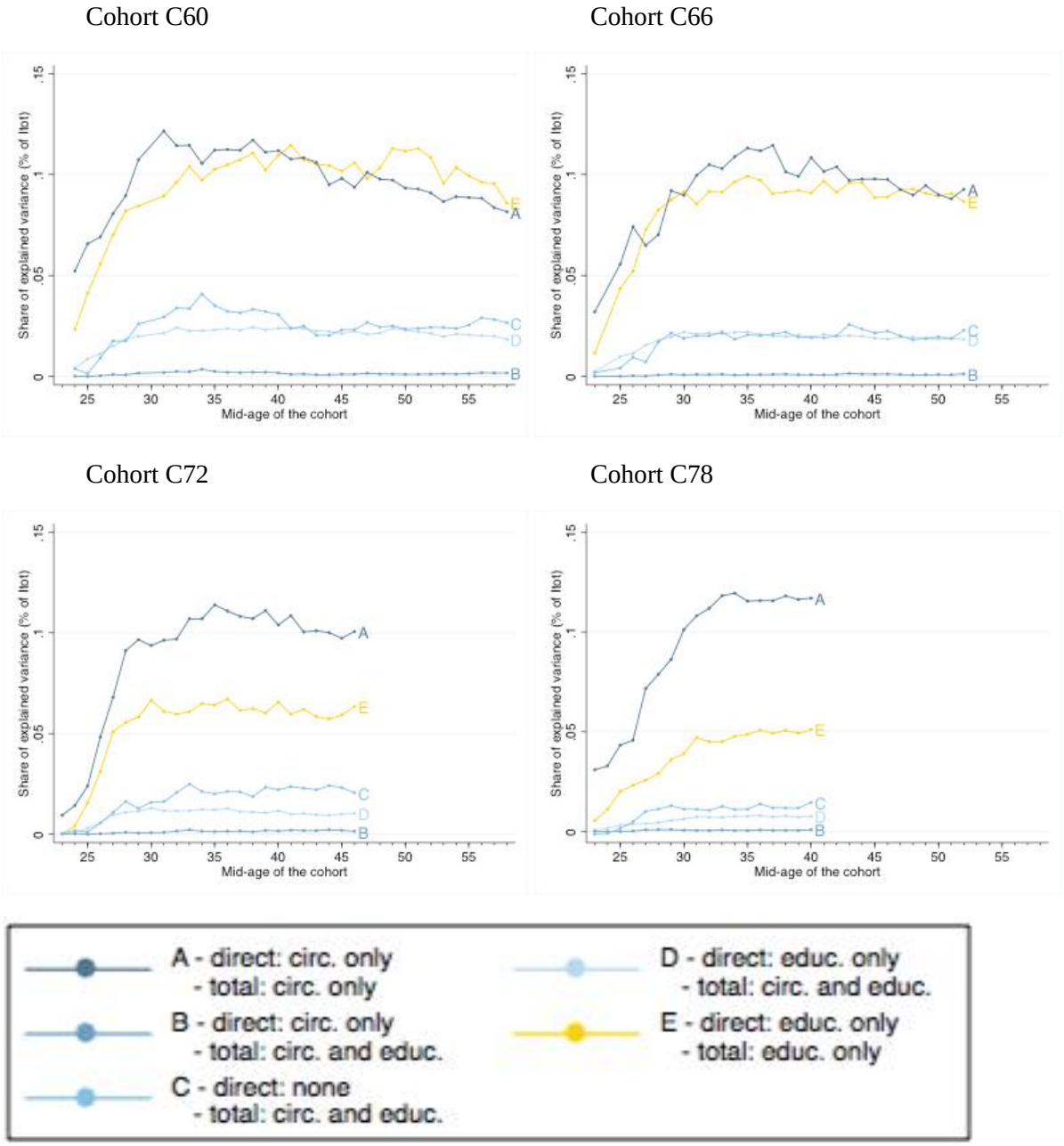
Cohort C72





Note: The figure reproduces Figure 5 for the first three cohorts with own education coded into four instead of two categories.

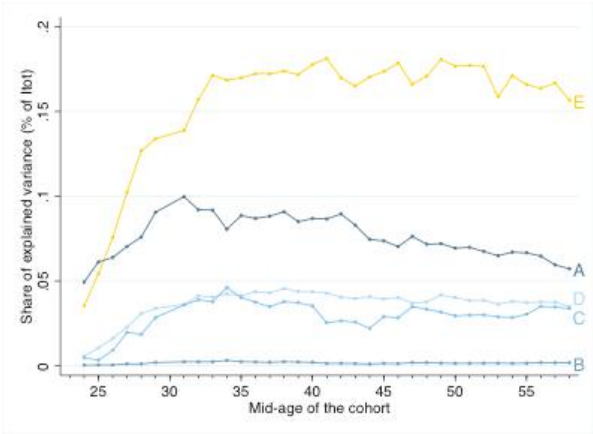
Figure A5.2: Shares of variance explained by circumstances and education: curves



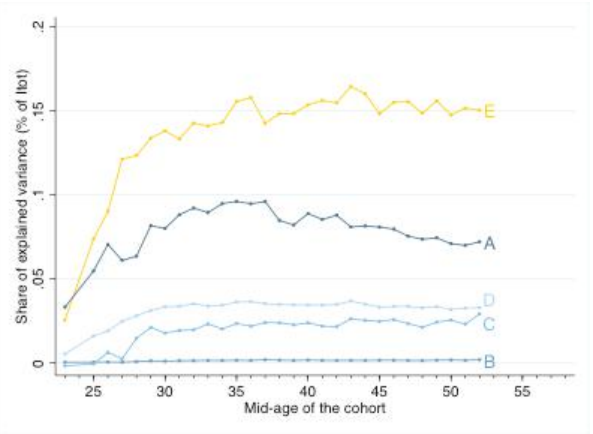
Note: The figure shows the same components *A* though *E* as Figure 5, but in the form of curves rather than cumulative areas, making it easier to follow the evolution of each.

Figure A5.3: Shares of variance explained by circumstances and education: quaternary education, curves

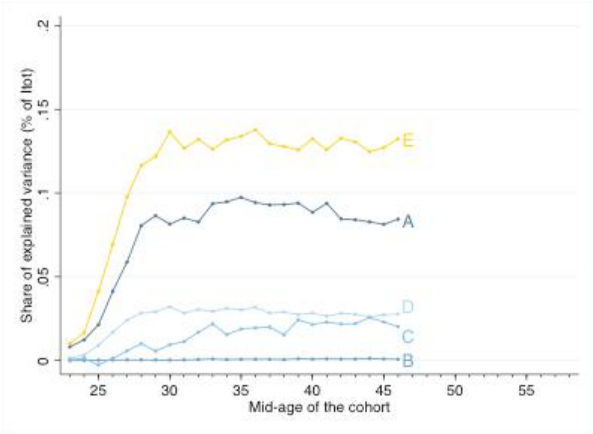
Cohort C60

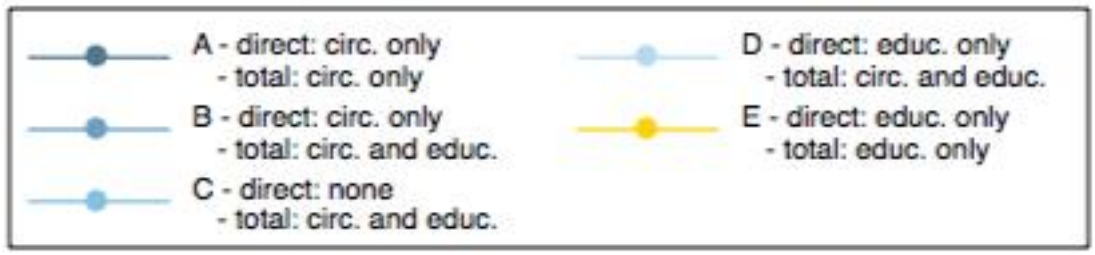


Cohort C66



Cohort C72





Note: The figure shows the same components *A* though *E* as Figure A5.1 (with own education coded into four categories) but in the form of curves rather than cumulative areas, making it easier to follow the evolution of each.

VI. Further data description

Table A6.1: Coding of father's occupation

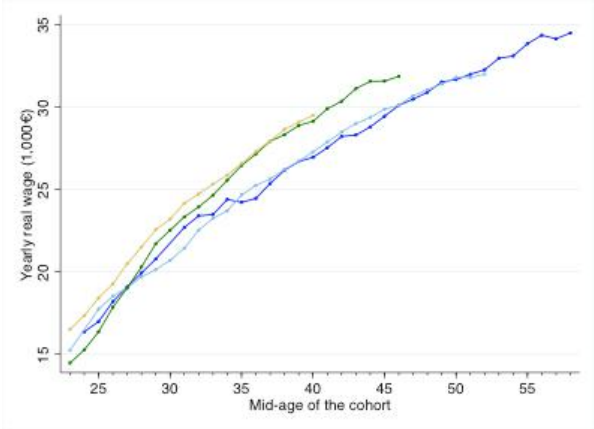
Employment status Socio-professional category	Employers and self-employed workers [1] *Employeurs et travailleurs indépendants*	Salaried workers [2] *Salariés*	Unemployed persons *Chômeurs*	Inactives *Inactifs*
Artisan entrepreneurs, shopkeepers, and business owners *Artisans, commerçants et chefs d'entreprise*	High	n.a.	High	n.a.
Executives and higher intellectual professions *Cadres et professions intellectuelles supérieures*	High	High	High	n.a.
Farm operators *Agriculteurs exploitants*	High	n.a.	High	n.a.
Agricultural workers *Salariés agricoles*	n.a.	Low	Low	n.a.
Manual workers *Ouvriers*	n.a	Low	Low	n.a.
Domestic workers *Personnels de service*	High	Low	Low	n.a.
Intermediate occupations *Professions intermédiaires*	High	Medium & others	Medium & others	n.a.
Employees *Employés*	n.a.	Medium & others	Medium & others	n.a.
Artists, clergy, army and police *Artistes, clergé, armée et police*	High	Medium & others	Medium & others	n.a.
Former workers *Anciens actifs*	n.a.	n.a.	n.a.	Medium & others
Other inactives *Autres inactifs*	n.a.	n.a.	n.a.	Medium & others

[1] Including family workers. *Y compris aides familiaux.*

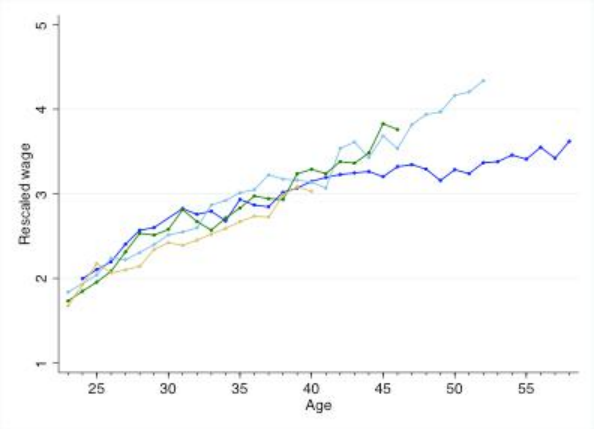
[2] Including workers under a training contract (e.g. apprentices, interns, and subsidized contracts for youth employment).

Figure A6.1: Average and top wages over the life course

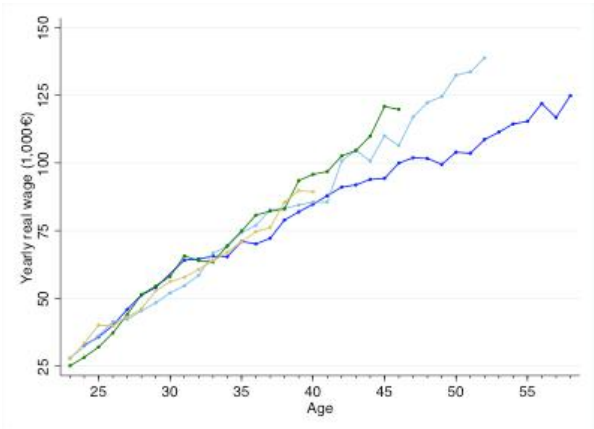
(a) Average wage, not rescaled



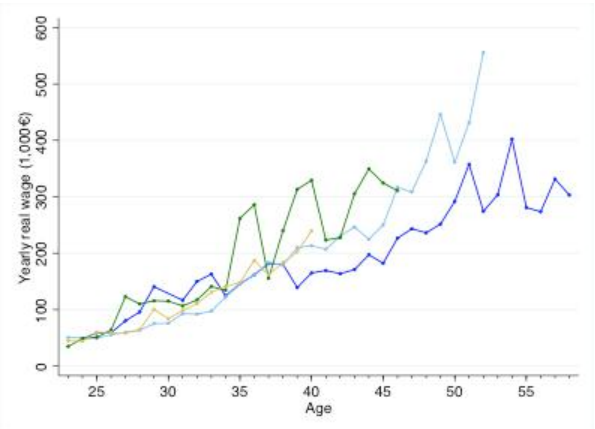
(b) 99th percentile, rescaled

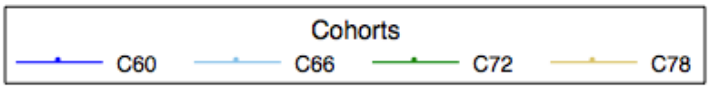


(c) 99th percentile, not rescaled



(d) 99.9th percentile, not rescaled





Notes: Panel (a) shows the average of our original individual wage variable, w_{it} in eq. (1), by year-cohort. It is estimated over the same sample used to compute inequality in Figure 2, after winsorization of the 99th percentile.

Panels (c) and (d) show the values of the 99th and 99.9th percentiles of w_{it}, respectively. The 99th percentile in panel (c) is the one we winsorize before computing inequality in Figure 2.

Panel (b) shows the value of the 99th percentile from panel (c) over the average from panel (a).

The percentiles in panels (b) to (d) are estimated on a slightly different sample than the average from panel (a): before casewise deletion on circumstances and before the exclusion of people whose full-year equivalent wage is strictly below the legal minimum.

The percentiles in panels (b) to (d) are estimated at the year-birth year level whereas the average from panel (a) is estimated at the year-cohort level. Accordingly, the "cohort" legend underneath the figure is an abuse of language: the curves in panels (b) to (d) correspond to the "middle" birth year of each cohort (e.g., 1960 for C60).

The values in all panels besides (b) are given in yearly real terms (thousands of euros from 2018).

VII. Literature review

Table A7.1: Review of the literature about the evolution of IO in income or earnings : time dimension

Studies	Countries	Data	Data period [a]	# data waves	# cohort-age obs. / country	Age groups	Birth cohorts	Comment about the temporal comparisons of inequality of opportunity
Teyssier & Toulemon (2024; this study)	France	EDP	1984-2018	34 waves (payroll)	105 obs.	36 age gps. (21-25, 22-26, …, 56-60)	4 cohorts (1958-62, 1964-68, 1970-74, 1976-80)	Tracks the cohorts explicitly across age gps., which delivers a cross-age analysis (holding cohort fixed) along the x-axis and a cross-cohort one (holding age fixed) along the y-axis. The cohorts are observed over the mid-age ranges: 24-58, 23-52, 23-46 and 23-40.
Dherbécourt & Flamand (2023)	France	EDP	2011-2019	2 waves (fiscal)	2 obs.	35-38	1 cohort per wave	Tracks the age gp. across waves (implicitly across cohorts).
Lefranc et al. (2011)	France	FQP	1977-1993	2 waves	2 obs.	30-40	1 cohort per wave	Tracks the age gp. across waves (implicitly across cohorts).
Lefranc et al. (2009, 2004)	France	BdF [a]	1979-2000	5 waves	5 obs.	30-50	1 cohort per wave	Tracks the age gp. across waves (implicitly across cohorts).
Bussolo et al. (2023) [b]	France	HBS [c] (BdF in French)	1978-2005	6 waves	Fig. 4: up to 48 obs. (UK)	Fig. 4: 2 age gps.. (30-44, 45-59)	Fig. 4: 1 cohort per wave-age gp.	Fig. 4 tracks the age gps. across waves (implicitly across cohorts). Fig. 6 tracks the cohorts across waves (implicitly across age gps.). Fig. 7 explicitly tracks the cohorts across age intervals of 5 years along the x-axis, ranging from 25 to 75 years old depending on the cohort or country. This delivers a cross-age analysis (holding cohort fixed) along the x-axis and a cross-cohort one (holding age fixed) along the y-axis. However, the predetermined 5-year intervals of the x-axis do not fit naturally with the actual age gps. of the cohorts in any given wave. [d] Fig. 8-9 show Deaton's decomposition. Fig. 8 tracks one average birth cohort across the age gps. 30-34 to 75-79. Fig. 9 tracks one average age gp. across the cohorts 1920-24 to 1975-79.
	Germany	SOEP	1984-2013	11 waves	Fig 6: up to 72 obs. (UK)	Fig. 6-7: 1 age gp. per wave-cohort	Fig. 6-7: 3 cohorts (1940-49, 1950-59, 1960-69)	
	Italy	SHIW	1993-2014	11 waves	Fig 7: up to 23 obs. (UK)	Fig.8: 10 age gps. (30-34, 35-39, …, 75-79)	Fig. 8: 1 average cohort of 5 years of birth	
	UK	BHPS-UKHLS	1991-2014	24 waves	Fig 8: 10 obs. Fig. 9: 13 obs.	Fig. 9: 1 average age gp. of 5 years	Fig. 9: 13 cohorts (1920-24, 1925-29,…, 1980-84)	
Carranza (2022)	32 Euro. countries	EU-SILC	2005-2019	up to 15 waves	up to 15 obs.	25-55	1 cohort per wave	Tracks the age gp. across waves (implicitly across cohorts). Relies on a fixed-effect approach to use additional waves of the EU-SILC.

Andreoli et al. (2021)	26 Euro. countries	EU-SILC	2005-19	up to 3 waves	up to 3 obs.	Fig. 2-3: 24-59 Fig.4-11: 1 age gp. per wave-cohort	Fig. 2-3: 1 cohort per wave Fig.4-11: 4 cohorts (1945-55, 1955-65, …,1975-85)	Fig.2-3 track the age gp. across waves (implicitly across cohorts). Fig.4-11 compare cohorts along the x-axis, separately by wave. But cohorts in a given wave are completely separated age gps., so these comparisons neither hold the age nor the cohort fixed.
Ramos & van de Gaer (2021)	24 Euro. countries	EU-SILC	2005-2011	2 waves	2 obs.	25-59	1 cohort per wave	Tracks the age gp. across waves (implicitly across cohorts).
Suárez-Á. & López-M. (2021, 2018)	26 Euro. countries	EU-SILC	2005-2011	2 waves	2 obs.	25-59	1 cohort per wave	Tracks the age gp. across waves (implicitly across cohorts).
del Valle-Inclán (2019)	31 Euro. countries	EU-SILC	2003-2016	up to 14 waves	up to 14 obs.	30-59	1 cohort per wave	Tracks the age gp. across waves (implicitly across cohorts). Relies on a "capital income approach" to use additional waves of the EU-SILC.
Andreoli and Fusco (2017)	19 Euro. countries	EU-SILC	2005-2011	2 waves	2 obs.	30-50	1 cohort per wave	Tracks the age gp. across waves (implicitly across cohorts).
Checchi et al. (2016)	26 Euro. countries	EU-SILC	2005-2011	2 waves	2 obs.	30-60	1 cohort per wave	Tracks the age gp. across waves (implicitly across cohorts).
Brzezinski (2015)	23 Euro. countries	EU-SILC	2005-2011	2 waves	2 obs.	26-50	1 cohort per wave	Tracks the age gp. across waves (implicitly across cohorts).
Wu et al. (2021)	Australia	HILDA	2001-2017	17 waves	17 obs.	15-65	1 cohort per wave	Tracks the age gp. across waves (implicitly across cohorts).
Martinez et al. (2017)	Australia	HILDA	2001-2013	13 waves	13 obs.	25-65	1 cohort per wave	Tracks the age gp. across waves (implicitly across cohorts).
Bourguignon et al. (2007)	Brazil	PNAD	1996	1 wave	7 obs.	1 age gp. per cohort	7 cohorts (1936-40, 1941-45, …, 1966-70)	Compares cohorts which are completely separated age gps. in a single wave, so these comparisons neither hold the age nor the cohort fixed.
Dai & Li (2021)	China	CGSS	2010-2015	4 waves pooled	8 obs.	1 age gp. per cohort	8 cohorts: (1950-54, 1955-59, …, 1985-89)	Compares cohorts which are completely separated age gps. in a single pooled cross-section., so these comparisons neither hold the age nor the cohort fixed.
Golley et al. (2019)	China	SWSSC	2010	1 wave	6 obs.	6 age gps. (26-30, 31-35, …, 51-55)	1 cohort per age gp.	Compares age gps. which are completely separated cohorts in a single wave, so these comparisons neither hold the age nor the cohort fixed
Zhang & Eriksson (2010)	China	CHNS	1989-2006	7 waves	7 obs.	20-50	1 cohort per wave	Tracks the age gp. across waves (implicitly across cohorts).
Brunori & Neidhöfer (2021)	Germany	SOEP	1992-2016	25 waves	25 obs.	30-60	1 cohort per wave	Tracks the age gp. across waves (implicitly across cohorts).
Moramarco et al. (2020)	South Korea	KLIPS	2001-2014	14 waves	14 obs.	20-65	1 cohort per wave	Tracks the age gp. across waves (implicitly across cohorts).

Marrero & Rodríguez (2016)	USA	PSID	1970-2009	34 waves	34 obs.	25-50	1 cohort per wave	Tracks the age gp. across waves (implicitly across cohorts).
Marrero & Rodríguez (2011)	USA	PSID	1969-2007	34 waves	34 obs.	25-50	1 cohort per wave	Tracks the age gp. across waves (implicitly across cohorts).
Pistolesi (2009)	USA	PSID	1968-2001	32 waves	32 obs.	30-50	1 cohort per wave	Tracks the age gp. across waves (implicitly across cohorts).

[a] We refer to the years when the data was collected. Incomes in the EU-SILC studies and in Dherbécourt & Flamand (2023) refer to the year preceding the data collection.

[b] The 2019 working paper version of Bussolo et al. (2023) also considers Switzerland over 6 waves of the SHP (1999-2014).

[c] The first wave of the BdF was run in 1978-79. Lefranc et al. (2004, 2009) refer to it as the 1979 wave and Bussolo et al. (2023) as the 1978 wave.

[d] Regarding France for instance, Fig. 7 stars at 25 years old for the cohort 1940-49 although its members are between 29 and 38 years old in the first 1978 wave of the BdF, and the same value of IO is then ascribed to this cohort at 30 years old. Another issue is that the authors appear to restrict their sample to people above 25 years old although they consider some cohorts in years where their youngest members are below this age limit (e.g. the cohort 1950-59 is between 19 and 28 years old in the 1978 wave of the BdF).

Note: Studies about France only are listed first, those about France and other countries come next, and those not about France come last by alphabetical order of the country. Within each group, they are listed by reverse chronological order of the study and by alphabetical order of the authors.

Table A7.2: Review of the literature about the evolution of IO in income or earnings: circumstance and income variables

Study	Parental education	Parental occupation	Migratory or ethnic origin	Sex	Age	Other	Income variable
Teyssier & Toulemon (2024; this study)	Father's ed.	Father's occ.	Father's citizenship	Sex		Age difference with father Area of residence during youth	Individual wage
Dherbécourt & Flamand (2023)		Parental occ.	Parental country of birth	Sex		Region of residence during youth Area of residence during youth	Individual active income
Lefranc et al. (2011)						Predicted father's income	Individual labor earnings (equivalent full-time)
Lefranc et al. (2009, 2004)		Father's occ.					Household equivalized income (before and after government transfers and taxes)
Bussolo et al. (2023)	Highest parental ed. (except in France)	Highest parental occ. (only in France)	Born abroad	Sex	Age and its square	Wealthy birth region (except in France)	Personal income
Carranza (2022) (lower-bound specification)	Father's ed. Mother's ed.	Father's occ. Mother's occ. Father's activity Mother's activity Father's managerial status (not in 2005) Mother's managerial status (not in 2005)	Father's country of birth (not in 2005) Mother's country of birth (not in 2005) Father's citizenship (not in 2005) Mother's citizenship (not in 2005)	Sex		Parental composition of the household Nb. of siblings (2005 only) Nb. of adults in the household (not in 2005) Nb. of children in the household (not in 2005) Tenancy status of the parents (not in 2005)	Household equivalized disposable income
Andreoli et al. (2021)	Father's ed. Mother's ed.	Father's occ. Mother's occ.		Sex		Presence of siblings in the household Parental composition of the household	Household equivalized disposable income
Ramos & van de Gaer (2021)	Parental education	Parental occupation	Born abroad	Sex		Whether lived with both parents when young	Household equivalized disposable income
Suárez-A. & López-M. (2021, 2018)	Highest parental ed.	Highest parental occ.	Born abroad	Sex	Age	Population density of the place of residence	Household equivalized disposable income

del Valle-Inclán (2019)	Highest parental ed. (or own capital income as a proxy instead)	Highest parental occ. (or own capital income as a proxy instead)	Born abroad	Sex			Individual wage
Andreoli and Fusco (2017)	Father's ed.						Individual labor earnings
Checchi et al. (2016)	Highest parental ed.		Born abroad	Sex	Age		Individual disposable income
Brzezinski (2015)	Father's ed. Mother's ed.	Father's occ.	Country of birth				Household equivalized disposable income
Wu et al. (2021)		Father's occ. Mother's occ.	Country of birth Parental countries of birth Migratory status and origin Refugee status Indigenous status	Sex	Age	Parental composition of the household Number of siblings Whether firstborn	Individual income (before and after government transfers and taxes)
Martinez et al. (2017)		Father's occ. Mother's occ.	Country of birth Father's country of birth Mother's country of birth Indigenous status	Sex			Household equivalized income (before and after government transfers and/or taxes)
Bourguignon et al. (2007)	Father's ed. Mother's ed.	Father's occ.	Race			Region of birth.	Individual labor earnings
Dai & Li (2021)	Father's ed	Father's occ. Father's empl. status		Sex		Hukou of birth (rural or not) Father's membership to the CCP Father's administrative rank	Individual income
Golley, et al. (2019)	Father's ed.	Father's occ.		Sex		Hukou of birth (rural or not) Region of residence (when adult)	Individual labor earnings
Zhang & Eriksson (2010)	Father's ed. Mother's ed.	Father's employer Mother's employer		Sex	Age	Urban or rural area of birth Coastal or inland province of birth Parental household income Parental household size	Individual income

Brunori & Neidhöfer (2021)	Father's ed. Mother's ed.	Father's occ. Father's training	Migration background	Sex		Resident in East or West Germany in 1989 Number of siblings Whether disabled	Household equivalized disposable income
Moramarco et al. (2020)	Highest parental ed.			Sex		Birth place Whether received material or financial support from the parents in the preceding year	Household equivalized disposable income
Marrero & Rodríguez (2016)	Father's ed.		Race				Household head's labor income plus per capita household capital income
Marrero and Rodríguez (2011)	Father's ed.		Race				Household head's labor income plus per capita household capital income
Pistolesi (2009)	Father's ed. Mother's ed.	Father's occ.	Race		Age	Birth region	Household head's labor income

Note: Studies are listed in the same order as in Table A7.1.