

How long do situations of single parenthood last? An estimation based on French data

Vianney Costemalle *

Abstract – Single-parent families currently account for over 20% of families with minor children in France, in line with the European average. Single parenthood is often associated with greater risks of insecurity and exclusion, to which social policies must respond. It is thus important to know how long such situations last. In this paper, we present an original method for estimating the duration of such periods based on a sample of single parents for whom only the length of time spent in the situation at the time of the survey is observed (stock sampling). It combines a calculation of the likelihood function of the observations using the methodology proposed by Nickell and the introduction of proportional instantaneous probability of exiting the situation based on the Cox model. Several simulations replicating a variety of observed scenarios confirm the reliability of this method. Applying this method to the data from the 2011 *Family and Housing Survey* allows us to estimate that single parenthood ends after 3 years for half of the single parents.

Codes JEL / JEL codes: J12, C14, C15, C41

Keywords: single parenthood, survival models, duration models

Reminder:

The opinions and analyses in this article are those of the author(s) and do not necessarily reflect their institution's or Insee's views.

* Insee, Division of Demographic Surveys and Studies (vianney.costemalle@insee.fr).

Received October 10, 2016; accepted after revision March 31, 2017

DOI: 10.24187/ecostat.2017.493s.1914

Translated from the original version: « *Combien de temps durent les situations de monoparentalité ? Une estimation sur données françaises* »

Single parenthood, i.e. the situation characterised by a parent – most often a mother – living alone with one or more dependent children, has become increasingly common in recent decades¹ both in France and in other countries. In France, according to Insee's population census², there were 953,000 single-parent families in 1990 and 1,687,000 in 2010, that is an increase of 77% in 20 years. In parallel, the number of families with minor children has not changed significantly (7,944,000 in 2010, i.e., 3.8% more than in 1990); hence the proportion of single-parent families as a share of all families with minor children has significantly increased from 12% in 1990 to 21% in 2010 and has been constantly on the rise for several decades. In Europe, this trend has been observed in Great Britain since the 1970s (David et al., 2004) and then spread to all the other countries in the 1980s. By 2012, the share of single-parent families in Europe had reached 19%. There are stark contrasts between Northern European countries and Eastern and Southern European countries; traditional family structures are still well established in the latter, even though single parenthood has increased significantly there (Le Pape et al., 2015).

Single-parent families are, more often than other types of families, facing difficult situations: they are the category of family with the highest poverty rate, standing at 33% in France (Boiron et al., 2016), and they are more likely to live in poor living conditions (Chardon et al., 2008); although single mothers' activity rate is a little higher than that of mothers living with a partner, their unemployment rate is twice as high (Rabier, 2014) and they are confronted with specific difficulties in terms of striking a work-life balance (Algava et al., 2005). In order to address these risks of professional and social exclusion, various social policies have been implemented in Europe since the 1950s (Eydoux & Letablier, 2009). In France, the family support allowance (ASF) and the family supplement allowance (CF) have been created to supplement the various other family allowances. It is thus important to know how long periods of single parenthood last.

Single parenthood is necessarily a temporary situation, since it comes to an end, either when the single parent finds a new partner, or when the children reach an age where they are no

longer regarded as dependents or leave home. However, there are still few statistics relating to the duration of periods of single parenthood. The aim of this paper is to provide an estimation of the duration of single parenthood, which, to our knowledge, has never been done before in France. To do this, we propose an original method for estimating this duration based on the time spent as a single parent at a given time of observation ("seniority" of the situation) and on existing duration modelling approaches, in particular the research by Nickell (1979) and Cox (1972).

Before proceeding, it should be pointed out that there is no universal definition of single parenthood. All the approaches refer to a single parent who raises dependent children, but several different criteria are used for each of these two dimensions. For many years, the single parent criterion was based on the legal marital status of the parent: unmarried women (or men) with children were regarded as single parents. This approach is problematic nowadays, as increasing numbers of couples are not married. Moreover, some couples, referred to as LAT couples (living apart together), do not live under the same roof. For example, 10% of single parents in France in 1999 reported being a couple with a partner who did not live in the same dwelling (Algava, 2002). Marital status is thus no longer a good indicator of whether a parent is single or not and the notion of 'living together as a couple' has become more difficult to grasp (Toulemon, 2011). There are also several definitions of a dependent child: the most common is any child under the age of 25 (for example, Algava, 2002) or under the age of 18 (Buisson et al., 2005), but some research does not set an age limit (David et al., 2004). When analysing family types, Insee generally uses the notion of minor children (under the age of 18), as does the High Council for the Family (Haut Conseil de la famille, 2014). We will also use this age limit of 18, and, for the purposes of this paper, we define single parenthood as the situation in which a parent is not living with a partner and is living with at least one child, who himself or herself neither lives with a partner, nor has any children living in the dwelling.

The first part of the paper sets out the estimation method. We first explain how the seniorities observed at a given time among the stock of single parents differ from the total durations of single parenthood: seniorities can then be regarded as doubly biased durations. We then explain how it is possible to infer the

1. Single-parent families appeared for the first time as a statistical category in France in the 1982 population census.

2. By this, we mean single-parent families with dependent children aged under 18.

distribution of durations from the distribution of seniorities using information on the flows of individuals entering a situation of single parenthood. The approach here is based on numerical simulations, generating random variables based on the model presented, in order to illustrate different specific scenarios and to test the robustness of the proposed method. In the second part, this method is applied to the data from the *Family and Housing Survey (FHS)*, using in addition the data from the *Survey of Family and Intergenerational Relations (Erfi)* (see Box 1 for a presentation of these sources); we then present our estimation of the distribution of the durations of single parenthood.

Determining the duration of single parenthood on the basis of seniority

Little is known about the duration of periods of single parenthood because measuring it presents many difficulties: the periods of single parenthood are deduced by comparing the periods when the individuals are living alone (i.e. without a partner) and the periods when they have dependent children living with them. In order to know the dates on which these periods started

and ended, longitudinal (or retrospective) data are required, but such data are seldom available. Mean durations have been measured using such retrospective data in the United States, based on the 1987 *National Survey of Families and Households*: on average, mothers who entered situations of single parenthood between 1970 and 1974 remained in that situation for 4.5 years, compared to 3.4 years for those who entered such a situation 10 years later (Bumpass & Raley, 1995). In the United Kingdom, the retrospective *Survey of Family and Working Lives* from 1994 was used to estimate median durations of single motherhood: this median was 5.8 years on average, 4.6 years for single mothers, 4.7 years for divorced mothers, 6.8 years for separated mothers and 10.5 years for widowed mothers (McKay, 2002).

More standard cross-sectional surveys provide information about the stock of single-parent families at a given point in time. On this basis, it is possible to measure the seniority of the situation, but not its total duration. In France, the *Study of Family History* was used to measure the mean seniority for 1999: 6 years and 3 months for women and 5 years and 9 months for men (Algava, 2002). Based on the 2011

Box 1

THE FAMILY AND HOUSING SURVEY AND THE SURVEY ON FAMILY AND INTERGENERATIONAL RELATIONS

The *Family and Household Survey (FHS)* was conducted by Insee in early 2011 in conjunction with the population census. Within each household selected for the survey, all the men and all the women over the age of 18 were surveyed. In total, 359,770 people aged 18 or over, living in private households in mainland France were surveyed about their family life and life at home (Breuil et al., 2016). This survey can be used to identify single-parent families on 1 January 2011. It is an example of stock sampling. The persons who declared that they did not live with a partner at the time of the survey, but who had already previously lived with a partner, were asked to indicate the year the relationship ended, as well as the reason for the termination (separation or death of the spouse). This information is also used to determine the length of time already spent in the situation of single parenthood at the time of the survey (seniority) and the reason (separation, child born outside a relationship, death of spouse) for the situation.

The *Survey of Family and Intergenerational Relations (Erfi)* is the French contribution to the international *Generations and Gender Survey (GGS)* programme, the goal of which is to establish comparable country-by-country statistics at global level, and primarily at

European level (Régnier-Loilier, 2012). It was conducted jointly by Ined and Insee in three successive waves in 2005, 2008 and 2011. For the first wave, there were 10,079 respondents; for the second, 6,534 and, finally, for the third wave, 5,781 (including some responded to the first and third waves, but not the second). The respondents were aged 18 to 79 on 31 December 2005, with only one person interviewed per household. The added value of this survey is that it asks retrospective questions, which allows for longitudinal information about family and marriage histories to be obtained. The respondents describe all of the relationships (that lasted for at least three months) during which they lived with a partner. They also provide information about all the children they have had, including the date on which each child left home. Based on this information, it is possible to determine the periods during which the respondents were the head of a single-parent family comprising children under the age of 18. The drawbacks of this survey are the relatively small sample of persons who have been a single-parent at least once and possible inaccuracies in the dates provided by the respondents – as the questions sometimes referred to the distant past – which could lead to a lack of accuracy in terms of the periods of single parenthood.

Family and Housing Survey, it was ascertained that 1.5 million single-parent families in France had, on average, been formed 5.5 years earlier: 4.5 years for a parent separated from his/her partner, 5.5 years for a widowed parent and 10 years for a parent who had never lived with a partner (Buisson et al., 2015).

The two types of approach - longitudinal and transversal - relate to what we call respectively *flow sampling* and *stock sampling* within the framework of duration modelling.

From seniority to duration: the censoring bias and the selection bias

The seniority of a situation is the length of time from the beginning of this situation to the time when the situation is observed³, whereas

3. The time the observation is made will, in our case, be identical for all the individuals, i.e., the date of the survey.

the duration is the total time between the beginning and end of the situation (see Box 2 for the presentation of duration modelling). Seniority and duration are thus *a priori* two different concepts that address respectively the following questions: ‘How long have you been in this situation on date t ?’ and ‘How long did this situation last?’ Seniority can be regarded as a right-censored duration, since we do not know the date on which the situation will end, but only the date when it began. Seniorities are consequently shorter than durations, which is what we refer to here as the censoring bias. There is another problem with stock sampling: the probability of taking part in a survey while in a situation of single parenthood increases with the duration of single parenthood. This implies that the single parents observed at the time of the survey will, on average, experience longer periods of single parenthood (longer durations) than all of those who have experienced one spell of single parenthood during their

Box 2

MODELLING OF THE DURATIONS

The random duration variable T is discrete and the law of T is given by its density $f(t) = P(T = t)$.

Survival at time $t \in \mathbb{N}$, written $S(t)$, corresponds to the proportion of the population whose situation of interest lasted t units of time or more: $S(t) = P(T \geq t)$. Consequently, S is a decreasing function and $S(0) = 1$.

The instantaneous probability at time $t \in \mathbb{N}$, written $h(t)$, corresponds to the share of persons that exit the situation at time t from among those persons who were still in the situation at time t : $h(t) = P(T = t | T \geq t)$. i.e., $f(t) = h(t)S(t)$.

We thus have a relation between the survival function and the hazard function: for any $t > 0$,
$$S(t) = \prod_{u=0}^{t-1} (1 - h(u)).$$

It should be noted that $f(t) = S(t) - S(t+1)$ can be deduced from the expected value of the duration variable which is equal to $E[T] = \sum_{t \geq 0} t f(t) = \sum_{t \geq 1} S(t)$.

The median is defined here as $Med(T) = \frac{u(1 - S(u) - S(u+1)) + 0,5 - S(u)}{S(u) - S(u+1)}$, where u is the greatest integer with $S(u) \geq 0,5$. Which does indeed give us $Med(T) = u$ when $S(u) = 0,5$.

Knowing the law of probability, the survival function or the hazard function amounts to the same thing and provides all the information about the duration distribution. It is therefore sufficient to know one of these three functions in order to be able to calculate any indicator, such as the mean or the median.

For the continuous function, we use the Weibull distribution to simulate duration variables. In such a case, we can define the survival function and the hazard function in the same way as for the discrete function.

This gives us for any $t \in \mathbb{R}$,
$$S(t) = \exp\left(-\int_0^t h(u) du\right).$$

The Weibull distribution is set using two true positive parameters: a scale parameter λ and a shape parameter k .

The survival is expressed as $S(t) = \exp\left(-\left(\frac{t}{\lambda}\right)^k\right)$ and the instantaneous risk $h(t) = \frac{k}{\lambda} \left(\frac{t}{\lambda}\right)^{k-1}$.

When the shape parameter k is equal to 1, we get an exponential distribution: the survival is exponentially decreasing and the instantaneous risk is constant and equal to $1/\lambda$. If k is smaller than 1, the instantaneous risk is decreasing whereas if it is greater than 1, the instantaneous risk increases over time t . Moreover, the variance increases when k decreases.

lives⁴. We call this the selection bias. This bias has the opposite effect to the censoring bias, which means that, in theory, it is impossible to tell whether seniorities are, on average, longer or shorter than durations. What is the link between the distribution of duration and the distribution of seniority? In order to clearly understand the potential consequences of these censoring and selection biases, we will then randomly generate several data sets simulating various real situations. The parameters of these simulations can then be adjusted in various ways in order to highlight the two biases in chosen scenarios.

We illustrate these censoring and selection phenomena using simulations that randomly generate observations of durations and seniorities. Based on a uniform distribution, we randomly generate a date on which the situation of single parenthood started between 1950 and 2010 for 10,000 individuals. For each of these individuals, we also generate a random variable of duration using a Weibull distribution (cf. Box 2). We then deduce a stock of individuals who are still in a situation of single parenthood in 2011. For the individuals in this stock, we can calculate the length of time already spent

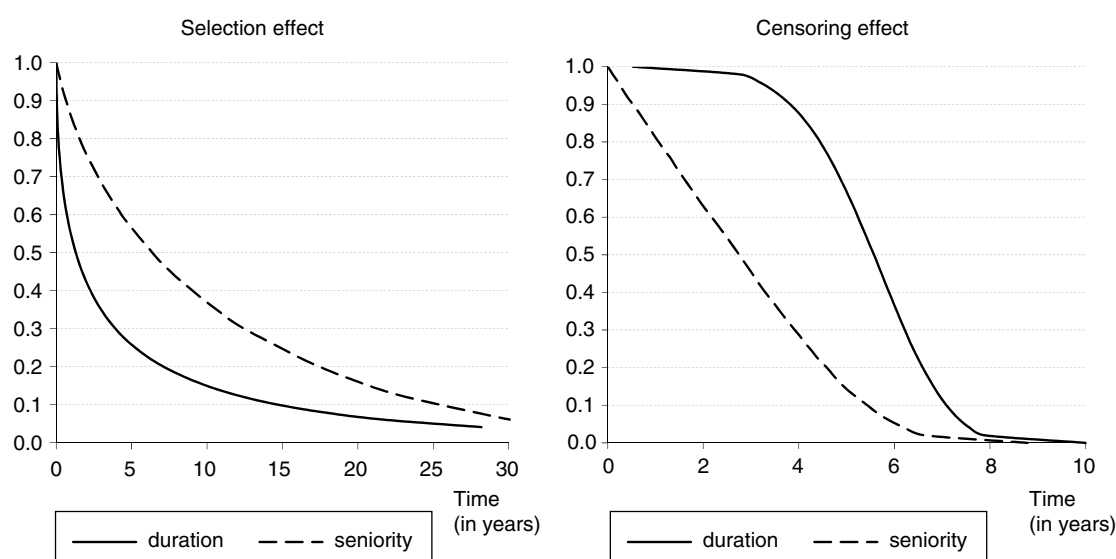
in the situation (i.e. the seniority) in 2011. The Weibull distribution shape parameter is the key parameter: if it is smaller than 1, then the selection bias is dominant, whereas, if it is greater than 1, the censoring bias is the strongest (Figure I).

This result can be understood intuitively. If the shape parameter is below 1, single parenthood durations are very short for most of the individuals and only a small number of them have experienced very long periods (there is a very wide range of different durations). Consequently, at the time of the survey, all the individuals who have experienced very long durations are in a situation of single parenthood, whereas only a fraction of those who have experienced very short durations are in such a situation (the individuals who entered the situation just prior to the survey). Those individuals affected by long durations of single parenthood are thus overrepresented within the stock of single parents at the time of the survey. Consequently, the observed seniorities of the situation are longer than the actual durations of the periods of single parenthood.

Conversely, if the shape parameter is greater than 1, the durations vary little: the duration is the same for most single parents, save for a few variations. Those who are in a situation of single parenthood at the time of the survey

4. To illustrate this, we can use the example of the 'battleships' game: when choosing a position on the board at random, the large ships are more likely to be hit than the small ones.

Figure I
Survival curves associated with the distribution of durations and the distribution seniorities and illustrating the effect of selection and the effect of censoring using simulations



Note: simulated data. For the left-hand figure, the duration is generated in accordance with a Weibulldistribution(0.5 , 2.75) and for the right-hand figure according to a Weibull distribution(5 , 6). The date of entering the period of single parenthood is generated according to a uniform distribution.

are thus more or less representative of the full set of individuals (as the durations vary very little). As the seniorities are always shorter than the overall durations of single parenthood, the result is that, overall, the seniorities for those individuals in a situation of single parenthood at the time of the survey will be shorter than the set of durations.

If the selection bias is stronger than the censoring bias, the seniorities observed will, on average, be longer than durations, whereas if the censoring effect is stronger, the seniorities will, on average, be shorter than the durations. If the comparison of two groups shows that, on average, the seniorities are shorter for the first group than for the second group, this does not mean that the underlying durations of single parenthood in the first group are, on average, shorter than the durations of the second group. The ordering of durations will not follow the ordering of seniorities if, in the first group, the censoring bias is very strong and, in the second group, the selection bias is very strong, as illustrated in Figure II.

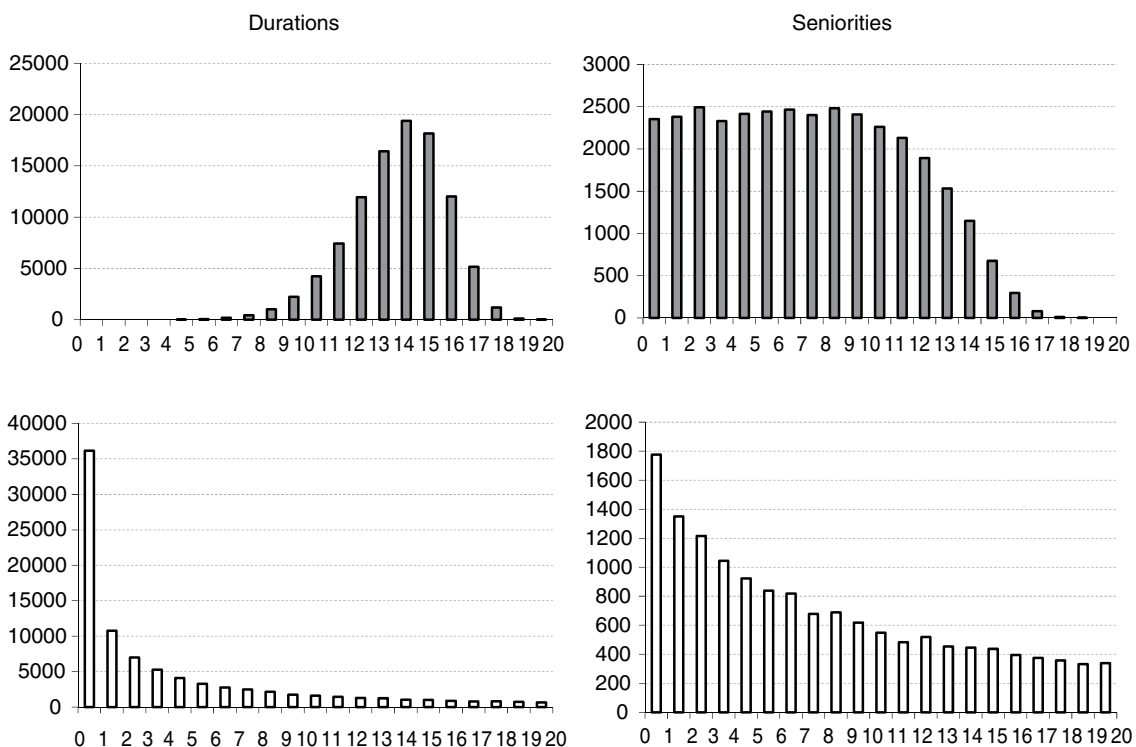
Influence of the flows of parents entering a period of single parenthood on the seniorities

The distribution of durations in the population thus has a direct impact on the seniorities. Another factor also influences the distribution of seniorities observed at a given point in time: these are the flows of parents entering a period of single parenthood. If, for example, increasing numbers of single-parent families are formed each year (increasing flow), then the seniorities observed will mechanically tend to be short, since in such a case, most of the single-parent families observed in the survey will have been formed shortly before the survey. Hence an increasing flow of individuals entering a period of single parenthood thus weighs in favour of the censoring bias (a decreasing flow increases the selection bias). The simulations presented in Figure III illustrate this phenomenon clearly.

By way of summary, the distribution of seniorities observed at the time of the survey depends both on the flow of individuals entering the situation of single parenthood and on the distribution of real durations. We can thus intuitively

Figure II

Comparison of the distribution of durations and seniorities between two groups



Note: the durations of group 1 (dark grey), T1, were simulated using a Weibulldistribution(8 , 15) and those of group 2 (white) using a Weibull distribution(0.5 , 5). The duration distribution is represented on the left and the seniorities distribution on the right. This figure illustrates the fact that durations can, on average, be longer in group 1 than in group 2 (14 years compared to 10 years), whereas seniorities are, on average, shorter in group 1 (7 years) than in group 2 (12 years).

say that if the distribution of seniorities and the flows of individuals entering the situation of single parenthood are known, it is possible to deduce the distribution of durations. This is the focus of the next section.

Inferring durations from seniorities and entry flows

Relation between flows, durations and stocks

We first consider the hypothetical situation in which the flows of parents entering the situation are constant over time, and where the mean duration of single parenthood does not change over time either. In such a case, the flows of parents exiting the situation of single parenthood offsets exactly the flows of parents entering such a situation and the stock does not change. In this precise case, there is a simple relation between the stock, the flows and the durations:

$Stock = Flows * E[T]$, where $E[T]$ is the expected duration.

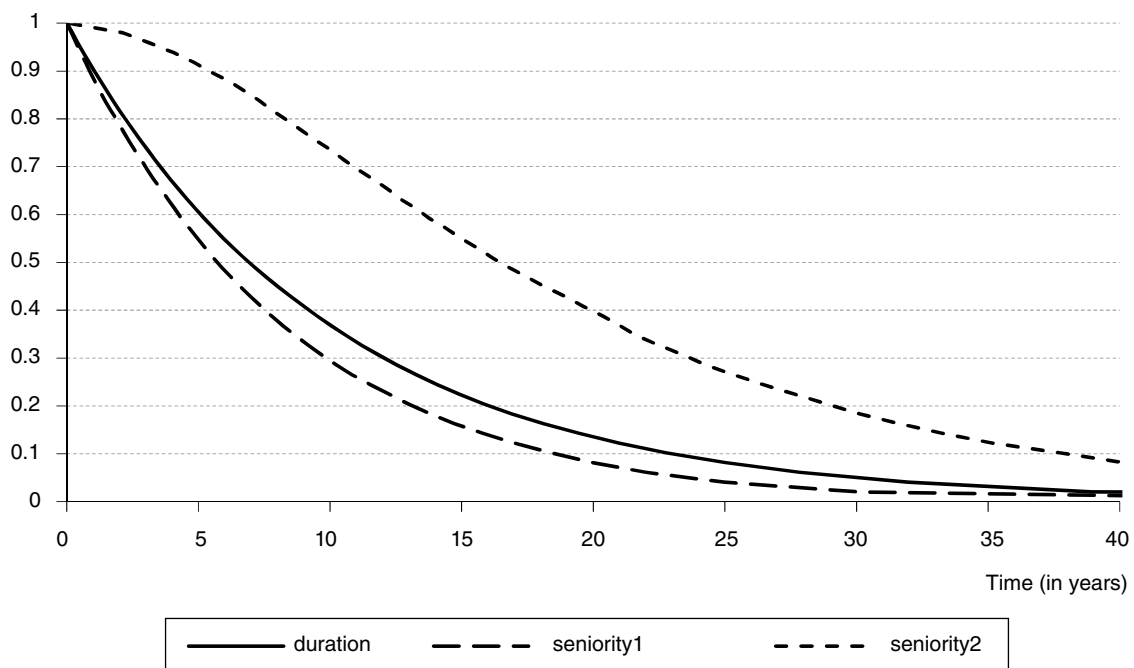
This relation is easy to understand: the greater the flows, the greater the stocks observed

at a given time, too; the longer the individuals remain in a situation of single parenthood, the higher the probability of their being part of the stock, and the greater the stock size. Knowing the flows and the stock, this relation thus directly shows the mean duration spent in a situation of single parenthood $E[T]$. In a stationary regime (constant flows), the mean duration of single parenthood can thus be directly deduced from the flows of parents entering the situation of single parenthood and the size of the stock of single-parent families.

Outside a stationary regime, the equivalence is no longer verified. If the flows increase regularly, we should arrive at a situation where, for a given year, the stock is lower than the last flow multiplied by the mean duration: $Stock < Last\ Flow * E[T]$, which gives us a lower boundary for the mean duration. We have the opposite relation in the case of a decreasing flow.

The relation between flows, durations and stocks does not, however, allow us to deduce the distribution of durations. To do this, we also need to know the distribution of seniorities. As we want to know the duration distribution (and not just the mean duration) and because,

Figure III
Influence of the flows of parents entering a period of single parenthood on the observed seniorities



Note: the dates of beginning of the situation of single parenthood and their durations are simulated for both groups. In both cases, the duration variable is generated following an exponential distribution with a mean parameter of 10. The difference between the two groups comes from the distribution of the dates on which the period of single parenthood started. In the first group, the flow of parents entering the situation is increasing, whereas in the second group it is decreasing. Consequently, the observed seniorities A1 et A2 have a different distribution.

in the case of single parenthood, flows are not constant over time, we go on to present a more sophisticated method that takes account of all the available information in order to determine the distribution of durations irrespective of the flows of parents entering the situation.

Flow sampling and stock sampling

Traditional duration models, such as the Cox model (1972), generally make use of flow sampling: this entails observing the persons who enter the situation of interest on a given date and then monitoring their situation over time through to a final date. This requires repeated observations of the same individuals and that is why, in the field of social sciences, this type of study is generally conducted on small samples, although these models can be applied to larger data where available. Some durations are thus observed in full while others are right censored in the case of individuals who have not exited the situation of interest prior to the final date of observation.

Conversely, stock sampling allows for the study of a sample of individuals who are in the situation of interest at a given time, regardless of the date on which they entered the situation. Even though this type of sampling is less common for the purpose of studying durations, several authors have, nonetheless, developed duration models on this basis. Wooldridge (2002) refers to a model where the individuals sampled from the stock are monitored over a certain period of time; thus all the durations are not right censored. Kiefer (1988) describes the difficulty of estimating unemployment periods using the *Current Population Survey* conducted in the United States due to the fact that the durations are both right and left censored. Very often, in order to estimate the probabilities of coming out of unemployment within the framework of job search theory (Atkinson et al., 1984), or within the field of epidemiology (Keiding, 2006), stock sampling models have been used. When all the data are right censored, it is always possible to estimate the probabilities of exiting the situation of interest, as shown by Nickell (1979), in a paper to which we will refer frequently in the rest of this article, provided we know the probabilities of entering the situation on the dates preceding the survey. To this end, he developed an entirely parametric model, with estimations based on the principle of maximum likelihood. Lancaster places the issue of stock sampling in

the more general context of renewal processes (Lancaster, 1990).

The method presented here uses to a very large extent the solution proposed by Nickell (1979), but several changes have been made in order to better address the specific issue of the duration of single parenthood. We pick up on the idea of the proportional hazard contained in the Cox model, which we incorporate into the likelihood calculation set out by Nickell, thus giving us the model proposed here, which, to the best of our knowledge, is unique in its kind. Moreover, in order to estimate the parameters of this model, we will need information that allows us to make an estimation of the flows of individuals entering this situation, which will prove to be essential, as we will see later in this paper. Here we use the data from the *FHS (Family and Housing Survey)* for the stock and the data from the *Erfi (Survey of Family and Intergenerational Relations)* for the flows.

General principle of the method

In the case of the *FHS*, we only know the year in which the single parenthood situation commences for persons living in a single-parent family at the time of the survey. We thus only know the seniorities to within one year. This uncertainty will also lead to uncertainty in terms of the inferred durations. Here, we have chosen to model the durations discretely (see Box 2): we consider the random duration variable corresponding to a whole number of years (0, 1, 2, etc.). For example, a person who entered the situation in 2010 and exited it in 2012 will have a discrete duration of 2 years, even though the true duration could be between 1 year (starting at the end of 2010 and ending at the start of 2012) and 3 years (starting at the start of 2010 and ending at the end of 2012).

The method of inference consists firstly in calculating the likelihood function of the observations (we observe the seniorities). We show that this likelihood function depends only on the seniorities, the flows of people entering the situation and the hazard function (see Box 3). Here we have chosen to model hazard using a piecewise constant function. Even though hazard is a discrete function, this choice allows us to reduce the number of parameters that need to be estimated and, thus, to increase the accuracy of the estimations. We thus need to find a piecewise constant function that maximises the likelihood. We can make the model more

sophisticated by making the instantaneous probability dependent on individual variables, as in a Cox model (Box 3). The inference consists thus of finding regression parameters and the piecewise constant function that maximise the likelihood.

Other modelling solutions could be used to establish a link between a set of covariates and the hazard function. For example, the accelerated failure time model is an interesting alternative to the Cox model. In this model, instantaneous probabilities are no

Box 3

METHOD FOR INFERRING DURATIONS BASED ON THE SENIORITIES AND THE FLOWS OF PARENTS ENTERING A PERIOD OF SINGLE PARENTHOOD

We take into consideration m individuals, born between the year a_0 and the year a_1 , who experience during their life one or more periods of single parenthood. For each period, D is the start year, F the year in which the situation ends and $T = F - D$ the duration (measured discretely) of the situation. If the start of the period of single parenthood and the end of the period of single parenthood are uniformly distributed over the year, the expected value T is thus equal to the expected true duration. D , F and T are discrete random variables that values are integers.

Assumption 1: we assume that the random duration variable follows the same distribution regardless of the ranking of the period of single parenthood (there could be several periods of single parenthood during a lifetime).

The objective is to estimate the distribution of the duration variable T conditionally to individual characteristics x . To do this, we will estimate the instantaneous probability of this variable, written $h(u, x) = P(T = u | T \geq u, x)$, for $u \in \mathbb{N}$.

At the start of year e , a survey is conducted and it is observed that n of these individuals ($n \leq m$) are in a situation of single parenthood. We can also observe the dates on which the situation started, which we express as d_1, \dots, d_n . It is based on these start dates that it will be possible to infer the distribution of T . To do this, we first calculate the likelihood of the observations based on the following assumption:

Assumption 2: the random variables T and D are independent of each other. This assumption is based on the fact that the distribution of the duration of single parenthood does not change over time: all generations are subject to the same durations. The contribution to the likelihood of an individual i having the characteristics x_i and having started a period of single parenthood in year d_i is expressed as:

$$P(D_i = d_i | D_i < e \leq D_i + T_i, x_i) = \frac{P(D_i < e \leq D_i + T_i | D_i = d_i, x_i) A(d_i, x_i)}{P(D_i < e \leq D_i + T_i | x_i)}$$

where $A(d, x) = P(D = d | x)$ is the conditional distribution of D where x is known.

According to assumption 2, we have:

$$P(D_i < e \leq D_i + T_i | D_i = d_i, x_i) = 1_{(d_i < e)} S(e - d_i, x_i) \text{ and}$$

$$P(D_i < e \leq D_i + T_i | x_i) = \sum_{u < e} S(e - u) A(u, x_i).$$

$$\text{But, } S(t, x) = \prod_{u=0}^{t-1} (1 - h(u, x)).$$

We propose here to model the hazard function the same way as the Cox proportional hazard model:

$$h(t, x) = \min(h_0(t) \exp(\beta x), 1)$$

Having 1 as the minimum ensures that the instantaneous probability is equal to a value between 0 and 1, as has to be the case with the discrete variable. Finally, we model the baseline hazard h_0 using a piecewise constant function, which will have to be estimated. We choose a constant 3-year step: h_0 is thus constant: from 0 to 2 years, from 3 to 5 years, etc. up to 18 years, then it is constant beyond 19 years. This thus entails estimating 7 parameters (one for each period of time over which the function is constant). A 3-year step was chosen in order to reduce the number of parameters and increase the accuracy of the estimations. The drawback is that this restrains a little the shape of the function h_0 .

The likelihood of the model is finally expressed as follows:

$$L(d, x) = \prod_{i=1}^n \frac{A(d_i, x_i) \prod_{k=0}^{e-d_i-1} (1 - h_0(k) \exp(\beta x_i))}{\sum_{u < e} A(u, x_i) \prod_{k=0}^{e-u-1} (1 - h_0(k) \exp(\beta x_i))}$$

where $d = (d_1, \dots, d_n)$ et $x = (x_1, \dots, x_n)$.

We thus seek the piecewise constant function h_0 and the parameter β , which maximises this likelihood function. We note that, to do this, it is sufficient to know - with x being considered as fixed - the function $A(d, x)$ to within one multiplicative function, which amounts to knowing the flows of parents becoming single parents each year for the persons born between year a_0 and year a_1 and with characteristics x . Unlike the Cox model, we are obliged here to jointly estimate the baseline hazard h_0 and the parameters β , even though, for h_0 , no parametric form is stipulated.

longer proportional to each other. The idea is that each person is confronted with the same baseline hazard, but that time does not pass by at the same speed for each of them: it is either accelerated or slowed down depending on whether the related coefficient is greater or less than 1. All in all, it has a multiplier effect on the expected duration variable. The advantage of this model is thus its highly intuitive interpretation. The drawback is that the baseline hazard has to be parametrically determined. However, this is precisely what we avoid doing by not modelling the baseline hazard using a specific function and by letting the data ‘freely’ estimate this risk without any particular constraints. Any other parametric model, such as Nickell’s logit modelling, would also be suitable for modelling the instantaneous probability based on a set of covariates. Thus, although we have a more or less clear idea of how hazard changes over time and how it is modified by certain covariates, it is wiser to develop a fully parametric model that reflects this knowledge *a priori*. We have chosen the Cox model here because it does not require that a shape be given *a priori* to the baseline hazard and also because the interpretation of the estimated coefficients is straightforward.

Three assumptions were made as part of this modelling. The first is that the duration does not depend on the ranking of the period of single parenthood during the individual’s life, i.e., we assume that, providing the characteristics are similar, the fact of experiencing a first or a second period of single parenthood does not influence the duration of that period. It can happen that an individual has experienced this situation a second time, for example after entering a relationship with a new partner, thereby bringing an end to a period of single parenthood, which is then followed by a separation, thereby recreating another situation of single parenthood. The second assumption is that the duration of the situation does not depend on the date on which the person entered the situation of single parenthood, which amounts to saying that there is no generational effect. This is a strong assumption and is probably not verified in reality. For example, the time taken to enter a new relationship after a separation decreases over the generations (Costemalle, 2015). Nonetheless, we have not been able to create a model that enables us to estimate this link in a robust manner. Finally, the third assumption is the same as in a Cox model, i.e., that the instantaneous probabilities are proportional. This assumption allows for the calculation

of the effects of a covariate on the instantaneous probabilities regardless of time. We can get an idea of its validity by independently comparing the hazard function among the sub-populations studied.

The method developed here is thus original, resulting from the combination of two known models (the Nickell and Cox models) to make a third one. The Cox model allows the estimation of a baseline hazard and parameters that indicate how the instantaneous risk changes in relation to this baseline. Unfortunately, it cannot be used in the case of stock sampling. The Nickell model, however, does not allow the estimation of a baseline hazard, or *a posteriori* a duration distribution. Its purpose is to determine if a given external variable has a positive or negative bearing on the instantaneous probability of exiting a situation (initially unemployment in Nickell’s research). Its strength is the simple manner in which it expresses the instantaneous probability (a logistic function). The drawback of such modelling is that we force the hazard function to adopt a certain shape, which significantly restricts the model. Another difference between the modelling proposed here and Nickell’s model is that, with the former, the durations follow discrete distributions whereas, with the latter, they follow continuous distributions. The discrete aspect of the problem poses an additional difficulty.

Results of the simulation-based inference

If we have a perfect knowledge of the flows of individuals entering a situation of single parenthood, the method presented functions properly and allows the duration distribution to be determined. This enables us to deduce several values relating to the durations, such as the mean, median and other quantiles (see box 2). Figures IV and V show, in respect of the simulated examples, that, regardless of the dominant bias (censoring bias or selection bias), the inference method allows for the right duration distributions to be determined. There are still, however, three sources of uncertainty. The first stems from the fact that we estimate discrete (instead of continuous) durations. The second stems from the sample size: the smaller it is, the greater the uncertainty. Finally, the third relates to the fact that we do not have perfect knowledge of the flows of individuals entering the situation, which are estimated based on the *Erfi* survey (see Box 4).

Box 4

DETERMINING THE FLOWS OF PARENTS ENTERING A PERIOD OF SINGLE PARENTHOOD BASED ON THE *ERFI* SURVEY

For fixed individual with characteristics x an estimation of $A(d, x)$ is made for those persons born between a_0 and a_1 , to within one constant (see box 3). This thus amounts to estimating the number of persons with characteristics x that entered a period of single parenthood in year d . To do this, we use the *Erfi* survey, the respondents to which were born between 1926 et 1987. As for the respondents to the FHS, they could have been born up 1992. *Erfi* can thus not provide us with the data relating to the flows of parents entering a period of single parenthood for persons born between 1988 and 1992. In order to take account of this, we first estimate the data flows for person entering a period of single parenthood for the year 2005 or preceding years based directly on the third wave of the *Erfi* survey. The individuals born between 1988 and 1992 and who were under the age of 18 during these years thus contribute hardly at all to the flow. Then, for the period 2006-2010, we correct the flow for each year using a multiplicative coefficient in order to estimate the flows corresponding to persons aged 18 or over. These coefficients, which are presented in the table below, are calculated using the distribution of the ages at which the period of single parenthood started (this distribution is itself an estimation based on the

Erfi data and the data for persons that entered a period of single parenthood between 2000 and 2005).

Correcting coefficients to estimate the flows for the period 2006-2010

	2006	2007	2008	2009	2010
Women	1.026	1.026	1.056	1.1	1.125
Men	1	1	1	1	1

Source : Ined-Insee, *Erfi*, wave 1, 2005.

If a person experiences several periods of single parenthood during his/her life, only the last period is taken into account for the contribution to the flow.

Finally, the flows that have thus been estimated are smoothed in order to reduce the statistical noise inherent in the small sample size of the *Erfi* survey. To do this, we first calculate a moving average over five years, which allows for the random variations caused by the small sample sizes to be reduced locally. We then perform a polynomial regression on the smoothed data. The flows of parents entering a period of single parenthood estimated according to this method are presented in the graphs in annex 1.

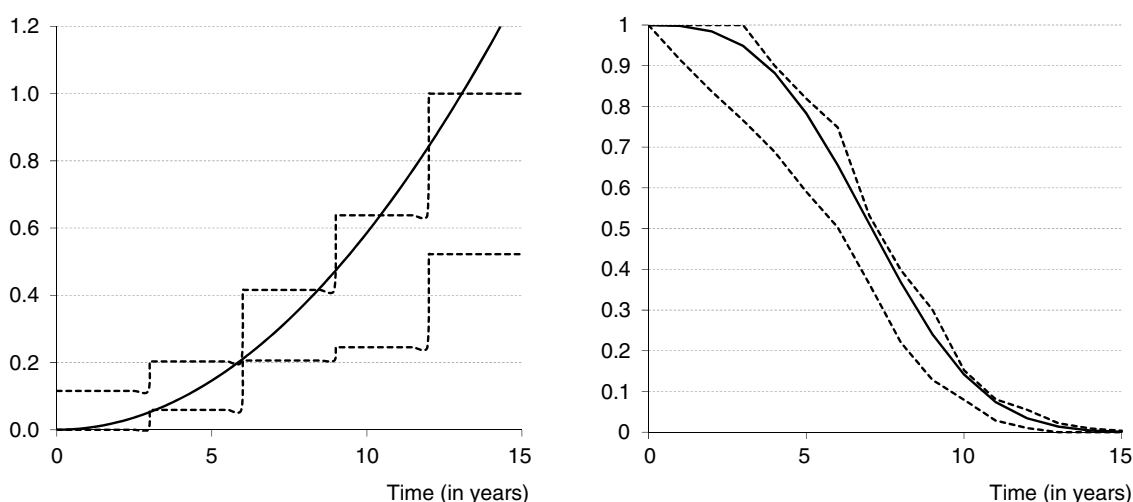
Influence of the sample size

In order to fully understand how the estimations change depending on the sample size, we conduct a series of experiments, each of which

consists of simulating flows of people entering the situation and durations for a given number of individuals, then deducing the distribution of discrete durations (based on the method presented) and lastly, estimating the mean and

Figure IV

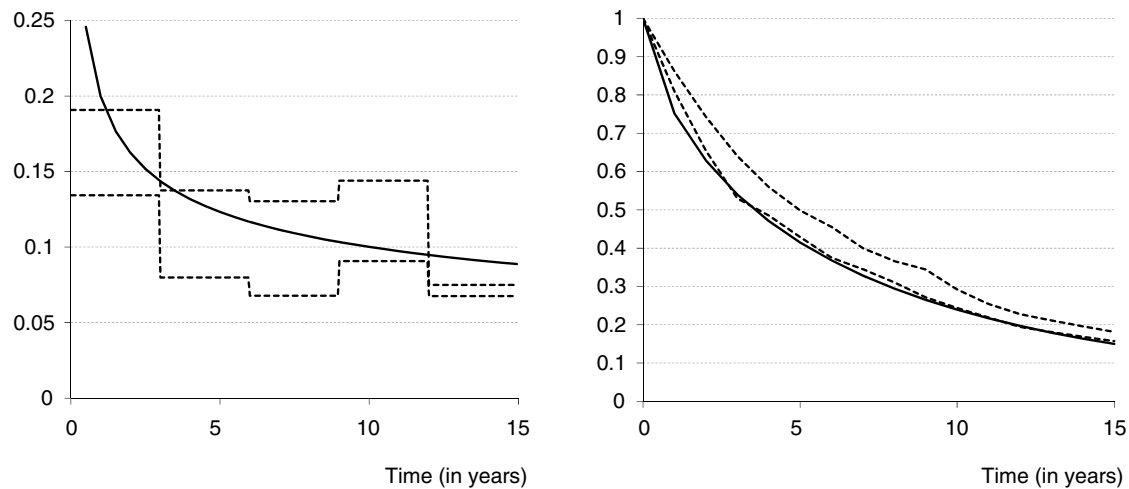
Comparison between the estimated and true hazard (left) and between the estimated and true survival (right) when the censoring bias is dominating



Note: we generated 100 samples of 1,200 observations of seniorities based on a Weibull distribution(3 , 8) for the durations and on a uniform distribution for the flows of parents entering a period of single parenthood. The solid line curves represent the true hazard and survival. The dotted lines show the intervals containing 95% of the estimations.

Figure V

Comparison between the estimated and true hazard (left) and between the estimated and true survival (right) when the selection bias is dominating



Note: we generated 100 samples of 12,300 observations of seniorities based on a Weibull distribution(0.7 , 6) for the durations and on a uniform distribution for the flows of parents entering a period of single parenthood. The solid line curves represent the true hazard and survival. The dotted lines show the intervals containing 95% of the estimations.

median durations. We test four measures of accuracy. The first is the difference between the estimated mean durations and the true expected continuous random duration variable; the second is the difference between the estimated median and the true median; the third is the difference between the estimated and the true survival curve; the fourth is the difference between the estimated and the true hazard function. When the sample size is large, the different accuracy measures barely fluctuate from one experiment to another, whereas this is not the case with small samples. The first result is that the estimated mean is around 0.5 years lower than the true mean (regardless of the underlying duration model). Thus, inference of the durations gives rise to a 0.5-year underestimation of the mean even when the sample is very large. However, the median is correctly estimated and, when the sample is sufficiently large, there is hardly any variance between the estimated median and the true median (Figure VI). The difference between the estimated survival function and the true survival and the difference between the estimated hazard and the true hazard moves towards zero the greater the sample size (Figure VII). These results, which were obtained on the basis of numerical simulations, indicate that inferring the duration distribution using the maximum likelihood method works and that the results are very accurate if the sample size is sufficiently large (around 1,500 people).

We will then add 0.5 years to the mean estimations in order to take account of the bias of the mean estimator. This correction is based on an empirical observation of simulated data, but has not been theoretically proven. We can, nonetheless, demonstrate mathematically that if the survival function is estimated accurately, as is the case here, the error term of the mean calculation will be around 0.5 years (see demonstration in Annex 4).

Estimating the risk factors

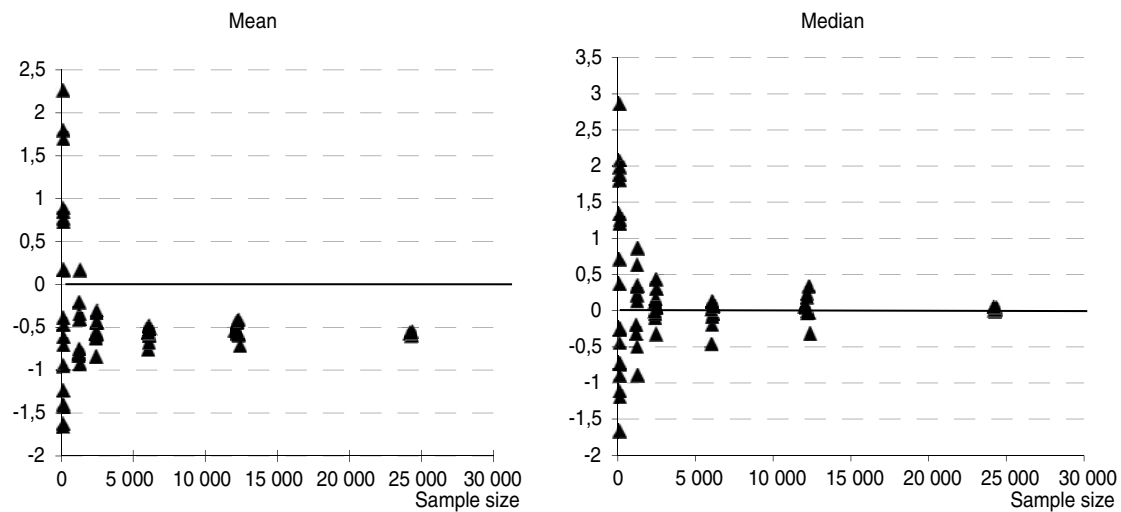
We now test whether the model accurately estimates the regression coefficients of the hazard function. We assume that we have a population comprising three distinct groups, each of which is faced with a different instantaneous probability (h_0 for the first group, which is the benchmark group, h_1 for the second group and h_2 for the third group):

$$h_1(t) = h_0(t)\exp(\beta_1) ; h_2(t) = h_0(t)\exp(\beta_2).$$

These probabilities are, as established in assumption 3 of our model, proportional to each other. It thus suffices to estimate only β_1 and β_2 to determine the differences in the instantaneous probabilities (and therefore in the durations) between the different groups. If the coefficient is positive, this means that the instantaneous probability is greater and the mean

Figure VI

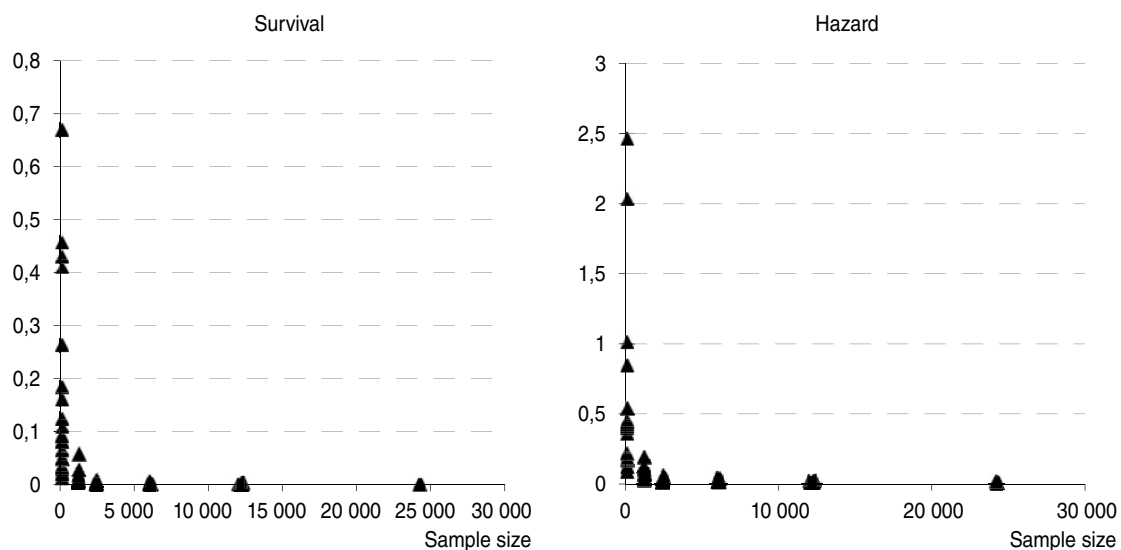
Differences between the estimated and true mean and between the estimated and true median depending on the sample size



Note: each point corresponds to an experiment, which consists of generating durations randomly, estimating the hazard according to the method presented and deducing the mean or median. The flow variable was generated according to a uniform distribution and the duration variable according to a Weibull distribution(5 , 7). What changes from one experiment to another is the size of the sample, which corresponds to the number of persons living in a single-parent family at the time of the survey.

Figure VII

Differences between the estimated and true survival and between the estimated and true hazard depending on the sample size



Note: each point corresponds to an experiment. The flow variable was generated according to a uniform distribution and the duration variable according to a Weibull distribution(5 , 7). If f is the estimated function (survival or hazard) and if g is the true function, we define the difference between these two functions as $\sum_{t \geq 0} (f(t) - g(t))^2$.

duration is thus lower. The opposite is true if the coefficient is negative.

In order to check if our estimation method allows us to obtain these two parameters, we simulate 200 times a total population of 50,000 individuals distributed evenly across

the three groups, with $\beta_1 = 0.5$ and $\beta_2 = -0.5$. For each of these 200 simulations, we estimate β_1 and β_2 using the maximum likelihood method. We then calculate the mean of these estimations for the 200 simulations. We can also calculate the rate of coverage, i.e., the proportion of estimations, such that the true

Table 1

Mean and coverage rate of the estimations of the coefficients β_1 and β_2

True value		β_1	β_2
		0.5	- 0.5
Estimation	Mean	0.47	- 0.48
	Coverage rate	82.5	85.0

value of the coefficient is within the estimated confidence interval (95% confidence interval for the maximum likelihood estimator based on the asymptotic behaviour of the estimator, which in theory should follow a normal distribution).

The results of these simulations are presented in Table 1. They show that the model does indeed provide the values of the coefficients β_1 and β_2 , even though the estimations seem to underestimate the true values a little. The coverage rates are below 95%, which indicates that the estimated confidence intervals are slightly too narrow. These coverage rates are, however, quite high, at around 85%.

An estimation of the durations of single parenthood in France

Of the 359,770 respondents to the *FHS*, 12,519 were in a situation of single parenthood at the

time of the survey⁵, 1,073 of whom were men and 11,446 women. In light of the previous results, we have a sufficiently large sample to be able to infer the distribution of durations for women, but the sample of men appears to be too small to obtain robust results.

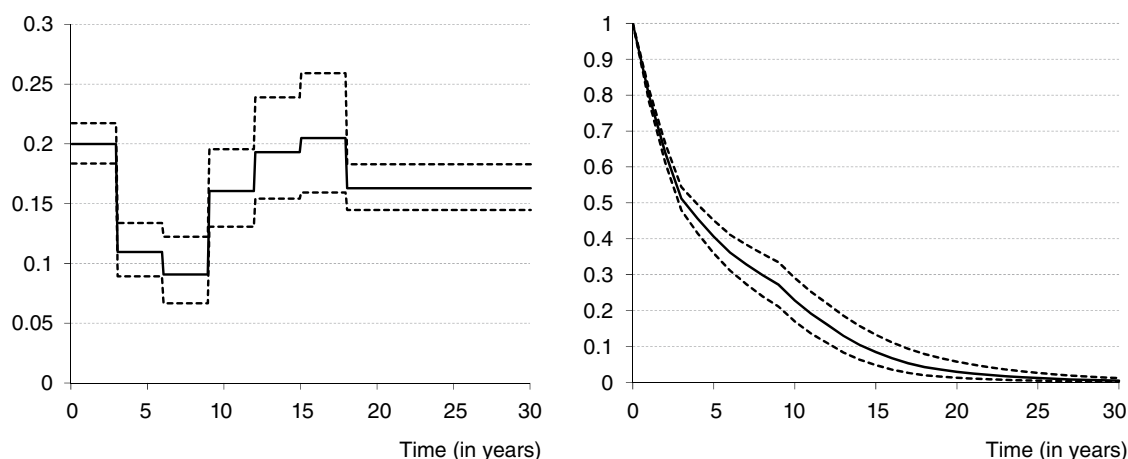
The estimated durations: a ‘U’ shape

We present here the results of the estimations of the instantaneous probabilities of exiting a situation of single parenthood without taking account of any covariates, i.e., taking into account only time, without introducing other factors that could influence the probability of exiting single parenthood. Figure VIII shows that the overall hazard function is not monotonous: initially it

5. We eliminated the 205 respondents who had entered the situation of single parenthood in the year of the survey (in 2011), as they do not contribute any information to our model, as well as the LAT's (parents living apart together), which is a situation whose seniority cannot be determined.

Figure VIII

Estimation of the hazard and the survival associated with single parenthood durations

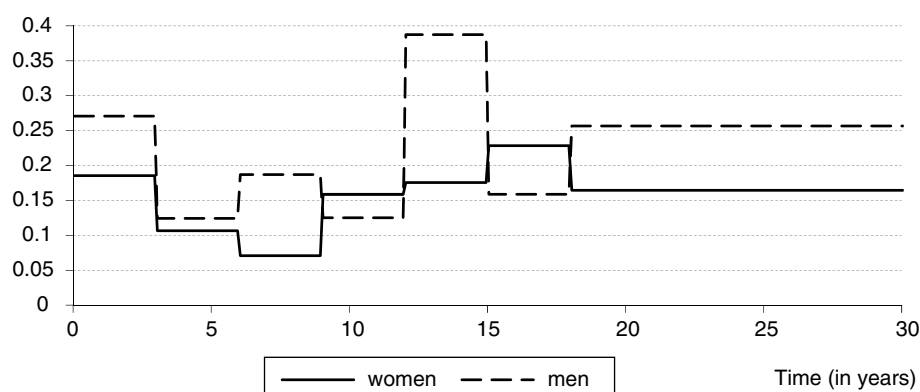


Note: the dotted lines represent the 95% confidence interval obtained with the maximum likelihood method.

Coverage: single parents with minor children, mainland France.

Sources: Insee, *Family and Housing Survey (FHS)* 2011 and Ined-Insee, *Survey of Family and Intergenerational Relations (Erfi)*, waves 1 and 3, 2005 and 2011.

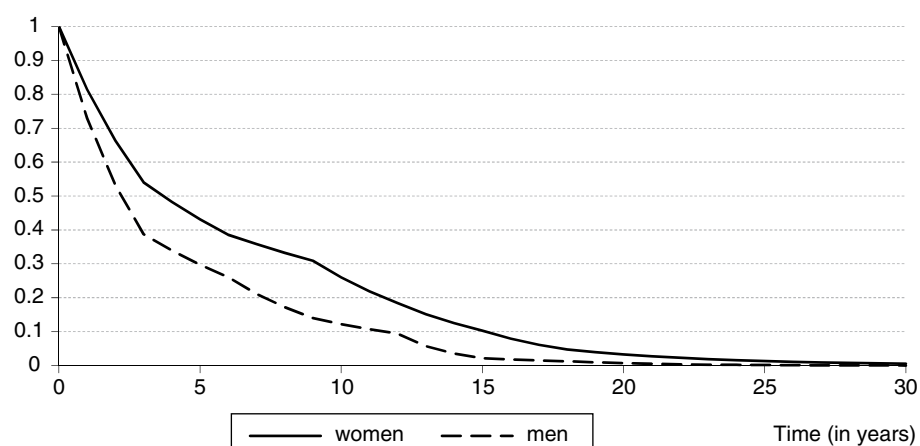
Figure IX

Instantaneous probability of exiting the situation of single parenthood for men and women

Coverage: single parents with minor children, mainland France.

Sources: Insee, *Family and Housing Survey (FHS)* 2011 and Ined-Insee, *Survey of Family and Intergenerational Relations (Erfi)*, waves 1 and 3, 2005 and 2011.

Figure X

Survival functions associated with the durations spent in a single-parent family for men and for women

Coverage: Single parents with minor children, mainland France.

Sources: Insee, *Family and Housing Survey (FHS)* 2011 and Ined-Insee, *Survey of Family and Intergenerational Relations (Erfi)*, waves 1 and 3, 2005 and 2011.

Table 2

Estimations of the mean and median duration of single parenthood

	Means			Medians		
	Estimated value	Confidence intervals		Estimated value	Confidence intervals	
		Maximum likelihood	Bootstrap		Maximum likelihood	Bootstrap
Total	5.7	[4.9 ; 6.6]	[5.5 ; 6.1]	3.2	[2.8 ; 3.9]	[2.8 ; 3.9]
Men	4.1	[2.8 ; 6.0]	[3.7 ; 4.7]	2.2	[1.8 ; 2.8]	[1.8 ; 3.0]
Women	6.1	[5.2 ; 7.2]	[5.9 ; 6.5]	3.7	[3.1 ; 4.6]	[3.0 ; 4.3]

Note: the mean is estimated here based on the estimator presented in box 2, adjusted for a bias of 0.5 years. The confidence intervals obtained by bootstrapping are based on 320 samples selected randomly (random sample with replacement) based on the distribution of observed lengths of time spent in a situation of single parenthood (seniority). For each sample, we estimated the hazard from which we deduced the mean and median durations.

Coverage: Mainland France.

Sources: Insee, *Family and Housing Survey (FHS)* 2011 and Ined-Insee, *Survey of Family and Intergenerational Relations (Erfi)*, waves 1 et 3, 2005 et 2011.

decreases, then stabilises and then increases. This 'U' shape suggests that either single parents exit their situation quickly or remain in it for a long time. The probability of exiting this situation is at its lowest between 3 and 8 years. The survival curve obtained for the hazard function shows that, after 3 years, half of the single parents have exited this situation; after 8 years, 30% are still in the situation, after 12 years, 16% and only 4% are still in the situation for longer than 18 years. We can also determine the mean duration of time spent in this situation (see Box 2): the mean duration here is 5.7 years. The mean seniority is 5.5 years. It is thus similar to the estimated mean duration. We are thus in a peculiar situation where the censoring bias and selection bias almost cancel each other out.

By estimating separately the instantaneous probabilities for women and for men, we observe that they do not have the same shape; consequently, the assumption that the instantaneous probabilities are proportional to each other is not appropriate for making comparisons between men and women (Figures IX and X). While for women the hazard function has the same 'U' shape as observed for the whole set of men and women, for men it fluctuates more and is higher.

This fluctuation suggests that there are not enough observations for men in the *EFL*: around 1,000. The fact that, overall, there are far more women than men (approx. ten times more) provides an additional explanation as to why we find the 'U' shape both for the whole set of single parents and for women only.

The mean estimated duration of single parenthood here is 6.1 years for women and 4.1 years for men (Table 2). Confidence intervals of 95% have been successively estimated using the maximum likelihood method, then by bootstrap. As regards the mean, the latter method results in narrower intervals than those obtained with the maximum likelihood method. For the median, however, we obtain almost the same confidence intervals. This is an additional indication as to the reliability of the estimations.

Determining the mean duration based on the flows and stocks

We have been able to determine mean durations of single parenthood based on the estimated duration distribution. As explained earlier, it is nonetheless possible to obtain an estimation of the lower boundary of the mean duration using the relation between the stocks, flows and durations.

We can apply this principle to the *FHS* as we know the size of the stock of single-parent families in 2011 (1,449,000), as well as the size of the last flow (254,000 parents entering the situation of single parenthood in 2010). We can thus deduce that the mean duration of single parenthood must be greater than 6 years for women and 4.4 years for men (Table 3), whereas our previous estimations were 6.1 years for women and 4.1 years for men.

The last flow is probably underestimated, as we do not take account of the parents who entered the situation in 2010 and exited it in the same year. Consequently, the estimation of the lower boundary of the mean duration spent in a situation of single parenthood is slightly underestimated as well. All in all, these different estimations appear to be completely consistent with each other.

Durations and seniorities correspond to two different concepts. However, in this case, the distribution of seniority is quite similar to that of duration. We are in fact in a situation where the censoring bias and the selection bias offset each other more or less, and the seniorities can thus provide an initial approximation of the durations.

Significant differences in the estimated durations depending on the reason for entering a situation of single parenthood

We now focus on the durations obtained when we introduce non time-dependent covariates. In

Table 3
Stock and flows of single-parent families

	Stock in 2011	Flow in 2010	Stock/Flow
Men	208 904	47 977	4.4
Women	1 239 843	206 067	6.0

Coverage: Mainland France.

Source: Insee, *Family and Housing Survey* 2011.

light of the small size of the sample of men, we consider only the results concerning women. The durations of single parenthood are estimated successively according to the reason for entering the situation, the level of qualification achieved at the time of the survey and the social category at the time of the survey. For each estimation, we present the estimated coefficients β of the proportional hazard model relative to a benchmark group; if β is positive, this means that the instantaneous probability of exiting the situation of single parenthood is higher than that of the benchmark group, and thus that its survival in the situation (i.e. duration) is shorter than that of the benchmark group. The mean durations are deduced from the estimated instantaneous probabilities for each group, themselves being deduced from the baseline hazard (see Annex 2) and the relative probabilities.

There are three reasons for entering a situation of single parenthood: separation from a spouse (78%), child born outside a relationship (16%), death of a spouse (6%). Significant differences of durations are observed between these reasons. Compared to a scenario where the mother entered the situation due to the birth of a child

outside a relationship, the instantaneous probability of exiting the situation of single parenthood is 1.8 higher when the reason is a separation and 1.7 times higher when the situation is entered as a result of the death of a spouse (Table 4.A). Women who have experienced single parenthood after a separation thus spend the shortest time in this situation (5.4 years), followed by those who become single parents after being widowed (5.7 years), while those who have had a child outside a relationship remain far longer in that situation (9.1 years).

Durations also differ between levels of education. We have distinguished four levels: no education, education below school leaver certificate level, education equivalent to a school leaver certificate and education above school leaver certificate level. The instantaneous probability of exiting the situation of single parenthood always comes up higher for women with a qualification, which means that those who do not have one remain, on average, longer in a situation of single parenthood than the others (Table 4.B). However, the probability of exiting the situation does not generally increase with the education level; thus, according to these

Table 4
Estimation of the relative probabilities of exiting single parenthood and of the mean durations spent in the situation

A. By reason for entering the situation

Reason for entering the situation	Relative risk	Value p	Mean duration
Child when not in a relationship with a partner (16%)	1	-	9.1
Separation (78%)	1.84	7.90E-77	5.4
Widowhood (6%)	1.73	6.00E-31	5.7

B. By qualification level

Qualification	Relative risk	Value p	Mean duration
No qualification (21%)	1	-	7.6
Below school leaving certificate level (34%)	1.40	2.10E-28	5.5
School leaving certificate level (19%)	1.38	6.70E-18	5.6
Above school leaving certificate level (27%)	1.20	4.70E-07	6.4

C. By social category

Social category	Relative risk	Value p	Mean duration
Workers (11%)	1	-	6.9
Craftsmen, merchants and business owners (3%)	1.02	7.70E-01	6.8
Senior managers (9%)	0.90	1.00E-01	7.5
Technicians and associate professions (23%)	1.05	2.90E-01	6.6
Employees and workers (54%)	1.11	7.60E-03	6.3

Coverage: women, mainland France.

Source: Insee, *Family and Housing Survey (FHS)* 2011 and Ined-Insee, *Survey of Family and Intergenerational Relations (Erfi)*, waves 1 et 3, 2005 et 2011.

estimations, it is women with an education below school leaver certificate level and those with a level equivalent to the school leaver certificate who spend the shortest time as single parents (5.5 years and 5.6 years respectively) compared to 7.6 years for those with no education and 6.4 years for those with a level above school leaver certificate level.

Lastly, we distinguish five social categories⁶: craftsmen, merchants and business owners, senior managers, technicians and associate professions, employees and workers. The social category seems to have very little bearing on single parenthood durations (Table 4.C). The only statistically significant deviation (at the 5% threshold) has been observed for employees, with an instantaneous probability of exiting the situation of single parenthood slightly higher than for workers. On average, craft workers, merchants or business owners remain in the situation for 6.8 years, senior managers 7.5 years, technicians and associate professions 6.6 years, employees 6.3 years and workers 6.9 years.

The biggest differences in duration are thus observed between the reasons for entering the situation of single parenthood. These differences associated with the reasons for entering the situation are due first and foremost to the age of the children at the beginning of the situation; the maximum duration of single parenthood is limited by the age of the youngest child at the time the family becomes a single-parent family. Women who have had a child outside a relationship are in the situation of single parenthood from the outset upon the birth of their child. If they do not form a couple subsequently, they may remain in this situation for 18 years, or longer, in the (rare) event that they have several children outside a relationship. Women who become single parents following a separation or the death of their spouse have older children at the start of this situation.

The reason for entering the situation could also contribute to the deviations observed between qualification levels: for women observed in the situation of single parenthood in 2011, entering the situation due to having a child outside a relationship occurs more frequently when they have no education than when they have a level

above school leaver certificate (20% and 11% of cases respectively).

While the age of the children at the beginning of the situation of single parenthood has a mechanical influence on the duration of the situation, another factor can also contribute to this duration: the parent forming a new couple, which instantly puts an end to the situation of single parenthood. We know, for example, that, after a separation, the more qualified women do not form a new couple more quickly than the less qualified women (Costemalle, 2015). But the influence of couple formation or re-partnering on the deviations in the duration of single parenthood cannot be assessed here, as the data do not provide any information on the reason for exiting the situation.

Some remaining sources of uncertainty must be mentioned. One of them relates to the duration of single parenthood among men; we have seen that the stock of men in this situation is too low to allow any accurate inferences (the simulations showed that accuracy requires a sample of some 5,000 individuals at least); the results are thus fragile. The other sources of uncertainty relate to the estimations for both men and women: one is that we only know seniorities rounded to one year. According to the simulations presented, this seems to give rise to a bias of approximately 0.5 years in the estimation of mean durations, but does not affect the estimation of the median. If we wish to take account of this result, which has been observed based on several simulated scenarios, we must therefore add 0.5 years to the estimation of mean durations. The second uncertainty relates to the estimation of the flows of individuals entering the situation: these have been estimated based on another source, *Erfi*, the population sizes of which are far smaller than those of the *HFS* and the impact of an error of estimation of these flows on the final results is not measured here.

* *
*

6. The retired and the unemployed who have already worked are placed in their former social category. The unemployed who have never worked, members of the inactive population under the age of 60, military staff, students and persons over the age of 60 not exercising an activity, as well as farmers, are disregarded.

Finally, let us go back to the method proposed for estimating the durations of the periods of single parenthood, and firstly to the three assumptions made for the purpose of modelling.

The first is that the duration of single parenthood does not depend on the ranking of the spell of single parenthood in the person's life. This assumption seems *a priori* somewhat unlikely, as it can be assumed that, in the event of a second experience of single parenthood, the children will be older, which mechanically limits the duration of the situation for those persons who do not form a new couple with another partner (according to the definition used here, the children must be under the age of 18). Nonetheless, most single parents only experience this situation once. According to the *Erfi* Survey, only 16% of the respondents aged between 18 and 72 in 2005 who had experienced one period of single parenthood during their lives experienced the same situation for a second time. The assumption thus appears to be reasonable.

The second assumption states that the distribution of single parenthood duration does not change over time. This seems unlikely, because the reasons for single-parenthood change over time: there are fewer and fewer widows and women who had a child outside a relationship and increasing numbers of separations (according to the *Erfi* Survey). We have seen that these different reasons for entering the situation of single parenthood give rise to varying durations of single parenthood, which means that the durations change along with the structure of the population of single-parent families and hence change over time. There is then a correlation between the date of entry in single parenthood and the duration of that situation. We have attempted to understand the effect of this correlation on the estimations on the basis of simulations (see Annex 3), which randomly generate durations that are either positively or negatively correlated to the date on which the situation started. Thus, if the mean duration of single parenthood decreases with the date of entering the situation, then the model underestimates the mean duration and, conversely, if the correlation is positive, the mean is overestimated.

Finally, the third assumption relates to the proportionality of the instantaneous probabilities. Unlike with the Cox model, we do not have a test that enables us to confirm or refute this assumption. It is, however, possible to get an idea of its validity by estimating separately the hazard function curves for each sub-population.

For example, we have observed that this assumption does not appear to be verified when men and women are distinguished.

More generally, certain limits to the estimation model developed in this paper should be mentioned. Firstly, large samples are required for reliable results to be obtained. This is a significant limit, since it means that this method cannot be applied to excessively small surveys or sub-populations. Then, in terms of the modelling, it would appear difficult to take into account several explanatory variables at the same time, as is the case with the Cox proportional hazard model. To do this, it would be necessary to estimate the flows of parents entering a period of single parenthood by crossing several variables, but the *Erfi* survey population does not allow such a high level of precision. Consequently, we did not take account of the interaction between the different explanatory variables and their influence on the instantaneous probability of exiting the situation of single parenthood. We thus cannot estimate 'all things being equal' effects. Moreover, even though modelling instantaneous probabilities using a piecewise constant function allows a great deal of freedom, we have constrained this instantaneous probability to be constant over 3-year periods for the needs of the estimation. Indeed, if there are too many parameters to be estimated for the instantaneous probability, this can undermine the accuracy of the estimations.

The main limit to the method presented remains the need to know the flows of individuals entering the situation of interest; yet these flows cannot be deduced from a survey that uses stock sampling. It is therefore necessary to have access to another source of information to determine these flows. This necessity is thus the weak link in the approach. Nevertheless, while we do not know these flows, it is still possible to get a presumptive idea or to develop several scenarios, or even to develop a Bayesian model which, based on a presumptive distribution of entry flows, would estimate an empirical duration distribution.

In many fields, durations are very difficult to measure because most of the data result from stock sampling. Despite its limits, the method developed in this paper still offers the advantage of providing a simple method for estimating durations on the basis of observed seniorities. □

BIBLIOGRAPHY

- Algava, E. (2002).** Les familles monoparentales en 1999. *Population*, 57(4-5), 733–758.
- Algava, E., Le Minez, S., Bressé, S. & Pla, A. (2005).** Les familles monoparentales et leurs conditions de vie. Drees, *Études et Résultats* N° 389.
<http://drees.social-sante.gouv.fr/etudes-et-statistiques/publications/etudes-et-resultats/article/les-familles-monoparentales-et-leurs-conditions-de-vie>.
- Atkinson, A. B., Gomulka, J. & Micklewright, J. (1984).** Unemployment benefit, duration and incentives in Britain. *Journal of Public Economics*, 23, 3–26.
- Boiron, A., Huwer, M. & Labarthe, J. (2016).** Inégalités de niveaux de vie et pauvreté en 2013. Insee Références *Les revenus et le patrimoine des ménages édition 2016*, 9–21.
<https://www.insee.fr/fr/statistiques/2017606?sommaire=2017614>.
- Breuil-Genier, P., Buisson, G., Robert-Bobée, I. & Trabut, L. (2016).** Enquête Famille et Logements adossée au Recensement de 2011 : comment s'adapter à la nouvelle méthodologie des enquêtes annuelles et quels apports ? *Économie et Statistique*, 483-484-485, 205–226.
<https://www.insee.fr/fr/statistiques/2017652?sommaire=2017660>.
- Buisson, G., Costemalle, V. & Daguet, F. (2015).** Depuis combien de temps est-on parent de famille monoparentale ? *Insee Première* N°1539
<https://www.insee.fr/fr/statistiques/1283845>
- Bumpass, L. L. & Raley, R. K. (1995).** Redefining single-parent families: cohabitation and changing family reality. *Demography*, 32(1), 97–109.
- Chambaz, C. (2000).** Les familles monoparentales en Europe : des réalités multiples. Drees, *Études et Résultats* N° 66.
<http://drees.social-sante.gouv.fr/IMG/pdf/er066.pdf>.
- Chardon, O., Daguet, F. & Vivas, E. (2008).** Les familles monoparentales – des difficultés à travailler et à se loger. *Insee Première* N° 1195.
<https://www.insee.fr/fr/statistiques/1281271>.
- Costemalle, V. (2015).** Parcours conjugaux et familiaux, des hommes et des femmes selon les générations et les milieux sociaux. Insee Références *Couples et familles*, 63–76.
<https://www.insee.fr/fr/statistiques/2017510?sommaire=2017528>.
- Cox, D. R. (1972).** Regression models and life-tables. *Journal of the Royal Statistical Society, series B*, 34(2), 187–220.
- David, O., Eydoux, L., Ouallet, A. & Séchet, R. (2003).** Les familles monoparentales : une perspective internationale. *L'essentiel* N° 15.
<https://www.caf.fr/sites/default/files/cnaf/Documents/Dser/essentiel/15%20ESSENTIEL%20MONOPARENTALITE.pdf>.
- David, O., Eydoux, L., Martin, C., Millar, J. & Séchet, R. (2004).** Les familles monoparentales en Europe. *Dossiers d'études de la CAF* N° 54.
https://www.caf.fr/sites/default/files/cnaf/Documents/Dser/dossier_etudes/dossier_54_-_familles_monoparentales.pdf.
- Eydoux, A. & Letablier, M.-T. (2009).** Familles monoparentales et pauvreté en Europe : quelles réponses politiques ? [L'exemple de la France, de la Norvège et du Royaume-Uni]. *Politiques sociales et familiales*, 98, 21–35.
http://www.persee.fr/doc/caf_2101-8081_2009_num_98_1_2487
- Haut Conseil de la Famille (2014).** Les ruptures familiales – État des lieux et propositions. Rapport du 10 avril 2014.
<http://www.ladocumentationfrancaise.fr/rapports-publics/144000594/index.shtml>.
- Keiding, N. (2006).** Event history analysis and the cross-section. *Statistics in Medicine*, 25, 2343–2364.
- Kiefer, N. M. (1988).** Duration data and hazard functions. *Journal of Economic Literature*, 26(2), 646–679.
- Lancaster, T. (1990).** *The econometric analysis of transition data*. Cambridge University Press.
- Le Pape, M.-C., Lhommeau, B. & Raynaud, E. (2015).** Les familles monoparentales en Europe : de nouvelles façons de faire famille pour de

nouvelles normes ? Insee Références *Couples et familles*, 27–40.
<https://www.insee.fr/fr/statistiques/2017504?sommaire=2017528>.

McKay, S. (2002). The dynamics of lone parents, employment and poverty in Great Britain. *Sociology and Social Politics*, 6-2, 101-124.

Nickell, S. (1979). Estimating the probability of leaving unemployment. *Econometrica*, 47(5), 1249–1266.

Rabier, R. (2014). Les familles monoparentales, souvent en situation de précarité. *Insee Analyses Languedoc-Roussillon* N° 2.
<https://www.insee.fr/fr/statistiques/1285832>.

Régnier-Loilier, A. (2012). Présentation, questionnaire et documentation de la troisième vague

de l'Étude des relations familiales et intergénérationnelles (Érfi-GGS 2011). Ined, Document de travail N° 187.

<https://www.ined.fr/fr/publications/document-travail/etude-relations-familiales-intergenerationnelles/>.

Toulemon, L. (2011). Individus, familles, ménages, logements : les compter, les décrire. *Travail, Genre et Société*, 26, 47–66.

Vespa, J., Lewis, J. M. & Kreider, R. M. (2013). America's families and living arrangements: 2012. *Current Population Reports*, P20-570. Washington, DC: U.S. Census Bureau.

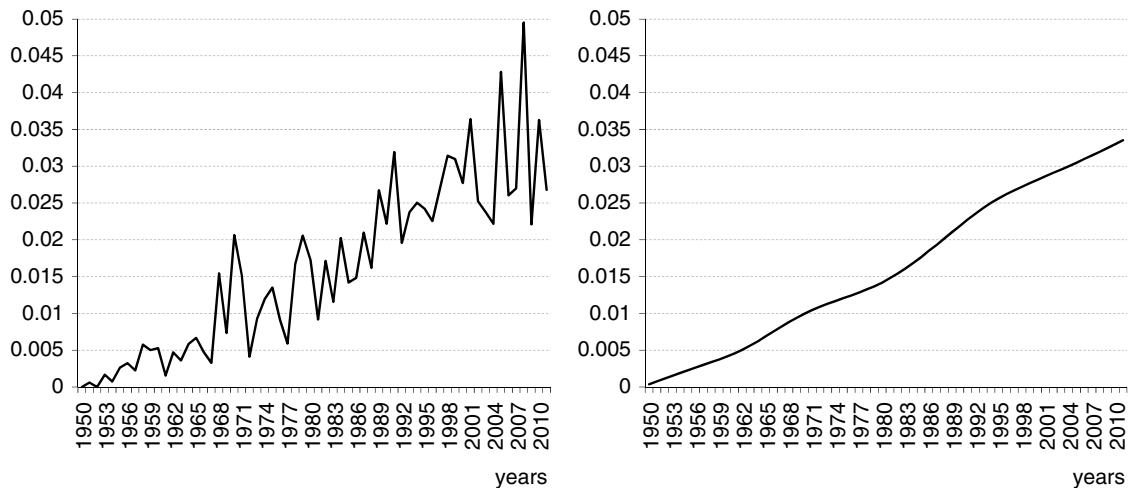
<https://www.census.gov/library/publications/2013/demo/p20-570.html>.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge (Ma): The MIT Press.

SMOOTHED ESTIMATIONS OF ENTRY FLOWS

Figure A1-I

Estimations of the annual flows of parents entering a period of single parenthood before and after smoothing



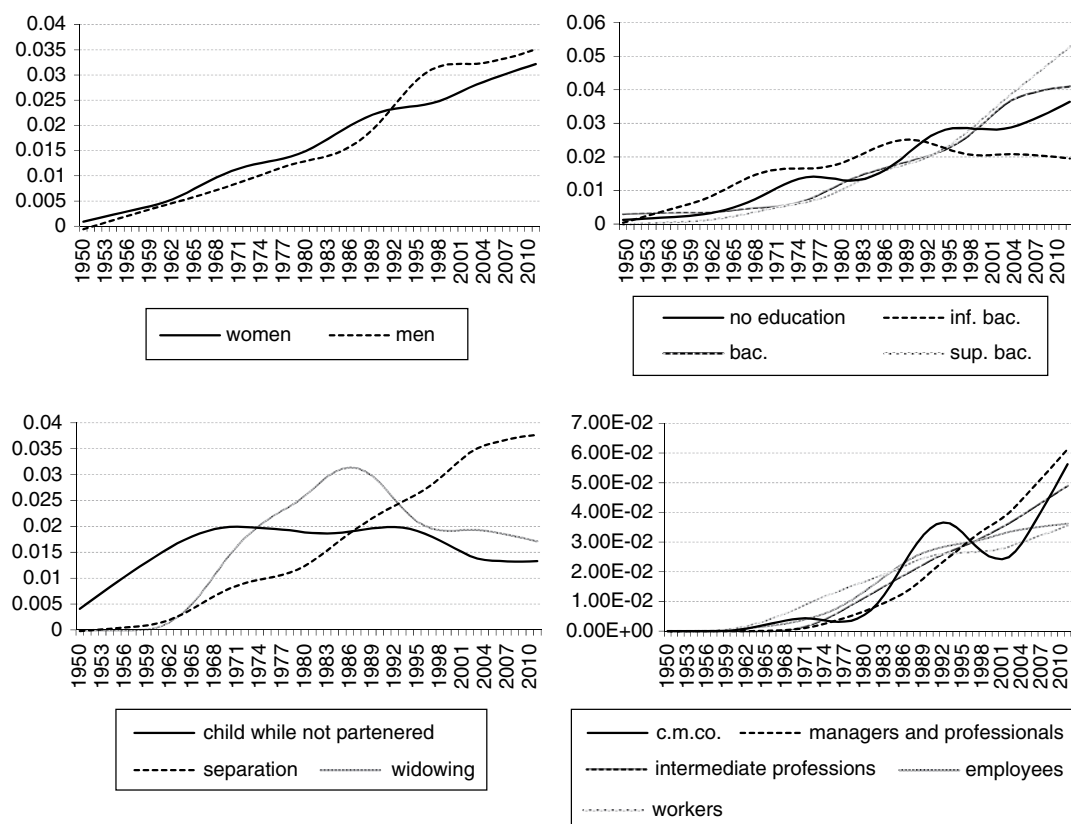
Note: the smoothing is obtained using a moving average followed by a polynomial approximation.

Coverage: mainland France.

Source : Ined-Insee, *Erfi*, waves 1 (2005) et 3 (2011).

Figure A1-II

Estimations des flux annuels d'entrée en famille monoparentale, pour quelques caractéristiques individuelles



Note: the smoothing is obtained using a moving average followed by a polynomial approximation. The social category c.m.co. corresponds to craftsmen, merchants and company owners.

Coverage: mainland France.

Source : Ined-Insee, *Erfi*, waves 1 (2005) et 3 (2011).

APPENDIX 2

ESTIMATION OF SURVIVAL FUNCTIONS FOR SINGLE PARENTHOOD DURATION

Table A2-1

Estimated survival functions in the model without covariates

	Total			Men			Women		
	Estimation	Lower boundary	Upper boundary	Estimation	Lower boundary	Upper boundary	Estimation	Lower boundary	Upper boundary
0	1	1	1	1	1	1	1	1	1
1	0.80	0.78	0.82	0.73	0.68	0.78	0.81	0.80	0.83
2	0.64	0.61	0.67	0.53	0.46	0.60	0.66	0.63	0.69
3	0.51	0.48	0.54	0.39	0.31	0.47	0.54	0.50	0.58
4	0.46	0.42	0.50	0.34	0.24	0.44	0.48	0.44	0.53
5	0.41	0.36	0.45	0.30	0.19	0.41	0.43	0.38	0.48
6	0.36	0.31	0.41	0.26	0.15	0.38	0.39	0.33	0.44
7	0.33	0.27	0.38	0.21	0.10	0.34	0.36	0.29	0.42
8	0.30	0.24	0.36	0.17	0.07	0.30	0.33	0.26	0.40
9	0.27	0.21	0.33	0.14	0.05	0.27	0.31	0.23	0.38
10	0.23	0.17	0.29	0.12	0.03	0.25	0.26	0.19	0.33
11	0.19	0.14	0.25	0.11	0.02	0.24	0.22	0.15	0.29
12	0.16	0.11	0.22	0.09	0.02	0.23	0.18	0.12	0.25
13	0.13	0.08	0.19	0.06	0.01	0.18	0.15	0.10	0.22
14	0.10	0.06	0.16	0.04	0.00	0.13	0.13	0.07	0.19
15	0.08	0.05	0.13	0.02	0.00	0.10	0.10	0.06	0.16
16	0.07	0.04	0.11	0.02	0.00	0.10	0.08	0.04	0.13
17	0.05	0.03	0.09	0.02	0.00	0.10	0.06	0.03	0.11
18	0.04	0.02	0.08	0.01	0.00	0.09	0.05	0.02	0.09
19	0.04	0.02	0.07	0.01	0.00	0.08	0.04	0.02	0.08
20	0.03	0.01	0.06	0.01	0.00	0.07	0.03	0.01	0.06

Coverage: Mainland France.

Source: Insee, *Family and Housing Survey (FHS)* 2011; Ined-Insee, *Survey of Family and Intergenerational Relations (Erfi)*, waves 1 et 3, 2005 and 2011.

Table A2-2

Survival functions estimated for women associated with the baseline hazard and a set of covariates (reason for entering situation, qualification, social category)

	Reason for entering the situation			Qualification			Social category		
	Estimation	Lower boundary	Upper boundary	Estimation	Lower boundary	Upper boundary	Estimation	Lower boundary	Upper boundary
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	0.89	0.87	0.90	0.85	0.87	0.90	0.84	0.87	0.90
2	0.79	0.76	0.81	0.73	0.76	0.81	0.71	0.76	0.81
3	0.70	0.67	0.73	0.62	0.67	0.73	0.60	0.67	0.73
4	0.65	0.61	0.69	0.57	0.61	0.69	0.55	0.61	0.69
5	0.61	0.57	0.65	0.52	0.57	0.65	0.50	0.57	0.65
6	0.57	0.52	0.62	0.48	0.52	0.62	0.46	0.52	0.62
7	0.54	0.48	0.60	0.45	0.48	0.60	0.42	0.48	0.60
8	0.51	0.45	0.57	0.42	0.45	0.57	0.38	0.45	0.57
9	0.49	0.42	0.55	0.40	0.42	0.55	0.34	0.42	0.55
10	0.44	0.37	0.51	0.35	0.37	0.51	0.30	0.37	0.51
11	0.40	0.32	0.47	0.31	0.32	0.47	0.26	0.32	0.47
12	0.36	0.29	0.44	0.27	0.29	0.44	0.23	0.29	0.44
13	0.32	0.24	0.39	0.23	0.24	0.39	0.19	0.24	0.39
14	0.28	0.20	0.35	0.20	0.20	0.35	0.16	0.20	0.35
15	0.24	0.17	0.32	0.17	0.17	0.32	0.13	0.17	0.32
16	0.21	0.14	0.28	0.14	0.14	0.28	0.10	0.14	0.28
17	0.18	0.12	0.25	0.11	0.12	0.25	0.08	0.12	0.25
18	0.15	0.09	0.22	0.09	0.09	0.22	0.06	0.09	0.22
19	0.13	0.08	0.19	0.08	0.08	0.19	0.05	0.08	0.19
20	0.11	0.07	0.16	0.07	0.07	0.16	0.04	0.07	0.16

Coverage: women, mainland France.

Source : Insee, *Family and Housing Survey (FHS)* 2011; Ined-Insee, *Survey of Family and Intergenerational Relations (Erfi)*, waves 1 et 3, 2005 and 2011.

APPENDIX 3

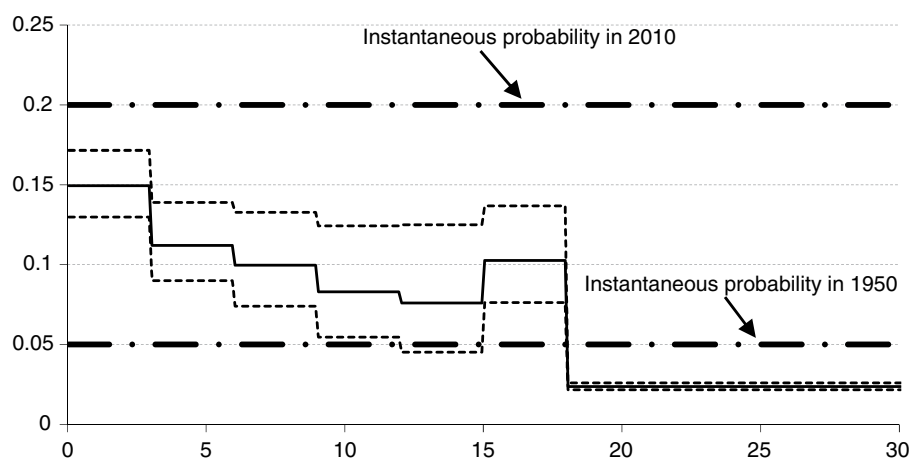
SIMULATION OF CORRELATIONS BETWEEN D AND T

To understand what happens when we release the assumption of independence between the variable of flow D and the variable of duration T , we simulate a population of 100 000 people entering the situation of single-parent uniformly between 1950 and 2010, but the duration of which is negatively correlated to the entry. For that purpose, we simulate T using a Weibull distribution of constant parameter of form of value 1 (a particular case for which the law is exponential) and a scale parameter of scale of value $(a-b \cdot D)$, where $a=507.5$ and $b=0.25$. So, for the people entered in the situation in 1950, the average duration is of 20 years, and 5 years for those entered in 2010 (of course this

situation is exaggerated and does not correspond to the reality). On the simulated population, the average duration is 12.4 years.

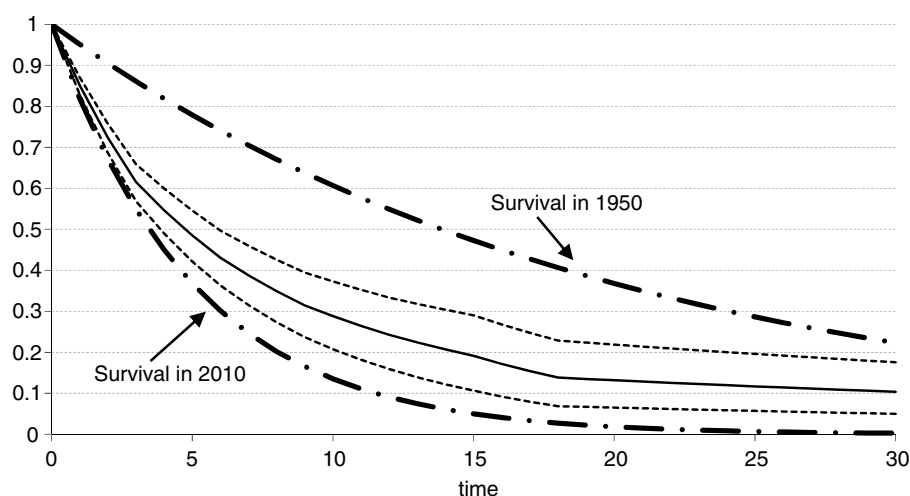
The estimated average duration is 8.2 years. We thus tend to underestimate the durations when there is a negative correlation between the duration T and the entry date in single-parenthood D . Nevertheless, the estimations actually provide estimates of an instantaneous probability and a survival which are situated between the instantaneous probabilities and the extreme survivals, i.e. those of 1950 and of 2010 (Figures A3-I and A3-II).

Figure A3-I

Estimation of the hazard function in presence of a correlation between variables D and T 

Note : for 100,000 individuals we simulated durations, the distribution of which depends on the date D of entering the period of single parenthood according to a Weibull distribution $(1, a-b \cdot D)$ where a and b are coefficients equal to 507.5 and 0.25. The dotted lines represent the 95% confidence interval obtained using the maximum likelihood method.

Figure A3-II

Estimation of the survival in presence of a correlation between variables D and T 

Note : for 100,000 individuals we simulated durations, the distribution of which depends on the date D of entering the period of single parenthood according to a Weibull distribution $(1, a-b \cdot D)$ where a and b are coefficients equal to 507.5 and 0.25. The dotted lines represent the 95% confidence interval obtained using the maximum likelihood method.

BIAIS OF THE MEAN ESTIMATE

Let us assume a continuous variable of duration noted T_c , and another discrete variable of duration, taking only integer values, noted T_d . Let us suppose furthermore that the survival function S_d of the discrete

variable is equal, for durations of integer values, to the survival function S_c of the continuous variable. In other words, $P(T_d \geq t) = P(T_c \geq t) \forall t \in \mathbb{N}$. Then we show that $E[T_d] \approx E[T_c] - 0,5$.

$$E[T_d] = \sum_{u \geq 1} S_d(u) = \sum_{u \geq 1} S_c(u) = \sum_{u \geq 1} \left[\int_{u-1}^u S_c(x).dx - \int_{u-1}^u (S_c(x) - S_c(u)).dx \right] = \sum_{u \geq 1} \int_{u-1}^u S_c(x).dx - \sum_{u \geq 1} R(u) = E[T_c] - R$$

where $R = \sum_{u \geq 1} R(u)$ and $R(u) = \int_{u-1}^u (S_c(x) - S_c(u)).dx$.

If $S_c(x)$ is approximated by a linear function between $u-1$ and u , then we have $S_c(x) - S_c(u) \approx (S_c(u-1) - S_c(u))(u-x)$ so that $R(u) \approx 0,5(S_c(u) - S_c(u-1))$ hence $R = \sum_{u \geq 1} R(u) \approx 0,5$.

It can be concluded that $E[T_d] \approx E[T_c] - 0,5$.