The French Welfare Distribution

Marcos Aurelio Díaz Ramírez *

August 27, 2014

Supervisor at the OECD:
Fabrice Murtin

Supervisor at Sciences Po:
Yann Algan

Dissertation for the
Master in Economics and Public Policy

École Polytechnique - Sciences Po - ENSAE ParisTech

Abstract

This paper estimates the distribution of welfare, defined as lifetime utility, at the national level for France. Taking stock of the French Labour Force Survey, the Statistics on Income and Living Conditions and a census-based mortality database, this work simulates the lifetime earnings and welfare of a hypothetical 1980's French birth-cohort with the help of Monte-Carlo methods and Copula functions, while accounting for gender and education-related differences in income, employment and longevity prospects. The results indicate that inequality in the monetary equivalent of lifetime utility is much higher than inequality in lifetime earnings and also higher than cross-sectional income inequality. This occurs because lifetime utility accounts for longevity, which is positively correlated with income, so that welfare inequality is the by-product of income and longevity inequalities that appear to add up rather than mitigate each other. Looking at the average equivalent income by educational attainment, it is found that the returns in welfare to education, which cumulate the earnings, employment and longevity premiums of higher schooling, are on average 65% higher than the standard income returns to education, suggesting that ignoring longevity and unemployment differential risks across educational groups may result in a non-negligible underestimation of the actual benefits of education.

*All the methodologies, results and ideas presented in this dissertation are extractions from the forthcoming paper of the same title by Díaz and Murtin (2014). References to the present work should be realised by quoting the aforementioned paper and including both authors.

Acknowledgements

I would like to express all my gratitude to Fabrice Murtin for giving me the opportunity to work with him on this challenging and rewarding project, and for all his patience, teachings and guidance along these past few months. I would also like to thank Romina Boarini for all her support and for opening to me the doors of the OECD's Statistics Directorate. To my family and to my best friend, for always being there for me unconditionally. To Tania, for helping me find peace whenever I most needed it.

Contents

1 Introduction

The recent years' consensus on the limits of GDP as a measure of Welfare has called for a radical reformulation on the study and measurement of welfare within and across societies. Different groups in academia, governments and international organisations have already embarked on the challenging task of creating more comprehensive measures of welfare and well-being (Stiglitz et al., 2010; OECD, 2013; OECD, 2014a; Boarini et al., 2013.) Although significant progress has been made on the subject, there are still important limitations to overcome. In this paper, two of the main difficulties are identified and addressed simultaneously: the aggregation of correlated material and non-material dimensions of welfare into one single monetary measure and the incorporation of the time dimension into the welfare analysis.

The recognition that a comprehensive measure of welfare should be multidimensional (Stiglitz et al., 2010) has generated the need for a reliable method to aggregate material and non-material dimensions altogether. For its solid theoretical foundations, this paper makes use of the equivalent income approach (see Fleurbaey and Gaulier, 2007) to value non-income determinants of living standards in monetary terms and to create a single money-metric measure that is fully comparable among individuals. This paper is the first attempt to account for the correlation of material and non-material dimensions at the individual level in the derivation of a welfare composite measure.

Authors such as Chau (2009), Bowlus and Robin (2004) and Bonhomme and Robin (2009) have stressed that looking at a cross-section to study individual income inequality might be inadequate for several reasons. The main one being that cross-sectional measures ignore the role of mobility in both labour market status and the individual's position in the income distribution. In fact, the aforementioned authors have documented that when taking into account the effect of mobility during the life cycle, income inequality tends to decrease. In other words, cross-sectional inequality in earnings is higher than lifetime earnings inequality.

The above result, even if true for disposable income, might not be valid for a comprehensive measure of welfare since it ignores a fundamental determinant of lifetime utility, namely lifetime duration. Actually, accounting for lifetime inequalities in health and job security through equivalent income results in a level of lifetime inequality even higher than the one observed in a cross-sectional analysis.

In this paper, the lifetime histories of a synthetic 25-year-old French cohort are generated. Lifetime histories are the series from 25 to 99 years old that describe the survival, labour force and unemployment statuses of the individuals as well as their disposable incomes over their lifetimes. Making use of all these variables and a CRRA Utility function with intercept, it is possible to calculate lifetime earnings and utility, as well as the distribution of disposable and equivalent income. The main findings are that lifetime inequality is significantly more severe when taking into account the values of health and job security during the life span. The distribution of equivalent income displays both a fat left tail due to premature deaths and a skewed right tail due to the positive correlation between disposable income and longevity.

The previous findings constitute an important contribution to the literature on inequality and welfare. They suggest that inequality in lifetime utility (or its monetary equivalent) is much higher than inequality in lifetime income as lifetime utility accounts not only for disposable income streams, but also for the value of health.

The structure of the paper is the following. Section 2 presents the Data. Section 3 explains the Model and derives the measures of income required for the analysis. Section 4 develops the Empirical framework used to simulate the hypothetical individuals and their lifetime histories. Section 5 is devoted to the Results and Section 6 for the Conclusions.

2 Data

This section presents the data for France in 2005 obtained from three main sources: the French *Labour Force Survey* (LFS), the *European Union Statistics on Income and Living Conditions* (SILC), and the data set on mortality rates by van Raalte et al. (2012) (MRD.)

2.1 The Labour Force Survey (LFS) and the European Union Statistics on Income and Living Conditions (SILC)

The French Labour Force Survey or *Enquête Emploi en continu 2005* contains information on labour income, social and unemployment benefits at the individual level. However, due to the lack of a representative sample for the older population and a reliable measure on pensions, the LFS doesn't provide sufficient information on the income for the retired. Hence, the sample extracted from this survey is limited to the labour force population (i.e. the employed and the unemployed). To overcome the aforementioned issue, the EU-SILC for France in 2005 has been incorporated to the analysis. This data set does not only offer data on labour income, social and unemployment benefits for the labour force population, but also on pensions and other sources of income for the retired individuals.

Both surveys provide information on age, gender, education level and the labour market situation of individuals, which enables the estimation of the distributions of income conditional on gender, education, labour market situation and age group. From these data sets it is also possible to obtain the probability of being retired and the unemployment risk by gender-education-age groups.

Merging the LFS and SILC

This study takes stock of and merges the LFS and SILC data sets. These two databases yield similar distributions of the log of disposable income for different education-labour status groups (from 25 to 60 years old) so that their pooling does not introduce any composition issue by data source. As shown in Figure 1, the kernel densities of the log of disposable income for the LFS and SILC are almost identical for the employed. There are slight differences between the distributions of the log of disposable income for the unemployed, however these disparities seem quite low to prevent the merging of the two data sets.

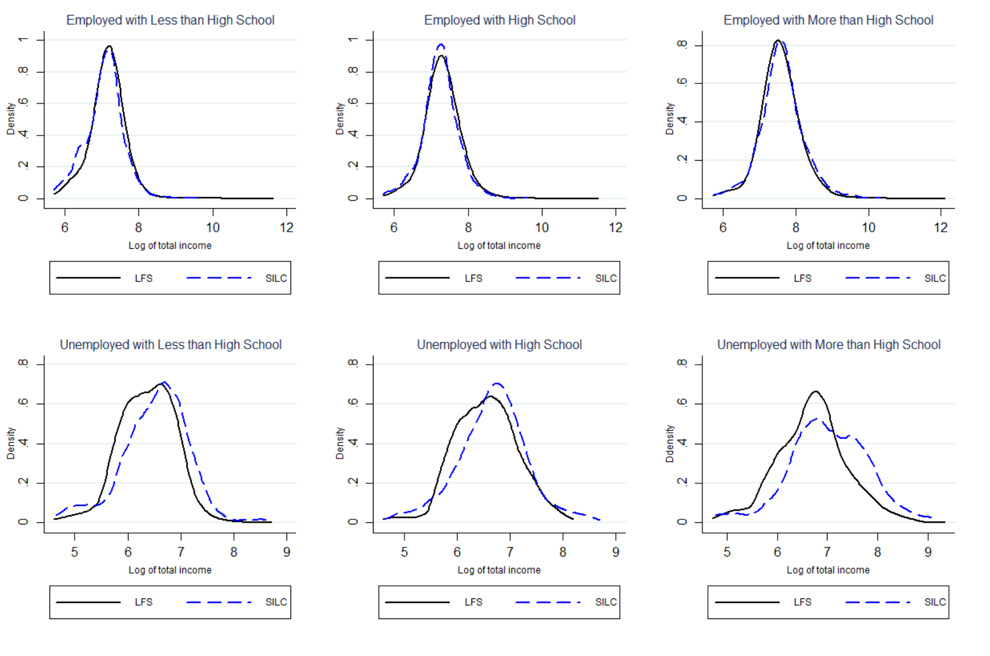


Figure 1: Kernel-Density of the Log of Disposable Income

The sample is restricted to the 25-99 year-old population and truncated for incomes lower than 300 euros per month for the employed, for incomes lower than 100 euros per month for the unemployed and the retired and for incomes higher than 10,000 euros per month for everyone. The final sample consists of 50,131 individual observations distributed across 2 genders (Males and Females), 3 educational levels (Less than High School, High School and More than High School), 3 labour market situation status (Employed, Unemployed and Retired) and 15 5-year age groups (from 25-29 to 95-99 years old.)

2.2 Mortality Rates Data set by van Raalte et al. (2012) (MRD)

The data on mortality rates has been obtained from the work of van Raalte et al. (2012). This data set was built using census-based data. It comprises death counts ($mx(a,g,s)$) and exposure ($nx(a,g,s)$) by age, gender and education groups for 11 countries, among which France is included. The data corresponding to France is extracted and mortality rates are constructed as $\pi(a,g,s) = \frac{mx(a,g,s)}{nx(a,g,s)}$. Then survival probability at age a is built as $\prod_{k=25}^{a}(1-\pi(a,g,s))$. The original data on mortality rates for France provides data for 2 genders, 3 education groups and 76 ages (from 35 to 110 years old). However, due to the interest in studying a life span that ranges from 25 to 99 years old, it is assumed that individuals with ages between 25 and 34 years have a mortality rate equal to zero and consequently, a survival probability equal to 1.

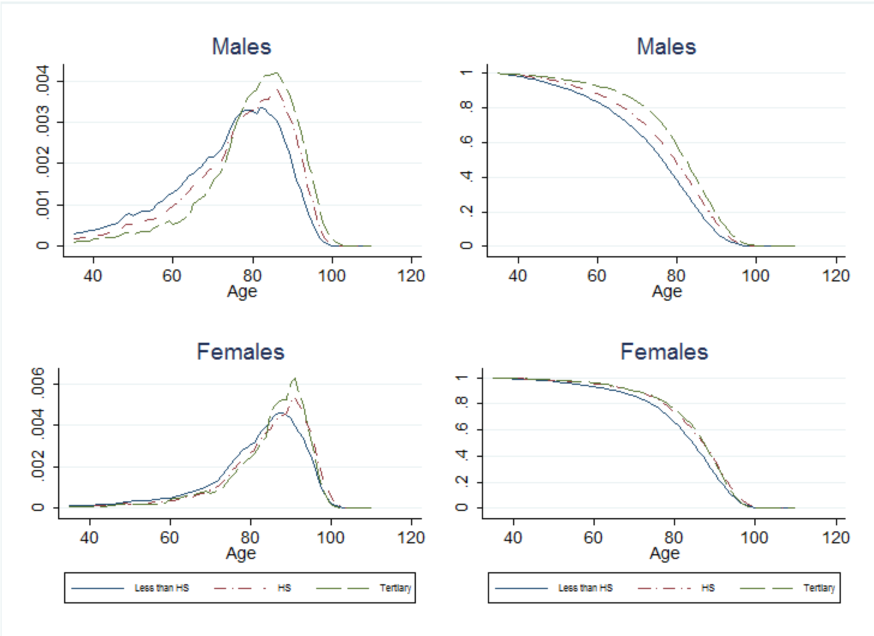


Figure 2: Mortality rates and Survival probabilities by Age

Figure 2 depicts the mortality rates (left quadrants) and survival probabilities (right quadrants) for different gender-education groups at different ages (from 35 to 110 years old). As shown on the left side of Figure 2, mortality rates are lower for the more educated groups at younger ages ($<$ 77 years old for males and $<$ 83 for females.) This implies that highly educated groups tend to have higher survival probabilities relative to the low educated (right side of Figure 2); and consequently, the more educated groups tend to have higher life expectancy.

Conditional on education and age, women have higher probabilities of survival and life expectancy than men. In addition, there is lower inequality in longevity among women than among men, as shown on the right side of Figure 2 where the survival probability gap between highly educated and low educated people is larger for men than for women. In the same vein, the gap between the survival probabilities for the highly educated women compared to the women with only high school is negligible.

2.3 Descriptive Statistics

Table 1 presents some descriptive statistics of the final data set. As expected, conditional on gender and age, average disposable income is higher for more educated groups, while average disposable income is higher for males in all age and education levels.

Table 1: Descriptive Statistics

	Males			Females		
Working Age (25-60 *years old*)	Primary	Secondary	Terciary	Primary	Secondary	Terciary
Average Total Income (€)	1477.334	1757.156	2515.345	1075.458	1333.856	1823.891
Employed (%)	88.24	90.01	93.11	84.53	88.53	93.7
Unemployed (%)	10.89	7.38	6.43	14.61	9.69	5.73
Retired (%)	0.86	2.61	0.46	0.86	1.78	0.57
Average Education (years)	9.04	12.00	15.66	8.80	12.00	15.41
Survival probability at 60	0.833	0.883	0.927	0.933	0.955	0.962
Observations	12173	5133	5891	10456	5049	6894
Seniors (> 60 *years old*)						
Average Total Income (€)	1270.568	1808.384	3146.685	873.4309	1312.174	1951.81
Employed (%)	5.21	6.05	22.93	13.83	8.2	19.39
Unemployed (%)	2.69	1.01	3.5	2.33	1.43	5.61
Retired (%)	92.1	92.94	73.57	83.84	90.37	75
Average Education (years)	5.54	12.00	16.14	5.68	12.00	15.94
Survival probability at 80	0.383	0.485	0.584	0.658	0.737	0.758
Observations	1228	595	314	1714	488	196

The average disposable income for the least educated is higher during working age than after 60 years old, while the reverse is true for the highly educated. The latter might be the consequence of the accumulation of capital income that is likely to be higher among people with tertiary education.

For all education and age groups, survival probability is higher for females than for males, and higher for more educated groups. The survival probability for the low educated is halved from 60 years to 80 years. This decrease is less pronounced for people with tertiary education. There is a lower proportion of retired people among seniors for those with higher education than for those with low education. This could be explained through the kinds of activities and occupations undertaken by those with lower education, as they typically perform jobs that are physically demanding and tend to retire earlier in the life cycle.

3 The Model

3.1 A Hypothetical Agent model

Due to the difficulty of having a panel data as long as to cover a 75-years lifespan (from 25 to 99 years old) of a representative sample of a French cohort, hypothetical agents and their lifetime histories are simulated. Even though these agents are hypothetical, the distribution of their characteristics and incomes are based on the actual distributions observed in the data.

More precisely, the observed distributions of age, gender and education of the actual 1980's French cohort (25 years old population observed in 2005) are taken as the baseline to generate a synthetic 1980's French cohort. Once this cohort has been generated, the lifetime histories in terms of life status, labour force situation, unemployment condition and disposable income are also simulated. Once lifetime histories are obtained for all the hypothetical individuals, the distributions of lifetime disposable income and lifetime utility can be estimated.

The model's hierarchy

The life status (alive or dead) and the labour market situation (retired, employed or unemployed) of an individual in a given year can be simplified in the below hierarchical diagram:

$$
\begin{tikzcd}
Individual \arrow[d] \arrow[dr] & \\
Alive \arrow[d] \arrow[dr] & Dead \\
Labour\,Force \arrow[d] \arrow[dr] & Retired \\
Employed & Unemployed
\end{tikzcd}
$$

In a given year, a hypothetical agent can be alive or dead; if he is alive, he can either be part of the labour force or belong to the retired population; then, conditional on being in the labour force, he can be employed or unemployed.

From this hierarchical model one can easily see that there is a dependency between statuses in a given year (e.g. one cannot be employed if he is not in the labour force, and one can not be in the labour force if he is dead.) Nevertheless, it is also important to take into account the dependency of statuses over time.

An obvious restriction is that if an agent dies in a given year $(a - 1)$, he will be dead in the subsequent years ($a + t$ for $t \in \mathbb{N}$). This constraint is also applied to the labour force-retirement status process since once an individual has retired it is very unlikely that he will re-enter the labour force in the following years. On the other hand, this restriction doesn't apply to the employment-unemployment status dynamics, for which an individual can move over time from being employed to be unemployed and vice versa. However, the latter doesn't imply that the employment-unemployment process is independent over time. In general, the dependency of these variables over time is taken into account using Copula functions (see section on Copulas for a detailed explanation.)

The variables and parameters that govern the model are defined following the aforementioned hierarchy. Hence, it is natural to start by defining the variables that determine the life status of the agent.

Life status

Let's denote the *mortality rate* faced by an individual of age a, gender g and schooling s as $\pi_{i,a}(g,s)$. Using this mortality rate as a proxy of the probability of death at a given age, the *probability of survival* is equal to $\prod_{k=25}^{a}(1-\pi_{i,k}(g,s))$, and then the *Survival status* can be expressed in the following way:

$$S_{i,a}(g,s) = \begin{cases} 1 & \text{with probability} \quad \prod_{k=25}^{a}(1-\pi_{i,k}(g,s)) \\ 0 & \text{otherwise} \end{cases}$$

Thus, Survival or Life status is a dummy variable taking value 1 if individual i is alive (which occurs with a probability equal to the probability of survival) and 0 if he is dead.

Labour Force status

Conditional on being alive, an individual has a *retirement probability* $r_{i,a}(g,s)$ that is a function of his age, gender and education. From this, the *Labour Force probability* at a given age a is expressed as $\prod_{k=25}^{a}(1-r_{i,k}(g,s))$ and, subsequently, the *Labour Force status* can be written in the following way:

$$LF_{i,a}(g,s) = \begin{cases} 1 & \text{with probability} \quad \prod_{k=25}^{a}(1-r_{i,k}(g,s)) \\ 0 & \text{otherwise} \end{cases}$$

From which *Retirement status* can be defined as $R_{i,a}(g,s) = 1 - LF_{i,a}(g,s)$.

Unemployment status

Conditional on being part of the Labour Force, the individual faces an unemployment risk equal to the *Unemployment rate* $u_{i,a}(g,s)$. *Unemployment status* is defined as a dummy variable taking value 1 if the individual is unemployed (which occurs with a probability equal to the unemployment risk) and 0 if he is employed (which happens with a probability equal to the *Employment rate* $e_{i,a}(g,s) = 1 - u_{i,a}(g,s)$):

$$U_{i,a}(g,s) = \begin{cases} 1 & \text{with probability} \quad u_{i,a}(g,s) \\ 0 & \text{otherwise} \end{cases}$$

Finally, from the Unemployment status variable it is possible to define the *Employment status* variable as $E_{i,a}(g,s) = 1 - U_{i,a}(g,s)$.

Annual Disposable Income of an alive individual

Once the variables that determine the labour market situation of the individual have been explained, it is suitable to define the Disposable income variables. The *Annual disposable income* of an alive agent of age a, gender g, education s and labour market situation l is denoted as $y_{i,a,l}(g,s)$.

This variable is stochastic since the processes that determine the labour market status and the ranking of the individual in the income distribution are obtained from random draws using Monte-Carlo methods (see next section for more details.)

For practical purposes, let's define Annual Disposable Income for each of the three possible labour market statuses. *Employment Disposable Income* $(y_{i,a,l}(g,s)|l = employed)$ is denoted as $y^e_{i,a}(g,s)$, *Unemployment Disposable Income* $(y_{i,a,l}(g,s)|l = unemployed)$ as $y^u_{i,a}(g,s)$ and *Retirement Disposable Income* $(y_{i,a,l}(g,s)|l = retired)$ as $y^r_{i,a}(g,s)$.

As it has been explained before, the labour situation (l) of individual i is a function of his age, gender and education due, among other things, to the dependency between these characteristics and the retirement probabilities and unemployment risks that he faces. Incorporating some of the previously defined variables, one can rewrite the *Annual disposable income of an alive individual* at a given year in the following fashion:

$$y_{i,a,l}(g,s) = \underbrace{(1 - LF_{i,a})}_{R_{i,a}} y^r_{i,a} + LF_{i,a}[U_{i,a}y^u_{i,a} + \underbrace{(1 - U_{i,a})}_{E_{i,a}} y^e_{i,a}] \quad (1)$$

3.2 Lifetime Disposable Income

Now that the measure of annual disposable income (of an alive individual) at a given year (and consequently at a given age for the agent) has been specified, it is possible to build a measure of lifetime disposable income. *Lifetime Disposable Income* is constructed as a discounted sum of annual disposable incomes during the life cycle:

$$Y_i(g,s) = \sum_{a=0}^{A} \beta^a S_{i,a}(g,s) y_{i,a,l}(g,s) \quad (2)$$

Since the initial and last possible age at which the individual can be observed is at 25 and 99 years old respectively, a goes from 0 to 74, where $a = 0$ represents the age of 25 years old and $a = 74 = A$ denotes the age of 99 years old.

Even though the lifetime of an individual can be up to 99 years old, it will be very unlikely to observe an agent alive at this age due to the low probability of survival he will find at older ages. Each individual in the model will have a different series $\{S_{i,a}(g,s)\}_{a=0,1,\cdots,74}$ that describes his years alive and his age at death. This variation in age at death will allow to capture part of the inequality in lifetime income and utility.

3.2.1 Rescaling Lifetime Disposable Income on a monthly basis

The definition of disposable income in the model so far has been annual. However, due to the nature of the data it will be more suitable to work with monthly measures of disposable income. In the following, lifetime disposable income is rescaled on a monthly basis to obtain alternatively measures of average monthly disposable income during the life cycle.

Average Monthly Disposable Income during Lifetime

As shown in Equation 3, which defines the *Average monthly disposable income during Lifetime*, the numerator (lifetime disposable income) is not only divided by the number of months ($12\sum_{a=0}^{A} S_{i,a}(g,s)$) the individual is alive, but also by a discount factor.

$$Y_i^d(g,s) = \frac{\sum_{a=0}^{A} \beta^a S_{i,a}(g,s) y_{i,a,l}(g,s)}{12\sum_{a=0}^{A} \beta^a S_{i,a}(g,s)} \tag{3}$$

Since lifetime disposable income is a sum that assigns less weight to future incomes, it is also pertinent to discount the value of the future years of life (i.e. the farther the year, the less value it has from the individual's current perspective.) More precisely, each year of the individual's possible lifetime ($a = 0, 1, \cdots, 74$) is discounted by β^a. Therefore, the denominator of Equation 3 could be thought as the discounted number of months the individual is alive.

Average Monthly Disposable Income during Life Expectancy

An alternative measure of monthly disposable is the normalization of the Lifetime disposable income by the Average discounted life expectancy at age 25 (in months.) The *Average discounted life expectancy at age 25* (in years) is defined as $\bar{T} = \frac{1}{N}\sum_{i=1}^{N}\sum_{a=0}^{A} \beta^a S_{i,a}(g,s)$. Using this measure as an approximation of the years lived by all the agents, it is possible to define the *Average Monthly Disposable Income during Life Expectancy* as:

$$Y_i^{lt}(g,s) = \frac{\sum_{a=0}^{A} \beta^a S_{i,a}(g,s) y_{i,a,l}(g,s)}{12\ \bar{T}} \tag{4}$$

Average Monthly Employment Disposable Income during Employed Lifetime

In order to study the effect of unemployment risk on lifetime disposable income, a measure of employment income during employed lifetime is proposed. *Average monthly employment income at full employment* is denoted as:

$$Y_i^e(g,s) = \frac{\sum_{a=0}^{A} \beta^a S_{i,a}(g,s) LF_{i,a}(g,s) E_{i,a}(g,s) y_{i,a}^e(g,s)}{12\sum_{a=0}^{A} \beta^a S_{i,a}(g,s) LF_{i,a}(g,s) E_{i,a}(g,s)} \tag{5}$$

3.3 Welfare as Lifetime Utility

In this work, individual welfare is indistinctly seen as lifetime utility. To define lifetime utility it is first necessary to specify the form of the utility function at a given age. As in previous literature (Murphy and Topel, 2006; Becker et al., 2005), utilities are modelled with the following Constant Relative Risk Aversion (CRRA) Utility function with intercept $v(y_i) = \frac{y_i^{1-\gamma} - D^{1-\gamma}}{1-\gamma}$, $\forall \gamma \neq 1$ where γ is the parameter of Relative Risk Aversion and D is a constant. The value of these parameters and the discount factor are chosen according to the literature, then $\gamma = 1.25$ (Murphy and Topel, 2006), $D = 267$ and $\beta = e^{-0.03} = 0.97$ (Boarini et al., 2013.)

Since this utility function is a monotonically increasing function of annual disposable income, the ranking in the distribution of disposable income in a given year remains the same after the transformation of annual disposable income into annual utils. For practical purposes, it is assumed that the utility function is constant over time (i.e. independent of the age of the agent) and the same for all individuals. Therefore, in a given year, individuals have different utilities only due to their differences in disposable income. On the other hand, when considering lifetime utility, inequalities in welfare among individuals also arise from their differences in life span, retirement probability and unemployment risk.

$$V_i(g, s) = \sum_{a=0}^{A} \beta^a S_{i,a}(g, s) v(y_{i,a,l}(g, s)) \tag{6}$$

Equation 6 defines *Lifetime Utility* as the discounted sum of annual utilities during the life cycle of the agent.

3.3.1 Monthly Equivalent Income

The idea of equivalent income is incorporated into this work as a way to value in monetary units the non-income dimensions that determine lifetime utility, such as health and job security. Once the monetary values of longevity and unemployment risk are taken into account, welfare inequality can be measured in a more comprehensive way.

Let's define the series of stochastic Life, Labour Force and Unemployment statuses, and the series of stochastic Disposable employment, unemployment and retirement incomes of agent i as the *Stochastic Lifetime History of individual i (First scenario)*:

$$SLH_i = \{S_{i,a}, LF_{i,a}, U_{i,a}, y^e_{i,a}, y^u_{i,a}, y^r_{i,a}\}_{a=0,1,\cdots,74}$$

Alternatively, denote the following series of deterministic Life, Labour Force and Unemployment statuses, and Disposable income conditional on the labour market situation of agent i as the *Deterministic Lifetime History of individual i (Second scenario)*:

$$DLH_i(y_i^{EV}) = \{S_{i,a} = \pi^*(a), LF_{i,a} = 1, U_{i,a} = 0, y^e_{i,a} = y_i^{EV}, y^u_{i,a} = 0, y^r_{i,a} = 0\}_{a=0,1,\cdots,74}$$

The equivalent income of individual i is defined as the constant annual disposable income (in the DLH_i) that verifies that the lifetime utility reached with the stochastic lifetime history is equal to the lifetime utility obtained with the deterministic lifetime history, i.e. $V_i[SLH_i] = V_i[DLH_i(y_i^{EV})]$. By solving the previous equality it is possible to verify that the *Monthly Equivalent Income* is given by the following Equation:

$$y_i^{EV}(g,s) = \frac{1}{12}\left[\frac{V_i(g,s)}{\sum_{a=0}^{A}\beta^a\pi^*(a)}(1-\gamma) + D^{1-\gamma}\right]^{\frac{1}{1-\gamma}} \tag{7}$$

In this paper, monthly equivalent income can be thought as the deterministic monthly income an individual should get during the life span from 25 to 99 years old with probability of survival $\pi^*(a)$ (the highest gender-education group survival probability at a given age, i.e. the survival probability of highly educated females) to get the same lifetime utility he obtains when facing stochastic processes that determine his survival, labour force and unemployment statuses, as well as his lifetime earnings.

For instance, an individual dying at a very young age (e.g. at 29 years old, $S_{i,4} = 0$) and falling into unemployment for one year (e.g. at 28 years old, $U_{i,3} = 1$) would be indifferent between this scenario and an alternative one where his probability of survival becomes the maximal probability of survival for the population at a given age (i.e. $S_{i,a} = \pi^*(a) \;\forall\, a$), he is always employed (i.e. $E_{i,a} = 1 \;\forall\, a$) and, in compensation for these improvements in welfare, he renounces to a part of his first scenario stochastic income.

The income the individual will get in the deterministic scenario after quitting a part of his first scenario stochastic income (for the sake of the improvement of some of his non-income variables) would be his equivalent income. Therefore, individuals dying at early ages and/or facing high unemployment risk might be satisfied with low equivalent incomes, while individuals with high longevity and/or job security will be likely to have higher equivalent incomes.

4 Empirical Framework

In a cross section, such as the data set used in this paper, it is not possible to observe a specific individual aging from 25 to 99 years old (or until death), nor his labour force history or his annual earnings during lifetime; nevertheless, conditional on gender and education, it is suitable to estimate the survival probability and unemployment risk by age, as well as to observe the distribution of incomes (conditional on gender, education and labour market situation) for different age groups.

Taking advantage of the latter and assuming that a hypothetical individual will face the same currently observed probabilities of survival and unemployment, and distributions of incomes by age group throughout his life cycle, it will be possible to simulate individual lifetime incomes and utilities. For instance, conditional on gender, education and labour market situation, a 25-year-old individual in 2005 placed in whichever centile of the 25-29 year-old income distribution should expect to be in 20 years somewhere in the range described by the currently observed (in 2005) 45-49 year-old income distribution. In the same way, a 25-year-old highly educated woman should expect to have in 60 years the same probability of survival that is currently observed for the 85-year-old females with high education.

In general, the life, labour force and unemployment statuses during lifetime will be determined by comparing simulated probabilities (for mortality rates, retirement probabilities and unemployment risks) to the probabilities observed in the French data; while the annual earnings streams (conditional on the labour market situation) will be derived from applying the inverse of the Empirical Cumulative Distribution Functions (ECDF) estimated from the data to the simulated rankings in the income distribution for the hypothetical individual. In the following subsections, the previous ideas and everything concerning the empirical framework is explained in more detail.

4.1 Generating the 25-year-old Synthetic Cohort

4.1.1 Monte-Carlo Simulation: Drawing the 25-year-old synthetic cohort

The objective of this section is to generate a hypothetical representative sample (of size $N = 10000$) of the actual 1980's French birth cohort (i.e. the 25-year-old population in 2005). To do so, two main steps have to be taken. First, the subsample of the 25-year-old individuals (of size $n = 1194$) has to be extracted from the main sample and the ranking (whichever) of its n elements has to be normalized to the [0,1] range. Second, using Monte-Carlo simulation, a random vector M of size $N \times 1$ that contains N independently and identically distributed random variables from the continuous uniform distribution ($\mathcal{U}[0,1]$) has to be generated.

Each element of the M vector represents a ranking in [0,1] that can be matched to the ranking of the n individuals of the actual 25-year-old cohort. For example, if a drawn random variable from the M vector takes the value of 0.25, this would be equivalent as to be drawing the individual at the end of the first quartile in the ranking of n (i.e. the individual 299.) Thus, vector M contains a sample of 10,000 simulated individuals that has (statistically) the same characteristics (in terms of gender, education, labour market situation and disposable income) than the actual cohort extracted from the data.

4.2 Lifetime Histories

Once the 25-year-old synthetic cohort has been generated, it is suitable to simulate the Lifetime History of each of the 10,000 hypothetical individuals. The *Lifetime History* of individual i is defined as the series that describe the Survival, Labour Force and Unemployment statuses of the individual as well as his Disposable income over his lifetime.

$$\left\{S_{i,a}(g,s), LF_{i,a}(g,s), U_{i,a}(g,s), y^e_{i,a}(g,s), y^u_{i,a}(g,s), y^r_{i,a}(g,s)\right\}_{a=0,1,\cdots,74} \tag{8}$$

For explanatory purposes, Lifetime History can be separated in two subsets of lifetime series: *Vital-Labour Force-Unemployment history* $\{S_{i,a}(g,s), LF_{i,a}(g,s), U_{i,a}(g,s)\}_{a=0,1,\cdots,74}$ and *Annual Disposable Income history* $\{y^e_{i,a}(g,s), y^u_{i,a}(g,s), y^r_{i,a}(g,s)\}_{a=0,1,\cdots,74}$.

4.2.1 Vital-Labour Force-Unemployment Histories

To construct the Vital-Labour Force-Unemployment history of individual i it is necessary to compare simulated latent probabilities of having one life/labour market status at a given age to the corresponding probabilities or risks faced by the individual according to the real data. Monte-Carlo simulations are implemented to generate these simulated probabilities. The final life/labour market statuses of the hypothetical individual at a given age are determined in the following way:

Vital history

Individual i is alive at age a if and only if from the age of 25 until age a ($k = 0, \cdots, a$), all his simulated latent probabilities of death $\{\varepsilon^{\pi}_{i,k}(g,s)\}_{k=0,\cdots,a}$ have been strictly higher than (i.e. out of the range of) the mortality rates $\{\pi(k,g,s)\}_{k=0,\cdots,a}$ (which are age-gender-education specific) faced by the French population, the latter being obtained from the Mortality rates dataset of van Raalte et al. (2012).

$$S_{i,a}(g,s) = \prod_{k=25}^{a} \mathbb{1}[\varepsilon^{\pi}_{i,k}(g,s) > \pi(k,g,s)] \tag{9}$$

Labour Force history

Similarly to the construction of the Vital history, the hypothetical agent i is in the Labour Force at age a if and only if from the age of 25 until age a, all his simulated probabilities of retirement $\{\varepsilon^{r}_{i,k}(g,s)\}_{k=0,\cdots,a}$ have been strictly higher than the actual probabilities of retirement $\{r(\bar{a}(k),g,s)\}_{k=0,\cdots,a}$ (age group-gender-education specific) observed in the French population.

$$LF_{i,a}(g,s) = \prod_{k=25}^{a} \mathbb{1}[\varepsilon^{r}_{i,k}(g,s) > r(\bar{a}(k),g,s)] \tag{10}$$

Where the probability of retirement for individuals of age group $\bar{a}$, gender g and education s is estimated from the data as the number of retired individuals of the same age group, gender and education divided by the whole population with the same age group, gender and education characteristics ($r(\bar{a},g,s) = \frac{n^r(\bar{a},g,s)}{n^r(\bar{a},g,s)+n^{lf}(\bar{a},g,s)}$.)

Unemployment history

Individual i is Unemployed at age a if and only if his simulated latent probability of being unemployed $\varepsilon^u_{i,a}(g,s)$ is strictly higher than the Unemployment rate $u(\bar{a}(a),g,s)$ observed in the French data (which is age group-gender-education specific.)

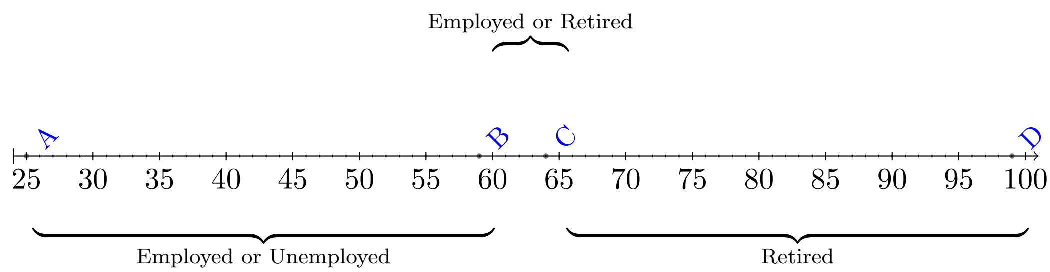
$$U_{i,a}(g,s) = \mathbb{1}[\varepsilon^u_{i,k}(g,s) > u(\bar{a}(a),g,s)] \tag{11}$$

The unemployment rate faced by agents of age group $\bar{a}$, gender g and education s is obtained from the data as the number of unemployed individuals (conditional on age group, gender and education) over the number of individuals in the labour force (of the same age group, gender and education category) ($u(\bar{a},g,s) = \frac{n^u(\bar{a},g,s)}{n^u(\bar{a},g,s)+n^e(\bar{a},g,s)}$.)

Precisions about the Labour market probabilities

Due to the lack of representative samples for some specific age group-gender-education-labour market categories, and following the observed structure of the French labour market, the following assumptions (depicted in the below timeline) are introduced in the model:

From 25 to 59 years old, individuals can be either employed or unemployed but not retired, $\{r(\bar{a}(a),g,s)\}_{a=0,1,\cdots,34} = 0$.



From 59 to 64 years old, agents can be either employed or retired, but unemployment probability is assumed to be zero, i.e. $\{u(\bar{a}(a),g,s)\}_{a=35,36,\cdots,39} = 0$. Finally, from 65 to 99 years old, people can only be retired, thus employment and unemployment probabilities are equalized to zero or $\{e(\bar{a}(a),g,s)\}_{a=40,41,\cdots,74} = 0$ and $\{u(\bar{a}(a),g,s)\}_{a=40,41,\cdots,74} = 0$.

4.2.2 Annual Disposable Income Histories

The Disposable income history of individual i conditional on his labour market status are determined by simulating the ranking of the hypothetical agent in the Distribution of Incomes at each possible age and for the three labour market situations. The inverse of the ECDF of Disposable income (which is age group-gender-education-labour market specific) is applied to each of these simulated rankings to obtain the Income of agent i at each a.

Empirical Cumulative Distribution Functions

As explained in section 3.1, for a given age, gender and education, and conditional on being alive, an individual can be in one of the three mutually exclusive labour market situations, i.e. the agent can be either employed, unemployed or retired. If the agent is employed, he will draw his income out of the distribution of employment income; if he is unemployed, he will do so out of the distribution of unemployment income (mainly composed of unemployment benefits); and following the same logic, if he is retired he will draw his income out of the distribution of retirement income (mostly constituted of pensions.)

Therefore, it would be desirable to estimate the distributions of employment, unemployment and retirement income for each of the 6 gender-education groups and for each of the 75 possible ages at which the agent can be alive (age goes from 25 to 99 years old). However, dividing the sample among 1,350 groups would produce very small (non-representative) subsamples from which it is not possible to estimate reliable ECDFs. To overcome this issue twelve 5-year age groups (from 25-29 years old to 80-84 years old) and one 15-year age group (85-99 years old) are generated (age groups are denoted by $\bar{a}$.) 234 ECDFs that correspond to each one of the 234 possible *age group-gender-education-labour market situation groups* are estimated (with an average sample of 214 observations per group.)

The *Empirical Cumulative Distribution Functions* for the Employment, Unemployment and Retirement Disposable income conditional on age group $\bar{a}$, gender g and education s are represented by the following notation: $\left\{F_{y^e(\bar{a},g,s)}, F_{y^u(\bar{a},g,s)}, F_{y^r(\bar{a},g,s)}\right\}$.

Disposable Employment Income history

$\varepsilon^{y^e}_{i,a}(g,s)$ is the simulated ranking of individual i at age a with gender g and education s in the distribution of disposable employment income. $F_{y^e(\bar{a}(a),g,s)}$ is the empirical cumulative distribution function of disposable employment income for the individuals of age group $\bar{a}(a)$ (which is a function of age), gender g and education s.

Therefore, the disposable income of individual i at age a with gender g and education s when he is employed is equal the inverse of the empirical cumulative distribution function of disposable employment income applied to the modelled ranking of the individual in the distribution of employment income.

$$y^e_{i,a}(g,s) = F^{-1}_{y^e(\bar{a}(a),g,s)}(\varepsilon^{y^e}_{i,a}(g,s)) \tag{12}$$

Disposable Unemployment Income history

In the same line, the disposable unemployment income of individual i at age a with gender g and education s is determined by applying the inverse of the empirical cumulative distribution function of disposable unemployment income $F^{-1}_{y^u(\bar{a}(a),g,s)}$ (which is age group-gender-education specific) to the generated ranking of the individual in the distribution of unemployment income denoted by $\varepsilon^{y^u}_{i,a}(g,s)$.

$$y^u_{i,a}(g,s) = F^{-1}_{y^u(\bar{a}(a),g,s)}(\varepsilon^{y^u}_{i,a}(g,s)) \tag{13}$$

Disposable Retirement Income history

Finally, conditional on gender g and education s, the disposable retirement income of individual i at age a is equal to the inverse of the empirical cumulative distribution function of disposable retirement income ($F^{-1}_{y^r(\bar{a}(a),g,s)}$) applied to the modelled ranking of the individual in the distribution of retirement income ($\varepsilon^{y^r}_{i,a}(g,s)$.)

$$y^r_{i,a}(g,s) = F^{-1}_{y^r(\bar{a}(a),g,s)}(\varepsilon^{y^r}_{i,a}(g,s)) \tag{14}$$

4.2.3 Monte-Carlo Simulation: The beginning of Lifetime Histories

It is now necessary to explain how the simulated latent probabilities and rankings over the lifetime ($a = 0, 1, \cdots, 74$) of individual i are determined, which are denoted as:

$$\left\{\varepsilon^{\pi}_{i,a}, \varepsilon^{r}_{i,a}, \varepsilon^{u}_{i,a}, \varepsilon^{y^e}_{i,a}, \varepsilon^{y^u}_{i,a}, \varepsilon^{y^r}_{i,a}\right\}_{a=0,1,\cdots,74}$$

Or, in terms of the simulated cohort:

$$\varepsilon^{\phi} = \begin{bmatrix} \varepsilon^{\phi}_{1,0} & \varepsilon^{\phi}_{1,1} & \cdots & \varepsilon^{\phi}_{1,A} \\ \varepsilon^{\phi}_{2,0} & \varepsilon^{\phi}_{2,1} & \cdots & \varepsilon^{\phi}_{2,A} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon^{\phi}_{N,0} & \varepsilon^{\phi}_{N,1} & \cdots & \varepsilon^{\phi}_{N,A} \end{bmatrix} \text{ where } \phi \in (\pi, r, u, y^e, y^u, y^r)$$

From random draws out of the uniform distribution, the first latent probabilities and ranks of the life cycle of the agent can be can easily generated and defined as:

$$\left\{\eta^{\pi}_i, \eta^{r}_i, \eta^{u}_i, \eta^{y^e}_i, \eta^{y^u}_i, \eta^{y^r}_i\right\} = \left\{\varepsilon^{\pi}_{i,0}, \varepsilon^{r}_{i,0}, \varepsilon^{u}_{i,0}, \varepsilon^{y^e}_{i,0}, \varepsilon^{y^u}_{i,0}, \varepsilon^{y^r}_{i,0}\right\}$$

where

$$\eta^{\phi}_i \sim \mathcal{U}[0,1] \text{ and } \phi \in (\pi, r, u, y^e, y^u, y^r)$$

Or, for the whole cohort:

$$\eta^{\phi} = \begin{bmatrix} \eta^{\phi}_1 \\ \eta^{\phi}_2 \\ \vdots \\ \eta^{\phi}_N \end{bmatrix} = \begin{bmatrix} \varepsilon^{\phi}_{1,0} \\ \varepsilon^{\phi}_{2,0} \\ \vdots \\ \varepsilon^{\phi}_{N,0} \end{bmatrix} \text{ where } \phi \in (\pi, r, u, y^e, y^u, y^r)$$

Where η^{ϕ} is a vector of independent identically distributed variables from the continuous uniform distribution on [0,1].

Nevertheless, for the subsequent years ($a > 0$) the simulation becomes more complicated due to the fact that Life, Labour Force and Unemployment latent probabilities, as well as the rankings on the distribution of Disposable income are dependent over time. For example, think of an individual in the first decile of the income distribution at age a, the probability that this individual stays in the first decile at age $a + 1$ is higher than the probability of him moving to the ninth decile due to unobserved individual heterogeneity.

Precisions about the Lifetime Histories modelling

In practice, the dynamics of the simulated latent mortality rates and latent retirement probabilities are simply generated by drawing two random matrixes of dimension $N \times A$ of independently and identically distributed random variables from the continuous uniform distribution on the [0,1] interval:

$$\varepsilon^{\pi} = \begin{bmatrix} \varepsilon^{\pi}_{1,0} & \varepsilon^{\pi}_{1,1} & \cdots & \varepsilon^{\pi}_{1,A} \\ \varepsilon^{\pi}_{2,0} & \varepsilon^{\pi}_{2,1} & \cdots & \varepsilon^{\pi}_{2,A} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon^{\pi}_{N,0} & \varepsilon^{\pi}_{N,1} & \cdots & \varepsilon^{\pi}_{N,A} \end{bmatrix}, \ \varepsilon^{r} = \begin{bmatrix} \varepsilon^{r}_{1,0} & \varepsilon^{r}_{1,1} & \cdots & \varepsilon^{r}_{1,A} \\ \varepsilon^{r}_{2,0} & \varepsilon^{r}_{2,1} & \cdots & \varepsilon^{r}_{2,A} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon^{r}_{N,0} & \varepsilon^{r}_{N,1} & \cdots & \varepsilon^{r}_{N,A} \end{bmatrix}$$

On the other hand, Copula modelling will be used to simulate the dynamics of the rankings in the employment, unemployment and retirement income distributions and to model the latent probabilities of unemployment over time:

$$\varepsilon^{\psi} = \begin{bmatrix} \varepsilon^{\psi}_{1,0} & \varepsilon^{\psi}_{1,1} & \cdots & \varepsilon^{\psi}_{1,A} \\ \varepsilon^{\psi}_{2,0} & \varepsilon^{\psi}_{2,1} & \cdots & \varepsilon^{\psi}_{2,A} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon^{\psi}_{N,0} & \varepsilon^{\psi}_{N,1} & \cdots & \varepsilon^{\psi}_{N,A} \end{bmatrix} \quad \text{where } \psi \in (y^e, y^u, y^r, u)$$

4.2.4 Copula Functions Modeling: The Dynamics of Lifetime Histories

As mentioned before, Copula Functions are used to tackle the issue of dependency over time of both the latent probability of unemployed and the ranking in the distribution of incomes. Denote any of these variables in t as ε_t (observed) and the same variable in $t+1$ as ε_{t+1} (unobserved). This section models ε_{t+1} given ε_t and a dependency structure between these two variables defined by a Copula.

Copula Functions

Copulas are very useful tools to model Joint Cumulative Distribution Functions (JCDF.) A Copula is a joint cumulative distribution function of the ranks of n random variables in their respective marginal distributions. A bivariate Copula can be defined as a mapping $C : [0,1]^2 \rightarrow [0,1]$ (see Bonhomme and Robin (2003); and Chau (2009) for details.)

As shown in the expression below (Equation 15), the JCDF of two random variables (e.g. y_{t+1} and y_t) can be represented as the Copula of the marginal CDFs ($F_{Y_{t+1}}$ and F_{Y_t}) applied to their corresponding variables.

$$F(y_{t+1}, y_t) = C(F_{Y_{t+1}}(y_{t+1}), F_{Y_t}(y_t)) = C(\mathcal{U}_{y_{t+1}}[0,1], \mathcal{U}_{y_t}[0,1]) = C(\varepsilon_{t+1}, \varepsilon_t) \qquad (15)$$

where

$$\varepsilon_{t+1} = F_{Y_{t+1}}(y_{t+1}) \sim \mathcal{U}[0,1] \text{ and } \varepsilon_t = F_{Y_t}(y_t) \sim \mathcal{U}[0,1]$$

From the data, y_t and F_{Y_t} are known (assume $t = 2005$ and l=employed, then $y_t = y^e_{i,a}(g,s)$ and $F_{Y_t} = F_{y^e_{(\bar{a}(a),g,s)}}$) and ε_t can be easily obtained as $F_{Y_t}(y_t)$, while y_{t+1} and $F_{Y_{t+1}}$ (and consequently ε_{t+1}) are unknown. Nevertheless, using the property of Copulas described by Equation 16, and under certain conditions, it is possible to estimate ε_{t+1}.

$$C(\varepsilon_{t+1}|\varepsilon_t) = \frac{\partial C(\varepsilon_{t+1}, \varepsilon_t)}{\partial \varepsilon_t} \sim \mathcal{U}[0,1] = w \tag{16}$$

To apply this property it is necessary to choose the Copula (with its specific parameters) that better describes the data and to generate w by random draws from the continuous uniform distribution in [0,1]. After doing this, it will be straightforward to obtain ε_{t+1} as $C^{-1}(w|\varepsilon_t) = \varepsilon_{t+1}$. Once ε_{t+1} has been estimated and under the assumption that $F_{Y_{t+1}} = F_{y^e_{(\bar{a}(a+1),g,s)}}$ it will be possible to obtain y_{t+1} (or $y^e_{i,a+1}(g,s)$.)

Plackett's Copula

Bonhomme and Robin (2009) tested several copulas using short panels from the French LFS data (from 1990 to 2000) and found that the one-parameter Plackett's copula was the most appropriate function to model French mobility in employment earnings. Even though they use this copula only to model employment earnings over time, in this paper Plackett's copula is also used to model unemployment and retirement incomes over time, as well as the dynamics of the latent probabilities of unemployment (i.e. $\{\varepsilon^{\psi}_{i,a}\}_{a=0,1,\cdots,74}$ where $\psi \in (y^e, y^u, y^r, u)$.) The parameter of dependency τ of the Plackett's copula will be different for the unemployment process and the income dynamics. Nevertheless, it will be assumed to be the same for the employment, unemployment and retirement income processes (i.e. $\tau^u \neq \tau^{y^e} = \tau^{y^u} = \tau^{y^r}$.) The *Plackett's one-parameter Copula* is defined by Equation 17 as:

$$C(\varepsilon_{t+1}, \varepsilon_t; \tau) = \frac{1}{2}\tau^{-1}\{1 + \tau(\varepsilon_{t+1} + \varepsilon_t) - [(1 + \tau(\varepsilon_{t+1} + \varepsilon_t))^2 - 4\tau(\tau+1)\varepsilon_{t+1}\varepsilon_t]^{\frac{1}{2}}\} \tag{17}$$

The parameter τ represents the dependency between rankings over 1 year. In other words, a high τ implies low mobility in earnings over 1 period, while a low τ represents high mobility in disposable income over one year. The parameter τ can also be related to other rank correlation measures. For example, Spearman's ρ, which is a popular coefficient of rank correlation, is an increasing function of τ.

By applying the property described in Equation 16 to the Plackett's copula, the following expression is obtained:

$$\frac{\partial C}{\partial \varepsilon_t}(\varepsilon_{t+1}, \varepsilon_t; \tau) = \frac{1}{2}\left[1 - \frac{1 + \tau(\varepsilon_{t+1} + \varepsilon_t) - 2(\tau+1)\varepsilon_{t+1}}{[(1 + \tau(\varepsilon_{t+1} + \varepsilon_t))^2 - 4\tau(\tau+1)\varepsilon_{t+1}\varepsilon_t]^{\frac{1}{2}}}\right] = w \sim \mathcal{U}[0,1] \tag{18}$$

Which, with a given ε_t, w and τ, can be solved to obtain ε_{t+1}. As explained before, ε_t and w can easily be obtained; hence, to solve 18 only τ remains to be determined. The following section explains how the parameters of dependency for the income and unemployment dynamics, τ^{y^e} and τ^u respectively, are estimated.

4.3 Estimation

4.3.1 A Minimum Distance Estimation (MDE)

The Plackett's copula parameter of dependency for the income dynamics τ^{y^e} is estimated using information on the Spearman's ρ over 1 and 2 years observed in the data (denoted $\rho_{t+1,t}$ and $\rho_{t+2,t}$ respectively), while the parameter of dependency for the latent probability of unemployment over time τ^u is obtained making use of previous estimates on average unemployment duration (UD). The Minimum Distance Estimation uses a search algorithm that starts with arbitrary low and positive values for τ^{y^e} and τ^u and iterates until the simulated data minimizes the quadratic distance between the values of $\rho_{t+1,t}$, $\rho_{t+2,t}$ and UD predicted by the model and the ones calculated with the real data:

$$arg\, min_{\tau^{y^e},\tau^u}[(\hat{\rho_{t+1,t}} - \rho_{t+1,t})^2 + (\hat{\rho_{t+2,t}} - \rho_{t+2,t})^2 + (\hat{UD} - UD)^2]$$

4.3.2 Estimation results on Copula parameters

Table 2 presents the estimated parameters of the Plackett's copulas for the income (τ^{y^e}) and the latent probability of unemployment dynamics (τ^u.) It also shows the Spearman's rank correlation coefficients over 1 and 2 years and the unemployment duration for both the observed and the simulated data.

The estimated value for τ^{y^e}, the parameter of dependency between rankings in the employment earnings distribution, is equal to 103.41. This parameter generates a rank correlation over 1 year (estimated with the simulated data) around 0.928, which suggests a very low mobility on the rank of the income distribution over 1 year. This estimate is in line with the 0.892 average rank correlations over 1 year found by Bonhomme and Robin (2009) using the French Labour Force Surveys from 1990 to 2000. The rank correlation over 2 years estimated by the model is 0.861, implying that mobility on the income distribution is higher over 2 years than over 1. This value is also very close to the 0.876 obtained by Bonhomme and Robin (2009). However, the gap between the correlation coefficients over 1 year and 2 years is higher in the model here presented than in the one by the aforementioned authors.

Table 2: Estimation of Copula Parameters

Average unemployment duration (years)		Correlation of ranks in earnings - 1 year lag		Correlation of ranks in earnings - 2 years lag		Copula Estimates	
Observed (UD)	Predicted ($\hat{UD}$)	Observed ($\rho_{t+1,t}$)	Predicted ($\hat{\rho_{t+1,t}}$)	Observed ($\rho_{t+2,t}$)	Predicted ($\hat{\rho_{t+2,t}}$)	Unemployment (τ^u)	Income (τ^{y^e})
1.824	1.826	0.892	0.928	0.876	0.861	13.578	103.141

The estimated parameter of dependency for the latent probability of unemployment τ^u is calculated around 13.578, this value (along with the other parameters of the model) predicts an average unemployment duration of 1.826 years, consistent with the average unemployment duration of 1.824 found by Murtin et al. (2014). This value implies that every time an individual falls into unemployment it takes him in average 22 months to get out of this labour market situation.

The first panel of Figure 3 shows the dependency between the ranks on the disposable income distribution in t and $t+1$. The mass of points is concentrated around the 45-degree line, which denotes a positive and strong correlation. This can also be thought as low mobility in the rank of the income distribution over 1 year. For example, an individual in the first decile of the income distribution at age 25 is more likely to remain in the same decile at the age of 26 than to move to a higher decile.

The second panel presents the dependency between the latent probability of being unemployed in t and in $t+1$. As the graph shows, the correlation between these variables is positive and close to 1.

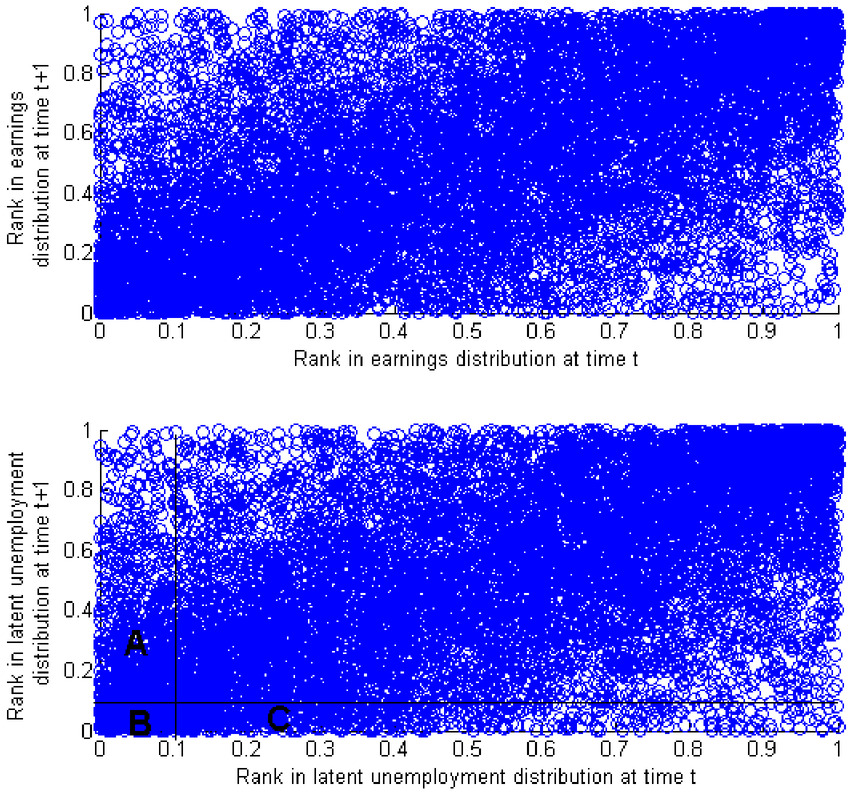


Figure 3: Dependency between the ranks in t and $t+1$

To illustrate this dependency, the second panel of Figure 3 shows three areas that denote the mobility on unemployment status over 1 year. Area A, comprehends the individuals that were unemployed in t and got out of unemployment in $t+1$. Area B shows the agents that were unemployed in year t and remained unemployed in $t+1$. Finally, area C encloses the individuals that were employed in t and fell into unemployment in $t+1$.

5 Results

After generating the Lifetime Histories of the hypothetical individuals, the average values and distributions of the measures of lifetime disposable and equivalent income (defined in Sections 3.2 and 3.3) for the whole cohort and by gender-education group are estimated. The main results of these calculations are summarized in the following subsections.

5.1 Monthly Equivalent and Disposable Income

5.1.1 Average Monthly Equivalent and Disposable Income

As shown in Table 3, groups with higher education tend to have higher lifetime utility and disposable income than the groups with lower education (regardless of the measure of lifetime disposable income it's used.) Conditional on the educational level, males have higher welfare and lifetime earnings than women. This result might not only be explained by the differences in years of schooling across genders, but also by the differences in unemployment risk between males and females. Conditional on education being lower than tertiary, the unemployment risk faced by women is higher than the one faced by men.

Table 3: Average Equivalent and Disposable lifetime incomes by group (monthly)

	Males			Females			All
	Primary	Secondary	Tertiary	Primary	Secondary	Tertiary	Sample
Average employment income during employed lifetime (€)	1501.1	1756.8	2579.1	1105.8	1351.6	1909.7	1777.1
Unemployment rate (%)	11.9	7.7	6.4	16.0	9.9	5.3	8.8
Average disposable income during lifetime (€)	1385.3	1686.7	2549.7	985.1	1258.8	1834.5	1696.7
Life Expectancy at age of 25 (years)	48.2	51.7	54.9	56.2	58.8	59.1	54.9
Average disposable income over average life expectancy (€)	1299.4	1644.5	2563.6	995.2	1298.0	1901.5	1700.2
Average equivalent income (€)	901.1	1243.7	2038.3	830.6	1159.1	1763.2	1401.0

On the other hand, women have higher life expectancy at the age of 25 than men. The gain in lifetime utility and disposable income for women due to their higher average life expectancy can be partially seen when comparing disposable income during lifetime to disposable income during average life expectancy by gender. For women, disposable income over average life expectancy is higher than disposable income during lifetime, whereas for men the opposite is true. However, the gains in lifetime income due to a higher life span seem to not be enough to outweigh the negative effects of higher unemployment risks bear by women with less than tertiary education.

Conditional on gender-education group, monthly lifetime employment income while being employed is always higher than monthly lifetime disposable income, this is due to the lower (in average) income the individual gets when he falls into unemployment or when he is retired during his life cycle.

Finally, for all gender-education groups, monthly equivalent income is always lower than monthly lifetime disposable income. This means that individuals could reach the same lifetime utility predicted by the model, with less monthly income if unemployment risk were suppressed and if probability of survival became deterministic instead of being stochastic. In other words, the fluctuations in monthly earnings due to the probability of falling into unemployment and/or the probability of dying at early ages generates very low lifetime utilities that can be satisfied with low but constant long-term streams of earnings.

5.1.2 Inequality in Monthly Equivalent and Disposable Income

Two measures of inequality are estimated to study inequality in equivalent and lifetime disposable income: the Gini coefficient and the Theil index. Both measures and the decomposition of the latter into its between and within-group components, and into the Theil indexes by gender-education group are presented in Table 4.

Table 4: Inequality Measures

	2005 Cross-sectional Disposable Income	Disposable Income during lifetime	Disposable Income over life expectancy	Equivalent Income
Gini	0.306	0.211	0.226	0.330
Theil	0.164	0.072	0.082	0.180
Between-group component	0.040	0.041	0.043	0.050
Within-group component	0.124	0.031	0.039	0.130
Males - Less than High School	0.103	0.024	0.036	0.170
Males - High School	0.121	0.029	0.041	0.160
Males - More than High School	0.155	0.038	0.046	0.155
Females - Less than High School	0.129	0.028	0.035	0.102
Females - High School	0.123	0.028	0.032	0.092
Females - More than High School	0.115	0.031	0.034	0.098

The 2005 Cross-sectional disposable income variable refers to the disposable income of the sample extracted from the LFS-SILC data. This sample is restricted to the active people (i.e. employed, unemployed or retired) from 25 to 99 years old observed in 2005. The Gini coefficient for this variable is 0.306, very close to the Gini of disposable income for France in 2005 (equal to 0.288) reported by the OECD (2014b.)

The common result from the literature showing that inequality in lifetime income is lower than cross-sectional inequality in income can be corroborated by the results presented in Table 4. For example, the Gini coefficient for the cross-sectional disposable income is approximately 45% higher than the Gini of disposable income during lifetime and 35% higher than the Gini of disposable income over life expectancy.

According to the Theil index decomposition, the previous result is mainly explained by a sharp (75%) decrease in the value of within-group inequality, since between-group inequality remains constant around 0.04. In other words, the life cycle process contributes to the reduction of inequality within groups, but it doesn't seem to have an impact on inequality across them.

On the other hand, when looking at equivalent income both within and between-group inequality, these seem to increase due to the life cycle with respect to cross-sectional income inequality. This is a striking result since it challenges the previous notion that lifetime inequality should be lower than cross-sectional inequality. As shown in Table 4, while inequality in disposable income decreases (31% for Gini and 56% for Theil) due to the inclusion of the life span (with respect to cross-sectional income inequality), inequality in equivalent income increases (7% for Gini and 9% for Theil) when taking into account the lifetime dimension. Hence, inequality in the monetary equivalent of lifetime utility, namely equivalent income is not only higher than inequality in lifetime earnings, but also than cross-sectional income inequality.

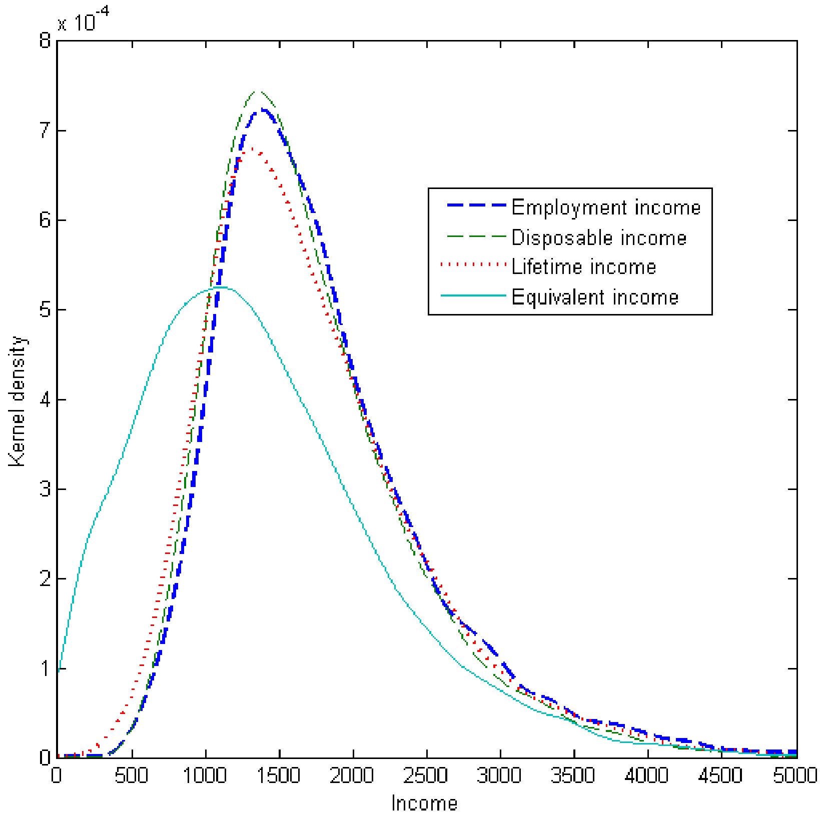


Figure 4: Distribution of Equivalent and Disposable lifetime incomes

Although all of the distributions depicted in Figure 4 are right-skewed, the distribution of equivalent income is by far the most positive-skewed. There is also a concentration of individuals very close to the zero equivalent income (depicted by a fat left tail), the mode is around the 1,100 euros per month and there is a long right tail that almost reaches the same high values as the other distributions. The probability density function (PDF) of the equivalent income also has a higher variance than the PDFs of the different measures of lifetime disposable income.

Figure 4 clearly shows that the distribution of equivalent income is the most unequal of all presented distributions. This is due to the fact that the equivalent income integrates into one measure the monetary values of disposable income, longevity and job security, which are (the three of them) positively correlated (see Table 1.) In other words, while the distribution of disposable income during lifetime only shows the inequality in lifetime earnings, the equivalent income also takes into account inequality in health and in job security and represents all these dimensions into one single (monetary) dimension. The fact that these variables are positively correlated (e.g. the low educated population not only has lower average incomes compared to the highly educated population, but also lower average life expectancy and higher average unemployment risk) intensifies disparities among individuals and across groups.

Inequality in Monthly Equivalent and Disposable Income by group

Figure 5 presents the PDFs of equivalent and disposable income for each of the 6 gender-education groups. For both measures of earnings, men's distributions seem to be more dispersed. As expected, conditional on gender, the higher the education level of the group the further to the right its distribution is placed.

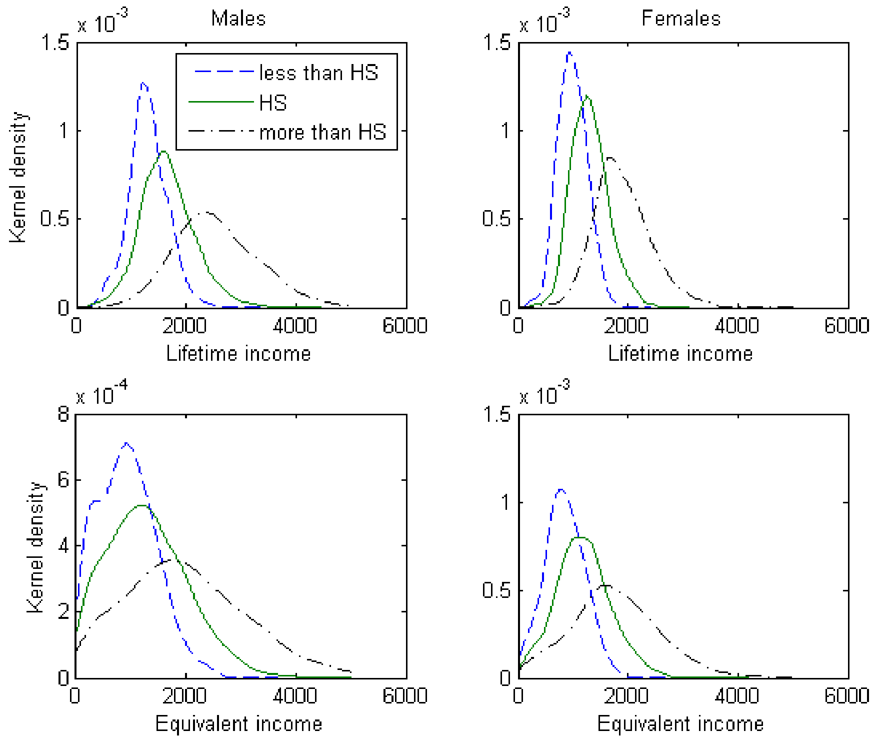


Figure 5: Distribution of Disposable income over life expectancy and Equivalent income by Gender-Education group

From the decomposition of the distribution of equivalent income by groups it is also possible to see that most of the variance, the fat left tail and the right-skewedness of the whole distribution of equivalent income (presented in Figure 4) is due to the distribution of equivalent income for males. In fact, the lowest values of equivalent income are associated to individuals who died at very early ages, which is more likely to happen to men due to their lower survival probabilities (compared to women) and to the less educated (conditional on gender.) The lifetime utility of the individuals who died at very young ages is so low that these individuals would have been indifferent between the simulated stochastic (hypothetically current) situation and an alternative scenario where they get very low deterministic monthly incomes but for a longer life span.

From Figure 5 it is possible to infer that most of the inequality in equivalent income it's being generated by inequality in both disposable income and longevity. In sum, the distribution of equivalent income displays both a fat left tail due to premature deaths and a skewed right tail due to a positive correlation between disposable income and longevity. Inequality in equivalent income is much higher than inequality in lifetime disposable income mainly due to inequality in longevity.

5.2 Returns to Education

Assuming that most of the variation in lifetime earnings across education groups, conditional on gender, is due to the effect of schooling, it is possible to estimate the pseudo Mincerian returns to education as $\frac{100}{k}\left[\ln y_{t+k} - \ln y_t\right]$. The average returns in employment income to education are 6.28% for males and 6.62% for females; whereas, the average returns in equivalent income to education are 10.77% for males and 9.9% for females.

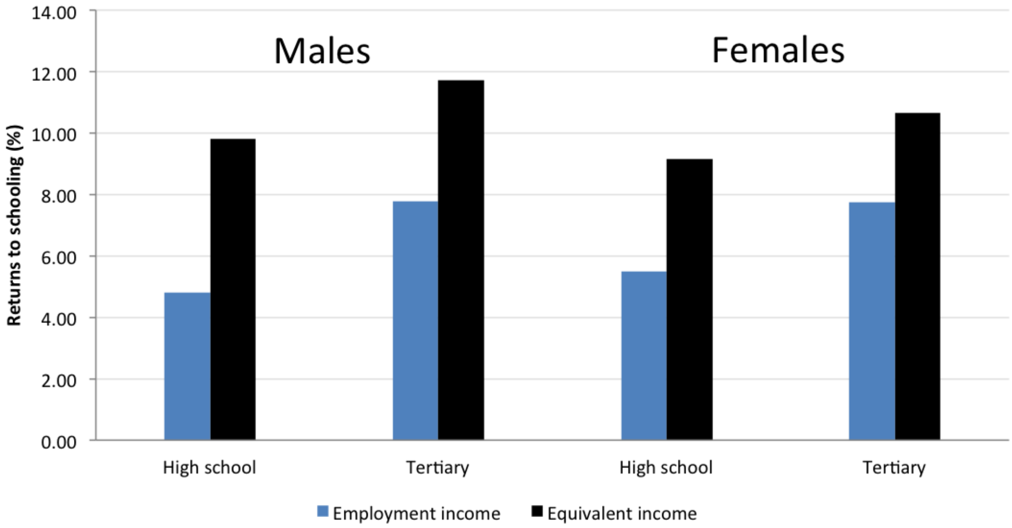


Figure 6: Returns to Schooling

In other words, for males (values for females are in parentheses), 1 extra year of education would increase monthly disposable income in 6.28% (6.62%) and monthly equivalent income in 10.77% (9.9%.) This would represent a 4.48 (3.28) percentage point gap on the returns to schooling between equivalent and disposable income, which suggests (assuming no confounding variable) that the benefits in welfare to education are higher for men than for women due to the higher improvements in survival probability that education generates on males compared to females.

These results are merely an illustration of the likely underestimation of the benefits to schooling that arises when ignoring the value of longevity and job security. When working with these estimates it is important to keep in mind the assumption in which they rely (i.e. no significant omitted variable bias.) Another relevant remark is that in the model, education only impacts lifetime utility (indirectly) through the effect it has on lifetime earnings; however, it is also possible to think of a framework where education enters the utility function directly (Stiglitz et al., 2010) in which case, the gap on the returns to schooling between equivalent and disposable income would be even higher.

6 Conclusions

This paper has simulated the lifetime histories form 25 to 99 years old of a synthetic 1980's French birth cohort to construct different measures of lifetime earnings as well as a measure that incorporates the value of longevity and unemployment risk, namely equivalent income. The results indicate that inequality in the monetary equivalent of lifetime utility is much higher than inequality in lifetime earnings and also higher than cross-sectional income inequality. This occurs because lifetime utility accounts for longevity, which is positively correlated with income, so that welfare inequality is the by-product of income and longevity inequalities that appear to add up rather than mitigate each other. Looking at the average equivalent income by educational attainment, it is found that the returns in welfare to education, which cumulate the earnings, employment and longevity premiums of higher schooling, are on average 65% higher than the standard income returns to education, suggesting that ignoring longevity and unemployment differential risks across educational groups may result in a non-negligible underestimation of the actual benefits of education.

On the other hand, it is important to remark some of the drawbacks of this work for the purposes of future research. First, lifetime utility, which is determined by income, education, gender, mortality and unemployment risk, has been used synonymously as welfare. However, there are other characteristics and dimensions that determine welfare and that have not been taken into account in the model. Second, the non-material dimensions incorporated into this work, namely education, longevity and unemployment risk have been valued in equivalent income terms only through the indirect effect they have in utility by the means of determining lifetime earnings, while in fact, some of these variables might have a direct impact in utility. Third, the model also disregards part of the dependency over time between different simulated variables (e.g. the dependency over time between the modelled mortality risk and the simulated unemployment risk.) Finally, in this study welfare has been obtained only at the individual level, while a measure of social welfare (i.e. a Social Welfare Function) should be considered for a thorough welfare analysis.

7 Bibliography

Becker, G. S., Philipson, T. J., and Soares, R. R. (2005). The Quantity and Quality of Life and the Evolution of World Inequality. *American Economic Review*.

Boarini, R., Cordoba, J., Murtin, F., and Ripoll, M. (2013). Beyond GDP: Is There a Law Of Shadow Price? *OECD Working Paper*.

Bonhomme, S. and Robin, J.-M. (2003). Modeling Individual Earnings Trajectories using Copulas with an Application to the Study of Earnings Inequality: France, 1990-2012. *Working Paper*.

Bonhomme, S. and Robin, J.-M. (2009). Assesing the Equalizing Force of Mobility Using Short Panels: France, 1990-2000. *Review of Economic Studies*.

Bowlus, A. J. and Robin, J.-M. (2004). Twenty Years of Rising Inequality in U.S. Lifetime Labour Income Values. *Review of Economic Studies*.

Chau, T. W. (2009). Estimating Mobility and Lifetime Inequality of Labor Income in the United States and Germany Using Extensions of the Copula Approach. *Job Market Paper*.

Díaz, M. and Murtin, F. (2014). The French Welfare Distribution. *OECD Working Paper*. Forthcoming.

Fleurbaey, M. and Gaulier, G. (2007). International Comparisons of Living Standards by Equivalent Incomes. *CEPII Working Paper*.

Murphy, K. M. and Topel, R. H. (2006). The Value of Health and Longevity. *Journal of Political Economy*.

Murtin, F., de Serres, A., and Hijzen, A. (2014). Unemployment and the coverage extension of collective wage agreements. *European Economic Review*.

OECD (2013). How's Life? 2013: Measuring Well-being. *OECD Publishing*. `http://www.oecd-ilibrary.org/economics/how-s-life-2013_9789264201392-en`.

OECD (2014a). All on Board: Making Inclusive Growth Happen. `http://www.oecd.org/inclusive-growth/all-on-board-making-inclusive-growth-happen.pdf`.

OECD (2014b). OECD Income Distribution database. `www.oecd.org/social/inequality.htm`.

Stiglitz, J. E., Sen, A., and Fitoussi, J.-P. (2010). Report by the Commission on the Measurement of Economic Performance and Social Progress. *Technical Report*. `http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf`.

van Raalte, A. A., Mackenbach, J. P., , Kunst, A., Lundberg, O., Leinsalu, M., Martikainen, P., Artnik, B., Deboosere, P., Stirbu, I., and Wojtyniak, B. (2012). The Contribution of educational inequalities to lifespan variation. *Population Health Metrics*.