N° 493 - 2017

Economie et Statistique

Economics and Statistics

Mélanges

Varia

Insee

Economie et Statistique

Economics and Statistics

Directeur de la publication / Director of Publication:
Jean-Luc TAVERNIER
Rédactrice en chef / Editor in Chief: Laurence BLOCH
Rédactrice en chef adjointe / Deputy Editor in Chief:
Sophie PONTHIEUX
Assistante de rédaction / Editorial Assistant: Véronique ALEXANDRE
Maquette PAO et impression / CAP and printing: JOUVE
1, rue du Docteur-Sauvé, BP3, 53101 Mayenne

OÙ SE PROCURER
Economie et Statistique / Economics and Statistics

Les numéros sont en accès libre sur le site www.insee.fr. Il est possible de s'abonner aux avis de parution sur le site.

La revue peut être achetée sur le site **www.insee.fr** via la rubrique « Acheter nos publications »

La revue est également en vente dans 200 librairies à Paris et en province.

HOW TO GET
Economie et Statistique / Economics and Statistics

All the issues and articles are available in open access on the Insee website: www.insee.fr. Publication alerts can be subscribed on-line.

The printed version of the journal (in French) can be purchased on the Insee website www.insee.fr and in 200 bookshops in Paris and province.

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES
Directeur Général : Jean-Luc TAVERNIER
Direction Générale : 18, boulevard Adolphe Pinard, 75675 PARIS Cedex 14
Tél : 01.41.17.50.50 - Télécopie : 01.41.17.66.66

N° 493 - 2017

Economie et Statistique

Economics and Statistics

Economie et Statistique / Economics and Statistics

Issue 493 - 2017

EVALUATION OF PUBLIC SUPPORT TO SMALL AND MEDIUM-SIZED FIRMS

The effect of R&D subsidies and tax incentives on employment: an evaluation for small firms in France

Vincent Dortet-Bernadet * et Michaël Sicsic **

Abstract – Between 2003 and 2010, the amount of tax incentives and subsidies granted by French public authorities to finance the R&D activities of SMEs increased fourfold. This very sharp increase is due to the research tax credit (RTC) reforms, particularly in 2008, the creation in 2004 of a young innovative business status and an increase in subsidies over the period. Based on exhaustive employment data for France, this paper presents the first ever evaluation of the effect of the increase in these aids on small firms.

Using a method that combines matching and a labour demand model, we show that the effect of public support on R&D employment is positive and increased during the period 2004-2010. Nonetheless, the increase in aid, particularly subsequent to the wide ranging reform of the RTC in 2008, was accompanied by a significant crowding-out effect: according to our estimates, only between 18 and 34% of the supplementary aid obtained by businesses between 2008 and 2010 was used to finance new jobs for highly qualified workers.

JEL codes: O38, H25, C33, C36
Keywords: R&D subsidies, R&D tax incentives, public policy evaluation, matching, labour demand model

** Insee (vincent.dortet-bernadet@insee.fr).*
*** Insee et Cred, Université Panthéon-Assas Paris II (sicsic.michael@gmail.com).*
The authors would like to thank Louis de Gimel, Claire Lelarge, Corinne Prost, Loriane Py, Fabienne Rosenwald, Frédérique Sachwald, three anonymous referees, as well as the participants at Insee's D2E seminars, AFSE/DG Treasury's public policy evaluation, and Cred's Working Group on fiscal economy. The authors would also like to thank the Directorate General for Research and Innovation at the Ministry for Education, Higher Education and Research for allowing them to use the research credit management database. The authors are sole responsible for any error or omission in this document.

Received August 27, 2015; accepted after revision March 29, 2017 DOI: 10.24187/ecostat.2017.493s.1914
Translated from the original version: « L'effet des aides à la R&D sur l'emploi : une évaluation pour les petites entreprises en France »

Research and development (R&D) activities are designed to promote the emergence of new manufacturing materials, products or processes and improvements to them. Stimulating innovation and technical progress, R&D is an important source of economic growth (Griffith *et al.*, 2003; 2004). In France, the ratio of domestic business expenditure on R&D (BERD)[1] remains quite weak compared to other major countries: 1.45 % of GDP in 2014 compared to 1.6 % for OECD countries, almost 2 % in Germany and 2.8 % in Japan (OECD, 2017). Nonetheless, during the 2000s, public authorities significantly developed support schemes in order to boost private R&D expenditure. The State allocated to firms some 8 billion Euro in financial support to R&D in 2013, i.e. close to 0.4 % of GDP (compared to less than 0.2 % in 2003). In 2013, France was the 3rd biggest public funder of R&D in the world and the leading country in terms of tax incentives for R&D (OECD, 2016).

This paper studies the impact on employment of the significant increase in public support to R&D received by small businesses during the period 2004-2010. Small firms have, for example, benefited from the creation in 2004 of the "Young innovative business" status (*Jeune Entreprise Innovante*, JEI) for firms younger than eight years old and specialised in R&D activities. More significantly, the various research tax credit (*RTC*) reforms from 2004 onwards, and particularly in 2008, allowed for a significant increase in the number of small companies benefiting from this scheme. Lastly, small firms have also been given the opportunity to receive subsidies from Oséo[2], a body set up in 2005 to support innovation projects undertaken by small and medium enterprises (SMEs).

Few studies have focused on evaluating the effect of R&D public support received by small businesses. These firms, nevertheless, receive the highest support rates: in 2010, 50 % of R&D expenditure declared by very small businesses[3] (VSBs) benefiting from RCT was funded through aid compared to 42 % for other SMEs, 36 % for intermediate-sized enterprises (*Entreprises de Taille Intermédiaire*, ETI) and 34 % for the biggest firms (Dortet-Bernadet & Sicsic, 2015, p. 15).

Why help small firms to perform R&D?

Public support for firms to fund their R&D expenditure is justified by the fact that, without this aid, they would tend to perform less R&D than the level desirable for the whole economy (Jones & Williams, 1998; Bloom *et al.*, 2013). By reducing the private cost of R&D activities, public funding is likely to increase R&D expenditure to a socially optimal level. Providing specific support to small and young firms can be justified by the fact that these firms are more affected by financing restrictions than other firms. These restrictions would give rise to excessively low R&D investments with procyclical evolutions[4] (Aghion *et al.*, 2012). Financial public support can also help small firms that would otherwise not have undertaken R&D activities (González *et al.*, 2005).

Helping the youngest businesses can also be effective, as it is argued that these firms are behind breakthrough innovations (Schneider & Veugelers, 2010; Cincera & Veugelers, 2012; Akcigit & Kerr, 2010). Based on simulations using a theoretical model, Acemoglu *et al.* (2013) find that it would be more effective to subsidise firms entering the market (especially young and small firms) to undertake R&D than already established firms. However, it should be noted that, in an empirical analysis, Garcia-Macia *et al.* (2016) find that most of the growth in productivity does not come from young businesses, but rather from improvements to the products of incumbents.

Although public support for R&D can serve to boost private funding of R&D (amplifying effect of aid), it can also simply have an additive effect, or even be a substitute for privately funded R&D (deadweight loss or crowding-out effect) (David *et al.*, 2000). In this latter case, the firms use the public funds to finance projects that they would have performed anyway, even without public support.

1. Gross domestic expenditure on research and development (GERD) corresponds to the research and development (R&D) activity performed on the national territory regardless of the funding source. Some of this activity is performed by public administrations, while the rest is carried out by firms (Business expenditure on research and development, BERD). This includes current expenditure (wage bill of R&D staff and operating costs) and capital expenditure (purchase of equipment needed for domestic R&D activity and real estate transactions completed over the course of the year).

2. Oséo was set up in 2005, bringing together the Anvar, BDPME and Sofaris, and was then incorporated into BPIFrance in 2013.

3. See Box 2 for the definition of the categories of companies.

4. R&D as a share of investments is said to fall during periods of recession and it is argued that this fall is not fully offset during periods of economic recovery.

Differing results relating to the effect of R&D depending on the size of the company

There is a significant body of literature on evaluations of the impact of public support to business R&D (Ientile & Mairesse, 2009; Kohler *et al.*, 2012; Zuñiga-Vicente *et al.*, 2014, for literature reviews). It points to the differing results of research into the effectiveness of R&D aid received by small firms. Some of these differences can be attributed to the variety of support schemes in place in the various countries of the OECD. For example, for Busom *et al.* (2014), young Spanish businesses without any experience of R&D mainly use subsidies and tax credit schemes are less suited to firms embarking on an R&D activity. In Spain once again, Corchuelo and Martinez-Ros (2009) demonstrate that tax incentives to perform R&D are more effective for large firms than for SMEs. However, according to Lokshin and Mohnen (2012), who examined a tax credit scheme proportionate to expenditure volumes in the Netherlands, tax incentives are particularly effective for small firms, the only category of firms where crowding-out effects do not come into play. Hægeland and Møen (2007) reach a similar conclusion for tax credits in Norway. However, like Lokshin and Mohnen (2013), they also show that support has a significant effect on the increase in salaries paid to researchers. In Italy, Bronzini and Iachini (2014) have highlighted the additive effect of an R&D subsidy programme for small businesses, but not for large firms. Finally, based on an existing system in Québec, Baghana and Mohnen (2009) point to the fact that a tax credit that is proportional to the volumes of R&D expenditure is not effective for large firms, but is for small businesses: for the latter, the increase in R&D expenditure outstrips the amount of financial support.

Studies based on French data use partial data relating to small businesses

In France, evaluations are mostly based on data from the R&D Survey conducted by the Ministry for Education, Higher Education and Research (MENESR). This survey provides very detailed information about the expenditure of major producers of R&D. However, it only partially covers the population of young or small firms and its changing coverage makes it difficult to use only this survey to evaluate the impact of R&D public support on small firms (Box 1). Research conducted to assess the effectiveness of R&D support using the data from this survey has thus focused on medium-sized and large enterprises. It is mostly the research tax credit (RTC) that has been evaluated: amongst the most recent research, we can cite Duguet (2012), Mulkay and Mairesse (2013), Bozio *et al.* (2015). These evaluations reach the overall conclusion that RTC has an additive effect or a slight amplifying effect. Duguet (2004) obtains a similar result for direct financial support paid out over the period 1985-1997. Lhuillery *et al.* (2013)

Box 1

THE R&D SURVEY: PARTIAL COVERAGE OF SMALL FIRMS AND CHANGES DURING THE 2000s

Evaluating the effectiveness of R&D financial public support for small firms based solely on the MENESR's R&D Survey poses statistical difficulties: the survey only provides partial and changing information about young and small firms.

Firstly, the survey only partially covers the population of small firms receiving indirect financial support (tax credits and tax breaks). Indeed, using only the survey sample and weightings leads to systematically underestimating the amount of indirect aid received by small firms. For example, between 2003 and 2010, it only covered 61% of the amount of research tax credits (RTC) received by very small firms (see online complement C1).

Secondly, the survey provides fairly unrealistic estimations of the change in R&D expenditure by small firms, as its coverage changed over the course of the 2000s. As the survey database is updated based on the lists of firms applying for aid, the increasing number of SME's benefiting from the RTC automatically expanded the coverage of the survey, which gave rise to very volatile estimations of changing levels of R&D employment: - 18% in 2005, + 40% in 2006, - 10% in 2007, + 22% in 2008.

The survey is also not well suited to individual monitoring of the R&D expenditure of the smallest firms in the panel. The youngest (under two years old) and the smallest firms have a very low probability of being surveyed over two consecutive years (Bellégo & Dortet-Bernadet, 2014). The survey is not designed either to observing firms that are just starting an R&D activity, as the updating of the database only takes account of the firms that have already carried out R&D.

point broadly to the additive effect of subsidies and research tax credits, but they also observe crowding-out effects for firms that receive low or moderate levels of support[5].

Studies that exclude the largest firms from their coverage arrive at more mixed results. According to Serrano-Velarde (2008), obtaining subsidies from Anvar is accompanied by a fall in private R&D expenditure (R&D net of aid) among SMEs and intermediate-sized enterprises that have received support. However, within a comparable coverage, Bellégo and Dortet-Bernadet (2014) show that the supplementary public funding received for being involved in competitive clusters' did not lead to a reduction in private expenditure. Nonetheless, these two studies were based on the R&D Survey, which excludes the smallest companies. Lelarge (2009), who does not use solely the R&D Survey, but also data similar to ours, shows that during the early years of the JEI scheme (2004-2005), firms increased the wages paid out, which enabled them to retain their most qualified staff.

Studying the effect of R&D public funding using exhaustive employment data

In order to avoid the difficulties associated with using the R&D Survey for small firms (Box 1), we propose studying only part of the R&D expenditure: that relating to R&D jobs. We evaluate the effect of R&D public support on the employment of highly qualified staff[6], for which exhaustive data are available through the *Déclarations annuelles de données sociales* (*DADS* - annual declarations submitted by employers, Box 2). The effect of support on the employment of R&D personnel is then deduced from this.

Only the effect of the total amount of support is studied: although this option does not allow for

5. They also show that the most effective aid is either very low or very high levels of aid.

6. Defined as total employment in the categories of senior managers, higher intellectual professions and company managers (professional categories 2 and 3 in the DADS*).*

Box 2

DATA

Data relating to R&D employment and public funding of R&D

Several databases are used in this study in order to measure the amounts of financial support received by the firms and to estimate the number of R&D-related jobs:

- The database used to manage research tax credits (GECIR, source: MENESR). In addition to the amount of the tax credit, this database provides information about all of the subsidies received by the firms in order to finance their R&D activities.
- The register of participants in the "Young Innovative Business" scheme (JEI, for *Jeunes Entreprises Innovantes*) (source: Acoss). This register provides the total amount of exemptions from employer payroll taxes granted to the participating firms. Tax exemptions (of the research tax credit type) are not taken into account, but they only represent 10% of the total amount in 2010.
- The list of MENESR accreditations (source: MENESR). An accredited firm performs R&D for other firms, which are thus entitled to benefit from the RTC.
- The R&D survey database (source: MENESR): the survey is used to measure the amount of direct support and estimate the number of R&D jobs between 2008 and 2010.

Other sources of information

In order to reconstruct categories of firms and to estimate the number of R&D posts, various Insee databases are used: tax data (*Ficus/Ésane*, Insee), the *Déclarations de données sociales* (DADS - annual declarations submitted by employers), the database of financial ties (*Lifi*) and the national register of enterprises and establishments (*Sirene*, Insee). It should be noted that the figures related to turnover, valued added, wages and aid are deflated using the value-added price index for each branch of activity (based on a reference date of July 2000).

The DADS provide an exhaustive description of salaried employment situations by professional category. Staff levels by professional category were recalculated in 2009 and 2010 in order to control the influence of a change in the method of coding professional categories. The labour cost has been estimated based on gross wages, to which have been added estimations of the levels of employer payroll taxes as proposed by Cottet *et al.* (2012).

The study covers small and medium enterprises (SMEs), as well as very small businesses (VSBs); the former count less than 250 employees, do not have an annual turnover in excess of 50 million Euro or a total balance sheet that does not exceed 43 million Euro; the latter have less than 10 employees, an annual turnover or a total balance sheet of 2 million Euros at most (see definition in Béguin *et al.*, 2012). 'Firms' relate solely to independent legal entities or groups: legal entities belonging to large groups were excluded from the coverage of the study.

a comparison of the respective merits of each R&D support scheme, it does allow for the study of the large number of cases where firms make use of several support schemes simultaneously[7]. In order to estimate the amount of R&D support received by small firms, we use lists of firms benefiting from indirect support mechanisms (RTC and JEI scheme), as well as the R&D Survey, which allows us to take account of direct regional, national and European support (Box 2).

The paper is structured as follows: In the first section, we recall the main developments in the R&D public support to SMEs during the period 2003-2010. We then estimate the aggregated changes in employment levels in the field of R&D and we show that, for SMEs, employment not financed by public support has fallen significantly. In the second section, we restrict the focus to small firms from R&D-intensive sectors and evaluate the effect of the R&D support on employment for the years 2004-2010, based on a panel of firms that received aid or did not receive aid. The results indicate that public funding had positive effects on the employment of highly qualified staff (and R&D staff), but there were also windfall effects, particularly by the end of the period.

Aggregated changes in R&D public funding and R&D-related employment during the period 2003-2010

Changes in the R&D support schemes used by SMEs

R&D support schemes include indirect and direct aids. Indirect aid primarily include the RTC and reductions in employer payroll social contributions that are part of the JEI status, while direct subsidies are allocated by various bodies responsible for promoting R&D in firms.

The RTC is a tax break granted to firms of all sizes that perform R&D. Between 1983 and 2003, the mechanism basically kept the same structure: the amount of the tax credit depended on the increase in R&D expenditure from one year to the next and was capped at a certain amount (MENESR, 2014). The first major reform of the RTC took place in 2004 when a supplementary tax credit share was introduced based on the volume of R&D expenditure. This represented 5% of expenditure in 2004 and 2005, and then 10% in 2006. The tax credit calculated on the basis of the increase in this expenditure was, however, reduced gradually (Table 1) and the tax credit cap was increased to 16 million Euro in 2007.

The reform of the RTC in 2008 abolished the tax credit based on the increase in R&D expenditure, increased the rate applicable to the expenditure volume to 30% for amounts up to 100 million Euro, then 5% above this amount, and abolished the cap on the amount of the tax credit. Higher rates have also been applied to firms applying for the first time for the RTC (a rate of 50% for the first year and 40% for the second).

Following the successive reforms of the RTC, the amount of this aid increased eleven-fold between 2003 and 2010, reaching 5 billion Euro.

The JEI status created in 2004 entitles SMEs that are less than 8 years old and whose R&D activities account for at least 15% of their charges to pay lower employer payroll contributions. The total amount of support linked to the JEI status is much lower than for the CIR (some 140 million Euro in 2010), but doubled

7. *In 2010, over 80% of the companies that received a direct aid also benefited from an indirect support.*

Table 1
Changes to the parameters of the research tax credit (RTC) between 2003 and 2010

	from 1991 to 2003	2004-2005	2006	2007	From 2008 to 2010
Rate (%) volume-based		5	10	10	30 % up to 100 million Euro 5 % above 100 million Euro Higher rates 50 % for the1st year and 40 % for the 2nd year (*)
Rate (%) increment-based	50	45	40	40	///
Cap (in millions of Euro)	6.1	8	10	16	Cap removed

(*) For firms applying to RTC for the 1st time
Source: based on MENESR documents

between 2004 and 2010. Moreover, for VSBs, it amounts to around 20% of all indirect subsidies.

Direct public aids are subsidies aimed at specific projects or covering a specific type of expenditure. These subsidies include refundable advances (refunds depend on the success of the project receiving support), premiums, subsidised loans, guarantees and public procurement orders. They are granted by local authorities, various national bodies, such as Oséo or the *Fonds unique ministériel* (FUI)[8], or by the European Union[9]. As of 2005, Oséo-Innovation was specifically charged with financing R&D performed by SMEs. Direct aids increased in total by 64% for SMEs between 2003 and 2010, whereas they remained stable for intermediate-size and large firms (Dortet-Bernadet & Sicsic, 2015).

Finally, the total amount of (direct and indirect) support to R&D received by SMEs increased by 300% between 2003 and 2010, reaching nearly 2 billion Euro, 26% of which, or some 500 million Euro, was received by VSBs (Dortet-Bernadet & Sicsic, 2015).

Aggregated estimation of changes in R&D employment in SMEs not financed by public funding

In this section, we present a three-stage estimation of changes between 2003 and 2010 in total R&D-related employment (hereafter referred to as R&D employment) in SMEs and the amount of support received to fund it. We first estimate the change in expenditure on R&D employment in SMEs (stage 1), and then the change in support for R&D spent on employment (stage 2). Lastly, in the 3rd stage, we compare these two results in order to estimate the change in the numbers of R&D jobs that have not been funded by public support.

Stage 1: estimation of R&D employment

To estimate the change in R&D employment, we make a two-step calculation. Firstly, we calculate this employment in SMEs (including VSBs) during the period 2008-2010 using data from the R&D survey, the Gecir database[10], the list of JEI, and the list of MENESR accreditations (see Box 2). We work on the assumption that, for the period 2008-2010, the development of public support to R&D enables us to obtain an almost exhaustive list of SMEs that undertake R&D activities. We then calculate the change in R&D employment for the period 2003-2010 based on the assumption that, for each sector (level 5 of the French classification of activities, NAF) and category of firm, the ratio[11] of the number of R&D jobs to the number of 'highly-qualified' jobs ($HQ\ jobs_t^{sector,size}$, source: *DADS*) is stable over time:

$$Estimated\ \ R\,\&\,D\ jobs_t^{sector,\,size} = \frac{R\,\&\,D\ jobs_{2008-2010}^{sector,size}}{HQ\ jobs_{2008-2010}^{sector,size}} \times HQ\ jobs_t^{sector,size}$$

This assumption of stability may appear to be a strong one, as this ratio may have increased as the support for R&D increased. However, a calculation based on the European innovation Survey shows that the ratio of firms domestic expenditure on R&D (BERD) to expenditure on highly qualified employment remained broadly stable for SMEs, excluding VSBs, between 2004 and 2008 (falling from 11% to 9%).

According to our estimations, R&D employment in SMEs (excluding VSBs) increased more than R&D employment in VSB between 2003 and 2008 and the 2008/2009 crisis had a greater impact on the latter (Figure I). Expenditure on R&D employment follows a similar trend. Nonetheless, unlike with R&D employment levels, expenditure on R&D employment in VSBs did not fall between 2003 and 2010: it increased by 5% (16% for other SMEs). These trends are very different from those obtained on the basis of the R&D Survey, but they seem more realistic (see discussion in the online complement C1).

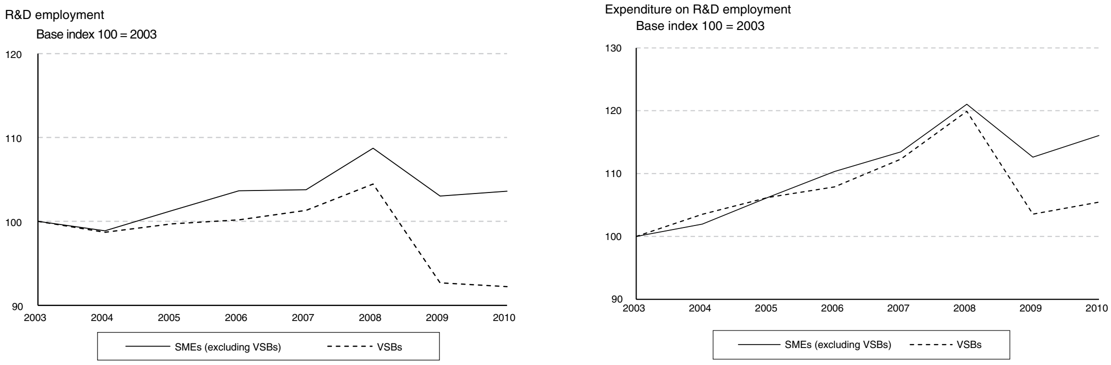
Stage 2: estimation of public funding used to finance R&D jobs

To estimate the amount of public support used to finance R&D employment, different rules are applied depending on the support scheme. For

8. For example, for projects conducted within the framework of competitive clusters (Dufau, 2017 ; Bellégo & Dortet-Bernadet, 2014).
9. Financing can be secured under the Framework Programme for Research and Technological Development or the European Regional Development Fund.

10. We take the amount of expenditure on R&D staff recorded in the database divided by the mean labour cost of an engineer (source: DADS*).*
11. Only part of highly-qualified jobs are allocated to R&D but, for the SMEs that responded to the R&D Survey, we get a correlation of 62% between real R&D employment and the estimation made using our method and a correlation of 72% for the companies in the panel used in the last section of the paper.

Figure I
Evolution of R&D employment and expenditure on R&D employment



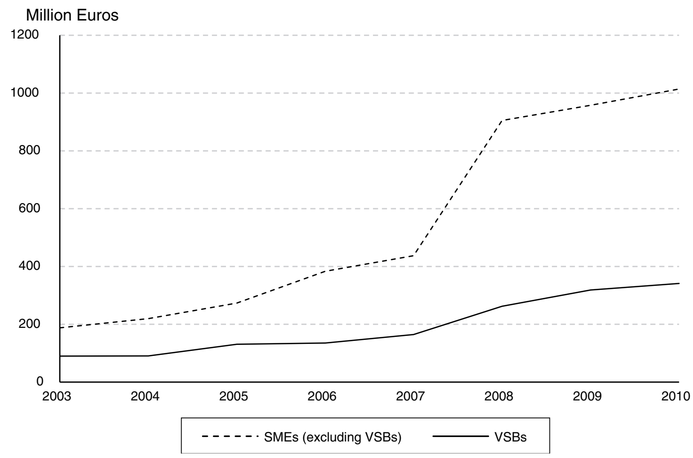
Reading note: between 2003 and 2010, R&D-related employment fell by 8% in VSBs, and rose by 4% in other SMEs. R&D-related employment expenditure rose by 5% in VSBs and by 16% in other SMEs.
Coverage: France, trade, manufacturing and market services.
Source: MENESR, GECIR database, *R&D Survey*; Acoss, JEI database; Insee, Lifi, Ficus/Ésane, *DADS*. Authors' calculations.

JEIs, the exemptions from employer payroll contributions are entirely considered in their entirety as aid for R&D employment. The share of the RTC that serves to finance employment corresponds to the share of staff and operating expenditure in the RTC tax base. To calculate the CIR tax base, the operating expenditure is set at 75% of staff expenditure: the resulting tax credit can thus be interpreted as an employment aid.

Unlike with indirect aid, no database provides exhaustive information on direct aids: their amount must be estimated. For each company, the estimation of support used to finance employment is made on the basis of the information reported in the CIR database, supplemented by, where necessary, information from the R&D Survey. As these sources sometimes differ, the amount of support retained is the highest amount reported by one of these sources.

According to our estimations, R&D support devoted to employment accounts for around three quarters of R&D support received by VSBs and other SMEs. Between 2003 and 2010, R&D public support devoted to employment increased by 280 % in VSBs and 440 % in SMEs, with a particularly large increase in 2008 due to the reform of the RTC (Figure II).

Figure II
Public funding of R&D employment



Reading note: in 2010, 341 million Euros and 1.01 billion Euros of R&D support were devoted to R&D employment in, respectively, VSBs and the other SMEs.
Coverage: France, trade, manufacturing and market services.
Source: MENESR, GECIR database, *R&D Survey*; Acoss, JEI database; Insee, Lifi, Ficus/Ésane, *DADS*. Authors' calculations.

Stage 3: change in R&D employment not financed by public funding

By taking the ratio of the amount of aid (estimated during stage 2) to the mean labour cost of R&D employment (estimated on the basis of the results of the first stage), we can estimate the amount of R&D-related employment that is 'funded by public support'. On the basis of the estimation of R&D employment in the second stage, we can then deduce R&D employment 'not funded by public support', i.e., the share of R&D employment that would not have benefited from public funding. According to our estimations, this share fell sharply (by 46 % between 2003 and 2010 for VSBs) and less significantly (- 16% over the same period) for other SMEs (Figure III). For very small firms, the fall was 9% between 2004 and 2007, followed by a more significant fall in 2008, the year in which the RTC was reformed (fall of 41% between 2007 and 2010). Overall, considering all SMEs (VSBs included), the fall was around 20% over the period.

These trends in R&D employment not funded by public support suggest deadweight effects, particularly from the significant increase in support to R&D in 2008. Nonetheless, this period was also marked by the financial crisis, which may have had an impact on the employment of researchers by small firms. The size of firms is also likely to have changed over time, including as a result of receiving support. For example, the most dynamic VSBs that received aid may have become SMEs (excluding VSBs), which could explain part of the fall in R&D employment not financed by aid within the category of VSBs (the same applying to the transition from SME to intermediate-sized enterprise).

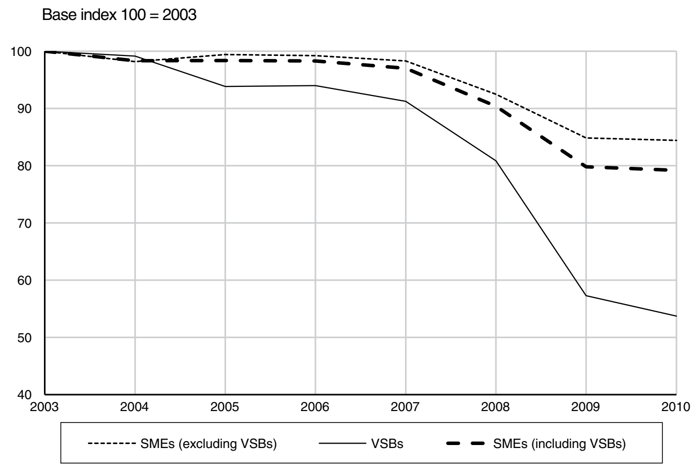
Evaluation of the effect of R&D public funding on employment

In this section, we seek to evaluate the effect of public funding on R&D-related employment in small firms. This evaluation monitors small firms, regardless of any changes in their size category, by comparing them to firms that were initially 'similar' and operating in the same economic environment.

To be more precise, the effect of R&D support on employment is estimated on the basis of a panel of small firms monitored over several years (2003-2010). As in the previous section, any financial public support that can be associated with R&D employment is taken into account.

The firms that receive public funding a given year cannot, however, apply for aid the following year, the effect of the aid already received continuing over time. In order to take this lagged effect into account, the firms 'treated' in any given year include those firms having

Figure III
Evolution of R&D employment not financed by public funding



Reading note: R&D employment 'not funded by public support' in 2003 is the reference (index=100). In 2010, this index reaches 54 for VSBs, i.e. a fall by 46% compared to 2003.
Coverage: France, trade, manufacturing and market services.
Source: MENESR, GECIR database, *R&D Survey*; Acoss, JEI database; Insee, Lifi, Ficus/Ésane, *DADS*. Authors' calculations.

received support that year and those that had already received it during the previous years.

The estimations are obtained first of all for highly qualified employment, then converted into R&D-related employment using the method presented in the previous section.

Construction of the panel used for the estimations

The evaluation is made using a panel of small firms from the 75 most highly R&D-intensive sectors of the economy (Dortet-Bernadet & Sicsic, 2015, p. 48). These firms are regarded as small because they have all been VSBs for at least one year during the period 2000-2010. They continue to be followed even if they grow and become an SME with 10 or more employees or an intermediate-sized enterprise[12]; however, most of the firms covered remain VSBs. The coverage also includes the large majority of young enterprises.

The coverage of the evaluation is restricted to firms that have highly-qualified staff. We calculate the effect of the support for firms present in 2003 (i.e., those that had highly qualified staff in 2003) and the effect for firms present in 2007[13]: these two reference years were chosen in order to observe the firms prior to the two major reforms of the RTC in 2004 and 2008. The panel is not balanced: some firms ceased to exist prior to 2010 or were created after 2003. However, each company must have had highly qualified employees for a period of at least two years (including the reference year, 2003 or 2007).

The panel includes firms that received support and others that did not, but which are similar to those that received it. The firms that never received support were selected on the basis of their age and a propensity score that estimates the probability of a company receiving support at least once between 2004 and 2010 on the basis of different variables (see online complement C2). This model indicates that the support schemes are more frequently used by young enterprises that make investments, export and have a lot of qualified staff. These results seem to be consistent with the idea that support schemes are used more by young, developing firms that do not yet produce very much or nothing at all (turnover has a negative effect and is barely significant). In total, the panel contains 15,128 firms, 4,597 of which received support at least once between 2003 and 2010 (Table 2).

The sectors are grouped together in three main categories: industrial, information and communication (IT, publishing, telecommunications, etc.) and a third category bringing together sectors comprising specialised, scientific and technical activities (R&D, engineering, etc.). Firms from services sectors form clearly the biggest sector (79%). Almost all the firms having received support benefited at least once from the RTC. The sectoral breakdown of subsidised firms is similar to that of firms that benefited from the CIR. A very large share of the firms granted the ‘Young innovative business’ status (JEI) are IT service firms.

12. In 2010, a third of the firms in the panel were SMEs and 3% intermediate-sized firms. Some were bought out by large groups, but these cases are very few. As they can give rise to ambiguities about the continuity of the initial activity, they were removed from the databases used to make the estimations.

13. These two treatment groups are not disjointed, as some companies have highly qualified workers in both 2003 and 2007.

Table 2
R&D public funding received by the small firms in the panel between 2003 and 2010

Firms that have	… received R&D public support at least once	… benefited from the RTC	… received a subsidy	… had JEI status	Firms that never received any aid
Number of firms	4,597	4,064	2,334	1,348	10,531
Breakdown (in %)					
Industry	20	20	19	9	22
IT services	42	42	41	55	33
Scientific and technical activities	38	38	40	36	46

Reading note: of the panel of 15,128 firms, 4,597 firms received aid at least once between 2003 and 2010, 4,064 benefited from the RTC, 2,334 received a subsidy, 1,348 enjoyed JEI status and 10,531 received no aid.
Coverage: panel comprising small firms present in 2003 and/or 2007 from 75 R&D-intensive sectors, having received R&D aid between 2003 and 2010 or which are ‘similar’ to the firms receiving aid (matched on their propensity score).
Source: MENESR, GECIR database, *R&D Survey*; Acoss, JEI database; Insee, Lifi, Ficus/Ésane, *DADS*. Authors’ calculations.

Identification strategy

Problems arising from the use of the difference-in-difference method

The effect of support on employment corresponds to the difference between the number of highly qualified jobs observed in the firms receiving support and the number of jobs there would have been if no aid had been received. To compute this effect, we must consider a fictitious situation where the firms that received support (treated firms) do not receive any support (or supplementary support). This estimation can be made by the difference-in-difference method based on the assumption that, without support, treated firms would have behaved in the same way as those that never received any support (non-treated firms) and which have a similar propensity score (see the results in the online complement C3).

This method gives rise, however, to several difficulties. The first relates to the choice of the control group. Indeed, the assumption of common trend for the treated and non-treated firms is not respected: before even receiving initial support, employment was more dynamic in the firms that received aid than in those never having received it. The firms that have not yet received support are then a more satisfactory control group: firms that received support as early as year *t* and those that only receive aid after *t* show similar trends in terms of highly-qualified employment through to *t*-1[14].

The second difficulty relates to the interpretation of the 'treatment' received by each generation of firms receiving aid. The simple difference-in-difference method does not allow us to account for the heterogenity within each generation of firms receiving support in terms of the amount received and the changes in the support rate.

Lastly, the simple difference-in-difference method does not allow us to take into account the firms having received support as of the reference year (or earlier) taking account of the amounts already received at that time.

An evaluation method combining labour demand and matching models

In order to overcome these various problems, we add all the firms that have not yet received support, but are going to receive some before the end of 2010 to the control group, and we estimate a labour demand model. This model allows us to calculate a level of employment based on the labour cost minus support (considered as a labour cost reduction) and the turnover. This allows us to compare between treated and non-treated firms not in terms of changes in employment levels but in terms of changes in the labour cost and the turnover, for which the assumption of common trend is better verified[15].

Using the labour demand model, where the demand for labour depends on its cost, also enables us to control for the initial amount of aid received and, therefore, to take account satisfactorily of all the firms receiving aid rather than just those that receive their first lot of aid after the reference year (2003 or 2007). We thus obtain results on the extensive margin (effect on the new firms receiving aid) and on the intensive margin (increase in the rate of aid).

The labour demand model for highly qualified labour

The labour demand model for highly qualified labour (l_{it}[16]) that is used is derived from a business costs minimisation programme with a Cobb-Douglas-type production function, close to the model presented in the paper by Bresson *et al.* (1992). It assumes that firms choose their level of highly qualified labour based on their turnover (y_{it}) and the relative mean cost of the highly-qualified labour (c_{it}) compared to other forms of labour.

The model is estimated only for firms receiving support. For these firms, support is equivalent to a reduction in the cost of highly qualified labour (rather than just the cost of R&D-related

14. By using as the control group the generation of companies that only received support as of 2010, we obtain a negative effect of support on highly qualified employment not financed by aid as of 2008 (online complement C3, Tables C3-4 and C3-5).

15. Before receiving their first support at time t*, the companies resemble the companies that have not yet received any support: they experience similar changes in the labour cost of highly-qualified staff and the turnover. For the companies that have never received any aid, the cost of labour is subject to similar changes, but the turnover remains less dynamic (online complement C4).*

16. All non-dichotomous variables mentioned in the model definition are expressed in logs.

jobs [17]) : the cost of labour considered here is a final cost after deducting the amount of R&D aid[18]. The model takes account of the lagged adjustment in firms' demand for labour and is expressed in the form of a autoregressive model:

$$\begin{aligned} l_{it} &= \rho . l_{it-1} + \alpha_1 . y_{it} + \alpha_2 . y_{it-1} \\ &+ \beta . c_{it} + \mu_i + \delta_t + \varepsilon_{it} \end{aligned} \quad (1)$$

where the level of highly qualified employment at time t depends on the level achieved the previous year, the turnover at t and $t\text{-}1$ and the final relative mean cost of labour at t.

Highly qualified labour is diverse in nature: it includes jobs devoted to R&D and jobs with no link to this type of activity. Bresson *et al.* (1992) recommend, in the event of diverse labour forms, to supplement the model by adding the cost variable measured at t-1, but this variable was not used here as it turned out to be too correlated to the cost at t and not significant [19].

17. In the case of small firms, this assumption appears to be fairly realistic because senior managers that do R&D only devote part of their working time to this. The fact that many small firms use R&D aid intermittently (between 2004 and 2010, around 25% of the VSBs receiving aid one year were no longer doing so in the next year) makes this likely, as it seems to indicate that these firms do not do R&D every year.

18. For some firms (especially those having obtained subsidies for a multi-annual project), the amount of support received may exceed the total cost of labour. In this case, the labour cost is cancelled out and the excess support is carried over to bring down the cost in the following year.

19. The model should also include a term to measure the cost of highly qualified labour relative to the capital. The cost of capital is difficult to evaluate: it can be approached by using different interest rates, which vary depending on the firms' level of debt. However, the firms covered differ little in terms of their debt level and the estimated values for the cost of capital are too homogeneous to be used for the estimation.

Unobserved heterogeneity of firms is taken into account by introducing a fixed effect (μ_i) specific to each firm: the autoregressive model then enables us to take account of the heterogeneity of changes in employment (rather than the employment levels). Time-related effects (δ_t) were added for each year of observation. Moreover, different models have been estimated depending on the classification of the firms in one of the three major types of activity (industry, IT services and scientific and technical activities).

Model estimation

To estimate the model, we take account of the endogeneity of the relative cost of highly qualified labour. At least two arguments back up this assumption. Firstly, bodies in charge of direct support allocate their grants based on the dynamism of the firms or the innovative dimension of their activities. These two unobserved characteristics explain the growth in employment within firms but, as they dictate the granting of support, they are also correlated to the reduction in the relative cost of labour. Secondly, during the period 2003-2007, the RTC was still partially calculated based on the increase in R&D expenditure, which implies endogenous changes in the costs of labour.

In order to correct the endogeneity of the relative cost of labour (see the test in the online complement C4, Table C4-2), an instrumental variable is developed on the basis of the different

Box 3

COMPUTATION OF THE INSTRUMENT USED TO CORRECT LABOUR COST ENDOGENEITY

The change in the relative cost of highly qualified labour after R&D support (c_{it}) can be decomposed based on the change in the rate of support (τ) and the change in the relative labour cost before deducing support (c_{it}*). Based on the assumption that the change in the rate of aid has no bearing (in the short term) on the share of highly qualified employment (d_i) devoted to R&D, we obtain the following decomposition:

$$\Delta c_{it} \approx d_i \,.\, \Delta \log\left(1 - \tau_{it}^{aid}\right) + \Delta c_{it}^{*}.$$

This decomposition enables us to find an instrument that is correlated to the change in the relative labour cost by replacing each term with an exogenous variable:

- For the term $\Delta \log\left(1 - \tau_{it}^{aid}\right)$, we use the different reforms of the RTC in 2004, 2006 and 2008, which correspond to exogenous variations in the support rate (16 rate variations are used over the period 2004-2010, see Table C4-1 of the online complement C4).

- To determine R&D employment as a share of highly qualified employment (d_i), the estimations are based on the characteristics of firms prior to the different reforms of the RTC in order not to take account of any modifications (increase in the share of R&D) due to the increase in the support rates.
- The variable Δc_{it}^{*} is simply replaced by the lagged variable Δc_{it-1}^{*}.

The instrument thus obtained is well and positively correlated to the change in the cost of labour and the regression of the change in labour cost on the exogenous variables and the instrument provides a positive coefficient, which is highly significant for the instrument.

exogenous variations of the RTC scheme stemming from the 2004, 2006 and 2008 reforms (Box 3).

The parameters of the fixed effect autoregressive model can be estimated by focusing on the change in the endogenous variable between two dates and using the lagged explanatory variables as an instrument: the aim is to control the endogeneity linked to the autoregressive term (Arellano & Bond, 1991). However, the coefficient ρ is particularly high and the employment change over the course of a year is barely correlated with the change during the following year. The instruments commonly used are thus weak, undermining the quality of the estimation. We prefer to use the solution proposed by Blundell and Bond (1998): based on an assumption of stationarity of the initial population of firms, they estimate a labour demand model using the lagged change in employment as an instrument of the lagged employment level[20].

Finally, to estimate the model, firms are also assumed to be faced with constrained demand: they cannot decide directly the level of their turnover, regarded as exogenous in the short term. This assumption is partially justified by the modest size of the firms under consideration.

The estimation is made using the generalised method of moments: the orthogonality assumptions apply to the residual terms $\mu_i+\varepsilon_{it}$ (for the lagged employment change, the turnover change and the instrumental variable) and to the residual changes ε_{it} - ε_{it-1} (for the same variables, except the lagged employment change).

---

20. On average, over the years and across the sectors, the correlation between the change in employment over a year and the change over the following year is slightly negative (-6%). The correlation between the lagged employment change and the employment level is higher and positive (+27%).

Box 4

METHOD FOR ESTIMATING THE EFFECT OF R&D PUBLIC SUPPORT ON EMPLOYMENT

The effect of R&D public support on highly qualified labour is calculated in two stages.

Stage 1: Calculation of the change in employment levels for fixed labour cost and turnover

The labour demand model (1) enables us to estimate, using the recurrence method, the logarithm of employment levels using an initial employment value (on date t_0), and of the change in the turnover and the cost of labour:

$$\begin{aligned}\hat{l}_{it} &= \hat{\rho}\hat{l}_{it-1} + \hat{\alpha}_1\, y_{it} + \hat{\alpha}_2\, y_{it-1} + \hat{\beta}\, c_{it} + \hat{\mu}_i + \hat{\delta}_t \\ &= f\left(l_{it_0}, y_{it_0}, \cdots, y_{it}, c_{it_0+1}, \cdots, c_{it}, \hat{\mu}_i, \hat{\delta}\right)\end{aligned} \qquad (2)$$

In formula (2), it is possible to separate what depends on the initial value of employment and the fixed effect from what depends on the change in the cost of labour and the turnover:

$$\begin{aligned}\hat{l}_{it} = {} & g_t\left(l_{it_0}, \hat{\mu}_i, \hat{\delta}\right) + \\ & h_t\left(y_{it_0}, \cdots, y_{it}, c_{it_0+1}, \cdots, c_{it}, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}\right)\end{aligned}$$

For a firm i receiving support at time T (this date is not identical for all firms), being granted support will modify both the cost of labour and the turnover. We can estimate the change in the employment level (between T-1 and t) due to this change in the cost of labour and the turnover by setting these two variables at their value at time T-1. This change in the employment level is proportional to:

$$\begin{aligned}\Delta_{it} = {} & exp\left(h_t\left(y_{iT-1}, \cdots, y_{it}, c_{iT}, \cdots, c_{it}\right)\right) \\ & - exp\left(h_t\left(y_{iT-1}, \cdots, y_{iT-1}, c_{iT-1}, \cdots, c_{iT-1}\right)\right)\end{aligned}$$

Stage 2: Comparison with firms that have not yet received aid

The previous calculation assumes that, in the absence of the receipt of aid, the turnover would not have changed, which appears to be a particularly strong assumption. Working on the basis of differences between treated and non-treated firms (and therefore on the basis of the difference-in-difference method) enables us to revert to a more realistic scenario by comparing the change measured for a firm i that has received aid with the mean change for firms that did not receive aid during the period between T to t and which have a propensity score close to that of firm i (this mean change is denoted by the exponent C). The effect of the aid on firm i is finally estimated as follows:

$$ATT(i,t) = exp\left(\frac{\hat{\sigma}_t^2}{2}\right) . exp\left(g_t\left(l_{it_0}, \hat{\mu}_i, \hat{\delta}\right)\right) . \left(\Delta_{it} - \Delta_{it}^C\right)$$

It should be noted that the bias induced by the transition to the exponential is controlled by a correction using an estimation of the standard deviation σ_t predicted by the labour demand model at time t.

Calculation of the effect of public funding on highly qualified labour

The effect of public funding on highly qualified labour is estimated using the labour demand model and calculating differences-in-differences. For each firm receiving support, we estimate the change in employment due to changes in the cost and the turnover by setting these two variables at the level achieved in the year preceding the year in which the first aid is received[21] (Box 4). The labour demand model allows us to decompose this change into two terms: one depending on the fixed effect, time effects and the initial employment level, the other depending on the cost of labour and the turnover. The group of non-treated firms is thus used to control only the change in the second term[22].

Results

The estimation of the model of demand for highly qualified labour by small firms having received support at least once and present in 2003 (Table 3) shows that demand in a given year depends greatly on demand in the previous year (the employment coefficients at *t*-1 are fairly high at around 0.8) and that firms increase their workforce if their turnover increases and the cost of labour falls. Differences across sectors are fairly limited, except for the labour cost effect: its coefficient is not significantly different from 0 in the IT services sector, whereas it is negative and significant (between - 0.16 and - 0.18) in the two other sectors.

The effect of the supplementary support received by the firms in the panel relative to the reference year is presented below.

The supplementary aid for R&D relative to 2003 received by the small firms present in 2003 follows a clear upward trend, increasing – in constant 2000 Euro – from 1 million in 2004 (last column in Table 4) to 106 million in 2010. According to our estimations, this supplementary aid for R&D led to an increase of 1,160 FTE highly qualified jobs in 2010 (Table 4, first column). The effect of the supplementary aid on highly qualified employment increases each year: after being close to 0 in 2004 and 2005, it increases from 2006 onwards. For each firm receiving aid, the number of jobs likely to have been funded by the supplementary support can be determined by dividing the amount of this aid by the average cost of an R&D job. Lastly, the effect on the number of highly qualified jobs not financed by aid (or financed by the firms themselves, Table 4, third column) equates to the difference between the effect on highly qualified employment and the number of highly

21. For firms already receiving support in the reference year (2003 or 2007), we use the levels from the reference year, which amounts to estimating the effect of the supplementary aid obtained since this date.

22. Firms that have not yet received any support (at a given date) are divided into 10 groups based on their propensity score. The mean results obtained for each of these groups serve as a reference for changes among firms receiving support.

Table 3

Model of demand for highly qualified labour for small firms having received aid at least once between 2003 and 2010 (equation (1))

Variables	Estimated coefficients		
	Sectors		
	Industry	IT services	Scientific and technical activities
Highly qualified employment at *t*-1 (log.)	0.76***	0.86***	0.77***
Turnover at *t* (log.)	0.08***	0.1***	0.07***
Turnover at *t*-1(log.)	- 0.02*	- 0.05***	0
Relative mean cost of highly-qualified employment at *t* minus aid (log.)	- 0.16***	- 0.04	- 0.18**

Note: model estimated using the generalised method of moments (GMM) and an instrument for labour cost. Employment is measured in full-time equivalent. The coefficients differ significantly from zero for level tests at 10% (*), 5% (**) and 1% (***). Confidence intervals are obtained by *bootstrap*.

Reading note: for the industry sectors, the level of highly-qualified employment at time t is explained by the level of highly-qualified employment at time *t*-1 (estimated coefficient of 0.76), the turnover in *t* and *t*-1 and the relative mean cost of highly-qualified labour in relation to the costs of other types of labour. The specification also include controls for each year of observation (estimated coefficients not presented).

Coverage: Small firms on the panel that had highly qualified staff in 2003 and which received aid at least once over the period 2003-2010 (2,261 firms, unbalanced panel).

Source: MENESR, GECIR database and *R&D Survey*; Acoss, JEI database, Lifi, Ficus/Ésane, *DADS*, authors' calculations.

Table 4
Estimated effect of the supplementary R&D aid obtained by small firms in reference to 2003 on total highly qualified employment and comparison with the supplementary aid received

	Effect on highly qualified employment	Effect on R&D employment	Effect on highly qualified employment not financed by aid	Effect on R&D employment not financed by aid	Effect on R&D employment not financed by aid excluding 'RTC operating expenditure'	Supplementary aid in reference to 2003 (in millions of constant 2000 Euro)
2004	- 20	- 20*	10	10	- 90*	1
2005	40	10	- 270**	- 290***	- 290***	18
2006	140**	100*	- 180*	- 220***	- 190***	22
2007	340***	240**	- 120	- 220**	- 150**	32
2008	530***	370***	- 700***	- 860***	- 520***	86
2009	810***	570***	- 400**	- 640***	- 220*	93
2010	1 160***	810***	- 140	- 480***	- 30	106

Note: effects in full-time equivalent employment (FTE); supplementary aid in millions of constant 2000 Euro. The results differ significantly from zero for level tests at 10% (*), 5% (**) and 1% (***). These tests are obtained through bootstrap.
Reading note: in relation to 2003, the supplementary aid received in 2005 amounted to 18 million Euro (column 6). The effect of this supplementary aid on highly qualified employment is estimated at +40 FTE posts in 2005 (column 1) and +10 FTE posts for R&D employment (column 2). The effect on employment not financed by aid is an estimated fall of 270 FTE posts for highly qualified employment (column 3), 290 FTE posts for R&D posts (column 4) and 290 FTE posts if the 'RTC operating expenditure' is not counted as aid (column 5).
Coverage: small firms on the panel that had highly qualified staff in 2003 and which received aid at least once over the period 2003-2010 (2,261 firms, unbalanced panel).
Source: MENESR, GECIR database and *R&D Survey*; Acoss, JEI database; Insee, Lifi, Ficus/Ésane, *DADS*, authors' calculations.

qualified jobs likely to have been financed by the supplementary aid: it is significant and negative, except in 2004, 2007 and 2010. In 2010, the estimated effect on highly qualified employment not financed through aid improves, but remains negative; the estimation also becomes more imprecise and, in the end, not significant.

Only some of the highly qualified jobs are really R&D jobs: to estimate how many (second column), we once again make an assumption that, for each sector and category of company, the ratio of the number of R&D jobs to the number of highly qualified jobs is equal to that estimated for the period 2008 to 2010. Based on this assumption, the deficit in terms of R&D jobs not financed through support (fourth column) is 220 FTE jobs in 2006 (compared to 180 FTE highly qualified jobs). As with highly qualified employment, we observe clearly more negative effects as of the 2008 reform of the RTC; moreover, they are significant and negative, except in 2004.

At last, adding together the estimations obtained in the various years for the firms present in 2003, only 63 %[23] of the supplementary aid paid out between 2004 and 2010 would have served to finance new highly qualified jobs (44 % if we only take into consideration R&D jobs).

This evaluation is based on an increase in employment aid, due to the fact that the share of the RTC linked to operating costs is taken into account[24]. If we assume that this part of the RTC is not taken into account, the number of R&D jobs financed by the firms themselves falls by 30 FTE jobs only (a change not significantly different from zero) in 2010 (fifth column in Table 4) instead of 480 FTE jobs (fourth column) and 58% of the supplementary support paid out between 2004 and 2010 would have served to finance new R&D jobs.

The results above were established for firms present in 2003. In order to measure the effect of the aid on more firms, we now focus on firms present in 2007. This new estimation enables us to better take account of the reform of the RTC in 2008. The total aid received by these firms was (in constant 2000 Euro) 344 million in 2010, whereas it was only 135 million for the earlier group (the supplementary support received in 2010 increases from 106 million compared to the reference year 2003 (last column Table 4) to 171 million compared to the year 2007 (last column in Table 5). For this expanded group of firms, the effect on the employment level of highly qualified staff of the supplementary support is at its maximum in 2010 with 830 FTE jobs (column 1); in parallel, the supplementary aid received that year equates

23. This result is based on a 95% confidence interval [42% ; 84%].

24. They are set at 75% of R&D staff expenditure (see above*). Not taking it into account would reduce the amounts of the RTC used to finance employment by 43%.*

to financing 2,140 FTE jobs, which equates to a fall of 1,310 FTE highly-qualified jobs not financed by aid (column 3). For R&D-related employment, there is a bigger fall of 1,520 FTE jobs financed by the firms themselves. These falls are statistically significant for all years. This is also the case if we do not take account of the part of the RTC linked to operating expenditure (column 5). Lastly, among firms present in 2007, only 24% (between 18 and 34% based on the 95% confidence interval) of the supplementary aid paid out between 2008 and 2010 compared to 2007 served to finance new highly qualified jobs (19% if we only consider R&D-related jobs and 29% if we do not take account of the RTC linked to operating expenditure).

The detailed results by sector of activity show that the fall in the employment of highly qualified staff financed by the firms themselves applies to all sectors, but was greater in the industrial sector and more moderate in the scientific and technical sectors (see online complement C5, Tables C5-1 and C5-2). For the firms present in 2007, only 9 % of the supplementary support received by firms from the industrial sector served to finance new highly-qualified jobs compared to 15 % for IT service firms and 43 %[25] for firms in the scientific and technical activities sector.

Discussion of the results and their robustness

In this section, we discuss the assumptions used to construct the control group and estimate the model and their influence on the results of the evaluation.

For the two populations of firms studied (present in 2003 or in 2007), we observe an increase in the effect of support on the employment of highly qualified staff at the end of the period. This effect is, in part, linked to the lower quality of the control group in 2010[26], which may have led to an overestimation of the effect of the aid in the latter years.

The coverage of Tables 4 and 5 only includes firms belonging to the 75 most R&D-intensive sectors and which had one or more highly qualified employee in 2003 or 2007. These two restrictions enable better quality counterfactuals to be found. However, these restrictions are more technical than actually reflecting economic reality and it seems reasonable to expand the estimations to include a broader coverage comprising all VSBs and other small SMEs.

The effect of R&D support is estimated by comparing firms receiving support with firms not receiving aid without taking account of a possible effect of the aid on firms' survival. If R&D support enabled the firms to prolong their activity (or to retain their highly-qualified employees), the effect of the aid would have been underestimated. However, a comparison of the firms from the panel receiving support prior to 2004 and the firms having similar propensity scores that have never received aid indicates that the firms receiving aid are observed for a

25. The 95% confidence intervals are respectively for each sector [6% ; 21%], [6% ; 24%], and [31% ; 61%].

26. It only includes firms that have never received support, which constitute a lower quality counterfactual than firms that have not yet received aid (online complement C4).

Table 5
Estimation of the effect of the supplementary aid obtained by small firms in reference to 2003, on total highly qualified employment and comparison with the supplementary aid received

	Effect on highly qualified employment	Effect on R&D employment	Effect on highly qualified employment not financed by aid	Effect on R&D employment not financed by aid	Effect on R&D employment not financed by aid excluding 'RTC operating expenditure'	Supplementary aid in relation to 2007 (in millions of constant 2000 Euro)
2008	210**	160**	- 1,710***	- 1,760***	- 1,150***	131
2009	440***	360***	- 1,660***	- 1,740***	- 980***	151
2010	830***	620***	- 1,310***	- 1,520***	- 720***	171

Note: effects in full-time equivalent employment (FTE); supplementary aid in millions of constant 2000 Euro. The results differ significantly from zero for level tests at 10% (*), 5% (**) and 1% (***). These tests are calculated using a bootstrap method
Reading note: in relation to 2007, the supplementary aid received in 2008 amounted to 131 million Euro (column 6). The effect of this supplementary aid on highly qualified employment is estimated at +210 FTE posts in 2008 (column 1) and +160 FTE posts for R&D employment (column 2). The effect on employment not financed by aid is an estimated fall of 1,710 FTE posts for highly qualified employment (column 3), 1,760 FTE posts for R&D employment (column 4) and 1,150 FTE posts if the 'RTC operating expenditure' is not counted as aid (column 5).
Coverage: small firms on the panel that had highly qualified employees in 2007 and which received aid at least once over the period 2007-2010 (4,117 firms, unbalanced panel).
Source: MENESR, GECIR database and *R&D Survey*; Acoss, JEI database, Insee, Lifi, Ficus/Ésane, authors' calculations.

slightly longer period, but that the 'survival' gap between the two groups is not statistically significant (online complement C5, Table C5-4).

To define the control group, we do not have exhaustive information about direct support: the control group may contain some firms that have received subsidies. This problem can lead to underestimating the effect of aid, as the change in the employment levels of the firms not receiving aid has perhaps been overestimated. This bias should, however, decrease as use of the RTC increases, since the subsidised firms are increasingly registered in the Gecir database.

The firms in the control group that have never received aid are selected based on observable characteristics that determine their propensity score. However, unobserved variables could have influenced both the participation of firms in aid mechanisms and their demand for highly qualified labour: not taking them into account could have biased our estimations. However, among these unobserved variables, the cost of seeking aid is significantly higher for small firms than for large firms (Arqué-Castells et Mohnen, 2015) and knowledge of the support schemes probably depends on the age of the firms: these variables are thus partially controlled with the propensity score, which takes account of various characteristics relating to firms' size and age.

In the model of demand for highly qualified labour, support is equated to an immediate reduction in the cost of labour. While the subsidies and reductions in employer payroll taxes provided under the JEI scheme are actually received by the firms as soon as the R&D expenditure is made, this is not the case with the RTC, especially for firms that do not declare any corporation tax: these firms hold a receivable, which sometimes only has to be reimbursed in full by the State after 3 years. However, since 2007, derogations have allowed small and young enterprises[27] to secure the reimbursement of the RTC receivables as early as the following year, which enhances the credibility of the assumption used for the labour demand model.

27. To be more precise, derogations were granted to dynamic JEIs and SMEs in 2007 and 2008, to all companies in 2009, and only to SMEs from 2010 onwards.

The results presented may be partially biased, as MENESR-accredited firms may belong to the control group. These may be firms classified in the R&D-intensive sectors, which, though still not having received any aid[28], benefit indirectly from the RTC received by their clients. However, if we exclude all the accredited firms from the coverage of the study, we obtain results that are very close to those presented in Tables 4 and 5 (online complement C5, Tables C5-5 and C5-6).

* *
*

The econometric analyses conducted (using the difference-in-difference method and a labour demand model combined with a matching model) confirm the crowding-out effect suggested by the aggregated analysis: they show that the effect of R&D public support on expenditure on highly qualified staff and highly qualified R&D staff has been positive, but well below the increase in the aid received, especially from 2008 onwards. It would appear that the very strong increase in the rates of R&D support during the 2000s did not lead to effectively increasing the employment of R&D staff in small firms. This result differs starkly from those obtained up until present in most of the empirical research based on French data. The result is obtained from a coverage including the VSBs, which are usually disregarded in research conducted on the basis of the data from the R&D Survey.

However, it should be noted that our results are interpretable essentially over the short term and not as an indication of long-term effects of R&D public funding. Lastly, this study does not take account of recent changes in the rules used to calculate the RTC. Thus, the reduction in 2010, followed by the abolition of the increased rates of 50% and 40% in 2013[29], which significantly reduced the support rate for small firms, may have mitigated the crowding-out effect highlighted in this study. ☐

28. A very large majority of the accredited companies within the coverage of the study have, however, received aid at least once (75% of the VSBs and 87% of the other SMEs that were accredited in 2010 received aid between 2005 and 2010).

29. It should also be noted that the operating expenditure base was lowered in 2011 (from 75% to 50% for staff expenditure).

BIBLIOGRAPHY

Acemoglu, D., Akcigit, U., Bloom, N. & Kerr, W. R. (2013). Innovation, Reallocation and Growth. NBER *Working Paper* 18993.

Aghion, P., Askenazy, P., Berman, N., Cette, G. & L. Eymard (2012). Credit Constraints And The Cyclicality Of R&D Investment: Evidence From France, *Journal of the European Economic Association*, 10(5), 1001−1024.

Akcigit, U. & Kerr, R. W. (2010). Growth Through Heterogeneous Innovations. NBER *Working Paper* 16443.

Arellano, M. & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equation. *Review of Economic Studies*, 58(2), 277−297.

Arqué-Castells, P. & P. Mohnen (2015). Sunk costs, extensive R&D subsidies and permanent inducement effects. *Journal of Industrial Economics*, 63(3), 458−494.

Baghana, R. & Mohnen, P. (2009). Effectiveness of R&D tax incentives in small and large enterprises in Québec. *Small Business Economics*, 33(1), 91−107.

Béguin, J.-M., Hecquet, V. & Lemasson, J. (2012). Un tissu productif plus concentré qu'il ne semblait. *Insee Première* N° 1399.

Bellégo, C. & Dortet-Bernadet, V. (2014). L'impact de la participation aux pôles de compétitivité sur les PME et les ETI. *Économie et Statistique*, 471, 65−83.

Bloom, N., Schankerman, M. & Van Reenen, J. (2013). Identifying technology spillovers and market rivalry. *Econometrica*, 81(4), 1347−1393**.**

Blundell, R. & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1), 115−143.

Bozio, A., Irac, D. & Py, L. (2014). Impact of research tax credit on R&D and innovation: evidence from the 2008 French reform. Banque de France, *Document de travail,* 532.

Bresson, G., Kramarz, F. & Sevestre, P. (1992). Dynamic labour demand Models. In: L. Matyas and P. Sevestre (Eds), *The Econometrics of Panel Data: Theory and Applications,* pp. 360−387. Kluwer Academic Publishers.

Bronzini, R. & Iachini, E. (2014). Are incentives for R&D effective? Evidence from a regression discontinuity approach. *American Economic Journal : Economic Policy*, 6(4), 100–134.

Busom I., Corchuelo B. & Martínez-Ros E. (2014). Tax incentives… or subsidies for business R&D? *Small Business Economics*, 43(3), 571−596.

Cinceraa, M. & Veugelers, R. (2013). Young leading innovators and the EU's R&D intensity gap. *Economics of Innovation and New Technology*, 22(2), 177−198.

Corchuelo, B. & Martinez-Ros, E. (2009). The Effects of Fiscal Incentives for R&D in Spain. *Business Economics Working Papers* wb092302.

Cottet, V., Quantin, S. & Régnier, V. (2012). Coût du travail et allègements de charges : une estimation au niveau établissement de 1996 à 2008. Insee, *Document de travail* G2012/12.

David, P. A., Hall, B. H. & Toole, A. A. (2000). Is Public R&D a complement or substitute for private R&D? A review of the econometric evidence. *Research Policy*, 29(4-5), 497−529.

Dortet-Bernadet, V. & Sicsic, M. (2015). Effet des aides publiques sur l'emploi en R&D dans les petites entreprises. Insee, *Document de travail* G2015/11.

Dufau, M. (2017). Les pôles de compétitivité, qui rassemblent les acteurs de l'innovation dans les territoires, comportent une proportion importante de jeunes entreprises. DGE, *Le 4 pages de la DGE* N° 67.

Duguet, E. (2012). The effect of the incremental R&D tax credit on the private funding of R&D an econometric evaluation on french firm level data. *Revue d'économie politique*, 122(3), 405−435.

Duguet, E. (2004). Are R&D subsidies a substitute or a complement to privately funded R&D? *Revue d'économie politique*, 114(2), 245−274.

Garcia-Macia, D., Hsieh, C.-T. & Klenow, P. J. (2016). How Destructive is Innovation? *NBER Working Paper* 22953.

González, X., Jaumandreu, J. & Pazó, X. (2005). Barriers to innovation and subsidy effectiveness. *Rand Journal of Economics*, 36(4), 930−950.

Griffith, R., Redding, S. & Van Reenen, J. (2003). R&D and Absorptive Capacity: Theory and Empirical Evidence. *The Scandinavian Journal of Economics*, 105(1), 99–118.

—— (2004). Mapping the Two Faces of R&D: Productivity Growth in a Panel of OECD Industries. *The Review of Economics and Statistics*, 86(4), 883–895.

Hægeland, T. & Møen, J. (2007). Input additionality in the Norwegian R&D tax credit scheme. *Statistics Norway*, report 47.

Ientile, D. & Mairesse, J. (2009). A policy to boost R&D: Does the R&D tax credit work? *European Investment Bank Papers* 6/2009.

Jones, I. C. & Williams, C. J. (1998). Measuring the social return to R&D. *The Quarterly Journal of Economics*, 113(4), 1119–1135.

Köhler, C., Larédo, P. & Rammer, C. (2012). The impact and effectiveness of fiscal Incentives for R&D - compendium of evidence of the effectiveness of innovation policy intervention. *Manchester Institute of Innovation Reseach and NESTA*.

Lelarge, C. (2009). Soutenir l'effort de R&D des structures entrepreneuriales. Analyse de deux dispositifs ciblés récents. Thèse*, Université Paris X Nanterre*.

Lhuillery, S., Marino, M. & Parrotta, P. (2013). Évaluation de l'impact des aides directes et indirectes à la R&D en France. *Rapport pour le MENESR*.

Lokshin, B. & Mohnen, P. (2012). How effective are level-based R&D tax credits? Evidence from the Netherlands. *Applied Economics*, 44(12), 1527–1538.

Lokshin, B. & Mohnen, P. (2013). Do R&D tax incentives lead to higher wages for R&D workers? Evidence from The Netherlands. *Research Policy*, 42(3), 823–830.

MENESR (2014). Développement et impact du crédit d'impôt recherche : 1983-2011. *Rapport du MENESR*.

Mulkay, B. & Mairesse, J. (2013) The R&D Tax Credit in France: Assessment and Ex-Ante Evaluation of the 2008 Reform. *Oxford Economic Papers*, vol 65(3), 746–766

OCDE (2016). Incitations fiscales à la recherche-développement. In : OCDE (Ed.), *Science, technologie et industrie : Tableau de bord de l'OCDE 2015. « L'innovation au service de la croissance et de la société »*. pp. 172–187. Paris: OCDE.

OCDE (2017). Statistiques de la recherche et développement.
http://www.oecd.org/fr/sti/inno/srd.htm

Schneider, C. & Veugelers, R. (2010). On young highly innovative companies: why they matter and how (not) to policy support them. *Industrial and Corporate Change*, 19(4), 969–1007.

Serrano-Velarde, N. (2008). How Do Firms React to R&D Subsidies? Evidence from regression discontinuity design. mimeo.
https://www.insee.fr/fr/statistiques/2017510?sommaire=2017528.

Zuñiga-Vicente, J. A., Alonso-Borrego, C., Forcadell, F. J. & Galán, J. I. (2014). Assessing the effect of public subsidies on firm R&D investment: A survey. *Journal of Economic Surveys*, 28, 36–67.

Do public subsidies have an impact on start-ups survival rates? An assessment for four cohorts of firms set up by previously unemployed entrepreneurs in France

Dominique Redor *

Abstract – Business start-up assistance has been adopted as a tool for implementing proactive employment policies across most OECD nations. In France, the ACCRE start-up support programme for unemployed people creating or taking over firms has expanded strongly since its introduction in 1979. The number of people joining the ACCRE programme exceeded 80,000 in 2006 and peaked at 220,000 in 2010. We have studied the effect of the ACCRE system on the survival (measured after five years) of four cohorts of firms started by unemployed entrepreneurs in 1994, 1998, 2002 and 2006, based on survey data in INSEE's "new firms information system", SINE. According to descriptive statistics, the survival outlook for firms created by ACCRE beneficiaries is better than that of firms created by non-recipients. However, using simultaneous equations to model ACCRE approval and firm survival revealed evidence of ACCRE recipient selection based on the administrative approval process, as well as self-selection by entrepreneurs. Adjusted accordingly, ACCRE appears to have no effect on the survival of supported firms for most categories of unemployed people.

JEL codes: C26, D21, H25, J68, L38
Keywords: start-up, firm demographics, employment policy, support for the unemployed, public policy assessment

** Université de Paris-Est Marne-la-Vallée and CEET, Cnam (dominique.redor@cee-recherche.fr, domredor@free.fr).*

I would like to thank Richard Duhautois for his contribution to our joint research and discussions relating to business start-ups. I am also grateful to my colleagues at CEE, with whom we worked on processing Insee SINE surveys: Lionel Désiage (who died in May 2012), Géraldine Rieucau, Guillaumette de Larquier and Bernard Gomel. An earlier version of this document was presented at an internal CEE conference on 16/12/2014. The current version is informed by invaluable comments, criticism and suggestions from F. Guillaumat-Taillet (Dares), to whom I am deeply thankful. Thanks also to two anonymous referees, for their constructive comments and suggestions.

Received November 15, 2015; accepted after revision March 21, 2017 DOI: 10.24187/ecostat.2017.493s.1910
Translated from the original version: « L'aide à la création d'entreprises a-t-elle un impact sur leur survie? Une évaluation pour quatre cohortes d'entreprises créées par des chômeurs en France »

In recent decades, economists and public employment policy authorities have increasingly focused on business start-ups and closures. In developed economies, firm turnover (entry and exit) rates tend to be high. This rapid turnover of firms is generally associated with the process of "creative destruction" as developed by the economist Schumpeter in the early 1940's. According to this theory, creative destruction is a continuous process in contemporary economies, resulting in the simultaneous creation of innovative new activities and loss of obsolete activities. In this framework, start-ups are essential agents of the creative destruction process acknowledged as a key economic growth driver (Aghion et al., 2014). A large body of empirical research on this theme has been documented since the 1980-1990's, revealing the considerable impact of new firms – including both start-ups (less than three years old) and young firms (less than five years old) more generally – on employment dynamics in developed European countries and in the United States (cf. Audretsch & Mahmood, 1994; Davis et al. 2007; Haltiwanger, 2011; Haltiwanger et al. 2013; Mata & Portugal, 1994). Drawing on a database produced for the purpose by the OECD (http://www.oecd.org/fr/sti/dynemp.htm), Criscuolo et al. 2014 demonstrate that, in small and medium-size enterprises (up to 250 employees), young firms (created in the past five years) "make a disproportionately high contribution to job creation" in the 18 studied countries[1], corroborating recent research using American data (Haltiwanger et al., 2013). Furthermore, most of these jobs are created by the entry of new firms, and to a lesser extent by growth within start-ups.

These research findings have significant implications for employment policy. Public support to start-ups has gradually been incorporated into proactive labour market policies across the European Union[2] and in most OECD countries, generally aimed at enabling the unemployed to re-enter the labour market. Start-up support policies must empower unemployed individuals, who often have few qualifications and are in some cases subject to discrimination, to set up a firm and hence create their own employment, by helping them to overcome the initial hurdles associated with entrepreneurship. If this aim is met, such policies should improve entrepreneurs' employability and human capital. Other arguments also plead in favour of such policies: they can yield a "double dividend" when these new firms go on to create additional jobs. Lastly, they can positively impact economic growth, by contributing to innovation and the spill over of new technologies. Nevertheless, such start-up support policies are subject to serious criticisms. Firstly, they may have deadweight effects in cases where entrepreneurs would have started their firm with or without a subsidy. The survival and success of the firm are unrelated to the subsidy in such cases. They can create crowding out effects, by distorting competition between existing firms and subsidised start-ups. They may also lead to inverse selection effects in the context of high economic uncertainty, for example by facilitating entrepreneurship among individuals who lack the ability to manage their firm over the short-to-medium term. Conversely, by lowering barriers to entry, such policies may reveal to individuals that they do indeed possess the necessary capabilities to run their own firm. Lastly, such subsidies may lead to moral hazard. Subsidised entrepreneurs may be tempted to invest less effort, as they would not bear the costs and/or lost income associated with the failure of their firm.

Caliendo (2016) highlights that most studies relating to the impact of publicly-funded start-up support programmes aimed at the unemployed population in OECD countries are descriptive in nature, with only a small body of research devoted to medium- to long-term assessment. Caliendo notes that studies tend not to be convergent, due to the heterogeneous nature of the institutional provisions in different countries, and the wide range of statistical and econometric methods used[3]. Regarding Germany, Caliendo and Künn (2011) compared two start-up support programmes (concerning unemployment benefit and subsidies) over the period 2003-2008. Using propensity score matching methods, the authors show that the two programmes had significant positive impacts on participant employment and revenue after five years, particularly for the previously long-term unemployed. The two programmes were merged into one in August 2006; a recent assessment (Caliendio et al.,

1. Austria, Belgium, Canada, Costa Rica, Spain, USA, Finland, France, Hungary, Italy, Japan, Luxembourg, Norway, New Zealand, Netherlands, Portugal, United Kingdom and Sweden, and Turkey for the period 2001-2011.

2. See expenditure on each type of initiative, and in particular incentives for starting new firms, on the Eurostat website, in the table [lmp_expsumm].

3. Regarding assessments of the various national programmes, refer to Caliendo & Künn (2011), Caliendo et al. (2015) and Pfeiffer & Reize (2000) for Germany, Deidda et al. (2015) for Italy, Gu et al. (2008) for the USA, and the summary by Caliendo (2016).

2015) indicates that, based on a propensity score matching method, the 19-month survival rate of subsidised firms was higher than other firms. However, they delivered weaker performance in terms of growth, innovation and income, due to negative inverse selection and moral hazard effects. In France, following the introduction in 1994 of Insee's information system on new firms, *SINE*, a survey which enables to analyse entrepreneurs' profiles and launch conditions, as well as the growth conditions of new firms, the implemented start-up public support programme was subjected to a variety of assessments. Most of these focussed on survival and economic performance. Crépon and Duguet (2003) studied the effect of public support (of all kinds) on business creation, considering a three-year period for a cohort of firms started in 1994. Using a selective matching method based on observable variables (propensity score matching), they revealed a significant positive impact of public support on the survival of firms created by the previously unemployed. Furthermore, access to bank loans greatly enhanced the probability of survival of these firms when combined with public support. Cabannes and Fougère (2012 and 2013) used data from the *SINE* survey to assess the effect of ACCRE (a support programme for unemployed people creating or taking over firms) on firm lifetimes, considering a five-year period for a cohort of firms established in 1998. They took into account endogeneity in ACCRE grant and estimated a model with random effects featuring two simultaneous equations, one relating to ACCRE grant (*logit*), the other formalising the life duration of the firms in the 1998 cohort. This approach revealed that the causal effect of ACCRE on the five-year survival rate of firms set up in 1998 by individuals who had previously been unemployed for a year or less was negligible (not significantly different to zero).

Désiage et al. (2010) and Duhautois et al. (2015) built a database merging the 1998 *SINE* survey with firm-related data from administrative files (*FICUS* unified accounting files) and examined the survival rate for firms during the first eight years of their lives. The population of entrepreneurs receiving ACCRE was larger than that studied by Cabannes & Fougère, as it included all ACCRE recipients, including short- and long-term unemployed individuals as well individuals who were not in the labour force before starting their business. Using a Rubin-like propensity score matching method, they found that ACCRE had a significant causal effect on the five- and eight-year survival rates of supported firms

Our study aims to assess the effect of ACCRE on the five-year survival rates of cohorts of firms created by recipients of this support in 1994, 1998, 2002 and 2006, respectively. These years correspond to the first four *SINE* surveys conducted by Insee. Our approach offers two advantages compared with previous studies. Firstly, as each survey covers all firms created during the first half of the reference year (Box 1), the characteristics of ACCRE recipients may be compared, using the same variables, against those of non-recipient entrepreneurs, which is not the case in many studies covering other countries, where such surveys are not carried out[4]. Secondly, in contrast to previous assessments conducted for France, the availability of four different cohorts enables us to assess any change over time in the causal effect of this support on firm survival rates, allowing to account for changing regulations. Lastly, our results, obtained using a methodology similar to that of Cabannes & Fougère (2012), corroborate and generalise theirs, highlighting the lack of a significant impact of ACCRE on three- and five-year firm survival rates.

Our article is organised as follows. The first section presents the regulations governing ACCRE in France, and the changes made to the programme since it was introduced in 1979; it also examines the numbers of recipient entrepreneurs for each cohort of new firms. The second section describes the variables included in the four cohorts of firms based on *SINE* databases, and identifies the survival indicators for the firms in those four cohorts, distinguishing whether or not they received assistance through ACCRE. The third section addresses the econometric assessment strategy for estimating the causal effect of ACCRE on the survival of recipient firms. The fourth section of the study estimates this effect for entrepreneurs who had been unemployed for less than a year when they started their firm. The next section focuses on assessing the effect for other categories of entrepreneur (i.e. not in the labour force or unemployed for more than a year). The final section tests the robustness of the estimates.

4. This point is underscored by Caliendo (2016) p. 9.

Changes to ACCRE regulations and the recipient population

ACCRE eligibility criteria

The ACCRE programme has seen multiple regulatory changes since it was first introduced (Table 1). These changes concern the eligible population, the nature of the support granted to entrepreneurs and the granting criteria. Initially, in 1979, it was an "over-the-counter" measure granted automatically to all job seekers receiving unemployment benefit (Mouriaux, 1995). With effect from 1987, the French labour authorities were empowered to reject projects that it considered to be non-viable. A departmental committee reporting to the Department of Work (*direction du travail*) was set up in the late 1980's to evaluate the authenticity and content of candidate projects.

The nature and extent of the available support have also varied in response to changing budget policy (figure). The five-year planning Act of December 1993, which was not effectively implemented until 5 April 1994, marked a break from the previous system. ACCRE eligibility was extended to all job seekers, regardless whether they were currently receiving unemployment benefit or not (although recipients were required to have been unemployed for six months). Most importantly, the fixed subsidy was made the same for all recipients, and was increased to FRF 32,000 with effect from the second quarter of 1994. The generosity and egalitarian nature of the system led to an immediate spike in the number of new ACCRE recipients, starting in Q2 1994 and continuing into 1995 and 1996 (see the figure in online supplement C1). However, the high budget cost of ACCRE prompted the government to strip the FRF 32,000 payment in its 1997 Finance Act (Daniel & Mandelblat, 2010). The new system – phased in beginning in 1998 – was only truly useful for unemployment benefit recipients who continued to receive part or all of their unemployment allowance for a period of up to 15 months (on condition that they were not remunerated by their new business). Furthermore, if their firm were to fail during that period, the entrepreneur would recover their unemployment benefit entitlement, calculated with effect from the date that they started their firm. Job seekers not receiving unemployment benefit qualified only for an exemption from national insurance contributions on any remuneration received during the first year, subject to a cap of 1.2 times the guaranteed minimum wage (SMIC). However, with effect from July 1998, new anti-discrimination legislation (*loi contre les exclusions*) extended eligibility for ACCRE support to new categories of entrepreneurs, without altering the nature or amount of the subsidy (Table 1).

Beginning in 2007, several root-and-branch regulatory reforms relating to business start-ups and related assistance were implemented. First, with effect from January 2007, ACCRE support has been awarded on the basis of purely administrative criteria relating to regulatory compliance (Daniel & Mandelblat, 2010; Ould Younes, 2010) and, beginning in September of the same year, the business registration centre (*Centre de formalités des entreprises*) took over application processing. The indicators assessing the economic viability of planned start-ups were discarded. The number of ACCRE recipients increased sharply as a result (Figure). Furthermore, the new "*autoentrepreneur*" self-employment regime came into effect in January 2009. The benefits of this regime include minimal paperwork when setting up a firm as well as special tax treatment. It appealed to large numbers of people with small-scale projects, with a knock-on effect on the characteristics of the entrepreneurs included in the 2010 SINE survey, compared with the cohorts initiated in 2002 and 2006 (Béziau & Bignon, 2017). As a result of these radical changes, we have chosen to end our analyses with the cohort initiated in 2006, i.e. before the viability criteria previously applicable to ACCRE grant applications were removed (in 2007), and before the *autoentrepreneur* self-employment regime was introduced (in 2009).

Entrepreneurs potentially concerned by ACCRE

For the purposes of this article we use the term "short-term unemployed" to refer to entrepreneurs who had been unemployed for less than a year prior to starting their firm; "long-term unemployed" to refer to those who had been unemployed for at least a year prior to starting their firm; and "out of the labour force" for those who reported that they had no job and were not seeking employment at the time (this category typically includes students and people receiving minimum welfare benefits who have reported that they were not seeking employment).

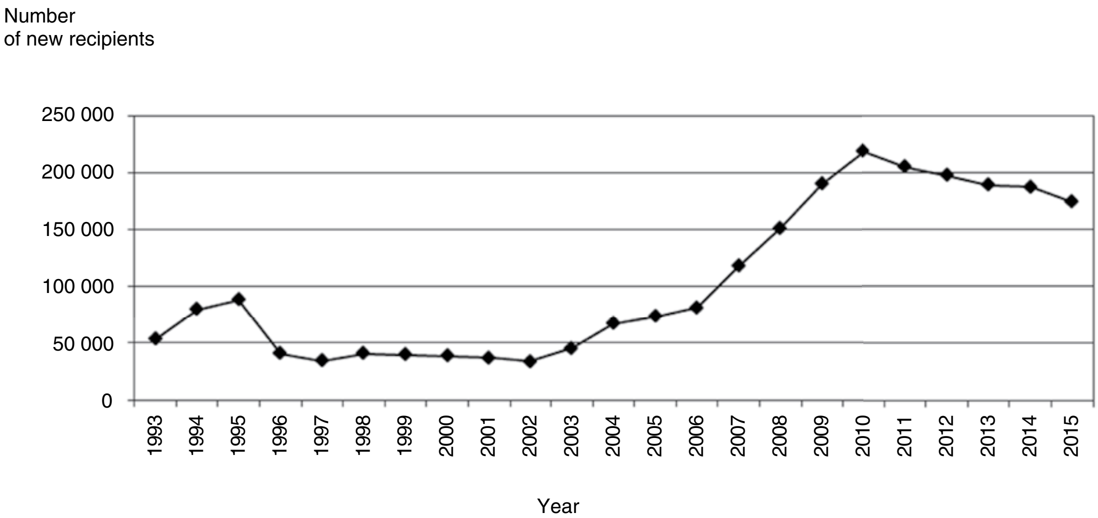
Table 1
The ACCRE programme: conditions of eligibility and support - Regulatory changes

Period	Eligible population	Nature of support	Award conditions and procedure
Prior to April 1994 (Act 80-1035 of December 1980; Decree 87-202 of 28 March 1987)	Unemployment benefit and minimum social welfare (RMI/ASS) recipients.	- Lump sum calculated based on the daily ASS social welfare payment for job seekers whose entitlement to unemployment benefits has expired. - Lump sum calculated based on residual unemployment benefit entitlement, capped at six months, for unemployment benefit recipients. - Six-month exemption from national insurance contributions for unemployment benefit recipients. ASS recipients: entitlement to ASS payments maintained for one year, in addition to any income generated by the firm, capped at 50% of the guaranteed minimum wage (SMIC). ASS payments are tapered if business income exceeds this level. RMI recipients: Only 50% of income generated by the firm is taken into consideration when calculating the resources on which RMI welfare payment amounts are based	Recipients must either start or take over a firm (regardless of activity sector or legal form) and effectively run that business. Application to be submitted to the departmental Department of Work (DDT) before the firm begins operating. A departmental committee evaluates the authenticity and content of candidate projects.
From April 1994 (entry into effect of the December 1993 five-year planning Act)	As above, plus job seekers registered as unemployed for more than six months but not receiving unemployment benefit.	- Same subsidy amount for all recipients: FRF 32,000. - Exemption from national insurance contributions for 1 year for unemployment benefit recipients and RMI and ASS minimum social welfare recipients; no exemption for job seekers not receiving unemployment benefit.	As above
First half of 1998 (1997 Finance Act).	As above, plus recipients of other social welfare payments for single parents (API) or the registered disabled.	- FRF 32,000 subsidy discontinued. - Exemption from national insurance contributions on the entrepreneur's pay for 1 year for unemployment benefit recipients and non-recipients, capped at 1.2 times guaranteed minimum wage (SMIC), and for RMI and ASS minimum social welfare recipients. - Unemployment benefit recipients and social welfare recipients continue to receive payments and allowances for 12 to 15 months if they are not remunerated by their new firm. In case of failure of their firm during that period, entrepreneurs recover their unemployment benefit entitlement, calculated with effect from the date that they started their firm. - All ACCRE recipients are issued with "consulting cheques" that can be used to pay for the services of approved experts.	As above
From H2 1998 to September 2007 (Planning Act of July 1998: anti-discrimination measures; Act 2003-721 of 1 August 2003 relating to business initiatives.	From July 1998: As above, plus holders of start-up support agreements, employees taking over their current employer in administration or liquidation proceedings, and under 26 year-olds eligible for the youth employment programme "*emplois jeunes*".	The ACCRE calculation method remained unchanged from H1 1998 to H1 2006. The Act of 1 August 2003 relating to business initiatives does not concern the ACCRE programme but made it easier to start a business by offering business owners additional guarantees (including protection against seizure of their home and tax relief).	As above
Decree 2007-1396 of September 2007	Same recipients as previously.	Nature of the assistance unchanged from previous provisions.	Radical changes to the ACCRE award procedure. With effect from January 2007, the award decision is based exclusively on administrative criteria, and since September 2007, applications are processed by business registration centres (CFE).

Source: Charpail (1995), Charpail (1996), Daniel & Mandelblat (2010), Guimiot & Mareau (2003), Mouriaux (1995), Ould & Younes (2010).

The share of ACCRE recipients among unemployed and non-working entrepreneurs (Table 2) appears to be strongly influenced by the changes in allocation rules analysed above (Table 1). The system introduced in 1994, at the time relatively generous to entrepreneurs, was subsequently greatly restricted, beginning in 1996-1997. The benefits of ACCRE support

Figure
Recipients of ACCRE start-up support



Note: Annual total of new recipients.
Coverage: Metropolitan France.
Source: DARES, *Poem* statistics database, available from http://poem.travail-emploi.gouv.fr.

were decreased, reducing the incentive to apply. However, for the cohorts of firms started in 2002 and 2006, the legislative changes that gradually expanded the population of eligible entrepreneurs resulted in recipients accounting for a significantly larger share of unemployed and out of the labour force.

The ACCRE programme accounts for the lion's share of the public support available to unemployed and out of the labour force entrepreneurs (Table 2). Apart from ACCRE, entrepreneurs may be eligible for local and regional subsidies, business tax exemptions and relief on national insurance contributions. All these forms of assistance are open to employed as well as unemployed and out of the labour force entrepreneurs. In certain cases, unemployed and out of the labour force entrepreneurs may be able to combine them with ACCRE. Ultimately, following a short break during the period extending from Q2 1994 to the end of 1996, when ACCRE payments were the same fixed amount for all categories of recipient, subsidies have been calculated based on the unemployment benefits or minimum social welfare payments made to unemployed or out of the labour force entrepreneurs. The revised legislation introduced great inequality between the financial benefits granted to the various groups of recipients.

Data and indicators used to assess survival rates among firms in the four cohorts

The data used in this study is drawn from Insee's "new information system on firms," *SINE*; it relates to cohorts of firms started in 1994, 1998, 2002 and 2006. Statistical methods and standardised concepts can be used to compare the four cohorts, subject to certain precautions, as described in Box 1.

Firm survival rates with and without ACCRE

Our chosen performance indicator is the survival rate of firms five years after creation. This rate is defined as the ratio of the number of firms created during the first half-year of the specified period and still trading five years later to the total number of firms created at the start of the specified period. Table 3 reveals that the firm survival rate for entrepreneurs who were working immediately prior to setting up their firm (whether as employees, traders, tradesmen, business owners or in the liberal professions) was an average of 5 to 7 percentage points

Table 2
Previously unemployed or out of the labour force recipients of the ACCRE programme in the four *SINE* surveys

In %

	1994	1998	2002	2006
Unemployed and out of the labour force as a share of all entrepreneurs	43.3	49.0	50.8	50.8
Among unemployed and out of the labour force entrepreneurs:				
- Recipients of public start-up support (of any kind)	51.7	38.8	47.7	65.4
- ACCRE recipients	n.d	30	40	59.0
Among short-term unemployed:				
- Recipients of public start-up support (of any kind)	69.2	49.5	58.5	76.2
- ACCRE recipients	n.d	40.1	51.6	70.8
Among long-term unemployed:				
- Recipients of public start-up support (of any kind)	59.7	47.5	59.7	75.6
- ACCRE recipients	n.d	39.4	52.8	69.8
Among out of the labour force entrepreneurs:				
- Recipients of public start-up support (of any kind)	6.7	13.4	18.2	27.9
- ACCRE recipients	n.d	5.4	11.2	18.3

Note: “short-term unemployed” refers to individuals who had been unemployed for less than a year when they started their business; “long-term unemployed” refers to individuals who had been unemployed for a year or more when they started their business; n.d: not determined. Calculations relating to the entrepreneur database; data weighted using the *poidsini* variable (Box 1).
Coverage: firms in non-farm private sectors established during the first half of the reference year, in metropolitan France except Corsica.
Source: Insee, *SINE* surveys 1994, 1998, 2002 & 2006.

Table 3
Mean survival rate among firms in the four surveys, according to the status of the entrepreneur immediately prior to start-up

In %

Year	Previously employed entrepreneurs	Previously unemployed or out of the labor force entrepreneurs		
		Total	ACCRE recipients	Not recipients of ACCRE
1994 cohort				
3-year survival	57.6 [57.0 ; 58.2]	52.7 [52.0 ; 53.4]	54.1 [53.1 ; 55.2]	51.8 [51.0 ; 52.6]
5-year survival	44.3 [43.6 ; 45.0]	38.3 [37.8 ; 39.0]	42.0 [41.0 ; 43.0]	36.3 [35.5 ; 37.1]
1998 cohort				
3-year survival	68.8 [68.3 ; 69.3]	62.1 [61.6 ; 62.6]	70.2 [69.3 ; 71.0]	58.6 [58.0 ; 59.2]
5-year survival	55.0 [54.5 ; 55.5]	49.2 [48.7 ; 49.7]	59.2 [58.3 ; 60.0]	45.0 [44.4 ; 45.6]
2002 cohort				
3-year survival	72.3 [71.9 ; 77.7]	66.1 [65.7 ; 66.5]	67.2 [66.5 ; 67.9]	65.5 [64.9 ; 66.0]
5-year survival	58.4 [57.9 ; 58.9]	51.1 [50.6 ; 51.6]	53.0 [52.3 ; 53.7]	49.7 [49.1 ; 50.3]
2006 cohort				
3-year survival	70.8 [70.4 ; 71.2]	64.6 [64.3 ; 64.9]	65.9 [65.5 ; 66.3]	62.7 [62.2 ; 63.2]
5-year survival	n.d.	n.d.	n.d.	n.d.

Note: Author’s calculations. Data weighted using the *poidsini* variable (Box 1). In square brackets: confidence interval at the 10% threshold calculated based on the studied rate’s standard deviation; n.d.: not determined. For the cohort of firms started in 1994, this includes ACCRE and any other forms of assistance (Box 1).
Coverage: firms in non-farm private sectors established during the first half of the reference year, in metropolitan France except Corsica.
Source: Insee, *SINE* surveys 1994, 1998, 2002 & 2006.

Box 1

THE *SINE* SYSTEM

Every four years since 1994, Insee has conducted a start-up survey via the new information system on firms, *SINE*. Using this system, it is possible to analyse entrepreneur profiles and the circumstances in which new firms are started, as well as growth and headcount changes during the first five years in the lives of new firms. The approach is based on a survey of a sample of approximately one third of firms started during the first half of the reference year. The sample is generated from the SIRENE register of firms. Each cohort is monitored over a five-year period. Firms respond to a survey during the first, third and fifth years in existence. The scope of these surveys extends across all non-farm private sectors. The vast majority of new firms are microenterprises operating in the trade, repair and other services sectors. In the studied cohorts, 80% of firms had only one employee at start-up.

Our study includes only firms created *ex nihilo*, not pre-existing firms that were taken over or reactivated, rendering the four surveys comparable from this perspective. We define an *ex nihilo* creation as the establishment of a new firm recorded in the SIRENE register of firms. For the 2006 survey, we adopted the same definition, although this survey also included a variable that treated pre-existing firms reactivated with a different activity as "start-ups". We preferred not to include such reactivated firms, in order to maintain a uniform definition across all four cohorts. We also stripped out firms that had ceased trading by the time of the first SINE survey (concerning the first year of existence). Lastly, we disregarded firms located in the French overseas departments and Corsica. This is because the special tax regimes applicable to firms in these two regions might impact attitudes to ACCRE.

The survey plan was designed to ensure that each survey would be representative with regard to the Region, Activity Sector (Nes16) and "Ex nihilo or takeover" criteria. The weight of each stratum (region x sector x ex nihilo criterion) depends on the dispersion of the five-year survival rates in each stratum. A firm's weight *i* (*poidsini* variable in the survey) in a particular stratum is equal to the inverse probability of drawing an observation from the same stratum in the sample population, relative to the probability of drawing a firm from a particular stratum in the population (Cabannes & Fougère, 2012).

The four surveys used contain essentially the same variables. There is one major exception for the 1994 survey, in which the "public support" variable made no distinction between the various forms of assistance (including ACCRE), unlike subsequent surveys. For 1994, we have used this "public support" variable, which may be treated as a proxy for ACCRE: in particular, it reflects the statutory break that prompted a larger number of job seekers and minimum social welfare recipients to apply for ACCRE subsidies (Tables 1 and 2).

higher than for previously-unemployed or out of the labour force entrepreneurs. The discrepancy is essentially the same at the three and five year marks.

Furthermore, the three-year and five-year survival rates of firms started by entrepreneurs who were working immediately prior to their establishment rose strongly from 1994 to 2002, before falling back slightly. The survival rates of firms started by previously unemployed or non-working entrepreneurs obey the same trend: a strong increase in the first three cohorts, followed by a slight decrease. Lastly, according to these descriptive elements, firms created by ACCRE recipients have a significantly higher survival rate than those of non-recipients, particularly in the first two cohorts (after five years, the difference was 5.7 points for the 1994 generation and 14.2 points for 1998). This survival rate for supported entrepreneurs increased strongly in 1998, to such an extent that the five-year figure exceeded 59%. This is more than four points higher than for firms created by entrepreneurs who were in employment prior to starting their firm. For the final two cohorts, on the other hand, the difference in survival rates between subsidised and non-subsidised firms shrinks to just over 3 points, while still remaining significant.

Econometric strategy

One cannot discount the hypothesis that ACCRE recipients are not chosen randomly. Firstly, concerning the four studied cohorts, the French labour authorities may have targeted projects proposed by entrepreneurs apparently best qualified to sustain and grow their firm. ACCRE recipients are chosen in a selection process where their personal characteristics as well as those of their business project are examined. Furthermore, self-selection may be an issue, if certain applicants are better informed and/or better able to complete the administrative formalities for obtaining the ACCRE subsidy. This would increase the probability of success of their application, and their breadth of information and ability to cope with complex administrative procedures may reflect personal characteristics that help to make a more capable business manager. In such cases, it may be these personal characteristics that account for the survival or life of an entrepreneur's firm, rather than the fact that they did or did not receive an ACCRE subsidy. However, adverse selection may also occur, by making it easier for capable individuals to set up and run a firm, by providing them with the necessary resources (Jovanovic, 1982).

Selection bias, choice and validity of instrumental-variables

In order to take such selection phenomena into consideration, using a methodology similar to that described by Cabannes & Fougère (2012, 2013), a firm survival rate or duration equation and an ACCRE distribution equation are estimated simultaneously. Inasmuch as any selection process would be partly based on non-observable variables (such as the detailed content of the project or the entrepreneur's personality and ability to manage a firm), a method based on instrumental variables is applied: one or more instruments are used that effect the probability that the entrepreneur receives the ACCRE, but have no effect on firm survival. We began by opting for a survival-oriented model rather than a duration model. Firm closure dates are not mentioned in the database for 1994, and are inaccurate for the 1998 cohort, whereas the annual data relating to cessation of activity after one, two, three, four and five years are included in all four bases (Box 2). One possible instrument is the indicator based on the quarter in which a firm was set up (Cabannes & Fougère 2012). The use of this instrument is supported by the following arguments:

- The ACCRE application must be submitted before the firm is created;

- The subsidy is deemed to be granted if the applicant has not received a notice of rejection within three months of their application;

- If the subsidy is granted, the recipient is required to create their firm within three months of approval (a firm is considered to have been created on the date on which it is recorded in the trade or firms registers).

The entrepreneur must therefore apply for ACCRE at least three months before they intend to start their firm. As the public funding allocation for ACCRE programme relates to both the State budget (payment of unemployment benefits for a year, funding for consultant advice relating to training for unemployed entrepreneurs) and the national social security budget (exemption from national insurance contributions), in a context of budgetary tightening, one may assume that the labour authorities are less restrictive during the first quarter of the current year (t) than in the final quarter of the preceding year (t-1). All other things being equal, one might expect firms started during the second quarter of the current year (t) to have a higher probability of receiving ACCRE support than those started during the first quarter. As the firms surveyed for each cohort were created during the first half of the year (Box 1), we initially tested the effect of creation during each of the six relevant months in the ACCRE distribution equation (equation 1 in Box 2). Grouping results by quarter proved to be relevant (showing that April, May and June mark a clear break with the preceding months). Ultimately, we adopted, as an instrumental variable, the dummy variable relating to start-ups born during the second quarter, with the expectation of a positive relationship between that variable and participation to the ACCRE programme.

The second instrument, used in several similar studies (Pfeiffer & Reize, 2000; Cabannes & Fougère, 2012) is an indicator of tension of the local employment market, defined as the ratio of the number of vacancies (V) to the number of unemployed (U). The geographical level chosen for France is the "*département*", which is the level on which ACCRE granting decisions are made by labour authorities. If the labour market in a *département* is slack (low V/U ratio), the probability of unemployed people finding paid employment is low, prompting the local administrative authorities to help them to move out of unemployment by encouraging firm creation. Accordingly, they tend to be less strict in granting ACCRE subsidies than in *départements* with a brighter employment situation. One can therefore expect a decreasing relationship between the V/U indicator and ACCRE granting. We have adopted the ratio of the mean monthly flows of new job vacancies and new job seekers over the course of the preceding year (t-1). This indicator takes into account possible delays by the administrative authorities in including labour market information.

Box 2

ECONOMETRIC METHOD

We have estimated a *biprobit* model featuring an equation of participation in the ACCRE programme and a five-year survival equation.

This model may be represented as follows, for an entrepreneur *i*.

$$ACCRE^*_i = \beta' X_i + \delta' instr_i + \varepsilon_{1i} \quad (1)$$

$$SURV^*_i = \alpha' X_i + \gamma ACCRE_i + \varepsilon_{2i} \quad (2)$$

$ACCRE^*_i$ and $SURV^*_i$ are latent variables respectively representing the scores of each entrepreneur *i*. These variables determine whether or not the entrepreneur received ACCRE support (1) and whether or not the firm was still trading after five years (2). The following selection rule applies:

$ACCRE_i = 1 \quad if \quad ACCRE^*_i > 0$ and $ACCRE_i = 0 \quad if \quad ACCRE^*_i \leq 0$

$SURV_i = 1 \quad if \quad SURV^*_i > 0$ and $SURV_i = 0 \quad if \quad SURV^*_i \leq 0$

X_i is a vector for the individual characteristics of the entrepreneur *i* and their project (e.g. age, nationality, gender, educational background, socioeconomic group prior to starting their firm, activity sector, start-up status, etc.).

$instr_i$ is a vector consisting of instrumental variables (two, in our model).

ε_{1i} and ε_{2i} are the error terms:

$$\begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{pmatrix} \longrightarrow N(0, \Sigma)$$

The error terms have a bivariate normal distribution (with a variance-covariance matrix formed with 1 along the principal diagonal, and the other elements in the matrix formed by the error term correlation coefficient). Where applicable, the correlation between error terms can be used to allow for unobserved heterogeneity.

The estimations were performed using the *Stata* software using the maximum likelihood method for the cumulative bivariate normal distribution function. Additionally, the regressions shown in the main text, appendices and online supplements were generated by weighting the observations *i* based on the *poidsini* variable (*pweight* procedure in *Stata*). As noted by Cabannes & Fougère (2012), this weight depends on the five-year survival dispersion in each stratum (Box 1), as a result of which, omitting the weight may introduce an endogeneity bias.

However, this instrument might be endogenous, having a proper effect on firm survival, or on an omitted variable that impacts survival rates. The reverse causality hypothesis, whereby surviving new firms would improve the local market conditions may be refuted. The chosen indicator of tension in the local labour market trails the actual creation of firms by a year. Furthermore, the possibility of a correlation between this instrument and the error term in the survival equation appears weak, even where entrepreneurs persevere in a barely profitable activity because of a lack of alternatives to persistent unemployment in the local labour market. Such a correlation could only occur if the labour market were durably degraded in the *département* over a period of five years following the creation date.

Inclusion of individual characteristics of entrepreneurs and their firms

We have used the wide array of variables available in the SINE survey, which relate not only to the observable characteristics of entrepreneurs but also to the economic characteristics of the firms they create (Box 1). These variables include the entrepreneur's gender, nationality, educational level, socioeconomic group prior to starting the firm, age, number of previous start-ups, as well as the firm's legal status, size and activity sector. We also used a dummy variable to reflect the effect (if any) of other forms of public assistance (Table 2). Regarding the financial resources invested in new firms, we created a dichotomous variable, assigning half of the firms in each cohort the "limited resources" value and the other half the "ample resources" value. Note that this variable does not include the ACCRE subsidy, which may be assimilated to either income (unemployment benefits paid to the entrepreneur for one year) or a cost saving (exemption from national insurance contributions) (Table 1). We declined to include the grant of bank loans to entrepreneur, due to the risk of selection bias with regard to survival rates. However, this article does examine the impact of incorporating this variable into our model on the results (see *infra*, Robustness of the estimation).

Impact of ACCRE on five-year survival rates of firms created by the short-term unemployed

To allow for the heterogeneous impact of ACCRE on the various categories of entrepreneurs (namely short-term unemployed, long-term unemployed or out of the labour force) at the time they start their firms, we initially focus on the short-term unemployed, a population in which entitlement to unemployment benefits (and hence ACCRE subsidies) is relatively uniform (Table 1). We then use a comparative approach to incorporate the other categories of recipients (long-term unemployed and people out of the labour force prior to starting their firms) into our estimations. Each *probit* equation was first estimated separately, and then the two were estimated jointly. Comparing the two sets of results provides insight into the existence of a selection bias in the participation in the ACCRE programme.

Estimation of a model of participation in the ACCRE programme (for the short-term unemployed)

Table A (in the appendix), which relates to entrepreneurs who had been unemployed for a short period prior to starting their firm, reveals that French citizens were more likely to participate in the ACCRE programme. It should also be noted that women are not discriminated against: in 2002 and 2006, all other things being equal, more women than men received the subsidy. Concerning education, unqualified applicants were markedly less likely to receive ACCRE assistance than those with basic general or vocational qualifications (BEP and CAP). However, applicants with baccalaureate or higher diplomas were not more likely to receive an ACCRE subsidy than the previous categories (except in 1998). Examining the entrepreneur's socioeconomic group prior to starting their firm clearly reveals the legal provisions governing the ACCRE granting process. Business owners, traders and tradesmen were less likely than executives to receive this support (except in 2006, although the results are hard to interpret as we were obliged to combine all non-employees in a single category). As they did not receive unemployment benefits, they had little incentive to apply for ACCRE. Students who had completed their studies and gone on to become entrepreneurs were in the same situation.

Furthermore, small projects with no employees were more likely to receive the subsidy than larger projects. Moreover, ACCRE support was often accompanied by other public subsidies (Table 2). In our estimations we allowed for this phenomenon by introducing a dichotomous variable (received/did not receive other public support) to account for the influence of such subsidies on firm survival. Concerning firm statuses, for all cohorts, firms structured as liberal professions, SARL or SA limited companies and partnerships, considered together, had a lower probability of receiving ACCRE subsidies than individual businesses. Two interpretations can be given for this. Firstly, the authorities may channel ACCRE support toward individuals with the lowest probability of finding employment, and/or who have fewer legal and financial resources than entrepreneurs setting up companies to implement their project, in accordance with the goals of the ACCRE legislation (see Table 1). However, a completely different hypothesis may also be advanced: individuals using complex legal arrangements are not company employees (for example, a non-salaried managing director paid out of operating profit), and as such have no incentive to apply for ACCRE (Daniel & Mandelblat, 2010). Concerning the financial resources invested when starting a business, firms (in all four cohorts) with relatively limited financial resources had a lower probability of receiving the subsidy than others.

Lastly, the variables reflecting local labour market tension and those indicating ACCRE awarded in the second quarter of the studied year, which are used as instruments when estimating the two-equation model, have an effect with the expected sign. (respectively < 0 in the first case and > 0 in the second) and are significant (at the 1% threshold in the first case, and 10% in the second). This result corroborates the rationale described above regarding the determining factors in the authorities' decision to grant ACCRE support.

Estimated five-year survival model (for the short-term unemployed)

Estimating the *probit* five-year survival equation (Table 4) shows that for the four cohorts of new firms, ACCRE support has a significant effect (at the 1% threshold) on five-year survival. This effect was the most pronounced in 1998, to such an extent that there was a significant difference (at the 1% threshold) between the estimated coefficient for the ACCRE (Yes/No) dummy variable in 1998, on one hand, and those estimated for the cohorts of firms started in 2002 and 2006, on the other hand. Comparing these results with the descriptive statistics referred to earlier (Table 3) reveals a smaller difference in survival rates between firms with and without ACCRE subsidies in 2002 and 2006, compared with 1998.

Furthermore, the level of education and the socioeconomic group prior to starting the firm had a weak and in most cases non-significant effect on five-year survival rates. On the other hand, firms structured as companies or as liberal professions had better five-year survival prospects than individual businesses. In addition, other forms of regional and local subsidies for start-ups had no effect on five-year survival

Table 4
Estimation of the *probit* model with the dependent variable: Survived five years (Yes/No) – short-term unemployed entrepreneurs

Cohort	1994	1998	2002	2006
ACCRE (Yes/No)	0.30***[0.06]	0.50***[0.06]	0.16***[0.04]	0.14***[0.05]
Limited financial resources (*ref.* Ample resources)	- 0.32***[0.06]	- 0.25***[0.06]	- 0.17***[0.04]	- 0.25***[0.05]
French nationality (*ref.* Foreign)	0.26*[0.11]	0.47***[0.11]	0.11 [0.07]	0.27***[0.08]
Male gender (*ref.* Female)	0.05 [0.07]	0.12*[0.07]	0.16***[0.050]	0.042 [0.050]
Age group >50 (*ref.* 16 - 50)	0.13 [0.16]	0.13 [0.12]	- 0.04 [0.07]	0.077 [0.067]
Education, qualifications (*ref.* CAP/BEP vocational diplomas)				
No qualifications	- 0.16*[0.08]	- 0.12 [0.09]	- 0.11 [0.07]	- 0.09 [0.08]
BEPC secondary school certificate	included in BEP	- 0.13 [0.12]	- 0.09 [0.08]	- 0.20**[0.09]
Vocational baccalaureate (bac pro)	- 0.01 [0.08]	- 0.01 [0.10]	0.06 [0.070]	0.02 [0.08]
General baccalaureate	(Voc.+gen. bac.)	- 0.16 [0.13]	- 0.04 [0.08]	- 0.18*[0.10]
Higher education	0.01 [0.08]	- 0.12 [0.09]	0.10*[0.05]	0.05 [0.07]
Previous job category (*ref.* Executive)				
Business owner	0.11 [0.26]	- 0.18 [0.30]	0.30* [0.17]	- 0.27 [0.20] (all
Tradesman or trader	0.13 [0.18]	- 0.19 [0.16]	inc. business owners	non-employees)
Supervisory worker	0.01 [0.13]	- 0.03 [0.15]		- 0.04 [0.08]
Intermediate profession	- 0.06 [0.13]	- 0.21*[0.13]	0.12 [0.11] (all employees	- 0.11 [0.08]
Clerical worker	- 0.06 [0.09]	- 0.23**[0.10]	except	- 0.03 [0.07]
Manual worker	0.12 [0.10]	- 0.10 [0.10]	executives)	- 0.02 [0.07]
Student	- 0.17 [0.18]	- 0.26*[0.15]	0.07 [0.16]	(inc. all
Other/non-working	- 0.03 [0.16]	- 0.51***[0.15]	- 0.02 [0.12]	non-employees)
No employees (*ref.* One or more employees.)	0.18***[0.07]	- 0.07 [0.08]	0.05 [0.05]	- 0.02 [0.07]
Other public subsidies: Received other (*ref.* Did not receive)	n.d.	0.05 [0.10]	- 0.06 [0.05]	0.03 [0.05]
Legal form: Companies and liberal professions (*ref.* individual business)	0.40***[0.08]	0.16**[0.07]	0.21***[0.05]	0.34*** [0.05]
Constant	- 0.58***[0.20]	- 0.44***[0.18]	- 0.45***[0.15]	0.06 [0.15]
Number of observations	4,230	3,355	5,588	7,300
Log likelihood	- 9,111	- 6,875	- 7,840	- 13,516
Pseudo R^2	0.061	0.068	0.031	0.036

Note: Observations weighted using the SINE survey's *poidsini* variable (Box 1). The standard deviations of the regression coefficients, shown between square brackets, and their significance were calculated using the *robust* procedure with the Stata software. Asterisks indicate significance thresholds: 10% (*), 5% (**) and 1% (***), respectively. For the 2006 cohort, the survival variable was calculated after three years, as the five-year data was unavailable. Some control variables are not shown in the table: the dummy variables used to categorise firms in 6 activity sectors (NES16), the number of firms created prior to the studied one, whether the firm was created by a single entrepreneur or several, whether or not market research was conducted prior to start-up, and whether or not the firm was located in the Paris region.
Reading note: Receiving ACCRE support had a significant positive impact (at the 1% threshold) on the probability of survival after five years in the cohorts of firms started in 1994, 1998, 2002 and after 3 years for the 2006 cohort, all observable characteristics being equal. Conversely, starting a business with limited resources had a significant negative impact (at the 1% threshold) on the same survival probability across all cohorts.
Coverage: firms in non-farm merchant sectors established during the first half of the reference year, in metropolitan France except Corsica; authors' database limited to short-term unemployed entrepreneurs (unemployed for less than a year prior to starting their firm).
Source: Insee, *SINE* surveys 1994, 1998, 2002 & 2006.

rates. Lastly, the probability of survival of a firm after five years is lower where the entrepreneur has limited financial resources.

Impact of ACCRE on survival rates in a simultaneous equation-based model (short-term unemployed)

When estimating the model based on two simultaneous equations, the interdependencies between the variables in the two equations are included, and if the model is correctly formalised, enables the impact on five-year survival rates of any selection in the participation in the ACCRE programme to be eliminated. It was estimated using the maximum-likelihood method with the *Stata* software (Box 2). According to this estimation, for the short-term unemployed (Table 5), the impact of ACCRE on five-year survival was not significantly different to zero (at the 10% threshold) for the four cohorts. The result is therefore totally different from the previous result for the basic survival equation (Table 4). This clearly indicates a selection effect, compromising the effectiveness of the ACCRE programme as an economic policy tool intended to increase the five-year survival rate for the four cohorts of firms.

The estimated coefficients of the instrumental variables have an expected sign and are significant (at the 1% threshold for the departmental employment market tension variable, and at the 10% threshold for the variable indicating the quarter in which a firm is created). However, estimating the effect of departmental labour market tightness on the participation in the ACCRE programme reveals a break after 1998. The estimated coefficient was lower (at the 1% threshold) in 2002 and 2006, than in 1998. This may be interpreted as reflecting a shift in the selection process that diminished after 1998, whereas the proportion of unemployed entrepreneurs receiving ACCRE support was significantly higher in 2002 and 2006 than in 1998 (Table 2). A similar shift was observed in the estimated coefficients for the ACCRE (Yes/No) variable; the coefficients were significant at the 20% threshold for the cohorts of firms started in 1994 and 1998, but not for subsequent cohorts. From this perspective, the decrease in the apparent effect of ACCRE in the descriptive data (Table 3) may be interpreted as reflecting decreased selection, in turn linked to a wider availability of ACCRE to unemployed entrepreneurs.

Table 5
ACCRE as a determining factor in five-year firm survival – Estimation of the model featuring two simultaneous equations – Short-term unemployed entrepreneurs

Cohort	1994	1998	2002	2006
5-year survival equation:				
ACCRE (Yes/No)	0.19 [0.14]	0.18 [0.13]	- 0.07 [0.24]	0.03 [0.15]
ACCRE participation programme equation:				
Departmental labour market tightness	- 0.65***[0.23]	- 0.93***[0.20]	- 0.26**[0.12]	- 0.30***[0.07]
Firm created in Q2. (*ref.* Q1.)	0.10*[0.6]	0.08*[0.05]	0.13**[0.05]	0.07 [0.05]
Number of observations	4,230	3,355	5,588	7,300
Corr. residues from equations	- 0.23 [0.33]	- 0.13 [0.37]	0. 23 [0.60]	0.01 [0.27]

Note: maximum-likelihood estimations for the model calculated using the Stata software (Box 2). Observations weighted using the *SINE* survey's *poidsini* variable (Box 1). For the 2006 cohort, the survival variable was calculated after three years, as the five-year data was unavailable. The database was limited to short-term unemployed entrepreneurs (unemployed for less than a year prior to starting their firm). The standard deviations of the regression coefficients, shown between square brackets, and their significance were calculated using the *robust* procedure in the Stata software application. Asterisks indicate significance thresholds: 10% (*), 5% (**) and 1% (***), respectively. Some control variables are not shown in the table: the dummy variables used to categorise firms in 6 activity sectors (NES16), the number of firms created prior to the studied one, whether the firm was created by a single entrepreneur or several, whether or not market research was conducted prior to start-up, and whether or not the firm was located in the Paris region.
Reading note: Receiving ACCRE support had no significant positive impact (at the 10% threshold) on the probability of survival after five years in the cohorts of firms started in 1994, 1998, 2002 and after 3 years for the 2006 cohort, all observable characteristics being equal.
Coverage: firms in non-farm private sectors established during the first half of the reference year, in metropolitan France except Corsica; authors' database limited to short-term unemployed entrepreneurs (unemployed for less than a year prior to starting their firm).
Source: Insee, *SINE* surveys 1994, 1998, 2002 & 2006.

Lastly, the correlation between the residuals of the two equations is never significant. Based on this result, no unobserved variables linked both to participation in the ACCRE programme and firm survival rates appear to exist.

Impact of ACCRE on five-year survival rates among firms created by various categories of recipients

This section extends the analysis for short-term unemployed to cover other categories of entrepreneurs eligible to the ACCRE programme. We estimate our model based on two simultaneous equations for the population of entrepreneurs, including not only the short-term unemployed but also the long-term unemployed and individuals out of the labour force prior to starting their firm. We introduce a dummy variable into the two equations for each of the latter two categories, relative to the short-term unemployed. The goal was to assess whether including the two additional entrepreneur categories modified the impact of ACCRE on firm survival (Table 6).

Including the whole population of entrepreneurs eligible to the ACCRE programme multiplies the size of each cohort by a factor of between two and three. This increase makes it possible to get more accurate estimations for the coefficients for the two equations. Considering the results obtained by estimating the equation of participation in the ACCRE programme confirms that entrepreneurs out of the labour force were much less likely to receive ACCRE support than the short-term unemployed, which corresponds to the results yielded by the descriptive statistics (cf. Table 2). The probability of receiving ACCRE support by long-term unemployed entrepreneurs varied according to the cohort, being higher than for the short-term unemployed in 2002 and lower in 1994. Concerning firm survival, estimating the corresponding equation shows that the probability of firms created by long-term unemployed and entrepreneurs out of the labour force to still be active after five years was not significantly different from the short-term unemployed (except in the 2006 cohort, when the probability was lower for those out of the labour force).

Table 6
ACCRE as a determining factor in five-year firm survival – Estimation of the model featuring two simultaneous equations – Short-term and long-term unemployed and out of the labour force entrepreneurs

Cohort	1994	1998	2002	2006
5-year survival equation:				
- ACCRE (Yes/No)	0.11 [0.09]	0.23***[0.06]	0.06 [0.08]	- 0.09 [0.07]
- Recipient category (*ref.* Short-term unemployed)				
Long-term unemployed	- 0.01 [0.05]	- 0.12 [0.08]	- 0.03 [0.03]	- 0.02 [0.03]
Out of the labour force	0.08 [0.20]	0.13 [0.09]	0.11 [0.11]	- 0.21* [0.12]
ACCRE programme participation equation:				
- Recipient category (*ref.* Short-term unemployed)				
Long-term unemployed	- 0.19***[0.05]	0.04 [0.04]	0.09***[0.03]	0.02 [0.03]
Out of the labour force	- 1.65***[0.07]	- 1.12***[0.06]	- 1.13***[0.04]	- 1.38***[0.05]
- Departmental labour market tightness	- 0.56***[0.19]	- 0.74***[0.13]	- 0.31***[0.07]	- 0.21***[0.04]
- Firm created in Q2 (*ref.* Q1)	0.15***[0.05]	0.09**[0.04]	0.091***[0.031]	0.06*[0.03]
Number of observations	8,256	8,269	13,792	18,416
Corr. residues from equations	- 0.14 [0.22]	- 0.33* [0.16]	- 0.07 [0.18]	0.29 [0.15]

Note: maximum-likelihood estimations for the model calculated using the Stata software (Box 2). Observations weighted using the *SINE* survey's *poidsini* variable (Box 1). For the 2006 cohort, the survival variable was calculated after three years, as the five-year data was unavailable. The standard deviations of the regression coefficients, shown between square brackets, and their significance were calculated using the *robust* procedure in the Stata software application. Asterisks indicate significance thresholds: 10% (*), 5% (**) and 1% (***), respectively. Some control variables are not shown in the table: see table 4.
Reading note: 5-year survival equation - Receiving ACCRE support had no significant impact (at the 10% threshold) on the probability of survival after five years in the cohorts of firms started in 1994 and 2002, and after 3 years for the 2006 cohort, all observable characteristics being equal. There was a significant impact (at the 1% threshold) for the cohort of firms started in 1998.
Source: Insee, *SINE* surveys 1994, 1998, 2002 & 2006.

Ultimately, considering all categories of entrepreneur, ACCRE had no effect on firm survival, as was the case for short-term unemployed entrepreneurs, with the notable exception of the cohort of firms started in 1998. For this cohort, the ACCRE programme had a significant positive impact on survival (at the 1% threshold). An initial hypothesis is that this effect would be positive for long-term unemployed and out of the labour force entrepreneurs for the cohort of firms started in 1998. However, this hypothesis can be discounted, as it was shown to be false when our model was estimated for each of these entrepreneur categories separately. The larger sample size was the key factor responsible for the more accurate estimation. However, despite much larger samples in 2002 and 2006, as a result in the increase in the number of recipients (compare Table 6 and Table 5), the impact of ACCRE for these two cohorts was not significant (and was indeed negative for the 2006 generation). There was a break in terms of ACCRE's impact between the 1998 cohort and subsequent generations.

With effect from 2002, the percentage of unemployed and out of the labour force ACCRE recipients increased considerably (Table 2), attracting new categories of entrepreneurs. The relationship detected between ACCRE support and firm survival was negative (but not significant) for certain categories of entrepreneurs. Specifically, this was the case for short-term unemployed entrepreneurs in 2002 (Table 5) and all ACCRE recipients in 2006. Both the size and the nature of the ACCRE recipient population has changed. Beginning in the early 2000's, ACCRE increasingly resembled an additional welfare benefit, granted automatically to entrepreneurs based on purely administrative criteria. This situation was officially acknowledged in the Decree published in 2007 (Table 1). This extension to the relevant legislation may have prompted individuals to create a firm who would not otherwise have done so without the incentive offered by the ACCRE programme. In such a scenario, self-selection prompts more people to apply for the subsidy, although they may be less well qualified to start a firm than when the award criteria were stricter. Analysing the recipients of minimum social welfare benefits who started firms as part of the 2002 and 2006 cohorts corroborates this finding (see additional information on the online supplement C2).

Robustness of the estimations

Inclusion of financial variables

Thus far, we have included the financial resources invested at start-up (Tables 4 and A1-1), without isolating any bank loans available to firms. One cannot exclude the possibility of this variable being endogenous with regard to survival. Comparing the estimations of the two equations for short-term unemployed entrepreneurs according to whether or not the variable "financial resources available to the firm at start-up" is included in the model's two equations reveals that the coefficients are very similar, and in all cases, not significantly different to zero regarding the impact of ACCRE on firm survival (Table C4-1 in the online supplement). The same applies to the coefficients of the instrumental variables, which were not significantly changed. Moreover, when the variable "financial resources at start-up" was replaced with the variable "approved bank loan" in the same two equations, the results of the estimations[5] were not significantly different to our original estimation (Table 5). These results are important inasmuch as they demonstrate that the financial resources deployed have a significant effect on five-year survival (Table 4) but, according to this analysis, are independent of the ACCRE programme's impact on survival.

Does the business cycle have a differentiated effect on survival of subsidised and non-subsidised firms?

The issue of the impact of the analysis period on survival results must be addressed. Firm survival is indeed sensitive to changes in the business cycle. Jacobson et al. (2011), and Fougère et al. (2013) both find that the 2008 economic crisis had a major impact on business failures, albeit after a considerable delay, according to these authors, who claim that the effect of the crisis on failures did not become significant until 2009, subsequently growing in severity until late 2010, when their study ended. In this case, the goal is to determine whether the cycle exerts a differentiated effect on the impact of ACCRE on firm survival. Note that the *SINE* survey only relates to firms created during the first half-year in each cohort.

5. *Not discussed herein; available from the author.*

The first lesson to be drawn from Table 7 is that the comparative effect of ACCRE on the survival of firms started by the short-term unemployed after three and four years (two years in the case of the 2006 cohort) was not significantly different to the effect on survival after five years (three years in the case of the 2006 cohort). This generalises the validity of our previous five-year analyses for shorter timescales. ACCRE had no effect for the 1994, 2002 and 2006 cohorts. The cohort of firms started in 1998 differs from the others, as this effect, which was not significant at the 10% threshold for five-year survival, was significant for the four- and three-year timescales. However, the coefficients and standard deviations associated with the corresponding three regressions reveal that these differences between coefficients are not significant. We stand by our original assessment that ACCRE had a weak effect, at all timescales, on the cohort of firms started in 1998 by short-term unemployed entrepreneurs.

In the light of these results, no significant influence of the unemployment cycle on the effect of ACCRE support on firm survival can be established. For example, for the cohort of firms started in 1998, the lower limit of the unemployment rate during the period from the first half of 2001 to the first half of 2002 was followed by a rise in early 2003 (see Figure C4-I in the online supplement). Can the slight fall in the estimated coefficient of the effect of ACCRE after five years, relative to previous years, be interpreted as a consequence of the deterioration in the employment market? We do not believe so, as this fall was not significant with respect to previous years. Furthermore, the aforementioned studies reveal a delay of months or half-years between the shift in the cycle and business failures, implying that if such an effect existed, it would be felt in subsequent years.

For the cohort of firms started during the first half of 2002, a surge in unemployment was observed in the second half of 2005 and the first

Table 7
Impact of ACCRE on firm survival after 3, 4 and 5 years - Estimation of a two-simultaneous equations model – Short-term unemployed entrepreneurs

	Survival equation: impact of ACCRE on survival	ACCRE participation equation: impact of labour market tension variable	ACCRE participation equation: impact of Q2 award variable
1994 cohort			
3 years	0.17 [0.18]	- 0.66***[0.23]	0.08*[0.05]
4 years	0.17 [0.13]	- 0.64***[0.23]	0.09*[0.05]
5 years	0.19 [0.14]	- 0.65***[0.24]	0.10*[0.06]
1998 cohort			
3 years	0.20*[0.11]	- 0.94***[0.20]	0.08*[0.05]
4 years	0.20*[0.12]	- 0.94***[0.20]	0.07*[0.04]
5 years	0.18 [0.13]	- 0.93***[0.20]	0.08*[0.05]
2002 cohort			
3 years	0.08 [0.32]	- 0.30**[0.14]	0.10*[0.06]
4 years	0.10 [0.23]	- 0.31***[0.11]	0.09*[0.05]
5 years	0.07*[0.24]	- 0.26**[0.12]	0.13*[0.05]
2006 cohort			
2 years	0.03 [0.11]	- 0.30***[0.06]	0.09*[0.05]
3 years	0.03 [0.15]	- 0.30***[0.07]	0.07*[0.05]

Note: maximum-likelihood estimations for the model calculated using the Stata software (Box 2). Observations weighted using the *SINE* survey's *poidsini* variable (Box 1). For the 2006 cohort, the survival variable was calculated after three years, as the five-year data was unavailable. The database was limited to short-term unemployed entrepreneurs (unemployed for less than a year prior to starting their firm). The standard deviations of the regression coefficients, shown between square brackets, and their significance were calculated using the *robust* procedure in the Stata software application. Asterisks indicate significance thresholds: 10% (*), 5% (**) and 1% (***), respectively. Certain control variables are not shown in the table: see table 4. Only the relevant variables are shown in the table.
Reading note: Receiving ACCRE support had no significant positive impact (at the 10% threshold) on the probability of survival after three, four and five years in the cohorts of firms started in 1994 and 2002. Neither did it impact the probability of survival of firms set up in 2006 after two and three years. For 1998, a positive effect on three- and four-year survival was observed (at the 10% threshold).
Source: Insee, *SINE* surveys 1994, 1998, 2002 & 2006.

half of 2006, followed by a slight improvement in 2007. Can the negative coefficient of the effect of ACCRE on five-year survival (2007) be interpreted as a delayed effect of the unemployment surge in the preceding two years? The coefficient is too weak to affirm such a relationship. For the cohort of firms started in 2006, unemployment reached a trough in the first half of 2008 before increasing sharply in the first half of 2009. Based on our estimations, this sudden upturn in unemployment did not affect the impact of the ACCRE programme, at least in the short term.

* *
*

Our research regarding the effectiveness of the ACCRE programme as a measure for improving the five-year survival prospects of firms set up by unemployed and out of the labour force entrepreneurs revealed a selection bias. For the cohorts of firms started in 1994, 2002 and 2006, adjusting for this bias reveals that ACCRE had no effect on the survival of subsidised firms in any of the entrepreneur categories (i.e. short-term unemployed, long-term unemployed or out of the labour force). For the cohort of firms started in 1998, this effect was weak for short-term unemployed entrepreneurs (significant at the 10% threshold after four years, and the 20% threshold after five years), but significant (at the 1% threshold) when the sample was enlarged to include all eligible individuals.

This difference may be accounted for by regulatory changes. In 1998, draconian budget restrictions impacted all categories of recipients (Table 2). Conversely, during the second quarter of 1994, the number of recipients increased following changes to make the subsidy more generous. For 2002 and 2006, the eligibility conditions were loosened. One cannot rule out the possibility that these regulatory changes may have acted as signals, impacting not only the number but also the personal characteristics of unemployed and out of the labour force entrepreneurs. In 1998, the dissuasive nature of the signal may have restricted the candidate population to only the most competent and best equipped to start a firm. This would have exerted a self-selection effect among entrepreneurs, which may account for our results regarding the effect of ACCRE for this cohort. The instrumental variables in our model were designed first and foremost to reflect the economic changes, at department level, taken into account by officials in their ACCRE granting decisions. These variables may less faithfully reflect the behaviour of potential entrepreneurs, who are more sensitive to national changes such as restrictions in national funding allocations for the ACCRE programme or the general macroeconomic situation.

The more accommodating legislation introduced in the second quarter of 1994, and from the early 2000s, may have prompted new categories of people (in particular minimum social welfare recipients) to set up a firm. This more generous legislation led to a considerable increase in the number of recipients, but had no impact on five-year survival rates among subsidised firms. There may have been a reverse self-selection effect, prompting individuals who were less competent and less well-equipped to run a firm to try their luck. Moral hazard effects may also have played a role. Individuals thus encouraged to start a firm took less risk and invested less, inasmuch as they were very likely to receive the subsidy. Accordingly, they may in some cases have been less well-prepared and motivated to make their project a success. Once again, these selection or moral hazard effects may have been imperfectly controlled in our model. The programme's limited or even non-existent effect on firm survival should not overshadow any qualitative aspect. Broadening the ACCRE eligibility conditions may have provided an incentive to groups with slim chances of employment to start their own business, hence improving their employability. □

BIBLIOGRAPHY

Aghion, P., Akcigit, U. & Howitt, P. (2014). What Do We Learn From Schumpeterian Growth Theory? In: *Handbook of Economic Growth*, vol. 2, chap. 1, pp. 515–563. Elsevier.

Audretsch, D. B. & Mahmood, T. (1994). Firm selection and industry evolution: the post-entry performance of new firms, *Journal of Evolutionary Economics*, 76(4), 243–260.

Béziau, J. & Bignon, N. (2017). Les entreprises créées en 2010, plus pérennes que celles créées en 2006, touchées par la crise. *Insee Première* N° 1639.

Boum Galiana, O. (2015). Le nouvel accompagnement pour la création et la reprise d'entreprise (Nacre). *Dares Analyses* N° 052.

Cabannes, P. Y. & Fougère, D. (2012). Une évaluation de l'effet de l'ACCRE sur la durée de vie des entreprises. Mimeo, présentation à la chaire « Sécurisation des parcours professionnels », 26/09/2012.

Cabannes, P. Y. & Fougère, D. (2013), *Aider à la création d'entreprise.* Paris: Presses de Sciences-Po, coll. Sécuriser l'emploi.

Caliendo, M. (2016). Start-up subsidies for the unemployed: Opportunities and limitations. IZA World of Labor N° 200.
(https://wol.iza.org/articles/start-up-subsidies-for-unemployed-opportunities-and-limitations/long)

Caliendo, M. & Künn, S. (2011). Start-up Subsidies for the unemployed: long term evidence and effect heterogeneity. *Journal of Public Economics,* 95, 311–331.

Caliendo, M., Hogenacker, J., Künn, S. & Weissner, F. (2015). Subsidized start-ups out of unemployment: a comparison to regular business start-ups. *Small Business Economics*, 45(1), 165–190.

Calvino, F. C., Criscuolo, C. & Menon, C. (2015). Cross country evidence on start-up dynamics. *OECD Science, Technology and industry working paper* N° 2015/06. Paris: OECD publishing.
(www.oecd.org/eco/growth/Cross-country-evidence-on-start-up-dynamics.pdf)

Charpail, C. (1995). L'aide aux chômeurs créateurs ou repreneurs d'entreprises en 1994. Dares*, Premières Informations, Premières Synthèses* N° 106.

Charpail, C. (1996). Aide aux chômeurs créateurs d'entreprise : Accre et chèque-conseil en 1995. Dares, *Premières Synthèses* N° 96-08-32-3.

Crépon, B. & Duguet, E. **(2003).** Bank loans, start-up subsidies and the survival of new-firms: An econometric analysis at the entrepreneur level, *Cahiers de la MESE, EUREQua*, N° 2003-77.

Criscuolo, C., Gal, P. N., & Menon, C. (2014). The dynamics of Employment Growth, New Evidence from 18 Countries. *OECD Science, Technology and industry Policy Paper* N °14. Paris: OECD Publishing.
(www.oecd-ilibrary.org10.1787/the-dynamics-of-employment-growth_5jz417hj6hg6-en)

Daniel, C., & Mandelblat, C. (2010). La sous-mobilisation de l'aide aux chômeurs créateurs d'entreprises en Ile de France. Dares*, Documents et Études* N° 154.

Davis, S. J., Haltiwanger, J., Jarmin, R. S., Krizan, C. J, Miranda, J., Nucci, A. & Sandusky, K. **(2007).** Measuring the Dynamics of Small Businesses, Working paper series N° 13226. NBER.
(www.nber.org/papers/w13226)

Deidda, M., Di Liberto, A., Foddi, M., & Sulis, G. (2015). Employment subsidies, informal economy and women's transition into work in a depressed aera: evidence from a matching approach. IZA Journal of Labour Economics, 4(7), 35–53.

Duhautois, R., Désiage, L. & Redor, D. (2015). Long Term Effect of Public Subsidies on Start-up Survival and Economic Performance: an Empirical Study with French data. *Revue d'économie industrielle*, 149, 11–42.

Désiage, L, Duhautois, R. & Redor, D. (2010). Do Public Subsidies have an impact on new firm survival? An Empirical Study with French Data. *Working paper TEPP* N° 2010/4.

Fougère, D., Golfier, C., Horny, G. & Kremp, E. (2013). Quel est l'impact de la crise de 2008 sur la défaillance des entreprises ? *Economie et Statistique*, 462-463, 69–97.

Gu, Q., Karoly, L. A. & Zissimopoulos, J. (2008). Small business assistance in the United States. *Working Paper Kauffman-Rand Insti-*

tute for Entrepreneurship Public Policy N° WR-603-EMKF.

Guimiot, A. & Mareau, E. (2003). Les chômeurs créateurs d'entreprises en 2001-2002. Dares, *Premières Informations* N° 37-3.

Haltiwanger, J. (2011). Job Creation and firm dynamics in the United States. In: *NBER*, Chapter: *Innovation Policy and the Economy*, pp. 17-38.

Haltiwanger, J., Jarmin, R. S. & Miranda, J. (2013). Who Creates Jobs? Small versus large, versus young. *The Review of Economics and Statistics,* 95(2), 347–361.

Jacobson, T., Lindé, J. & Roszbach, K. (2011). Firm default and agregate fluctuations. Mimeo, Board of Governors of the Federal Reserve System, International Finance Discussion Papers N° 0129.

Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica,* 50(3), 649–670.

Mata, J. & Portugal, P. (1994). Life duration of new firms. *Journal of Industrial Economics*, 42, 227–245.

Mouriaux, M. F. (1995). Croyance, expérience, évaluation : l'évolution de l'aide aux chômeurs créateurs d'entreprise. In: *Cahiers du Centre d'études de l'emploi : Les politiques publiques de l'emploi et leurs acteurs*, pp. 103–117. Paris : Presses universitaires de France.

Ould Younes, S. (2010). Les aides aux créations ou reprises d'entreprises en 2007 et 2008 : plus d'un tiers des créations a bénéficié de l'ACCRE. *Dares Analyses* N° 38.

Pfeiffer, F. & Reize, F. (2000). Business start-ups by the unemployed – an econometric analysis based on firm data. *Labour Economics*, 7, 629–663.

Richet, D. & Thomas, S. (2016). Les créateurs d'entreprises en 2014 : moins de chômeurs et des projets plus modestes qu'en 2010. *Insee Première* N° 1600.

Table A
Estimation of the *probit* model with the dependent variable: participation in the ACCRE programme (Yes/No) – short-term unemployed entrepreneurs

Cohort	1994	1998	2002	2006
Limited financial resources (*ref.* Ample resources)	- 0.44** [0.07]	- 0.18***[0.07]	- 0.16**[0.05]	- 0.26**[0.05]
French nationality (*ref.* Foreign)	0.47***[0.11]	0.26**[0.13]	0.28***[0.07]	0.19**[0.08]
Male gender (*ref.* Female)	- 0.01 [0.08]	- 0.11 [0.07]	- 0.12**[0.05]	- 0.09* [0.05]
Age group >50 (*ref*. 16 - 50)	- 0.03 [0.17]	- 0.33**[0.13]	- 0.02 [0.07]	- 0.10 [0.07]
Education. qualifications (*ref.* CAP/BEP vocational diplomas)				
No qualifications	- 0.11 [0.08]	- 0.30***[0.09]	- 0.28***[0.07]	- 0.30***[0.09]
BEPC secondary school certificate	included in BEP	- 0.10 [0.11]	- 0.02 [0.08]	- 0.22**[0.09]
Vocational baccalaureate (bac pro)	- 0.02 [0.08]	- 0.07 [0.10]	- 0.03 [0.07]	- 0.10 [0.09]
General baccalaureate	(Voc.+gen. bac)	- 0.02 [0.13]	0.06 [0.08]	- 0.15 [0.11]
Higher education	0.06 [0.10]	0.17** [0.09]	0.08 [0.06]	- 0.10 [0.07]
Previous job category (*ref.* Executive)				
Business owner	- 0.97***[0.23]	- 0.50* [0.33]	- 0.16 [0.17]	- 0.24 [0.26]
Tradesman or trader	- 0.91***[0.19]	- 0.51***[0.17]	with bus. owners	(all non-employees)
Supervisory worker	0.16 [0.14]	0.25 [0.16]	0.21*[0.12]	- 0.03 [0.08]
Intermediate profession	0.15 [0.14]	- 0.01 [0.13]	(all employees	- 0.06 [0.09]
Clerical worker	0.11 [0.10]	- 0.08 [0.10]	except executives)	- 0.16** [0.07]
Manual worker	- 0.06 [0.11]	- 0.04 [0.11]		- 0.18**[0.09]
Student	- 0.92***[0.18]	- 0.51***[0.14]	0.09 [0.17]	(inc. all
Other non-working	- 0.73***[0.17]	- 0.41***[0.15]	0.01 [0.12]	non-employees)
No employees (ref. One or more employees)	0.046 [0.072]	0.49***[0.08]	0.17***[0.05]	0.19**[0.06]
Other public subsidies: Received (ref. Did not receive)	d.m	0.62***[0.12]	0.38***[0.05]	0.28***[0.06]
Legal form: Companies and liberal prof. (*ref.* individual business)	- 0.36*** [0.08]	- 0.55***[0.08]	- 0.36***[0.05]	- 0.20***[0.06]
Firm created in Q2 (ref. Q1)	0.10* [0.06]	0.09* [0.05]	0.11***[0.04]	0.05 [0.05]
Departmental labour market tightness	- 0.62*** [0.23]	- 0.94***[0.20]	- 0.28*** [0.11]	- 0.30***[0.07]
Constant	1.04***[0.23]	- 0.21 [0.21]	- 0.03 [0.17]	0.297*[0.160]
Number of observations	4,230	3,355	5,588	7,300
Log likelihood	- 7,691	- 6,170	- 7,437	- 12,037
Pseudo R^2	0.116	0.135	0.087	0.077

Note: Observations weighted using the *SINE* survey's *poidsini* variable (Box 1). Coverage and control variables not shown in this table: see table 4 in the main text. For the 2006 cohort, the survival variable was calculated after three years, as the five-year data was unavailable.
Reading note: French citizens had a higher probability (significant at the 1% or 5% threshold, depending on the cohort of ACCRE recipients) than foreigners to participate in the ACCRE programme
Source: Insee, *SINE* surveys 1994, 1998, 2002 & 2006.

Effectiveness of public support for R&D and entrepreneurship

Comment *on the papers "The effect of R&D subsidies and tax incentives on employment: an evaluation for small firms in France"*[(i)] *by Vincent Dortet-Bernadet and Michaël Sicsic, and "Do public subsidies have an impact on start-ups survival rates? An assessment for four cohorts of firms set up by previously unemployed entrepreneurs in France"*[(ii)] *by Dominique Redor.*

Pierre Mohnen *

Abstract - The papers by Dortet-Bernadet and Sicsic and by Redor in this issue examine respectively the success of R&D financial support programs in stimulating private R&D and the success of subsidized start-ups for the unemployed in creating long-lasting firms. Both papers focus on small French firms. Both programs are found to suffer from a deadweight loss. This comment discusses the results obtained and the policy conclusions that can be drawn from them. It is argued that the deadweight loss is in part unavoidable but that there are ways to limit it, for instance by using a policy mix of R&D tax incentives and subsidies, favoring tax incentives for small firms and subsidies for large firms. It is also recalled that a policy ought to be evaluated from various perspectives. Besides R&D additionality and firm survival a full cost benefit analysis would also consider R&D externalities, firm retention and decrease in unemployment.

JEL codes: O31, J68
Keywords: R&D tax incentives, R&D subsidies, start-up subsidies, policy evaluation

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.

Translated from:
(i) « L'effet des aides à la R&D sur l'emploi : une évaluation pour les petites entreprises en France »
(ii) « L'aide à la création d'entreprises a-t-elle un impact sur leur survie ? Une évaluation pour quatre cohortes d'entreprises créées par des chômeurs en France »

** Université de Maastricht et UNU-MERIT (mohnen@merit.unu.edu).*

Received May 28, 2017; accepted after revision June 1st, 2017
DOI: 10.24187/ecostat.2017.493s.1911

The two excellent papers presented in this issue of *Economie et Statistique* provide some interesting and useful contributions to the discussion on the effectiveness of financial support towards R&D on the one hand and business start-ups for the unemployed on the other hand. Vincent Dortet-Bernadet and Michaël Sicsic examine the effectiveness of direct and indirect public support for investment in research and development by very small firms in France. Dominique Redor evaluates the success of another French policy aimed at small firms, namely the creation of start-ups by people who were previously unemployed.

Since the marginal productivity of capital is declining, a growth in GDP/capita can only be achieved by a more efficient utilization of resources, the introduction of new and more efficient technologies, or the production of new intermediate or final demand products that in the end achieve higher outputs (or consumer utility) with fewer resources. In this process of Schumpeterian creative destruction new products replace old products and newcomers replace incumbents. To some extent all this can happen endogenously by the mere forces of the market. However, irrespective of the social turmoil this process of creative destruction can cause, the market by itself might not reach the optimal growth and economic development because of market failures. Entrepreneurs following the invisible hand might not take externalities into account, for instance not spending sufficiently on R&D from a societal point of view or innovating in polluting technologies. Because of coordination failures, private entrepreneurs who fail to consult with each other may put unnecessary strains on some resources preventing other societal goals to be achieved, not speaking of moral hazard or intentional anticompetitive behavior. And finally, because of the public good nature of knowledge, innovators may be reluctant to provide fund providers with the required information to justify their lending, this being particularly the case for small firms and startups, which do not have the collateral or other guarantees to back up their financial requests.

A technical difficulty in the evaluation of the effectiveness of these policy interventions is the endogeneity of aid recipients and the self-selection into aid. Indeed, those firms that receive R&D subsidies or tax incentives as well as the unemployed who benefit from the ACCRE support might be inherently more hard-working, productive or efficient than those that receive no support. The former may also be more likely to apply for such aid in the first place. The superior economic performance of aid recipients might therefore not only - if at all - be due to the support itself. The econometric difficulty is to filter out these two sources of bias. Besides addressing similar issues and focusing on a similar sub-population of firms, both papers are careful and skillful in properly handling the endogeneity of public support. They slightly differ in the way they handle the endogeneity problem.

The comment is organized as follows. First, we summarize the two papers regarding their method of analysis and the results obtained. We then proceed to critically discuss them and compare them with other studies in the literature. We conclude with some policy recommendations in light of the conclusions reached in the two studies.

Summary presentation of the two papers

Vincent Dortet-Bernadet and Michaël Sicsic evaluate jointly the direct and indirect support for R&D employment in small and medium sized French enterprises (SMEs). Many papers have analyzed the effectiveness of R&D tax incentives and direct R&D subsidies in France and other countries (see the reviews by Ientile & Mairesse, 2009; Köhler et al., 2012; European Commission, 2014; Zuñiga-Vicente et al., 2014). This study has three particularities. First, it includes the very small enterprises, i.e. those with fewer than 10 employees and less than € 2 million of turnover and of assets, whereas most studies based on R&D survey data are biased towards large firms. The very small enterprises make up two thirds of the panel. Secondly, it merges many databases, namely those of the R&D tax credit (CRI), the young innovating enterprises program (JEI), the R&D survey, the list of accredited enterprises from the Ministry of Research, and various fiscal, social, financial and register data from the French Statistical Office. This data effort provides a unique, almost exhaustive, sample of SMEs and very small French firms. Thirdly, it examines at the same time direct and indirect R&D support measures, whereas most previous studies examined only one of the two types of support measures, thereby omitting a potential R&D determinant.

The authors carefully construct the estimation sample by first matching every firm that received financial support for R&D at least once between 2003 and 2010 with three firms of the same age that did not receive financial support in the corresponding period and that had similar probabilities of receiving support in 2003 or in 2007. They then estimate by system GMM a dynamic R&D labor demand equation as a function of present and past levels of turnover and the relative cost of highly qualified labor compared to other types of labor. The latter is then instrumented by multiplying official changes in the R&D tax credit by the R&D labor share just before the change. Finally, first differences in R&D labor due to changes in the relative R&D labor cost and changes in turnover brought about by changes in the R&D tax credit are computed and subtracted from those of matched firms that did not benefit from R&D financial support. Knowing the amount spent on supporting R&D and the average cost of R&D labor, the change in R&D labor supported by the incentive measures can be calculated and the change in R&D labor not supported by government can be obtained residually.

The authors come to the conclusion that the R&D support increased R&D labor but with partial crowding out. In other words, the number of R&D workers that were financed by the private sector decreased. Part of the financial support received by private firms from government was used to decrease their own investment in R&D labor. The statistically significant decrease is especially visible after the 2008 reform in the R&D tax credit, which replaced the increment-based by a volume-based R&D tax credit system. For the firms that had existed since 2007, only 24% of the increase in financial support for R&D in its various forms was devoted to new hires of R&D workers.

Dominique Redor examines the effect of the French policy ACCRE, which aims to assist the unemployed to start or take over a business on their survival five years later. The econometric analysis is based on four cohorts of such enterprises (1994, 1998, 2002 and 2006) obtained from a stratified sample survey on newly created enterprises. The underlying model is based on a simultaneous bivariate probit model estimated by maximum likelihood: the first equation estimates the determinants of getting financial support to start a new business, the second one explains the survival 5 years later as a function of, among others, the fact of having benefitted or not from the subsidy. Two exclusion restrictions are used in the selection equation: the fact that the enterprise was created in the second quarter (implying that the request for aid was introduced in the first quarter, where funds are typically more largely available than towards the end of the year), and the tension on the labor market (i.e. the ratio of job vacancies to the number of unemployed). Observable individual characteristics of the entrepreneur and the created enterprise are controlled for. Both exclusion restrictions are significant, although only at a 10% level of significance for the date of creation of the enterprise. When the two probits are estimated separately ACCRE has a positive effect on the probability of survival 5 years later; when they are estimated jointly the effect of ACCRE disappears. The endogeneity does not come from common unobservable determinants for the two endogenous variables as the correlation between the error terms of the bivariate normal distribution is not significant.

The conclusion is that ACCRE is not effective in creating enterprises that last at least 5 years. The result is found to be robust with respect to different definitions of financial means, different lengths of survival and different types of beneficiaries (inactive people, unemployed of less than one year and unemployed of more than one year). The only significant effect of ACCRE is for the year 1998 when all categories of unemployed are included in the sample. It is suggested that this exceptional result might be due to the smaller number of beneficiaries of ACCRE in 1998 due to less favorable terms of support offered in that year. The other cohorts benefitted from a more generous support. Maybe this could explain the different result for 1998, although on each subsample of beneficiaries the effect was insignificant even in that year.

Discussion of the results

Although Dortet-Bernadet and Sicsic's results contradict the findings of previous evaluations of the R&D tax credit system in France, the results are not entirely surprising.

First, the estimates combine the intensive margin (increase in R&D intensity for R&D performing firms) and the extensive margin (increase in the number of R&D performing firms). On the extensive margin more firms have decided to start doing R&D especially in the year of the 2008 reform of the R&D tax credit (Bozio et al., 2014). To enter the R&D game firms need to incur sunk costs in addition to the fixed and variable costs of R&D. Arqué-Castells and Mohnen (2015) estimate these sunk costs to be as large as 1% of total sales and to be higher for small firms than for large firms. The sample here is mainly composed of very small firms, which were not captured in previous studies based on the R&D survey database. Small firms may not make enough profits to claim any R&D tax credit. It is only since 2010 that SMEs are able to receive immediate refunds for unutilized credits (EU Commission, 2014, Annex Country Fiches). Finally, we should not ignore the compliance costs of applying for R&D tax credits. The 2008 reform facilitated the application for R&D tax credits, yet for very small firms these compliance costs, which on average have been estimated at 7% in Canada and the Netherlands, may be at least twice to three times as high for very small firms. All this to argue that starting to engage in R&D and applying for the first time for R&D tax credits, which probably occurred at a higher rate in the year of the reform, carries with it additional costs that reduce the amount left over to hire R&D workers. The decline in privately financed R&D labor is the highest in 2008. It would have been nice to show the difference in effectiveness of R&D support at the intensive and the extensive margins.

Second, the partial crowding out may also be related to the gradual introduction of volume-based R&D tax credits in France after 2004 and the full substitution of increment-based by volume-based R&D tax credits after 2008. The deadweight loss, i.e. the funding for R&D that would have been done anyway, is a typical phenomenon of volume-based schemes. R&D has been shown to be persistent (Peters, 2009, Arqué-Castells & Mohnen, 2015). Hence once in the R&D game, firms tend to remain in the R&D game. In that case, a good deal of the financial support for R&D could be done without it. Firms would continue spending on R&D anyway. In increment-based R&D tax credit schemes only an increase in R&D is eligible for R&D tax credits and only part of the increase is financed by the policy. It is therefore not surprising that Mairesse-Mulkay (2004) and Duguet (2012) for France find a strong additionality for the period prior to 2003, where France had only incremental R&D tax credits, while Mulkay and Mairesse (2013) report a bang for the buck (BFTB) of 0.7 under the regime of level-based R&D tax credits after 2008. The BFTB found in most other studies where level-based tax credits dominate is below 1 (see Ientile & Mairesse, 2009 ; Caiumi, 2011 ; European Commission, 2014). The deadweight loss holds especially for large firms ; but even for small firms, like those in the present sample, the phenomenon may occur.

Third, as the authors admit, there may be partial crowding out in the short run, as firms incur sunk costs and adjustment costs, but in the long run there may be additionality. Given the high autoregressive coefficient in the estimated dynamic labor equation, the long-run elasticity of R&D labor to the relative wage of highly qualified labor could be pretty high. This reversal could also explain the insignificant decrease in privately financed R&D in 2010 as opposed to significant negative signs prior to 2010 for R&D support increases with respect to 2003 (table 4).

Dortet-Bernadet and Sicsic combine matching, difference in differences and structural modeling approaches, but in all stages they only control for observables. It may be that unobservables drive the firms to apply for R&D tax credits and that those same unobservables influence the demand for R&D labor. The post 2008 world financial crisis is one of those variables that may have affected both the application for government support and the amount of R&D expenditure. Redor in his paper allows for the presence of such unobservables through the correlation in the error terms of the selection and survival equations. In his case the correlations are not significant in any of the four cohorts of firms, implying that the survival is conditionally independent of selection into ACCRE support. Given the positive evaluation of the ACCRE program obtained in the study by Duhautois, Désiage and Redor (2015) for the year 1998 it would be interesting to redo the propensity score based matching analysis on the four cohorts, where no assumptions are made regarding the functional form of the specification nor the distribution of the error terms. True, no account would then be taken of the presence of unobservables, but they do not seem to matter anyway.

International comparison

Most, if not all, countries have some policies in place to support R&D, see OECD (2017) for a recent review of these measures. Although there is a huge heterogeneity across countries in the way tax incentives and subsidies are organized, a few stylized facts emerge. Most countries have found ways to let firms use their tax credits even in the absence of payable taxes. Most countries give higher R&D tax credits to small firms. More and more governments shift to the volume-based R&D tax incentives because they are easier to manage, they do not encourage a see-saw behavior in R&D expenditure in order to capture the most out of the tax supports and they provide continuous support even in the absence of accelerating R&D investments. In principle, R&D tax incentives are neutral, although in many cases additional support is provided for collaborative research with universities. Some countries like Germany, Finland and Luxemburg have no expenditure-based R&D tax support, favoring subsidies to tax support. R&D subsidy programs are much more diversified and can to some extent be geared to projects with higher expected social returns.

The empirical studies on the effectiveness of R&D tax support (Ientile & Mairesse, 2009 ; European Commission, 2014) concur that R&D tax incentives are effective, i.e. they stimulate additional R&D. However, volume-based R&D tax incentives are rather inefficient in terms of cost-benefit analysis. The deadweight loss can be severe: under the assumption that firms will not cut their spending on R&D because of sunk costs, Lokshin and Mohnen (2012) evaluated the additional R&D originating from one Euro of R&D tax support in the Netherlands at 0.42 Euro. Mulkay and Mairesse (2013) for France found a long-run budget multiplier of 0.72. In their survey of the literature Zuñiga-Vicente et al. (2014) found mixed evidence regarding the crowding out versus additionality of direct R&D support, although the latest evidence seemed to be tilting more towards additionality.

Subsidized start-ups for the unemployed is a policy that has also been introduced in a number of other countries. As Caliendo (2016) reports in his survey of the literature, these policies have generally been effective in terms of job creation, but not so successful in creating long-lasting firms.

Policy recommendation

Direct and indirect R&D support might not be immediately successful in generating additional R&D employment because other costs need to be covered in the short run, and because support goes to R&D expenditure that would have been undertaken anyway. This deadweight loss is to some extent unavoidable, except if the subsidies are restricted to additional R&D. The increment-based R&D tax incentives have, however, proved to be costly to administer, for the firms and for the government, and limited in their ability to generate a lot of new R&D. The question is whether these inefficiencies are outweighed by the externalities generated by the additional R&D. There are also ways to limit the inefficiencies. Large R&D performers need less help because they have other ways to generate money to finance their R&D projects: retained earnings, easier access to external financing and to venture capital markets. Another way would be to do a smart policy mix, giving easily obtainable tax credits to small firms and start-ups, and direct grants and subsidies to big projects, possibly collaborative projects involving big and small actors, private firms and universities, where a sound cost-benefit analysis has indicated the presence of social benefits in the long run. The idea is that small firms are the most affected by the asymmetric information problem and the lack of financial capital, whereas large firms are more likely to create R&D spillovers (Bloom et al., 2013). Finally, these financial support policies could be complemented by public procurement, protection of intellectual property rights, the creation of a venture capital market and a readiness to take risks and to accept failure.

The other thing to keep in mind is that a policy may have several effects and ought to be evaluated from various perspectives. The generous R&D tax credit policy after 2008 was also intended to keep R&D facilities in France instead of seeing R&D labs and personnel move to other countries. Hence even if the policy was not very effective in stimulating private R&D it may have been effective in retaining R&D in France. Likewise, subsidizing start-ups for the unemployed might not be very successful in creating long-lasting firms, but it may give the beneficiaries the chance to gain experience and then be in a better position to find a new job or to start a new business. If the aim is to create new firms that have a chance to survive a long period of time, it would make sense to be more selective in providing subsidies and to accompany the subsidies with training and mentoring. But, the start-up subsidies for the unemployed are also geared towards fighting unemployment, probably even more than towards creating new firms. Instead of examining the survival of newly created firms 5 years after, the employment record of the erstwhile unemployed five years after they received the ACCRE support plus and of the newly hired workers in the process, might be an alternative performance worth examining. ☐

BIBLIOGRAPHY

Arqué-Castells, P. & Mohnen, P. (2015). Sunk costs, extensive R&D subsidies and permanent inducement effects. *Journal of Industrial Economics*, 63(3), 458–494.

Bloom, N., Schankerman, M. & Van Reenen, J. (2013). Identifying technology spillovers and market rivalry. *Econometrica*, 81(4), 1347–1393.

Bozio, A., Irac, D. & Py, L. (2014). Impact of research tax credit on R&D and innovation: evidence from the 2008 French reform. Banque de France, Document de travail N° 532.

Caiumi, A. (2011). The Evaluation of the Effectiveness of Tax Expenditures- A Novel Approach: An Application to the Regional Tax Incentives for Business Investments in Italy. OECD Taxation Working Papers 5, OECD.

Caliendo, M. (2016). Start-up subsidies for the unemployed: Opportunities and limitations. IZA World of Labor, 200.

Duguet, E. (2012). The effect of the incremental R&D tax credit on the private funding of R&D an econometric evaluation on French firm level data. *Revue d' économie politique*, 122(3), 405–435.

Duhautois, R., Désiage, L. & Redor, D. (2015). Long Term Effect of Public Subsidies on Start-up Survival and Economic Performance: an Empirical Study with French data. *Revue d'Économie Industrielle*, 149, 11–42.

European Commission (2014). A study on R&D tax incentives: Final report. DG TAXUD Taxation Paper 52.

Ientile, D. & Mairesse, J. (2009). A policy to boost the R&D: Does the tax credit work? European Investment Bank paper 14(1).

Köhler, C., Laredo, P. & Rammer, C. (2012). The impact and effectiveness of fiscal incentives for R&D. Compendium on evidence of the effectiveness of innovation policy intervention. University of Manchester, NESTA.
http://www.innovation-policy.org.uk/compendium/

Mairesse J. & Mulkay, B. (2004). Une évaluation du crédit d'impôt recherche en France, 1980-1997. *Revue d'économie politique*, 114 (6), 747–778.

Mulkay, B. & J. Mairesse (2013). The R&D tax credit in France: assessment and ex-ante evaluation of the 2008 reform. *Oxford Economics Papers*, 65(3), 746–766.

OECD (2017). OECD review of national R&D tax incentives and estimates of R&D subsidy rates, 2016. deliverable 3.1 for TAX4INNO project 674888.

Peters, B. (2009). Persistence of innovation: stylized facts and panel data evidence. *Journal of Technology Transfer*, 34, 226–342.

Zuñiga-Vicente, J. A., Alonso-Borrego, C., Forcadell, F. J. & Galán, J. I. (2014). Assessing the effect of public subsidies on firm R&D investment: A survey. *Journal of Economic Surveys*, 28, 36–67.

High turnover among nursing staff in private nursing homes for dependent elderly people (EHPADs) in France: impact of the local environment and the wage

Cécile Martin * and Mélina Ramos-Gorand **

Abstract – The high turnover among nursing staff working in nursing homes for dependent elderly people (EHPADs) in France has negative consequences in terms both of cost and of quality of care for the residents. We study the causes of this staff turnover using the estimate from a *probit* model estimated on two samples, one of 5,478 nurses and the other of 13,444 nursing auxiliaries working in private EHPADs under open-ended contracts. The probability of the nurses and nursing auxiliaries leaving is significantly influenced by factors related to the environment around the employee's place of residence, computed at a highly disaggregated geographical level, including closeness to a hospital, competition between residential care facilities for elderly people, shortage of nursing staff, and attractiveness of the self-employed professional sector for nurses. The wage level, corrected for endogeneity, has a positive effect on the retention of nursing auxiliaries working in EHPADs, but it does not seem to have an influence in the case of nurses.

JEL codes : C25, I11, J63.
Keywords: staff turnover, nursing homes for dependent elderly people, care quality.

* *LEDa-LEGOS, Université Paris-Dauphine, Paris and Académie de Caen (cecile.martin@ac-caen.fr).*
** *Lab'Urba, Université Paris-Est Créteil and Caisse nationale d'assurance vieillesse (Cnav) (melina.ramos-gorand@cnav.fr).*

Acknowledgements: the authors would like to thank the French Health Ministry's directorate for research, studies and statistics (Direction de la recherche, des études, de l'évaluation et des statistiques - Drees), and the Centre Maurice Halbwachs for access to the data used in this study. They would also like to thank the two anonymous reviewers for their suggestions and remarks.

Received April 18, 2015; accepted after revision March 3, 2017 DOI: 10.24187/ecostat.2017.493s.1912
Translated from the original version: « *Turnover* élevé du personnel soignant dans les Ehpad privés en France : impact de l'environnement local et du salaire »

In view of the increasing medicalisation of nursing homes for dependent elderly people (*Etablissements d'hébergement pour personnes âgées dependantes*, hereafter EHPADs), use of nursing staff in such facilities should intensify in the coming years. Nurse and nursing auxiliary positions already accounted for 4 out of 10 full-time equivalent jobs (FTEs) in nursing homes for dependent elderly people in 2007 and in 2011[1]. The sector suffers from difficulties in attracting and retaining these professions, in particular in private facilities, and this has detrimental effects on the quality of the care. The mean exit rates were 61% for nurses and 68% for nursing auxiliaries in EHPADs in 2008[2]. Those rates are higher than in the hospital sector: for nurses, the turnover rates[3] were only 22% in non-profit private hospitals and 30% in for-profit private hospitals in 2011 (Loquet & Nagou, 2014). There are a variety of causes for early leaving of nurses and nursing auxiliaries working in EHPADs. They might prefer to work in a competing EHPADlocated in a more attractive geographical zone. They might also change the mode of practice of their profession and choose to work at a hospital, for a home care service provider, or as a self-employed professional. Barlet and Cavillon (2011) thus show that nurses working in facilities for elderly people are the category that change mode of practice most (on average, 6% per year, as against 3% for all nurses in the period from 2004 to 2008).

This high turnover among nursing staff can degrade the quality of care in EHPADs. Large numbers of leavers among nurses or nursing auxiliaries can lead to the EHPAD being temporarily understaffed, even though it is known that staff-to-resident ratio or staffing level is positively correlated to quality (Martin, 2014). High employee turnover can also cause interruptions in the continuity of care (Cohen-Mansfield, 1997), generate iatrogenic accidents due to prescription errors, and degrade the state of health of residents (Lerner *et al.*, 2014; Antwi & Bowblis, 2016). Finally, it prevents residents from forming relationships of trust with their nursing carers (Wiener *et al.*, 2009). As a former professional from the sector says, "*the high staff turnover of the sector has consequences: the new staff do not know the residents, their pathologies, their assistance needs, or their habits, and do not have the necessary training*" (Nénin & Lapart, 2011, p. 51 – translated from the French). In addition to detrimental consequences on quality, high employee turnover can also generate additional costs for the facility due to the need to recruit replacement staff, sometimes temporary staff. It can also lead to a reduction in productivity, given the time necessary for training new staff (Brannon *et al.*, 2002).

In this study, we analyse the determinants of nursing staff departures in private EHPADs in France. Controlling for the characteristics of the job, of the facility, and of the employees, we analyse, in particular, the effects of the salary and of the environment from different angles – nursing staff shortage in the geographical area, presence of a hospital, competition between facilities for elderly people, attractiveness of the self-employed professional sector – on the probability of nurses and nursing auxiliaries leaving. Our analysis is performed on the basis of two samples of staff working under open-ended contracts in private EHPADs in 2008: a sample of 5,478 nurses and a sample of 13,444 nursing auxiliaries. This study is original on three counts. Firstly, there is no other econometric study that examines the causes of staff turnover in EHPADs in France. Secondly, we take into account the endogenous nature of the wage, which has rarely been done in non-French research devoted to this subject. Thirdly, we incorporate numerous environmental variables making it possible to analyse the reasons why staff leave. For this purpose, we use highly disaggregated geographical data, compiled on the basis of the home address of each employee and computed at the level of the "local area"[4].

In the following section, we present the imperfections of non-French studies analysing the causes of nursing staff turnover. We then describe the variables and the data used, and our approach. Finally, we analyse the results before giving some conclusions.

1. Source: http://www.data.drees.sante.gouv.fr (30 March 2017).
2. The exit rate for nurses (or for nursing auxiliaries) corresponds to the ratio of the number of nurses (or of nursing auxiliaries) having left the facility during the year to the number of positions for nurses (or for nursing auxiliaries) in the facility at the end of the year. The calculations were made on the basis of two samples of 1,393 nursing homes for dependent elderly people (exit rate for nurses) and of 1,392 nursing homes for dependent elderly people (exit rate for nursing auxiliaries). Source: DADS 2008 *(Insee) and* enquête EHPA 2007 *(Drees), authors' calculations.*
3. In this study, the turnover rates correspond to the averages between the entry rates and the exit rates. They can therefore be very different from the exit rates if the staff who leave the facility are not replaced within the same year.
4. See below.

Analysis of nursing staff turnover in non-French research

Various studies using American data have aimed to identify the causes of employee turnover in nursing homes, i.e. in nursing homes for elderly people. However, most of them suffer from two limitations.

Firstly, many of these studies are conducted at the scale of the facility, and the environment-related variables are therefore often poorly incorporated. For example, Castle and Engberg (2006) sought, using a negative binomial regression, to show lower turnover among nursing staff in nursing homes located in rural areas. Indeed, nurses and nursing auxiliaries enjoy fewer job opportunities in such areas. However, they observed the reverse effect, which was probably due to the fact that they took into account the nursing home's location rather than the location of the employees' place of residence. However, employees working in such rural nursing homes may live in towns, so that working in rural areas entails long distance commuting, which is a source of dissatisfaction. Using a multinomial logistic model, Brannon *et al.* (2002) also studied the factors associated with a low turnover of nursing assistants working in nursing homes (lower than 6.6% in 6 months) and the factors associated with a high turnover (higher than 64%). They incorporated some variables related to the market environment, such as the concentration of nursing homes and the local unemployment rate. However, since those variables were constructed again on the basis of the location of the facility and not that of the employees' place of residence, no significant effect was observed.

Secondly, the role of wages or salary is little or poorly studied. It is not always added to the variables explaining staff turnover, like in the work by Brannon *et al.* (2002) and by Castle and Engberg (2006). That omission skews the effects of the environmental variables if the wage adjusts to the local difficulties. On the basis of a qualitative survey conducted on 345 nursing assistants in 18 nursing homes in the USA, Dill *et al.* (2013) showed a positive impact on the employees' intention to stay at the facility when they felt financially rewarded. However, the actual impact of the wage level on staff turnover remains difficult to determine. Some authors have incorporated this wage variable, such as Temple *et al.* (2009) who studied the factors related to a low turnover and to a high turnover of nursing assistants (characterised by the first and last quartile of their sample) using a multinomial logistic regression. Wiener *et al.* (2009) also studied the effect of the wage and the environment on the seniority of nursing assistants working in nursing homes using an ordinary least squares regression. Those authors thus showed that a rise in salary apparently reduces the probability of having a high turnover rate and increases seniority in the nursing homes. However, they did not take into account the possible endogeneity of the wage. Indeed, unobserved factors, such as experience or the propension of the nursing staff to protest, can be correlated both to the wage and the exit rate. Furthermore, someone who changes jobs regularly may have a lower wage because they do not benefit from seniority bonuses.

In this study, we test whether the relationships empirically observed in those studies apply to the French case, by making corrections or adjustments to overcome those two stumbling blocks. Wage endogeneity is corrected for by instrumentation, as performed for the United States by Baughman and Smith (2012) who, using a duration model, analysed the determinants of the probability of nursing assistants leaving the facility at which they work; they observed that the instrumented wage had a negative effect on the probability of leaving, whereas it was not significant when it was not corrected for endogeneity. However, few variables related to the environment are taken into account in their study, and those that are (unemployment rate, average wage in other professions) are calculated at the State level. Unlike those authors, we include several environment variables constructed at a highly disaggregated geographical scale. We use individual data and construct the environmental variables on the basis of the employees' home addresses rather than on the basis of location of the facility.

Variables and data used

We analyse the probability of nurses and nursing auxiliaries leaving their facility (variable "*Départ*"). As with various theoretical models (Cohen-Mansfield, 1997), we assume that nursing staff leaving is related to two types of factors: factors that can generate job dissatisfaction, such as the employee's characteristics, and the characteristics of their job and of their facility; and factors that influence the decision to leave the facility, such as the local labour market. The variables used to study these determinants empirically are described below and summarised in Table 1.

Individual characteristics of the nurses and nursing auxiliaries

Barlet and Cavillon (2011) observed a relationship between the ages of the nurses and the probability of them changing mode of practice of their profession: this probability seems to be higher before the age of 35 and lower after the age of 45. Retirement, and sometimes early retirement, can however take place after the age of 50. To take account of these reasons for leaving, we include an age variable (*Âge*) in the form of a categorical variable: aged under 35, from 35 to 45, from 45 to 50, from 50 to 55, and over 55. We also add a gender variable (*Homme*). As in Wiener *et al.* (2009), we expect this variable to have a positive effect on the probability of leaving. Finally, the longer the journey (*Distance*) between the workplace and the employee's place of residence, the higher the job dissatisfaction. Since the effect of this variable is doubtless not linear, we categorised it into four modalities: distance less than 5 km, from 5 to 10 km, from 10 to 20 km, and greater than 20 km. We did not have any data on the employees' marital status or the number of children they have. However, Zhang *et al.* (2014), on the basis of data gathered from 1,589 employees working in nursing homes in the United States, did not observe any link between those variables and the intention of staff to leave their facility.

Characteristics of the job and of the facility

Various sources of dissatisfaction are related to employment conditions and can, as a result, constitute levers for nursing home directors who wish to reduce staff turnover (Anderson *et al.*, 1997). According to the wage efficiency theory (Stiglitz, 1974; Salop, 1979), a high wage can be efficient insofar as it encourages staff to stay at the facility where they work, and thus makes it possible to reduce employee turnover. In its Annual report for 2012, the Le Noble Age Group thus indicated it had "*put in place [...] favourable pay management [...] in order to limit the risk of understaffing and of increasing the staff turnover rate*" (p. 41 in French). We thus include the logarithm of the employee's net annual wage (*Log_Salaire*). The possibility of working nights (*Nuit*) in the nursing home can also have an effect on dissatisfaction; that effect can be positive if night work is an obligation that is not desired by the employee, or negative when it is a possibility that can be chosen by the employee. Care quality also probably has an impact on job dissatisfaction, and thus on the probability of employees leaving (Irvine & Evans, 1995). We chose to include, as a quality variable, a ratio related to the staffing level or staff-to-resident ratio of the EHPAD where the employee works. This variable is a good proxy for quality because it has a significant effect on the well-being and on the residents' health status (Spilsbury *et al.*, 2011). In addition, a low staffing level has an impact on working conditions because it necessarily induces a heavier workload. In order to take account of the diversity of the EHPADs' staffing needs, we compute an optimum theoretical staffing level (N^*) as a function of the dependency rating categories of the residents (these categories are related to the *groupes iso-ressources* or *GIR* (groups based on needs and dependence level and going from *GIR 1* for the most severely dependent to *GIR 6* for residents who are not dependent). For this purpose, we used the recommendations of the *Plan Solidarité Grand Âge 2007-2012* (a national solidarity plan for the very old): on a daily basis, a *GIR 1* dependent person needs 1 FTE, a *GIR 2* person needs 0.84 FTE, a *GIR 3* needs 0.66 FTE, a *GIR 4* needs 0.42 FTE, a *GIR 5* needs 0.25 FTE, and finally a *GIR 6* needs 0.07 FTE (Ratte & Imbaud, 2011). We then calculate the ratio of the actual staffing level (excluding administrative and corporate services employees) of the EHPAD (N) to the theoretical staffing level (N/N^*). Various authors have shown that staffing level could have a major impact on job dissatisfaction and on employee turnover in nursing homes in the United States (Temple *et al.*, 2009). However, some authors mention the problem of endogeneity that can arise by including this variable (Kash *et al.*, 2006). In the same way as for the wage, causality can work both ways: a large number of employees leaving can give rise to replacement difficulties and to a staffing level that is temporarily lower. However, the staffing level ratio (N/N^*) is calculated on the basis of all of the staff in contact with the residents, not only the nurses and nursing auxiliaries, and so the risk of endogeneity is lower. We check that the estimation of our model without this staffing level variable did not change the results obtained (cf. on-line supplement C1).

Certain factors specific to the facility's structure can have impacts on staff exits. Various American studies have thus shown that for-profit private nursing homes were confronted with higher staff turnover (Banaszak-Holl & Hines,

Table 1
Definition of the variables

Variable	Definition
Départ	Binary variable equal to 1 if the employee leaves the facility where they used to work during the year, 0 otherwise.
Individual characteristics	
Âge	Category variable indicating the age of the employee: aged under 35, from 35 to 45, from 45 to 50, from 50 to 55, or over 55.
Homme	Binary variable equal to 1 if the employee is male, 0 if they are female.
Distance	Category variable indicating the commuting distance between the employee's home and the EHPAD where they work: less than 5 km, from 5 to 10 km, from 10 to 20 km, or greater than 20 km.
Characteristics of the job and of the facility	
Log_Salaire	Logarithm of the employee's annual net salary, endogeneity corrected for by instrumentation.
Nuit	Binary variable equal to 1 if the employee may work nights in their nursing home, 0 otherwise.
*N/N**	Ratio of the actual staffing level of the EHPAD where the employee works to the calculated staffing needs (calculated on the basis of the degrees of dependency of the residents).
Statut	Binary variable indicating whether the private EHPAD where the employee works has non-profit or for-profit status.
N_Lits	Number of beds at the EHPAD where the employee works.
% GIR1	Proportion of residents with high dependency and rated as having the highest need for resources (*Groupe Iso-Ressources GIR 1*) at the employee's care home.
% GIR2	Proportion of residents with high dependency and rated as having the next highest need for resources (*GIR 2*) at the employee's care home.
% GIR3	Proportion of residents with medium dependency and rated as having the next highest need for resources (*GIR 3*) at the employee's care home.
% GIR4	Proportion of residents with medium dependency and rated as having the next highest need for resources (*GIR 4*) at the employee's care home.
% GIR5	Proportion of residents with low dependency and rated as having the next highest need for resources (*GIR 5*) at the employee's care home.
% GIR6	Proportion of residents with no dependency and rated as having the lowest need for resources (*GIR 6*) at the employee's care home.
Directeur > 2 ans	Binary variable equal to 1 if the director has been in office for more than two years in the employee's residential care home for dependent elderly people, 0 otherwise.
Environment	
Hôpital	Huff model coefficient measuring the attractiveness of hospital jobs on the employee.
‰Inf_libéraux_bv/pop	Ratio of number of self-employed nurses working in the local area of the employee's place of residence to its population (in thousands).
HHI_{Ehpa} *_bv*	Herfindhal-Hirschmann Index measuring the concentration of nursing home jobs in the local area in which the employee lives.
‰Inf_résid_bv/pop	Ratio of number of salaried nurses living in the local area in which the employee lives to the population (in thousands).
‰AS_résid_bv/pop	Ratio of number of salaried nursing auxiliaries living in the local area of the employee's place of residence to the population (in thousands).
Paris	Binary variable equal to 1 if the employee lives in Paris, and to 0 otherwise.
Île-de-France	Binary variable equal to 1 if the employee lives in the Ile-de-France Region (but outside Paris), and to 0 otherwise.
Instruments excluded	
EHPAD non régulé	Binary variable equal to 1 if the accommodation price of the facility at which the employee works is set freely, 0 if it is set administratively.
Taux d'occupation moyen_bv	Mean occupancy rate of the EHPADs in the local area in which the employee works.
Tarif dépendance moyen_dpt	Mean price for dependency care and support for heavily dependent persons (*GIR 1* and *GIR 2*) living in EHPADs in the *département* in which the employee works.
Groupe	Category variable indicating whether the facility where the employee works is independent or is part of a group of EHPADs comprising less than 5 care homes, from 5 to 20 care homes, or more than 20 care homes.

1996; Anderson *et al.*, 1997; Brannon *et al.*, 2002; Castle & Engberg, 2006). According to Wiener *et al.* (2009), "*some advocates argue that nonprofit nursing facilities are more mission driven than for-profit facilities and have higher staffing and other characteristics that may increase job tenure*" (p. 200). Staff may prefer to work in a facility whose purpose is purely societal rather than profit-motivated, i.e. in which the sole objective of the managers is to improve the residents' well-being. We therefore include a binary variable indicating whether the facility at which the employee works has a non-profit or for-profit private status (*Statut*). Facility size, i.e. the number of beds installed (*N_Lits*), seems to be less evident effect on staff leaving: while Castle (in 2005) observed a significant positive effect of the number of beds on the probability of having high turnover among nurses and nursing assistants, Wiener *et al.* (2009) did not obtain a significant effect of facility size on tenure of nursing assistants. We also add the proportions of residents in each GIR category at the employee's nursing home in order to take account of their degree of dependency (*%GIR1* to *%GIR6*). A binary variable indicating whether the director has been in office for more than two years (*Directeur > 2ans*) was also incorporated into the model. Like Castle (2005), we expect greater stability in management to reduce the probability of the employees leaving, for the three reasons mentioned by that author: a higher top management turnover could have a destabilising influence on the organisation, it could lower the nursing staff's commitment to the facility, and it could be detrimental to care quality and therefore increase dissatisfaction among staff.

Environmental factors

External job opportunities can encourage an employee who is not fully satisfied with their job to leave it. For nurses or nursing auxiliaries working in EHPADs, such opportunities can be of various types.

Firstly, they can change mode of practice of their profession by going to work in the hospital sector. Pay for nurses and nursing auxiliaries working in hospitals is close to what they earn in EHPADs. According to the French National Institute of Statistics and Economic Studies (Insee, 2011), in 2008, net annual wage of nursing, care, and social intermediate occupations (including nurses and also midwives, special needs professionals, medical technicians, and social workers) averaged 24,820 euros in for-profit private hospitals and 25,220 euros in non-profit private hospitals. In our sample, nurses working in private EHPADs earn annual net pay of 25,205 euros (cf. Table 2). If the hospital sector is attractive, then that would appear to be related more to the nature of the job. "*Working at a hospital is perceived as being more qualifying [...] and less limited because it seems easier to change department within the hospital than by coming from an EHPAD*" (Josse, 2012, p. 16 – translated from the French). The opportunity of finding a job at the hospital close to their place of residence could therefore be an encouragement to change jobs for nursing staff working at a EHPAD. We calculate Huff coefficients (cf. box 1) for measuring the attractiveness of a nearby hospital on each nurse and on each nursing auxiliary (*Hôpital*). We expect that this variable would have a positive effect on the probability of leaving. Nurses can also go to work in the self-employed sector. To take account of this reason for leaving, we include density per 1,000 people of self-employed nurses working within the area around each nurse's place of residence (*‰Inf_libéraux_bv/pop*). Another possible mode of exercising the profession is through home care services, but we did not have data enabling us to study the attractiveness of that sector.

Departures of staff can also be intra-sector, nurses and nursing auxiliaries then choosing to work in another nursing home for elderly people (EHPA) (cf. Box 2). The possibilities of being hired at another nursing home in the local area around the employee's place of residence can be assessed by the Herfindhal-Hirschmann concentration index (*HHI_{EHPA}_bv*). To compute this index, we use market share in terms of staff, i.e. share in terms of the theoretical staffing level needs (*N**) of each facility *i* relative to the needs of the other nursing facilities for elderly people in the local area. The Herfindhal index is then computed in the following manner: $HHI_{EHPA}\_bv = \sum_{i=1}^{n} s_i^2$, where $s_i^2 = N_i^* / \sum_j^n N_j^*$ and *n* is the number of nursing homes for elderly people in the local area of residence. The higher the Herfindhal index is and the closer it is to 1, the more the job offers are concentrated in a few nursing homes for elderly people, and the more difficult it then is for employees to change facility. Finally, we add a variable of presence of the nursing staff in the employee's local area of residence, computed as the ratio of the number of salaried nurses (or nursing auxiliaries) domiciled in the local area to the population (in thousands of people) (*‰Inf_résid_bv/pop et ‰AS_résid_bv/pop*). In France, the geographical distribution of nursing staff appears to be uneven. The densities of nurses are, for example, much higher in the southern regions than in the northern regions of France (Barlet & Cavillon, 2011). We expect that this variable would have a negative effect on the probability of nursing staff leaving, or, in other words, a positive effect on the shortage of nursing staff. Such localised staff shortages might indeed explain certain difficulties encountered for retaining staff working in EHPADs, for two reasons. Firstly, an overall shortage of nursing staff might reveal unattractive characteristics of the local area (rural nature, high cost of living, etc.) that might encourage staff to go and live and work in other areas. Secondly, on a market highly constrained by the labour supply, immediate job opportunities for nurses or nursing auxiliaries are numerous and can encourage them to leave their jobs if they are not satisfied with them.

Two binary variables indicate whether or not the employee lives in Paris (*Paris*) or in the Paris Region but excluding Paris (*Île-de-France*) were also included so as to take account of the specificities of those areas.

We choose to construct various environmental variables at the scale of the "local area of residence" ("*bassin de vie*"), which is defined by Insee (2003) as being "the smallest area within which the population has access both to infrastructures and amenities, and to jobs". From 1999 to 2012, there were 1,916 such local areas of residence in France. Although the scale of the "employment area" ("*zone d'emploi*") is more usual for this type of study, finer zoning per local area of residence seemed to us to be more appropriate here. 67% of the nurses and 70% of the nursing auxiliaries in our sample worked in a EHPAD that was located in the local area of their place of residence. This zoning was also constructed so as to qualify the predominantly rural area better. One of our hypotheses is that facilities located in rural areas have more difficulty in recruiting and retaining their staff, because of the small workforce in these areas. However, we tested the robustness of the results by comparing those obtained by regressions conducted using environmental data constructed at the scale of the employment area (cf. on-line supplement C2).

Box 1

THE HUFF MODEL COEFFICIENT

The Huff model is a gravity-based model commonly used in geography (Pumain & Saint-Julien, 2010). We use it in this study to measure the attractiveness of hospitals (public and private hospitals) on nurses and nursing auxiliaries. We consider that, for these professionals, the longer the distance between the place of residence and the hospital, the less the hospital is attractive. Furthermore, the higher the number of nurse or nursing auxiliary positions (*Postes*) at the hospital, the more the hospital is attractive. Formally, a professional living in a local district or "*commune*" *i* is attracted by the hospitals located in another *commune j* proportionally to the number of hospital positions corresponding to him or her in *commune j*, but inversely proportional to the square of the distance between *i* and *j*:

$$A_{ij} = \frac{Postes_j}{D_{ij}^2}$$

where A_{ij} is the attractiveness of the hospital positions of *commune j* on the professionals living in *commune i*; $Postes_j$ is the number of hospital positions in *commune j*; and D_{ij} is the distance between *communes i* and *j*.

By convention, when a *commune* has a hospital, the distance between the professionals living in that *commune* and the hospital is 1 km. When the distance between *communes i* and *j* is greater than 250 km, the attractiveness is considered to be zero.

For each professional living in *commune i*, we are thus able to compute the coefficient (PR_i) defined by Huff (1964), by summing all of the attraction indicators A_{ij} that we divide by 1,000 to reduce the magnitude. We thus obtain a synthetic indicator of the attractiveness of the hospitals of the surrounding *communes j* on the professionals living in *commune i*:

$$PR_i = \frac{1}{1000}\sum_j A_{ij}$$

Databases used

The data on turnover among nurses and nursing auxiliaries working at EHPADs, and on their wages and ages, come from the *Déclarations annuelles de données sociales* (*DADS*, which are administrative data based on annual declarations that employers are bound to fill in and return). All facilities must supply information about each of their employees annually to the *Caisse nationale d'assurance vieillesse* (Cnav, France's national old-age insurance fund): net wage, FTE, age, type of contract, whether they leave during the year, etc. Those data are then re-processed by Insee. They also make it possible to have precise information on the place of residence of each employee, which is a real asset for this type of study. We are thus able to construct environmental variables on the basis of the employee's place of residence, and not merely their place of work. We use the 2008 *DADS* declarations so that we can match up that data with the *EHPA 2007* database on nursing homes for elderly people compiled by the French Health Ministry's directorate for research, evaluation and statistics (*Direction de la recherche, des études, de l'évaluation et des statistiques – Drees*).

The *EHPA* survey (a survey on nursing homes for elderly people) is conducted every four years on all of the residential facilities for elderly people (EHPADs, retirement homes, long-term care units, etc.). 79% of all such facilities in France responded to that survey in 2007 (Perrin-Haynes, 2010). It includes various questions on how the facility operates (prices, number of places, status, etc.), on the employees, on the residents, and on the buildings at 31 December 2007. It therefore provides us with details about most of the variables related to the characteristics of the job and of the facility.

The identifier of the EHPAD that is used in the *DADS* data is the "*Siret*" registration number of the facility, whereas it is the "*Finess*" registration number in the *EHPA* survey. An extraction from the *Finess* database in 2007 enables us to have correspondences between these identifiers and to merge the two databases. However, several of these correspondences were missing, which led us to exclude facilities. The 2007 *Finess* extraction also enables us to obtain the codes of the "*communes*" of the facilities, such codes being necessary in order to measure the distance between the employees' place of residence and their place of work.

The data on the self-employed nurses working the area come from Insee's *base permanente des équipements (BPE) 2007* (2007 permanent facilities database).

Model estimated and study sample

The model estimated

We estimate the probability of a nurse or of a nursing auxiliary leaving their facility using a *probit* model, and correcting for the endogeneity of the wage variable. The model estimated is as follows:

$$y_i^* = \eta + x_i\alpha + \omega_i\beta + \mu_i$$

$$\omega_i = \delta + x_i\Pi_x + z_i\Pi_z + \nu_i$$

where $i \in [1\,;\mathrm{N}]$, N corresponding to the number of nurses or of nursing auxiliaries studied, ω_i is the logarithm of the wage of employee i, considered to be endogenous, z_i is a vector of $1{*}k_z$ instrumental variables, x_i is a vector of $1{*}k_x$ exogenous variables related to the individual characteristics of employee i, to environmental factors defined on the basis of their place of residence, to their employment conditions, and to the characteristics of the facility at which they work. The error terms (μ_i, v_i) follow a multivariate normal distribution of null expectation.

We do not observe the latent variable y_i^*, but rather we observe a dichotomous dependent variable:

$$Départ = \begin{cases} 0 & \text{si} \quad y_i^* < 0 \\ 1 & \text{si} \quad y_i^* \geq 0 \end{cases}$$

The model is estimated by maximising a likelihood function. η, α and β correspond to the constant term and to the vectors of the parameters of the model. δ, $\prod_x$ et $\prod_z$ are the constant term and the vectors of the parameters of the wage instrumentation equation. Those parameters were estimated jointly with η, α and β; if the instruments are specified correctly, this method of correcting for endogeneity makes it possible to obtain unbiased estimators (Cameron & Trivedi, 2009).

Various excluded instruments were incorporated into the first-stage estimation (cf. Table 1). They were assumed to have an effect on the wage variable, but not to have any direct effect on the probability of nursing staff leaving.

Firstly, we used a binary variable indicating whether or not the accommodation price of the facility where the employee works is set freely, i.e. not regulated (*EHPAD non régulé*). Freedom to set prices in private EHPADs, and thus to pass on the wage costs, varies depending on the mode of regulation of the facility[5] (cf. Box 2). The facilities whose accommodation prices are not regulated should be able to adjust the salaries in such a manner as to recruit and retain their staff more easily at a desired staffing level or staff-to-resident ratio. However, such EHPADs can be constrained in setting their accommodation prices by the degree of competition on the market and by the solvency of the demand in their geographical area. When the market share of an EHPAD in any given geographical area is small, that care home is a price taker: it cannot adjust its price in response to a necessary adjustment in the salaries. Conversely, it is easier for a single facility or for a facility that has a large market share to adjust its price and therefore the pay

5. Since regulation of accommodation prices concerns to a greater extent nursing homes for dependent elderly people that are not for profit (cf. Text Box 2), it is probable that the variable "EHPAD non régulé" *also has a direct effect on the probability of staff leaving as a result of the status. However, this effect was controlled because a status variable (*statut*) was included in the main regression. This made it possible to guarantee the condition for exclusion of the* EHPAD non régulé *variable.*

Box 2

ADJUSTMENT OF PAY AND REGULATION OF PRICES IN EHPADS

Residential care homes for elderly people (*Etablissements d'hébergement pour personnes âgées*, EHPAs) essentially comprise nursing homes for dependent elderly people (*Etablissements d'hébergement pour personnes âgées dependantes*, EHPADs), but they also include long-term care units (*Unités de soins de longue durée*, USLDs) which are highly medicalised, and, at the other end of the scale, non-medicalised facilities such as retirement homes or sheltered housing for elderly people who are not dependent or not very dependent but at which nursing staff work. In 2007, EHPADs (for dependent elderly people) accounted for 67% of EHPAs (for all elderly people) and for 75% of EHPA places (Prévot, 2009).

Unlike in public facilities where wages are set according to a national scale, directors of private nursing homes for dependent elderly people are constrained only by specific collective bargaining agreements, and even often only for a minimum pay and not for a maximum. In theory, they could therefore raise the pay levels of their employees so as to compensate for the benefits sought through changing jobs, thereby retaining them better. However, such a rise might be limited for budgetary reasons whenever prices cannot mirror changes in salaries, in particular because of being regulated by the public authorities.

EHPADs' pricing system is ternary: three daily rates are set corresponding to the three main activities of EHPADs, namely accommodation, dependency support, and nursing care. The costs are distributed between these three categories according to legally imposed distribution keys. The cost of nursing staff is, for example, in theory covered by the nursing care rate, while the cost related to the nursing auxiliaries is covered in part by the nursing care rate and in part by the dependency support rate. However, the partitions between the rates subdivisions is not always hermetically sealed. Martin (2014) observed that nursing auxiliaries' level of pay could have a positive effect on the accommodation prices in private EHPADs.

The dependency support and care prices are determined administratively by the *Département* Councils and by the Regional Health Authorities (*Agences Régionales de Santé*, ARS) Several rates are set, corresponding to the various categories of dependency of the residents (*groupes iso-ressources*, GIR, which are groups of equal dependency needs). Dependency ratings in EHPADs are defined by a national scale (*AGGIR*) making it possible to sort individuals into six GIRs depending on the activities that they are capable of doing alone, from the most severely dependent (*GIR 1*) to the least dependent (*GIR 6*). Up until 2016, those prices were set retrospectively, depending on the costs announced by the facilities (see Bozio et al., 2016). Since 1 January 2017, overall package prices for nursing care and dependency support are set on the basis of the *GIR* categories and of the residents' pathologies (Decree No. 2016-1814 of 21 December 2016).

Some EHPADs also have accommodation rates set by the *Département* Councils. Those are facilities authorised to accept residents who receive social support from the *Département*, that financial assistance making it possible to cover all or some of the expenses related to accommodation. The other facilities, that are not authorised to accept these residents, can set their accommodation rates freely when the residents arrive, but revaluation is then capped by a percentage set by ministerial order (Article L 342-3 of the Social Action and Families Code (CASF, *Code de l'Action Sociale et des Familles*). This concerned 15% of non-profit private EHPADs and 78% of for-profit private EHPADs in 2007 (Perrin-Haynes, 2010) and, respectively, 10% and 66% in 2011 (Volant, 2014).

for its staff. A variable for the mean occupancy rate of the facilities in the local area around the EHPAD at which the employee works was thus also included (*Taux d'occupation moyen_bv*). When the number of vacant places is high (i.e. when the occupancy rate is low), the possibility of increasing the prices in order to raise the wages is probably more limited. To reflect the degree of competition between EHPADs on the market of residents, zoning by local area is the most appropriate because most of the elderly people living in EHPADs used to live nearby. In 2011, 75% of residents were thus in EHPADs that were located less than 15 kilometres away from their previous place of residence (Martin 2014). Since the prices of dependency care and support are determined by the overseeing authorities (cf. Box 2), we add a variable measuring the choices made by the *Département* Councils in setting these prices; for this purpose, we use the mean of the dependency prices for *GIR 1* and *GIR 2* residents in the *Département* in which the employee works (*Tarif dépendance moyen_dpt*). The higher this price, the more the facilities can propose high wages to their nursing staff. Finally, we incorporate a category variable indicating whether the facility belongs to a group of EHPADs (*Groupe*). This variable is broken down into three modalities: fewer than 5 facilities, from 5 to 20 facilities, and more than 20 facilities. This variable can influence the wages in two opposite directions. Firstly, facilities belonging to a group can benefit from economies of scale, which can enable them to propose higher wages in the event of local difficulties in attracting and retaining staff. Secondly, certain facilities can come under financial pressure exerted by the group's parent company, which can constrain them to limit the pay level for their staff or to reduce the staffing level (the latter being incorporated as a control variable in our main model). Belonging to a group could thus have an indirect effect on employee turnover through the wages and the staffing levels, but it is unlikely for the effect to be direct. On the basis of American data, Castle and Engberg (2006) thus did not observe any significant impact of belonging to a chain on turnover among nursing staff in nursing homes. Brannon *et al.* (2002) and Castle (2005) observed a positive effect of this variable on employee turnover but they did not incorporate the wage as an explanatory variable in their model; the effect observed is thus probably an indirect effect related to the impact of belonging to a chain on the wage.

Study sample

Employee exits can be voluntary (resignations) or involuntary (redundancies, end of contract, retirement, resignations due to being constrained to move for extra-professional reasons). Unfortunately, the reasons for staff leaving are not given in the *DADS 2008* data. Since we are seeking to study the causes of voluntary turnover, we took into account only employees under open-ended employment contracts; and we therefore excluded students on training, temporary staff, and staff on fixed-term contracts. Only employees aged under 60 were included in our samples so as not to take employees retiring into account. However, some employees can retire before they are 60; we incorporated an age category of 55 or older as an explanatory variable in our estimates in order to isolate such exits.

In this study, we look only at turnover among nurses and nursing auxiliaries working in private EHPADs. Those private facilities accounted for 49% of all French EHPADs in 2007 and for 51% in 2011 (Volant, 2014). The mean exit rates are much lower in the public facilities, since they were 26% for nurses and 21% for nursing auxiliaries in 2008[6]. Since the vast majority of those exits were related to staff on fixed-term contracts[7] leaving (59% of the nurses leaving, and 68% of the nursing auxiliaries leaving), the number of voluntary exits from public EHPADs was too low to be able to study the causes of such exits. Two reasons might explain this low number of voluntary exits. Firstly, public nursing homes employ mainly civil servants; they accounted for 71% of nurses and for 75% of nursing auxiliaries in 2008. Such public service workers leave their facility only if they find another job in the public sector, and they therefore might encounter fewer opportunities. Secondly, it is possible that nurses and nursing auxiliaries who have chosen to work in public service might seek job stability more than those working in the private sector.

In Table 2, we present the descriptive statistics for each of the variables used for the nurses and for the nursing auxiliaries. There are few differences between these two categories of

6. *The calculations were made on the basis of two samples of 1,169 EHPADs (exit rate for nurses) and of 1,184 EHPADs (exit rate for nursing auxiliaries). Source:* DADS 2008 *(Insee) and* enquête EHPA 2007 *(Drees), authors' calculations.*
7. *This phenomenon is common to all of the organisations in the tertiary sector (Bourieau* et al.*, 2014).*

professional, except that age, wage, and distance from home to work are a little lower for nursing auxiliaries than for nurses.

Results: the reasons for nurses and nursing auxiliaries leaving

We present the effects on salaries of the various variables of the model and of the excluded instruments in Table 3.

Directors of EHPADs do not seem to adapt the wage levels to local difficulties. Only living in Paris or in Île-de-France and the presence of self-employed nurses have significant effects on salaries. The attractivenessof hospitals, shortage of nurses and nursing auxiliaries, and concentration in terms of positions do not have any impact on the pay level. The facilities might be constrained in setting wages by regulated prices, or by prices that are set freely but whose rise is limited by competition. The degree of competition on the market, as measured by the occupancy rate in the local area has a positive impact on the wages of nursing auxiliaries: the higher the occupancy rate, the more the directors of EHPADs can adjust their prices to enable the pay of their staff to be increased.

Various tests were conducted to validate the instruments. Fisher's tests for overall significance, conducted on the basis of ordinary least squares regressions of the wage variable on the exogenous variables and instruments of the model, made it possible to eliminate the null hypothesis of weak instruments. However, the Fisher's statistic was lower for the regression relating to the nurses, the correlation between the instruments and the explanatory variable was thus less strong, which might be detrimental to the accuracy of the results. Amemiya-Lee-Newey overidentification tests made it possible to verify the exogeneity of the instruments. We also conducted Wald tests for the exogeneity of the wage variable. The null hypothesis for exogeneity of wages was rejected for nursing auxiliaries, but not for nurses. The same regression without instrumentation of the wage led to similar estimates of the effects of the environmental factors, but it indicated a positive impact of the wage on the probability of nurses leaving that seems biased (cf. on-line supplement C3). The same impact was observed when correction was not made for the endogeneity of the nursing auxiliaries wage variable. It thus seems preferable to present and to analyse the results of the estimates obtained with the instrumental variable for both categories of staff.

In Table 4, we present the mean of the marginal effects, i.e. the mean impact of each variable on the probability of nursing auxiliaries and nurses leaving. Firstly, as regards the effects of personal characteristics, age seems to have little impact. Only nursing auxiliaries aged from 45 to 50 have a higher probability of leaving (by 4 to 5 percentage points) than nursing auxiliaries aged under 35. As regards the distance from home to work, the impact is, unsurprisingly, positive: the longer the distance, the higher the probability of them leaving. However, the effect is lower for the nurses: only those living more than 20 km away from their place of residence had a higher probability of leaving. Men are also more likely to leave their facility, the probability being higher than women by 7 percentage points.

The impacts of the employment conditions and of the characteristics of the facility varied depending on the profession (nurse or nursing auxiliary). The wage level thus had a highly significant effect on retention of nursing auxiliaries: a rise of 1% in the wage reduces their probability of leaving by from 1.2% to 1.3%. However, it has no significant impact on the probability of nurses leaving. Nurses' wages are higher than those of nursing auxiliaries; actually, they earn 40% more on average (cf. Table 2). They probably therefore face fewer financial difficulties in their daily lives and are more influenced by other dimensions of their working conditions, such as care quality, assessed here by the ratio related to staffing levels. The higher the staffing level, the lower the probability of nurses leaving, even more in facilities having a low staffing level ratio because the coefficient associated with $(N/N^*)^2$ is positive. Night work does not have any direct effect on employee retention, but it can have an indirect effect via its effect on the wage (cf. Table 3). A rise in the proportion of residents who are *GIR 2* and *GIR 3* dependent, compared with the reference category *GIR 1* increases the probability of nurses leaving. Nurses probably prefer to perform more technical care rather than ordinary nursing care, and thus remain longer at EHPADs at which they can put their know-how into practice (i.e. when the residents' state of health is severely degraded). This effect is also positive, but less strongly significant, for nursing auxiliaries. The larger the facility,

Table 2
Descriptive statistics relating to nurses and nursing auxiliaries working under open-ended contracts in EHPADs in France

Variables	Nurses			Nursing auxiliaries		
	Median	Mean	(σ)	Median	Mean	(σ)
Salaire (in euros)	24 449	25 205	(5 629)	17 379	17 848	(3 180)
N/N* (staffing level ratio)	0.783	0.798	(0.173)	0.782	0.801	(0.175)
N_Lits	80	82.94	(34.15)	80	84.48	(36.01)
%GIR1	0.178	0.183	(0.092)	0.176	0.183	(0.095)
%GIR2	0.327	0.328	(0.102)	0.321	0.325	(0.103)
%GIR3	0.140	0.148	(0.065)	0.138	0.145	(0.064)
%GIR4	0.186	0.192	(0.086)	0.186	0.192	(0.085)
%GIR5	0.061	0.073	(0.056)	0.063	0.075	(0.058)
%GIR6	0.051	0.076	(0.087)	0.052	0.080	(0.090)
Hôpital (coeff. de Huff)	0.068	0.860	(2.670)	0.047	0.556	(1.681)
‰Inf_libéraux_bv/pop	0.797	0.992	(0.606)	-	-	-
HHI_{EHPA}_bv (indice H-H)	0.092	0.200	(0.258)	0.110	0.213	(0.263)
‰Inf_{resid}_bv/pop	7.695	7.721	(2.228)	-	-	-
‰AS_{resid}_bv/pop	-	-	-	6.817	7.051	(2.167)
Taux d'occupation moyen_bv	0.949	0.945	(0.038)	0.953	0.948	(0.038)
Tarif dépendance moyen_dpt	17.35	17.45	(1.40)	17.35	17.46	(1.39)

	N	%	N	%
Départ = no	4 371	79.8	11 058	82.3
Départ = yes	1 107	20.2	2 386	17.7
Âge < 35	1 515	27.6	5 191	38.6
Âge [35 ; 45[	1 313	24.0	4 075	30.3
Âge [45 ; 50[	881	16.1	1 960	14.6
Âge [50 ; 55[	1 035	18.9	1 425	10.6
Âge ≥ 55	734	13.4	793	5.9
Homme = no	5 024	91.7	12 464	92.7
Homme = yes	454	8.3	980	7.3
Distance < 5km	1 974	36.0	5 375	40.0
Distance [5km ; 10km[	1 160	21.2	2 842	21.1
Distance [10km ; 20km[	1 360	24.8	3 207	23.9
Distance ≥ 20km	984	18.0	2 020	15.0
Nuit = no	615	11.2	1 457	10.8
Nuit = yes	4 863	88.8	11 987	89.2
Statut = non-profit private	3 514	64.1	9 368	69.7
Statut = for-profit private	1 964	35.9	4 076	30.3
Directeur ≤ 2 ans	1 116	20.4	2 606	19.4
Directeur > 2 ans	4 362	79.6	10 838	80.6
Paris = no	5 425	99.0	13 335	99.2
Paris = yes	53	1.0	109	0.8
Île-de-France = no	4 923	89.9	12 387	92.1
Île-de-France = yes	555	10.1	1 057	7.9
EHPAD non régulé = no	3 102	56.6	8 254	61.4
EHPAD non régulé = yes	2 376	43.4	5 190	38.6
Groupe < 5 EHPAD	4 093	74.7	9 891	73.6
Groupe [5 ; 20[EHPAD	882	16.1	2 425	18.0
Groupe ≥ 20 EHPAD	503	9.2	1 128	8.4

Note: the upper part of the table presents the median, mean, and standard deviation (σ) of the continuous variables used in the estimation. The lower part presents the proportions of each modality of the category variables.
Reading note: in the estimation of leaving, the median percentage of residents rated as GIR 1 *dependent was 17.8% and the mean percentage of those residents was 18.3% in the facilities studied. 64.1% of those facilities were non-profit private EHPADs.*
Coverage: 5,478 nurses and 13,444 nursing auxiliaries working under open-ended contracts in private EHPADs in France.
Source: Insee, DADS 2008*; Drees, enquête* EHPA 2007*; Insee, base permanente des équipements (BPE, permanent facilities database) 2007; Drees, répertoire Finess 2007; authors' calculations.*

Table 3
Effects of exogenous variables of the model and effects of excluded instruments on salary

	Nurses		Nursing auxiliaries	
	Coeff.	(σ)	Coeff.	(σ)
Individual characteristics				
Age < 35	*ref*	*ref*	*ref*	*ref*
Âge [35 ; 45[	0.094***	(0.007)	0.047***	(0.003)
Âge [45 ; 50[	0.140***	(0.008)	0.074***	(0.004)
Âge [50 ; 55[	0.155***	(0.007)	0.084***	(0.005)
Âge ≥ 55	0.169***	(0.008)	0.099***	(0.006)
Homme	0.032***	(0.009)	0.021***	(0.005)
Distance < 5	*ref*	*ref*	*ref*	*ref*
Distance [5 ; 10[	ns		ns	
Distance [10 ; 20[	ns		ns	
Distance ≥ 20	0.015**	(0.007)	ns	
Characteristics of the job and of the facility				
Nuit	0.017*	(0.010)	ns	
N/N*	- 0.118**	(0.060)	ns	
$(N/N^*)^2$	ns		- 0.014***	(0.003)
Statut = non-profit private	*ref*	*ref*	*ref*	*ref*
Statut = for-profit private	ns	ns	- 0.056***	0.004
N_Lits	0.001***	(0.0002)	0.001***	(0.0001)
N_Lits^2 (coeff. by 10^{-6} and σ by 10^{-7})	- 2.38***	(8.13)	- 1.74***	(4.04)
%GIR1	*ref*	*ref*	*ref*	*ref*
%GIR2	ns		ns	
%GIR3	ns		ns	
%GIR4	- 0.114***	(0.036)	- 0.054***	(0.019)
%GIR5	ns		ns	
%GIR6	ns		ns	
Directeur > 2 ans	0.014**	(0.006)	0.008**	(0.003)
Environment				
Hôpital (coeff. de Huff)	ns		ns	
‰Inf_libéraux_bv/pop	0.008*	(0.005)	-	-
HHI_{EHPA}_bv (indice H-H)	ns		ns	
‰Inf_résid_bv/pop	ns		-	-
‰AS_résid_bv/pop	-	-	ns	-
Paris	0.164***	(0.050)	0.048*	(0.027)
Île-de-France	0.085***	(0.011)	0.039***	(0.006)
Instruments excluded				
EHPAD non régulé	ns		ns	
Taux d'occupation moyen_bv	ns		0.088***	(0.031)
Tarif dépendance moyen_dpt	ns		ns	
Groupe < 5 EHPAD	*ref*	*ref*	*ref*	*ref*
Groupe [5 ; 20 EHPAD[	- 0.018**	(0.007)	- 0.020***	(0.004)
Groupe ≥ 20 EHPAD	- 0.029***	(0.009)	- 0.009**	(0.004)
δ	10.086***	(0.095)	9.616***	(0.039)
Tests				
Fisher's Test				
H0: weak instruments				
Test statistics	F(5.5446) = 3.25		F(5.13413) = 7.49	
p-value	0.0062		0.0000	
Amemiya-Lee-Newey Test				
H0: exogenous instruments				
Test statistics	$Chi^2(4) = 5.101$		$Chi^2(4) = 0.907$	
p-value	0.2771		0.9235	
Wald Test				
H0: exogenous Log_Salaire				
Test statistics	$Chi^2(1) = 1.18$		$Chi^2(1) = 28.35$	
p-value	0.2772		0.0000	

****: significant at 1%; **: significant at 5%; *: significant at 10%; ns: not significant.*

Note: the table shows the parameter estimates for the instrumentation equation used to adjust the Log_Salaire *variable for endogeneity. The coefficients were estimated by maximising likelihood functions (see above). Tests for exogeneity and validity of the instruments are presented in the lower part of the table.*
Coverage: 5,478 nurses and 13,444 nursing auxiliaries working under open-ended contracts in private EHPADs in France.
Source: Insee, DADS 2008*; Drees, enquête* EHPA 2007*; Insee, base permanente des équipements (BPE, permanent facilities database) 2007; Drees, répertoire Finess 2007; authors' calculations.*

the higher the turnover of nursing auxiliaries. The size of the EHPAD probably influences the relations between the management team and the staff. In small facilities, the director can be closer to the employees, and more in touch with their wishes in terms of work organisation and hours. Turnover of nurses is also higher in for-profit private EHPADs, but status does not have any impact on turnover of nursing auxiliaries. Finally, the presence of a director for more than two years at the facility does not have any direct impact on employee retention, but it

Table 4
Marginal effects on the probability of nurses and nursing auxiliaries leaving (*probit* model with adjustment for salary endogeneity)

	Nurses		Nursing auxiliaries	
	Coeff.	(σ)	Coeff.	(σ)
Individual characteristics				
Âge < 35	*ref*	*ref*	*ref*	*ref*
Âge [35 ; 45[	ns		ns	
Âge [45 ; 50[	ns		0.044*	(0.026)
Âge [50 ; 55[	ns		ns	
Âge ≥ 55	ns		ns	
Homme	0.072***	(0.023)	0.073***	(0.012)
Distance < 5	*ref*	*ref*	*ref*	*ref*
Distance [5 ; 10[	ns		0.021**	(0.009)
Distance [10 ; 20[	ns		0.026***	(0.009)
Distance ≥ 20	0.080***	(0.017)	0.095***	(0.015)
Characteristics of the job and of the facility				
Log_Salaire	ns		- 1.215***	(0.246)
Nuit	ns		ns	
N/N*	- 0.417***	(0.125)	ns	
$(N/N^*)^2$	0.200***	(0.073)	- 0.026***	(0.008)
Statut = privé associatif	*ref*	*ref*	*ref*	*ref*
Statut = privé lucratif	0.095***	(0.016)	ns	
N_Lits	ns		0.001**	(0.0003)
N_Lits^2 (by 10^{-6})	ns		- 2.72**	(1.12)
%GIR1	*ref*	*ref*	*ref*	*ref*
%GIR2	0.172*	(0.088)	0.084*	(0.050)
%GIR3	0.278***	(0.095)	ns	
%GIR4	ns		ns	
%GIR5	ns		ns	
%GIR6	0.254***	(0.095)	ns	
Directeur > 2 ans	ns		ns	
Environment				
Hôpital (Huff coefficient)	0.008*	(0.004)	0.008*	(0.004)
‰Inf_libéraux_bv/pop	0.029***	(0.010)	-	-
HHI_{Ehpa}_bv (indice H-H)	- 0.038*	(0.023)	- 0.024*	(0.014)
‰Inf_résid_bv/pop	ns		-	-
‰AS_résid_bv/pop	-	-	- 0.007***	(0.002)
Paris	- 0.154***	(0.060)	ns	
Île-de-France	0.117**	(0.050)	0.059***	(0.014)

****: significant at 1%; **: significant at 5%; *: significant at 10%; ns: not significant.*

Note: the table presents the mean of the marginal effects of the variables on the probability of nurses and nursing auxiliaries leaving within the year. For example, living in Paris reduces the probability of a nurse leaving by from 15% to 16% on average, with a degree of significance of 1%. Endogeneity of the variable Log_Salaire *is corrected by instrumentation.*
Coverage: 5,478 nurses and 13,444 nursing auxiliaries working under open-ended contracts in private EHPADs in France.
Source: Insee, DADS 2008*; Drees, enquête* EHPA *2007; Insee, base permanente des équipements (BPE, permanent facilities database) 2007; Drees, répertoire Finess 2007; authors' calculations.*

may have an indirect impact through the wage (cf. Table 3).

Finally, as regards the impact of the local environment, closeness of a hospital has a positive effect on nurses and nursing auxiliaries leaving. They are then more encouraged to change mode of practice of their profession and turn towards the hospital sector, which can offer more interesting career prospects and a greater diversity of positions and sectors. For nurses, the presence of self-employed nurses working in their local area of residence also had a positive and significant effect on their probability to leave. This is another opportunity for changing mode of practice. There is high regional inequality in terms of density of self-employed nurses. In areas where this density is high, self-employed nurses perform relatively fewer nurse medical treatment acts (AMIs) and more nursing care acts (AISs) (Barlet & Cavillon, 2011). In regions under-endowed with self-employed nurses, such nursing care and support is provided by home care services. This substitution of acts of different types enables self-employed nurses to maintain an acceptable level of activity even when supply is large, and this can attract salaried nurses who live in such areas[8]. The concentration variable has a significant negative effect: the more the market is concentrated in terms of nursing staff positions, the less staff are encouraged to leave their facilities. We also observe that the better their local area of residence is endowed with nursing auxiliaries relative to the size of the local population, the less the nursing auxiliaries left their jobs. The reasons might be twofold. Firstly, nursing auxiliaries stay in their jobs because job opportunities are rare in view of the high number of nursing auxiliaries on the local labour market. Secondly, they probably work in a local area that is relatively attractive, and therefore do not seek to work in another locality. Finally, staff are more mobile in terms of changing jobs in Ile-de-France, possibly because of the public transport network, that is better than in other regions of France. Nurses living in Paris have a lower propensity to leave their jobs. Such nurses have probably chosen to pay more for their accommodation in order to reduce their commuting time, and are therefore less willing to change facilities if that would affect their distance from home to work.

8. Since 2011, social security approvals for self-employed nurses in "over-endowed" zones can no longer be granted except for replacing a nurse who is leaving (amendment No. 3 to the national agreement for self-employed nurses). This effect is therefore probably strongly attenuated.

* *
*

We have highlighted the existence of ocal difficulties for retaining nursing staff at private EHPADs: closeness of a hospital, density of self-employed nurses, overall shortage of nursing auxiliaries, and increased pressure from competition between nursing homes for elderly people can encourage nurses and nursing auxiliaries to leave the facilities where they work.

The wage levels also have an effect on he behaviour of nursing auxiliaries. The higher they are, the lower the probability of leaving. Wage compensation could be a means of reducing the effect of regional disparities on turnover among this category of staff. However, their pay is not currently set according to such local difficulties because it is limited by the prices of the EHPADs. The nursing dependency care and support prices that, in theory, cover the costs of the nursing staff, are set administratively by the *Département* Councils and by the Regional Health Authorities (*Agences régionales de santé*, ARSs) The accommodation prices can sometimes compensate for the insufficient pay; however, even when such prices are not regulated, they can be constrained by price competition that we can observe in certain local areas. The wage does not have any impact on the probability of nurses leaving their jobs. Their labour supply thus seems inelastic to prices and to be determined by other factors, such as quality measured by staffing levels, or by the degree of dependency of the residents that influence the nature of the work to be done.

Nursing homes directors then have few levers available for reducing turnover of their nursing staff. Presence of a director for more than two years and night work do not have any impact on retention of such staff. Only a pay rise and reinforcement in staffing level would seem to reduce the probability of the employees leaving. However, such measures require an increase in the wage costs, and that then needs to be passed on to the nursing dependency care and support prices of the EHPADs. Since those prices are covered respectively by the French state health insurance scheme (*Assurance Maladie*) and by the *Département* Councils through payment of the personal independence allowance (*Allocation Personnalisée d'Autonomie - APA*), these measures would induce an extra cost for the public finances. However, they seem essential in view of the impact that a reduction in turnover among the nursing staff can have on the quality of care of the residents.

This econometric study is the first to analyse the reasons for nursing staff leaving EHPADs in France. It could be interesting, in further studies, to analyse in more detail the impact of certain variables, in particular care quality, on decisions by nursing staff to resign. Unfortunately, we only have information on the staffing level, which naturally does not make it possible to approach all of the multi-dimensionality of care quality. Also in this study, we lack data on the socio-demographic characteristics of the nurses and of the nursing auxiliaries, and on the organisational culture and the managerial policy of facilities directors. Various authors have shown that the involvement of the nursing staff in managing timetables or in administrative decisions can have a non-negligible impact on job satisfaction, and thus on the choice of staying at the facility (Donoghue & Castle, 2007). Finally, here we only have cross-sectional data that, unlike panel data, do not make it possible to adjust for endogeneity related to unobserved heterogeneity. And yet, those employees who decide to leave their facility may have particular characteristics that are not all taken into account by the explanatory variables of the model. ☐

BIBLIOGRAPHY

Anderson, R. A., Issel, L. M. & McDaniel, R. R. (1997). Nursing staff turnover in nursing homes: A new look. *Public Administration Quarterly*, 21(1), 69–95.

Antwi, Y. A. & Bowblis, J. R. (2016). The impact of nurse turnover on quality of care and mortality in nursing homes: Evidence from the Great Recession. *Upjohn Institute working paper*, 16–249.
http://research.upjohn.org/cgi/viewcontent.cgi?article=1267&context=up_workingpapers

Banaszak-Holl, J. & Hines, M. A. (1996). Factors associated with nursing home staff turnover. *The Gerontologist*, 36(4), 512–517.

Barlet, M. & Cavillon, M. (2011). La profession d'infirmière : situation démographique et trajectoires professionnelles. Drees, *Etudes et Résultats* N° 759.
http://drees.social-sante.gouv.fr/IMG/pdf/er759.pdf

Baughman, R. A. & Smith K. E. (2012). Labor mobility of the direct care workforce: Implications for the provision of long-term care. *Health Economics*, 21(12), 1402–1415.

Bourieau, P., Paraire, X. & Sanzeri, O. (2014). Les mouvements de main-d'œuvre en 2013. Forte augmentation des entrées en CDD dans le tertiaire. Dares, *Dares Analyses* N° 094.
http://travail-emploi.gouv.fr/IMG/pdf/2014-094.pdf

Bozio, A., Gramain, A. & Martin, C. (2016). Quelles politiques publiques pour la dépendance ? *Les notes du Conseil d'analyse économique* N° 35.
http://www.cae-eco.fr/IMG/pdf/cae-note035v2.pdf

Brannon, D., Zinn, J. S., Mor, V. & Davis, J. (2002). An exploration of job, organizational, and environmental factors associated with high and low nursing assistant turnover. *The Gerontologist*, 42(2), 159–168.
https://doi.org/10.1093/geront/42.2.159

Cameron, A. C. & Trivedi, P. K. (2009). *Microeconometrics using Stata*. Stata Press.

Castle, N. G. (2005). Turnover begets turnover. *The Gerontologist*, 45(2), 186–195.
https://doi.org/10.1093/geront/45.2.186

Castle, N. G. & Engberg, J. (2006). Organizational characteristics associated with staff turnover in nursing homes. *The Gerontologist*, 46(1), 62–73.
https://doi.org/10.1093/geront/46.1.62

Cohen-Mansfield, J. (1997). Turnover among nursing-home staff. *Nursing management*, 28(5), 59–64.

Dill, J. S., Morgan, J. C. & Marshall, V.W. (2013). Contingency, employment intentions, and retention of vulnerable low-wage workers: An examination of nursing assistants in nursing homes. *The Gerontologist*, 53(2), 222–234.
https://doi.org/10.1093/geront/gns085

Donoghue, C. & Castle, N.G. (2007). Organizational and environmental effects on voluntary and involuntary turnover. *Health Care Management Review*, 32(4), 360–369.

Huff, D. L. (1964). Defining and estimating a trading area. *Journal of Marketing*, 28(3), 34–38.

Insee (2003). Structuration de l'espace rural : une approche par les bassins de vie. Rapport de l'Insee pour la Datar.

Insee (2011). Emploi et salaires, *Insee Références*, édition 2011.
http://documentation.fhp.fr/documents/17274P2.pdf

Irvine, D. M & Evans, M. G (1995). Job satisfaction and turnover among nurses: Integrating research findings across studies. *Nursing Research*, 44(4), 246–253.

Josse, F. (2012). Recrutement en Ehpad : y'a quelqu'un ? *Mensuel des maisons de retraite*, 153, 15–20.

Kash, B. A., Castle, N. G., Naufal, G. S. & Hawes, C. H. (2006). Effect of staff turnover on staffing: A closer look at registered nurses, licensed vocational nurses, and certified nursing assistants. *The Gerontologist*, 46(5), 609–619.
https://doi.org/10.1093/geront/46.5.609

Le Noble Age (2012). *Rapport financier annuel* 2012.
http://www.lenobleage-groupe.com/sites/default/files/finance/rapport_financier_annuel_2012.pdf

Lerner, N. B., Johantgen, M., Trinkoff, A. M., Storr, C. L. & Han, K. (2014). Are nursing home survey deficiencies higher in facilities with greater staff turnover ? *JAMDA*, 15, 102–107.

Loquet, J. & Nagou, G. (2014). Le turnover du personnel infirmier en France : une analyse empirique à partir des déclarations annuelles de données sociales. *Journées des Économistes de la Santé français*, Collège des Économistes de la Santé, Bordeaux, décembre.
http://www.ces-asso.org/sites/default/files/Loquet_Nagou.pdf

Martin, C. (2014). Concurrence, prix et qualité de la prise en charge en Ehpad en France : Analyses micro-économétriques. Thèse de doctorat, Université Paris-Dauphine.

Nénin, F. & Lapart, S. (2011). *L'or gris*. Flammarion.

Perrin-Haynes, J. (2010). Les établissements d'hébergement pour personnes âgées : activité, personnel et clientèle au 31 décembre 2007. Drees, *Document de travail - Série statistiques* N° 142.
http://drees.social-sante.gouv.fr/IMG/pdf/seriestat142-2.pdf

Prévot, J. (2009). L'offre en établissements d'hébergement pour personnes âgées en 2007. Drees, *Études et Résultats* N° 689.
http://drees.social-sante.gouv.fr/IMG/pdf/er689.pdf

Prévot, J., Tugores, F. & Bertrand, D. (2007). Les établissements d'hébergement pour personnes âgées. Activité et personnel au 31 décembre 2003. Drees, *Document de travail - Série statistiques* N° 106.
http://drees.social-sante.gouv.fr/IMG/pdf/seriestat106.pdf

Pumain, D. & Saint-Julien, T. (2010). *Analyse Spatiale, les interactions* (2e éd.). Armand Colin.

Ratte, E. & Imbaud, D. (2011). Accueil et accompagnement des personnes âgées en perte d'autonomie. Rapport du groupe de réflexion n°3 sur la dépendance, Ministère des solidarités et de la cohésion sociale.
http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/114000334.pdf

Salop, S. C. (1979). A Model of the Natural Rate of Unemployment. *American Economic Review*, 69(1), 117–254.
https://www.researchgate.net/profile/Steven_Salop/publication/4722803_A_Model_of_the_Natural_Rate_of_Unemployment/links/5463626f0cf2837efdb3005a.pdf

Spilsbury, K., Hewitt, C., Stirk, L. & Bowman, C. (2011). The relationship between nurse staffing and quality of care in nursing homes: A systematic review. *International Journal of Nursing Studies*, 48(6), 732–750.

Stiglitz, J. E. (1974). Alternative theories of wage determination and unemployment in LDC's: The labor turnover model. *The Quarterly Journal of Economics*, 88(2), 194–227.
https://opendocs.ids.ac.uk/opendocs/bitstream/handle/123456789/480/dp125part1-318303.pdf?sequence=1

Temple, A., Dobbs, D. & Andel, R. (2009). Exploring correlates of turnover among nursing assistants in the national nursing home survey. *Health Care Management Review*, 34(2), 182–190.

Volant, S. (2014). L'offre en établissements d'hébergement pour personnes âgées en 2011. Drees, *Études et résultats* N°877.7.
http://drees.social-sante.gouv.fr/IMG/pdf/er877.pdf

Wiener, J. M., Squillace, M. R., Anderson, W. L. & Khatutsky, G. (2009)K. Why do they stay? Job tenure among certified nursing assistants in nursing homes. *The Gerontologist*, 49(2), 198–210. https://doi.org/10.1093/geront/gnp027

Zhang, Y., Punnett, L., Gore, R. & The CPH-NEW Research Team (2014). Relationships among employees' working conditions, mental health, and intention to leave in nursing homes. *Journal of Applied Gerontology*, 33(1), 6–23.

Does women's employment growth increase wage inequalities between couples? The case of France between 1982 and 2014

Milan Bouchet-Valat*

Abstract – It has often been argued that women's employment growth is a factor that contributes to the increase in inequalities between households due, in particular, to an alleged reinforcement of social homogamy. In contrast to this idea, an accounting approach to inequality decomposition, based on Insee's Labour Force surveys (enquêtes Emploi) shows that wage inequalities between couples aged 30 to 59 remained stable between 1982 and 2014 in France, whereas they would have increased had women's employment rate not risen.

This overall stability results from two converse developments, which are themselves linked to the strong growth in women's employment over this period: a fall in wage inequality between women and an increase in the correlation of partners' wages within couples. However, the almost uniform increase in women's employment rate, regardless of their partner's wage level, has limited the increase in the correlation of partners' wages and prevented an increase in wage inequalities between couples.

JEL codes : D10, D31, D63, J12, J22
Keywords : wage inequality, women's employment, couples, homogamy

**National Institute for Demographic Studies (Ined) and Crest, Quantitative Sociology Laboratory (milan.bouchet-valat@ined.fr).*
The tables and the code that enable the analyses to be reproduced are available on the author's personal web page at http://bouchet-valat.site.ined.fr.

Received 26 January 2016; accepted after revision 13 February 2017 DOI: 10.24187/ECOSTAT.2017.493s.1913
Translated from the original version: « Le développement de l'emploi des femmes augmente-t-il les inégalités de salaire entre couples ? Le cas de la France entre 1982 et 2014 »

Women's employment rate has increased significantly in France since the 1960s. This trend, which continues to this day at a regular pace, has had significant repercussions for the sources of household income. Thus, between 1982 and 2014, the contribution of all women's wages to the total earnings of couples aged between 30 and 59 (excluding the self-employed and the retired) increased from 27% to 38% (see Figure I below).

A fairly significant increase in wage and income inequalities between individuals or between households has also been observed in France, primarily due to the strong growth in income at the top of the social structure (Landais, 2007 [2008]; Amar, 2010; Solard, 2010; Godechot, 2012, 2013 [2014]; Piketty, 2013 [2014]). It is then interesting to investigate, as did the Organisation for Economic Cooperation and Development (OECD, 2008, 2011) recently, whether the increase in the employment rate of women, by reinforcing the association between partners' wages (consequence of social homogamy), has contributed to strengthening this trend.

In a somewhat counter-intuitive manner, this paper provides a negative response to this question. Wage inequalities between couples aged between 30 and 59 (excluding the self-employed and the retired) remained fairly stable in France between 1982 and 2014, and even declined slightly above the median. This trend results mainly from an equalising effect of women's employment growth, whereas the slight fall in men's employment rate has tended to exacerbate inequalities between couples.

This overall development masks contrasting trends depending on the part of the wage distribution of couples taken into consideration. Thus, although the increase in the employment rate of women has had an equalising effect on the overall distribution, the fall in the employment rate of men has led to more inequality below the median.

In France, the increase in women's wages has not been accompanied by a sufficient strengthening of the association between spouses' wages within couples to increase inequalities between couples, unlike in the United States (Cancian & Reed, 1998; Hyslop, 2001; Schwartz, 2010). This phenomenon can be explained by the fact that the increase in the employment rate of women has occurred at the same pace regardless of the spouse's wage decile (with the notable exception of the first decile, the inactive population and the unemployed). The association between spouses' wages has even weakened at the top of the distribution.

These trends are consistent with those also observed in France with respect to wage inequalities and the choice of spouse. Thus, contrary to popular belief, if we disregard the top of the distribution, we have seen a slight fall in wage inequalities since the 1960's (Verdugo et al., 2012; Charnoz et al., 2013 for the wages of men working full time in the private sector; Verdugo, 2014), whereas inequalities in the standard of living have been fairly stable since the 1990's (Pujol and Tomasini, 2009; Boiron et al., 2016). In addition, the similarity between spouses (social homogamy) has tended to weaken in terms of qualifications and occupation for several decades now (Vanderschelden, 2006; Bouchet-Valat, 2014).

The study is based on the annual data from the French Labour Force surveys (Insee's *enquêtes Emploi)* since 1982, and relates solely to wages (Box 1). Couples where one of the partners is self-employed or retired are excluded. In order to avoid excessively cumbersome formulations, we will from now on define 'employment' as receiving a wage; 'employment rate' will thus designate the proportion of individuals within this scope that receive a salary. No distinction is made between full-time and part-time wage earners, and the absence of a wage is regarded as a zero wage. The analyses thus focus on the wage actually received, which is the result of both decisions to engage in active employment that are more or less freely chosen or forced upon them (and frequently linked to family situations, see Meron & Maruani, 2012, chapter 2), and of a more or less favourable position in terms of hourly wages. Although not taking into account other types of income, including capital income, means disregarding a significant aspect of the increase in income inequalities (Landais, 2007 [2008]; Piketty, 2013 [2014]), this approach is fully relevant to the study, in particular, of the effects of the growth in women's employment on the link between partners' incomes and thus on inequalities between couples.

We first present an international and French literature review. We then show, using a decomposition of the coefficient of variation, that the growth in women's employment rate has not further exacerbated wage inequalities between couples. This result stems in particular from the fact that the social distribution of women's employment has remained fairly stable, thus limiting the increase in the correlation between partners' wages. Lastly, using a decomposition method based on different counterfactual scenarios developed using log-linear models, we highlight contrasted effects of women's and men's employment trends on inequalities depending on the area of the wage distribution under consideration, as well as the fairly limited role played by the change in the association between partners' wages.

Women's employment, homogamy and income inequalities between couples in the literature

The question of the changing association between partners' wages within a couple and its effects on inequalities has been addressed mainly in work relating to the United States: research initially focused on evaluating the effect of the increase in women's participation in the labour force on inequalities between households since the 1960s. The US context, which is characterised by a significant increase in educational homogamy frequently perceived as a risk for the cohesion of US society (Breen & Salazar, 2011), is clearly fairly different from the context in France. Women's growing participation in the labour market has been regarded as one of the factors of the increase in inequalities observed in many countries across the world (Blossfeld & Buchholz, 2009; Esping-Andersen, 2007). Contrary to this view, all studies conclude that women's employment growth has in fact tended to limit the increase in income inequalities in the United States (Cancian & Reed, 1998, 1999; Reed & Cancian, 2001; Devereux, 2004; Pencavel, 2006; Western et al., 2008; Daly

Box 1

SOURCE, SCOPE AND DEFINITION OF WAGE

This study is based on the series of French Labour Force Surveys (*Enquêtes Emploi*) conducted by Insee from 1982 to 2014, and covers all cohabiting couples (whether married or not) where both partners are aged between 30 and 59 and neither is self-employed or retired. Restricting the sample to individuals aged at least 30 allows us to limit as far as possible the effects of recent cohorts first cohabiting and entering the labour market at a more advanced age, which would require separate analyses. Over the whole period, around a third of men and half of women aged between 20 and 30 cohabit with a partner in any given year, while this rate exceeds three quarters for individuals of both sexes in their thirties; this share has fallen by 6 to 8 percentage points over the years (Bouchet-Valat, 2014, p. 331-332). Moreover, although the employment rate of individuals aged 20 to 30 has fallen significantly since 1982 both among men and women, it has been more stable among men in their thirties and has increased among women in their thirties or in otherage brackets (Insee, 2016).

The Labour Force Surveys provide annual data and homogeneous monthly wages based on large samples. However, monthly wages are self-reported. While this can give rise to a difference in wage levels compared to the wage data from administrative sources, a comparison with the wages reported by employers in the *Annual Declarations of Social Data (DADS)* shows that the trends are similar to a large extent.

The Labour Force Surveys questionnaire asks respondents to specify their monthly wage in the month preceding the survey, as well as any supplementary earnings received annually (bonuses, 13th month's salary, etc.). The respondents have the option of not responding, or either of indicating a wage bracket rather than an amount, in which cases Insee imputes a value computed on the basis of other available variables.

As wages were provided only in the form of brackets between 1982 and 1989, we have imputed the wages for these years using the simulated residual method (O'Prey, 2009, p. 17). The imputation model, which is applied separately to men and women (either cohabiting with a partner or not), takes account of the interval censoring linked to the wage brackets, and assumes a log-normal wage distribution. The variables taken into account are the regular working hours, seniority combined with the type of contract, age (and its square), the socio-professional category (PCS level 3), the level of qualifications, the urban area size and the region of residence.

Due to the significant fluctuations linked to the sampling at the top of the wage distribution, the wages in each year that exceed the 995th thousandth (i.e., 0.5% of cases) have been pegged back to that level. The sample size does not allow for this group (around 40 individuals per year) to be studied with precision from one year to the next.

Only those actually employed at the time of the survey are asked about their wages. We have attributed a wage equal to zero to the unemployed and the inactive population (of whom less than 3% declare a wage). The sample, restricted to individuals taking part for the first time in the survey, comprises between 5,300 and 7,000 couples per year prior to 2009 and between 7,700 and 8,700 since then, which makes a total of 217,000 couples.

& Valletta, 2006; Hryshko et al., 2014) and in all of the OECD countries, particularly in France (OECD, 2011, p. 226 [p. 207]; Harkness, 2013; these two comparative studies are based on the Luxembourg Income Study, which comprises the *Family Budget* survey for France).

The specific role of changes in homogamy is more disputed. In France, an analysis of all of the wages earned by couples during their lives (Courtioux & Lignon, 2015a) recently showed that educational homogamy only partially mitigates the equalising effect of couple formation. The Gini coefficient for wages earned by individuals from the same generation during the whole of their lives falls by 12% if we consider couples rather than individuals taken in isolation. This fall would be greater, by 3 percentage points for women and 7 percentage points for men, if couples were formed randomly (no homogamy). Based on a different method, another study (Frémeaux & Lefranc, 2015) estimated that educational homogamy causes a 3% to 10% increase in inequalities between the annual wages of couples. However, these studies disregard the issue of changes over time. Given the weakening of homogamy in terms of education and social class, as highlighted by the research on France (Vanderschelden, 2006; Bouchet-Valat, 2014), it seems unlikely that changes to this factor could have contributed significantly to an increase in inequalities between households in recent decades.

Most of the literature available for other countries confirms that, contrary to popular belief, educational homogamy only forms a fairly loose, albeit real, link with the association between partners' wages[1], and its increase has only had negligible - or even negative - impacts on the increase in wage inequalities in many countries (Worner, 2006 for Australia; Western et al., 2008 ; Breen & Salazar, 2010, 2011 for the United States and the United Kingdom; Breen & Andersen, 2012 for Denmark ; Eika et al., 2014 for the United States and Norway).

We thus focus here on measuring the effect of the increase in the association between partners' wages on inequalities between couples: we believe that this question must be addressed prior to the issue of the influence of variations in educational homogamy, which only has an indirect link with partners' wages. Existing studies on this issue have noted either that the increase in the association between partners' incomes explains between 15% and 30% of the total increase in inequalities in the United States, depending on the periods studied and the methods used (Karoly & Burtless, 1995; Burtless, 1999; Cancian & Reed, 1999; Hyslop, 2001; Schwartz, 2010), or that it makes an even more modest contribution to this (Cancian & Reed, 1998; Hryshko et al., 2014 for the United States; OECD, 2011, p. 226 [p. 207] for the member countries of the organisation; Funes Leal, 2015 for Argentina). To be more precise, this factor is said to have mainly contributed to the increase in wage inequalities in the United States in the 1980s, but only to a negligible extent since then (Larrimore, 2014).

In the case of France, it would appear that the equalising effect of the growth in women's employment outweighs the opposite effect of a potential increase in the association between partners' wages. Indeed, it is the former phenomenon which constitutes the major change over the period under consideration.

The contribution of this study is first and foremost to analyse the changes over time, whereas the existing research for France (Frémeaux & Lefranc, 2015; Courtioux & Lignon, 2015a) focuses on a single point in time (either a cohort or a survey year). This historical perspective is necessary to identify the impact of the growth in women's employment on inequalities. Like Frémeaux and Lefranc, we study the inequalities between the wages earned by couples, whereas the Courtioux and Lignon study focuses on the inequalities that can be attributed to educational homogamy (see Courtioux & Lignon, 2015b, for a presentation of the different methods used by this latter approach).

Courtioux and Lignon (2015a) attempted to reconstruct the income of members of the same cohort over the course of their lives using dynamic microsimulation models. Put more simply, the advantage of the analysis of wages at the time of the survey used here is that it describes the association between partners' wages actually observed each year, without having to make assumptions in order to reconstruct the composition of the couples taking account of such factors as educational homogamy. However, it does not take account of the individuals' income over the course of their lives, or unpartnered individuals, or the

1. Though a study asserting that the increase in educational homogamy explains a significant share of the growth in inequalities between households in the United States has been given some credence, the results have been subject to a corrigendum that significantly mitigates this assertion (Greenwood et al., 2014).

household size (which allows standards of living to be calculated).

Frémeaux and Lefranc (2015) focused on annual wages – actually received and in full-time equivalent – whereas this paper limits itself to monthly wages. This restriction enables us to cover a longer time span thanks to the Labour Force surveys. However, it introduces wage variations over a short period, which would be smoothed by using annual or multi-annual averages. This restriction probably leads to an underestimation of the association between partners' wages; however, the estimations of the contributions of the different factors to the inequalities and their changes over time are barely affected by this bias (Frémeaux & Lefranc, 2015, p. 11; Hryshko et al., 2014, p. 771).

Lastly, we should point out that the accounting approach (Courtioux & Lignon, 2015b) that we adopt is designed to provide a decomposition of the effects of the various factors on homogamy based on the assumption that the behaviour of the individuals remains unchanged. This is thus a descriptive and illustrative exercise, but not one designed to identify causal links.

Women's employment growth has not exacerbated wage inequalities between couples

The growth in women's employment and wages in France since 1982

The growth in women's employment, which has been under way since the start of the 1960s, has been very marked in France since 1982. The strong growth in the contribution of women to the total wages of couples mentioned in the introduction has resulted from a combination of two trends: firstly, an increase in the employment rate, i.e., from the viewpoint adopted here, the growth in the share of women who earn a wage; and secondly, wage growth among women in employment.

The growth trend of women's employment rate is well-documented: the share of women aged between 30 and 59 who live with a partner (excluding the self-employed and the retired) and earn a wage increased from 51% in 1982 to 78% in 2014 (Figure I). However, another development has been less visible: in parallel to the increase in women's contribution to couples' wages, the share of women earning a higher wage than their partner has doubled, increasing from 12% to 24% over the same period (see also Morin, 2014).

This last trend is of course attributable to the increase in the employment rate of women, as well as to a very slight increase in working women's wages compared to those of men (Minni, 2015): the mean wage of women living with a partner was 36% lower than that of men living with a partner at the start of the period, and 28% lower at the end of the period. This marginal development was more marked at the bottom of the distribution (a decrease from a 67% gap to a 49% gap for the first decile).

It should be noted that this change occurred despite a strong growth in the share of part-time work among women (Afsa Essafi & Buffeteau, 2006), which increased from 19% in 1982 to 32% in 1999, and has remained stable at that level since then (Insee, 2016). This rate has also increased among men, from 3% to 8%, although the overall level has remained low. Conversely, the rate of unemployment has increased more among men (for whom it doubled from 5% in 1982 to 10% in 2014) than among women (Cabannes, 2014; Insee, 2016). It was indeed already quite high among women (8 % in 1982), and though it reached almost 12% in the mid 1990s, it has since fallen back to the same level as for men (10% in 2014). The differences between men and women in terms of part-time work and unemployment have thus followed opposite trends, whereas the increase in the activity rate and the employment rate have significantly boosted growth in women's wages compared to those of men.

Growth in women's wages over the last thirty years have been very marked. However, even if the very stable pace of the increase observed since 1982 were to be maintained, the total amount of women's wages would account for half of the total wages of couples only by around 2045. It is moreover unlikely that this will happen by then, as the rate of employment among women living with a partner, which is the main factor driving his change, would (based on the same assumption) reach 100% a decade earlier, and would probably become stable well before then. For the same reasons, it is far from guaranteed that half of women will receive a higher wage than that of their partner some time during the 21st century: even based on the very optimistic projection that the past pace would continue, this event would only occur in around 2080.

Moreover, the growth in women's employment has not prevented their less favourable situation on the labour market from continuing (Meron, 2008; Albouy et al. 2012): greater job insecurity, part-time work, lower wages due to the structure of these jobs (segregation between sectors and professions), but also when they perform the same job as a man. These caveats and the uncertainty about future trends should not, however, mask the significance of the changes that have occurred since the 1980s.

A reduction in wage inequalities among women and the stability of inequalities between couples

Has the 10 percentage-point increase in women's wages as a share of the total wages of couples led to a reduction in inequalities between couples? In order to answer this question in this section, we decompose the inequalities between the wages of couples, and their changes over time, into three different sources. To do this, we measure the overall level of inequalities using the coefficient of variation, which offers definition which is easily interpretable in terms of the dispersion around the mean. It is defined as the ratio between the standard deviation σ and the mean μ of a distribution, i.e.:

$$CV = \frac{\sigma}{\mu}$$

The change in the inequality between the total wages of couples (with a wage of zero attributed to the unemployed and the inactive population) is presented in Figure II. We observe a fluctuation without a clear trend, which culminates, in 2014, in a level of inequality that is close to that observed in 1982. The decomposition of the coefficient of variation will enable us to understand this result.

The square of the coefficient of variation can be expressed as the sum of three terms, each of which corresponds to a clearly identified source of income:

$$\begin{aligned} CV^2 &= \left(1 - S_f\right)^2 CV_h^2 + S_f^2 CV_f^2 \\ &+ 2\rho_{hf} S_f \left(1 - S_f\right) CV_h CV_f \end{aligned}$$

Figure I
Growth in the wages of women living with a partner since 1982 via three indicators

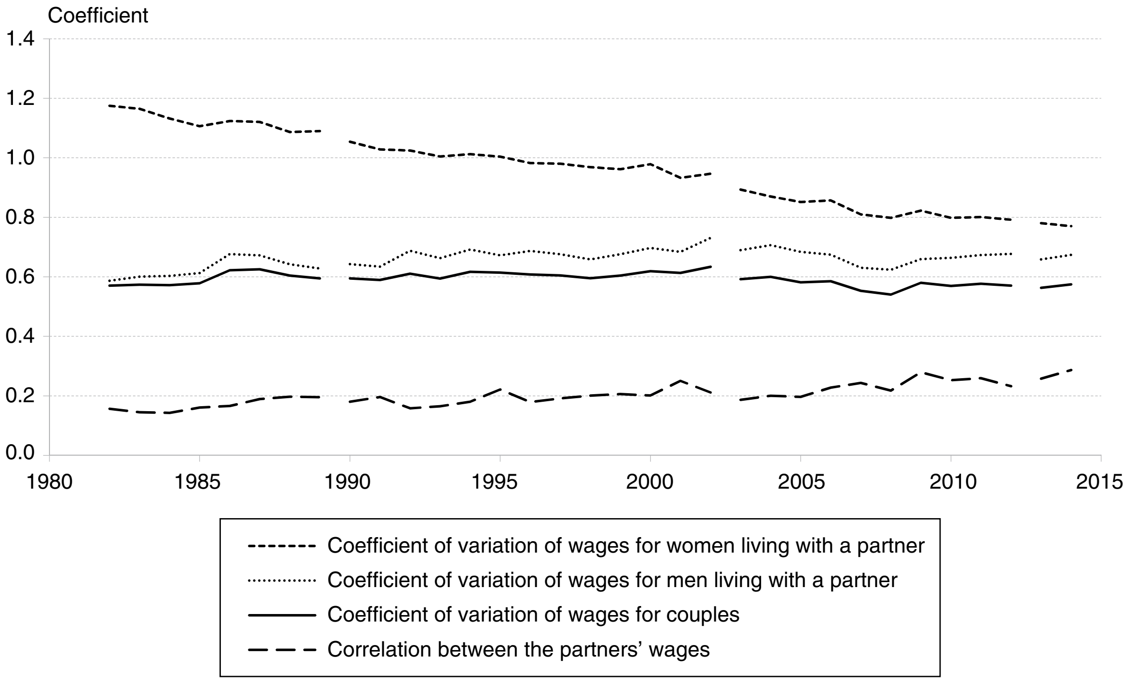
Note: the unemployed and the inactive are included and are attributed a wage equal to zero.
Reading note: in 1982, 51% of women aged between 30 and 59 and living with a partner were earning a wage, whereas these wages only accounted for 27% of the total wages of couples, and only 12% of them were earning more than their partner.
Coverage: cohabiting men and women, where the partners are aged between 30 and 59 and neither is self-employed or retired.
Source: Insee, Labour Force Surveys (*enquêtes Emploi*), 1982-2014.

S_f measures women's wages as a share of the total wages of couples, CV_h and CV_f (with h for men, f for women) are respectively the coefficient of variation for men's and women's wages, and ρ_{hf} corresponds to the correlation between partners' wages within couples. Inequality between couples is thus more pronounced when the inequality between individuals of the same sex is high; when the gender most affected by the inequality contributes significantly to the total wages of couples; and when the correlation between partners' wages is strong.

According to this equation, the contribution of each sex to inequality between couples can be evaluated in relation to (at least) three benchmark situations, known as counterfactual situations, which allow for three different causes of the changes in inequalities to be identified (Cancian & Reed, 1998, p. 74). First situation: if only the individuals from one sex contributed to the total wages of couples, as the members of the other sex would all be inactive or unemployed, the level of inequalities between couples would correspond respectively to CV_h or CV_f, depending on whether those actively working are men or women. Second situation: if the wages of the individuals from a given sex were all equal and, consequently, the inequality was attributable exclusively to the wages of the other sex, then the level of inequality between couples would be equal to $\left(1-S_f\right)CV_h$ or $S_f CV_f$, depending on whether the inequality stemmed from men or women. Third situation: if there were no correlation between partners' wages, the coefficient of variation would be equal to the square root of the sum of the terms $\left(1-S_f\right)^2 CV_h^2$ and $S_f^2 CV_f^2$.

Thus, the association between partners' wages potentially plays an important role in determining inequalities between couples. Based on a hypothetical scenario, which was unrealistic in 1982, but which has become increasingly credible over time, where wage inequalities are the same among men and among women and where both sexes contribute equally to the total wage volume, moving from no correlation to a perfect

Figure II
Change in wage inequalities and in the correlation between partners' wages since 1982 (all couples)



Note: the unemployed and the inactive are included and are attributed a wage equal to zero.
Reading note: the coefficient of variation of wages for women living with a partner (including those who have a zero wage) fell from 1.17 in 1982 to 0.77 in 2014.
Coverage: cohabiting men and women, where the partners are aged between 30 and 59 and neither is self-employed or retired.
Source: Insee, Labour Force Surveys (*enquêtes Emploi*), 1982-2014.

correlation between partners' wages would amount exactly to doubling the level of inequalities between couples. Such a radical change in the correlation is fairly improbable: conversely, the correlation between partners' wages is usually too weak to have such strong impacts on inequalities[2].

Thus, the stability of the overall inequality results from the fact that the different components of the equation either remained fairly stable or followed opposite trends, which cancelled each other out to a large extent (*cf.* Figure II). Firstly, despite a fairly clear increasing trend up until the 2000's, inequality between men's wages (CV_h) changed little, and fluctuated around a coefficient of variation of 0.7[3]. On the other hand, inequality between women's wages (CV_f) has significantly fallen, with the coefficient of variation falling from 1.17 to 0.77, and has clearly moved towards parity with the level among men. This change is attributable to the growth in women's employment, which has caused the wage of a significant proportion of the sample to increase from zero (inactivity) to a level that is very likely to be closer to the mean (Pasqua, 2002). This decrease in inequality between women, coupled with the significant increase in women's wages as a share of the wages of couples (S_f), has had a very clear equalising effect (*cf.* Figure I).

Although this change is a mechanical consequence of the increase in the employment rate of women, it should be noted that it has not been accompanied by a sufficient increase in the third term of the equation - the correlation between partners' wages - to reverse the equalising trend: the growth in this correlation (taking account of both active and inactive individuals) from 0.16 in 1982 to 0.26 in 2013 has just been sufficient to offset the effects of the lowering of the inequality between women[4]. This level of correlation, which is slightly higher than that already observed in existing studies for France (Frémeaux & Lefranc, 2015, p. 10), is significantly higher than that reported by several authors for the United States. In the US, the correlation was slightly negative[5] prior to 1980, and has stood at around 0.1 in recent years (Schwartz, 2010, p. 1540; Cancian & Reed, 1998, p. 76; Reed & Cancian, 2012, p. 10). This can be seen as a reflection of the French model of participation of women in the labour market, where full-time work is more prominent than in other countries (Meron & Maruani, 2012).

The correlation between partners' wages within dual-earner couples (who, excluding the unemployed and the retired, accounted for 48% of couples in 1982 and 68% in 2014) is higher than for all couples: for the period under consideration, it stands at around 0.35 to 0.40 (Figure III). This gap has also been observed in the United States (Schwartz, 2010), where there is once again a less strong correlation than in France. However, this correlation is still fairly modest. Thus, the growth in the correlation between partners' wages is not found among dual-earner couples: the correlation here has fluctuated without any clear trend since 1982 (Figure III). As a result of this relative stability, wage inequality between dual-earner couples has been fairly much in line with the changes in wage inequalities between men and women living in a couple: it increased by 19% through to 2002, then fell back down to its initial level. Very different mechanisms have thus given rise to fairly similar changes, whether we take account of all couples or only dual-earner couples.

Almost uniform growth in women's employment at all partner wage levels

The effect of the increase in the share of dual-earner couples on the correlation between partners' wages among all couples depends to a large extent on the link between women's employment and the partner's wage level (Pasqua, 2002). Thus, in the United States, the increase in the correlation between partners' wages is attributable to a great extent to the fact that the negative relationship between men's

2. *This mechanism corresponds to the standard phenomenon of regression towards the mean (Verbakel, 2008, p. 132), based on which an individual receiving a wage far removed from the mean (either very high or very low, or zero) is very unlikely to form a couple with a person who receives such an outlying wage.*

3. *The increase in wage inequalities between men from 1982 to 1986, which occurred exclusively above the median, has already been observed in other works based on administrative data (Charnoz et al., 2013, p. 73; Verdugo, 2014, p. 135). However, the existing studies point instead to a fall in wage inequalities between men since the 1960's, excluding unemployment and inactivity (Verdugo, 2014; Verdugo et al., 2012).*

4. *The reality of the sudden transition to a correlation of 0.29 in 2014 needs to be confirmed by future surveys.*

5. *This was due, in particular, to the low rate of employment of wives whose partner belongs to the top deciles.*

wages and employment of his partner has gradually disappeared, giving way to an inverted U curve according to which the middle classes have the highest levels of women's employment (Schwartz, 2010, p. 1541).

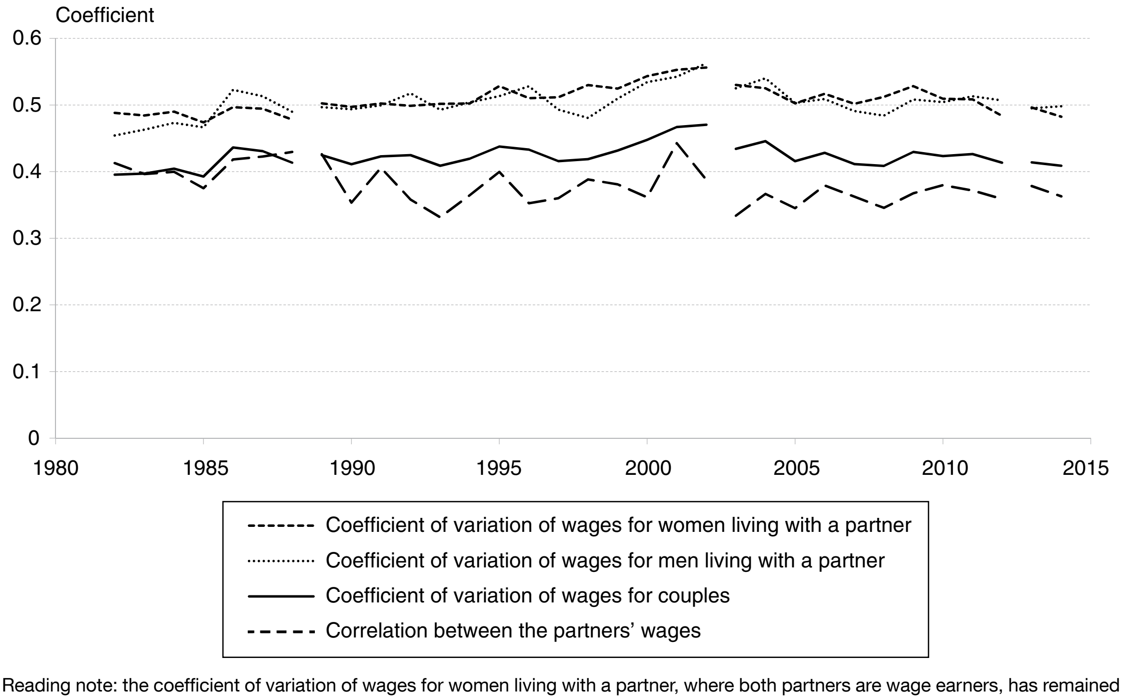
However, the French situation appears to be fairly different (Figure IV, left-hand graph). As early as 1982, the highest employment rate of women (at around 60%) could be found among the partners of men belonging to the seventh wage decile (confirming the results of Frémeaux & Lefranc, 2015, p. 15). Conversely, its level was almost as low for the top decile (11 percentage points below the maximum) as it was for the bottom decile (13 percentage points below). However, women whose partner has no wage stood out clearly from the rest with an employment rate that was 18 percentage points below the maximum.

In the United States at the same period, only the upper deciles stood out with an employment rate of women that was clearly lower than the others, with the lower deciles joining them gradually only after this date. France seems thus to have experienced the change observed in the United States at an earlier date: in France, women's employment already caused a higher increase in wages for couples in the middle of the distribution than for couples at the outer boundaries of the distribution.

In France, between 1982 and 2014, the employment rate of women increased at the same pace regardless of the partner's wage decile, with the notable exception the lower decile and the inactive population and the unemployed, for whom the gap with the highest employment rate increased to 18 and 28 percentage points respectively. Thus, the effects of the growth in women's employment on the correlation between partners' wages (with a wage of zero attributed to the inactive and the unemployed) continued to be limited and were not sufficient to increase inequalities between couples.

The picture is somewhat different for men (Figure IV, right-hand graph). Overall, we note a slight fall in the employment rate over time, which contrasts with the increase observed among women. Moreover, the employment rate of men increases along with their partner's wage up to the median, and becomes

Figure III
Change in wage inequalities and in the correlation between partners' wages since 1982 within couples where both partners are wage earners



Reading note: the coefficient of variation of wages for women living with a partner, where both partners are wage earners, has remained fairly stable, falling from 0.49 in 1982 to 0.48 in 2014 with a peak of 0.56 in 2002.
Coverage: cohabiting men and women, where the partners are aged between 30 and 59 and both earn a wage and neither is self-employed or retired.
Source: Insee, Labour Force Surveys (*enquêtes Emploi*), 1982-2014.

stable thereafter. However, as is the case among women, the group in which the female partners have no wage stands out quite clearly from the others. The gap between the employment rate of this group and the maximum rate increased significantly over time (from 4 to 16 percentage points), reflecting a certain trend towards the polarisation of employment between couples (Ravel, 2007).

Thus, unlike in the United States, the change in the employment rate of women has not constituted a significant cause of change in the correlation between partners' wages. The almost general increase in the employment rate of women, regardless of the partner's wage, has limited the increase in the correlation between partners' wages, thus avoided an increase in inequalities between couples. However, whether for women or men, we note a fall in the employment rate of the partners of individuals earning the lowest wages, or no wage, the effect of which is necessarily to increase inequality. In the next section, a more detailed decomposition of inequality will enable us to evaluate the effects of these trends at different points of the distribution of couples' wages.

A fall in inequalities that has not benefited couples situated just below the median

Diverging trends depending on the part of the distribution taken into consideration

The decomposition of the coefficient of variation described in the last section has the advantage of being very simple. However, it does not allow us to examine whether the effects highlighted were uniform throughout the whole wage distribution, or evaluate the change in inequalities that would have been observed in other counterfactual situations than those already mentioned. In this section, we draw on the approach adopted by Schwartz (2010), which consists in modelling the joint wages of both partners using log-linear models (Agresti, 2002). These models, the parameters of which are subjected to constraints, serve to simulate several counterfactual situations by imposing constraints on parameters, which makes it possible to evaluate the contribution of each of the different trends to changing inequalities between couples.

Figure IV
Change in the rate of employment of men and women by wage of their partner

Note: the unemployed and the inactive are included in the scope and are attributed a zero wage. The curves are smoothed using a local regression of degree 1 (LOESS).
Reading note: in 1985, within couples aged 30 to 59, 58% of women partnered with men in the 7th men's wage decile were in employment.
Coverage: cohabiting men and women, where the partners are aged between 30 and 59 and neither is self-employed or retired.
Source: Insee, Labour Force Surveys (*enquêtes Emploi*), 1982-2014.

This decomposition, which is far more flexible than the previous one, is no longer necessarily limited to the coefficient of variation. In order to study whether the changes differed depending on the part of the distribution of the couples' wages under consideration, we make use of three new measures: the ratio between the upper decile and the median (D9/D5), the ratio between the median and the second decile (D5/D2)[6], as well as the share of couples where neither partner has a wage.

The added value of this more detailed approach can be seen in Figure V, where we can observe the very different changes in the three indicators. Inequality above the median (D9/D5) increased a little in the 1980's, then decreased more significantly by 11% between 1986 and 2014. Inequality below the median (D5/D2) increased by 8% up until 1994 and then fluctuated or fell slightly thereafter. Lastly, the share of couples with no wage increased from 3.9% to 5.9% between 1982 and 1987, then peaked again during the 1990's and fell again a little in the 2000's, before increasing again after 2008.

Thus, the stability of inequality between couples observed above (Figure II) masks diverging trends. Although we have noted a fall in the ratio of the median to the upper deciles[7], the relative situation of the couples situated below the median first deteriorated a little, then improved, but not in a marked way. Below we will try to understand the factors behind this phenomenon.

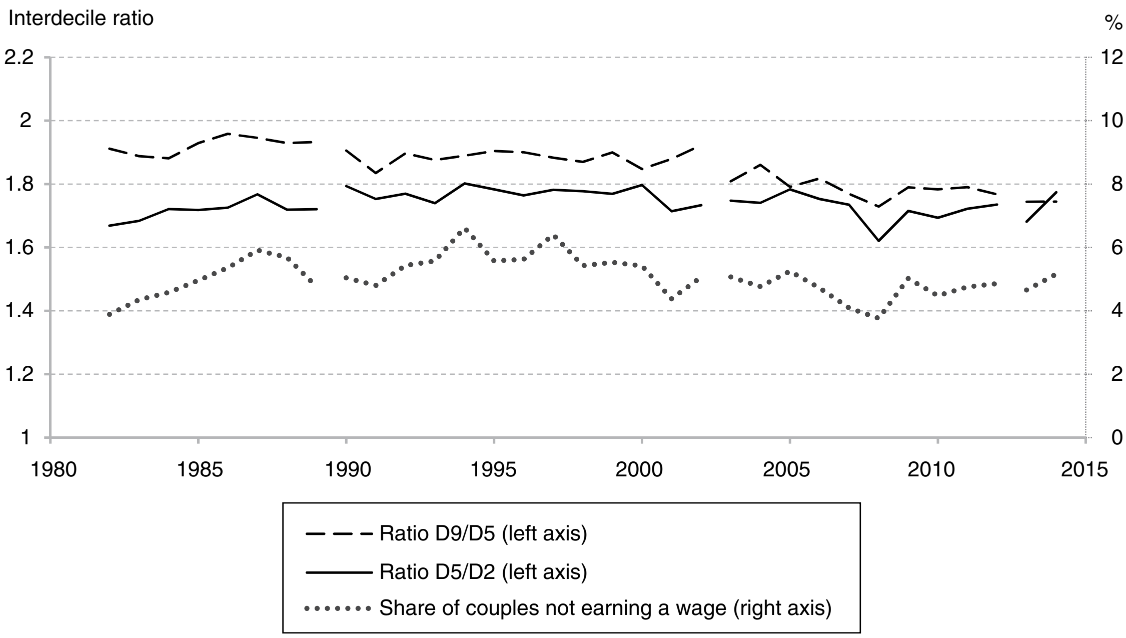
Decomposition method using log-linear modelling: six counterfactual scenarios

The decomposition method proposed by Schwartz (2010) consists in: dividing the wage distribution for each sex observed each year

6. *It is often the first decile (D1) that is considered rather than the second. This choice would be problematic here insofar as, for some years, around 10% of couples have no wage: the first decile is subject to irregular changes linked to the variations in the very low wages depending on the economic situation and the unemployment rate. Since inequality at the bottom of the distribution is already reflected in the share of couples with no wage, we deemed it preferable to make use of the second decile, where the changes are more regular.*

7. *This fall occurred for all the deciles above the median.*

Figure V
Change in three indicators of inequality between couples since 1982



Note: the unemployed and the inactive are included and are attributed a zero wage.
Reading note: in 1982, the wage separating the wealthiest 10% of couples from the poorest 90% was equal to 1.9 times the median wage of couples.
Coverage: cohabiting men and women, where the partners are aged between 30 and 59 and neither is self-employed or retired.
Source: Insee, Labour Force Surveys (*enquêtes Emploi*), 1982-2014.

into fairly fine quantiles, plus a group for the inactive and the unemployed; and constructing a three dimensional homogamy table, cross-classifying the quantiles to which each of the partners belong for each year. The coefficient of variation and the interdecile ratios can be re-calculated using this table by attributing to each individual the median wage for their quantile. In order to replicate fairly faithfully the values of the indices measured using raw data, we divide the distribution into 20 quantiles, each of which represents 5% of the individuals of each sex having earned a wage, plus a category comprising those that have not earned a wage.

The objective of this method is to use the table charting actually observed homogamy to model several tables, each of which corresponds to a counterfactual situation. We start with a model that imposes very strong assumptions on the association between partners' wages and on changes in this association; the assumptions are then relaxed gradually by introducing additional terms until we arrive at the changes actually observed: each model is an extension of the previous model (see Box 2). By comparing the values of the inequality indices obtained for each situation, we obtain an estimation of the contribution of the lifting of each of the assumptions to the inequalities and their changes over time.

The trends followed by the indicators between 1982 and 2014 for the different counterfactual scenarios are presented in Figure VI. The model-fit statistics for the log-linear models are listed in the appendix. It should just be noted that each more complex model constitutes a substantively and statistically significant

Box 2

LOG-LINEAR MODELS USED FOR THE DECOMPOSITION

The base model (model 1), which is the independence model, only includes the parameters needed to correctly reconstruct the share of couples in each wage quantile for men and women (including the zero wages) and for each year, but without any interaction between these three dimensions. With m_{hft} designating the counts predicted by the model for the cell at the intersection with line *h* (quantile for men), column *f* (quantile for women) and layer *t* (survey year) in a table with dimensions $H{\times}F{\times}T$, this model is expressed as follows:

$$M1 : log\, m_{hft} = \lambda + \lambda_h^H + \lambda_f^F + \lambda_t^T$$

In addition, model 2 takes account of the change in the marginal distribution for women over time.

Model 3 adds that for men. They are expressed as follows:

$$M2 : M1 + \lambda_{ft}^{FT} \qquad \text{and} \qquad M3 : M2 + \lambda_{ht}^{HT}$$

Given that the populations for both sexes are segmented into quantiles each year, the marginal parameters have, at this stage, little effect (to within the approximation linked to the segmentation), with the exception of those that reflect the marginal shares of the inactive population and the unemployed: that is why this model amounts primarily to allowing these shares to vary over time.

Model 4 adds on to the specifications of the previous model the association between the inactivity or unemployment of the woman and the man's wage quantile, based on the assumption that the association remains stable over time. It is expressed as follows:

$$M4 : M3 + \lambda_h^{HF_0} 1_{f=0}$$

with $f = 0$ indicating the absence of a wage for the woman (inactive or unemployed), and $1_{f=0}$ the corresponding indicative.

Model 5, which is the stability model, includes the full, but stable association between men's and women's wages. It thus incorporates the association between inactivity or unemployment and the partner's wage from the previous model (but which no longer appears as a specific term). It is expressed as follows:

$$M5 : M3 + \lambda_{hf}^{HF}$$

Although these five models assume that the association between partners' wages has remained stable in terms of odds ratios, the inequality indices may change over time, since they are not independent of the margins of the table.

Model 6 also allows the association between inactivity or unemployment of the woman and the man's wage quantile to vary linearly over time. It is expressed as follows:

$$M6 : M5 + \left(\lambda^{F_0} + t\lambda^{F_0T}\right)\lambda_h^{HF_0T} 1_{f=0}$$

with $f = 0$ indicating the absence of a wage for the woman.

Finally, the last model corresponds to the data actually observed (saturated model). A comparison with the previous model allows us to measure the effects on inequalities between couples resulting from the change in the association between partners' wages in couples in which the woman is a wage earner.

improvement on the previous one, as indicated by the *Akaike Criterion Information*, AIC[8].

The first scenario (cf. model 1 in Box 2) assumes that there is no association between partners' wages or between the wage of one of the partners and the fact that the other is employed, and that the employment rate does not change over time. As a consequence, the only factor of inequality and change in inequalities taken into account here corresponds to the wage inequalities between men in employment, on the one hand, and between women in employment on the other, based on the assumption that the couples are formed randomly. We can see that this factor has contributed very little to the changes in inequalities over time, as the indicators are stable (Figure VI, first row). The growth in women's employment and the changing association between partners' wages thus explain most of the changes.

In the second scenario, we once again assume that there is no link between the wage level of a partner and the status of the other partner, but we allow the employment rate to change over time. This specification allows us to evaluate the contribution of the increase in the employment rate of women to the change in inequalities between couples. The change in this sole factor brings about a very marked reduction in inequalities between 1982 and 2014, which shows very clearly the equalising role of women's employment.

8. However, the BIC (Bayesian Information Criterion) indicator, which is more parsimonious, does not give us reason to consider as worthy of note the respective contributions of models 3 and 6, for which the share of couples classified in the wrong cell (dissimilarity index) decreases very little.

Figure VI
Change in the variation coefficient, the ratios D9/D5 and D5/D2 and the share of couples not earning a wage between 1982 and 2014 in the different counter-factual situations

Note: the unemployed and the inactive are included and are attributed a zero wage. In order not to take account of the variations caused by the random sampling and the economic situation, the arrows do not directly represent the indices calculated based on the population stocks predicted by the models, but instead their value smoothed using a local regression of degree 2 (LOESS).
Reading note: if only the wage inequalities between men and between women had varied (first line), the overall level of the inequalities (coefficient of variation) would have remained stable (at 0.50) between 1982 and 2014.
Coverage: cohabiting men and women, where the partners are aged between 30 and 59 and neither is self-employed or retired.
Source: Insee, Labour Force Surveys (*enquêtes Emploi*), 1982-2014.

Thus, the inequalities (coefficient of variation) would have fallen by 14% overall, by 5% below the median (ratio D5/D2) and by 8% above the median (ratio D9/D5), whereas the share of couples with no wage would have decreased from 4.9% to 2.3%.

Compared to the second scenario, the third adds the change in the employment rate of men over time. The slight fall in the employment rate of men mitigates the fall in inequalities predicted under the previous scenario (*cf.* Figure IV). In particular, the overall inequality (coefficient of variation) would have fallen by only 7% (compared to 14% in the second scenario), and inequality below the median and the share of couples without wage would have remained stable. We can thus see that the growth in women's employment has offset the effects of the trend of increasing inequality occurring among men.

The fourth scenario is based on the assumption that the only association between partners' wages stems from the link between women's employment and the partner's wage, while also assuming that this link remains stable over time. It enables us to measure the effect on the inequalities and on their changes resulting from the differences in the employment rate of women depending on their partner's wage. This scenario does not substantially modify the changes over time predicted by the previous scenario. However, inequalities below the median (ratio D5/D2) and overall inequalities (coefficient of variation) are higher both in 1982 and 2014. This is due to the over-representation of women without wage among the couples in which the men are situated at the bottom of the wage distribution.

The fifth scenario expands on the previous one by assuming that there is an association between partners' wages (including no wage), but that this association remains stable over time. Contrary to expectations, this model (as well as the previous one) predicts a change in inequalities that is potentially different from model 3. Indeed, even though the association itself remains stable over time, the change in the distribution of men and women between employment and absence of employment moves the couples between areas of the homogamy table, which differ in terms of the intensity of the association between partners' wages. Thus, the comparison between the inequality predicted by this model and by the previous one enables us to evaluate the effect of the existence of an association between partners' wages on the changes in inequalities, even based on the assumption that this association remains stable. This scenario predicts greater inequalities that the previous one above the median (ratio D9/D5), as well as overall (coefficient of variation), due to the strong tendency for men and women earning the highest wages to form couples with each other. The deviation, however, is still modest: compared to the fourth scenario, the coefficient of variation is 5% higher in 1982 and 9% in 2014.

As for the time trends, they remain essentially the same as in the previous scenario. The lessening of overall inequalities (coefficient of variation) and those below the median (ratio D5/D2) is slightly less marked, since the women that enter the labour market earn a wage that is closer to that of their partner than in the scenario that assumes a total absence of association. However, at the end of the day, the existence of a tendency towards homogamy only has a weak impact on changes to inequalities over time.

The sixth scenario introduces an initial source of change over time to the association between partners' wages: that of the link between the woman's employment and the man's wage, based on the assumption that the association between partners' wages remains stable. As could be expected due to the relative stability of the link between men's wages and women's employment already mentioned earlier, this scenario does not give rise to any notable differences compared to the previous one.

Finally, the last scenario corresponds to the table actually observed (saturated model). The difference compared to the previous scenario thus stems exclusively from the change over time in the association between partners' wages among couples where the woman is a wage earner. This comparison gives rise to a negative result, which warrants highlighting: contrary to the observation made in the United States, the change in the association between the wages of dual-earner couples (consequence of social homogamy) has no notable effect on inequalities between couples. However, we can note that the fall in the level of inequality above the median (ratio D9/D5) is slightly more marked, sign of a weakening of the association at the top of the distribution.

The finer decomposition performed in this section has shown that the increase in the employment rate of women has been by far the main factor contributing to the change in wage inequalities between couples (where neither of the partners is self-employed or retired) since 1982. The slight fall in the employment rate of men has, on the other hand, exacerbated inequalities, but not in a very marked way. Lastly, the changes in the different components of the link between partners' wages have played a fairly minor role.

Although the growth in women's employment has had an equalising effect on the distribution as a whole, the slight fall in the employment rate of men whose partner earns a low wage or has no wage, had a particularly marked disequalising effect below the median. The association between partners' wages also seems to have weakened a little at the top of the distribution. The combination of these three phenomena explains the contrasting changes in the indicators of inequality mentioned earlier: decrease in inequalities between couples above the median, stability elsewhere.

* *
*

The results presented here, which contradict the idea that there is a general increase in both homogamy and wage inequalities, suggest extensions in three directions.

Firstly, inequalities are measured here in respect of wages only. However, the increase in economic inequalities stems for the most part from the change in the distribution of wealth and capital income (Landais, 2007 [2008]; Piketty, 2013 [2014]). Expanding the study to include all types of income and inheritance seems to be a promising approach, even though the data needed to study these are scarcer (Frémeaux, 2014). Moreover, it would appear essential to take account of transfer income - and particularly unemployment benefits - in order to refine the analysis of low incomes, which are only covered here in the form of the share of couples with no wage.

Secondly, we have adopted a purely cross-sectional approach, thus disregarding the significant variations in partners' incomes during their lives, including as a result of any periods of unemployment or inactivity (Courtioux & Lignon, 2015a). Future research will have to endeavour to combine a study of the inequality changes over time with a longitudinal perspective at individual level in order to better take account of the possible compensations between partners for fluctuating activity status (inactivity, unemployment, working time).

Thirdly, our analysis focused exclusively on the individuals living with a partner in a given year. Based on the research from the United States (Karoly & Burtless, 1995; Burtless, 1999; Cancian & Reed, 1999; Reed & Cancian, 2001; Daly & Valletta, 2006; Western et al., 2008; Larrimore, 2014), we might think that the decrease in the share of persons living with a partner has exacerbated inequalities among all households. For example, single-parent families, which are more often women with low qualifications who are marginalised from the labour market (Chardon et al., 2008), may have seen their situation worsen in respect of the median wage of couples. ☐

BIBLIOGRAPHY

Afsa Essafi, C. & Buffeteau, S. (2006). L'activité féminine en France : quelles évolutions récentes, quelles tendances pour l'avenir ? *Économie et statistique*, 398(1), 85–97.
https://www.insee.fr/fr/statistiques/1376348?sommaire=1376355

Agresti, A. (2002). *Categorical Data Analysis* (2e éd.). New York: Wiley.

Albouy, V., Djider, Z. & Mainguené, A. (2012). Activité, emploi, salaires et retraites : la convergence des situations entre hommes et femmes s'opère, mais parfois bien lentement. In: *Regards sur la parité 2012*, pp. 67–80. Paris: Insee.
https://www.insee.fr/fr/statistiques/1372769?sommaire=1372781

Amar, M. (2010). Les très hauts salaires du secteur privé. *Insee Première* N°1288.
https://www.insee.fr/fr/statistiques/1281185

Blossfeld, H.-P. & Buchholz, S. (2009). Increasing Resource Inequality among Families in Modern Societies: The Mechanisms of Growing Educational Homogamy, Changes in the Division of Work in the Family and the Decline of the Male Breadwinner Model. *Journal of Comparative Family Studies*, 40(4), 603–616.

Boiron, A., Huwer, M. & Labarthe, J. (2016). Inégalités de niveaux de vie et pauvreté en 2013. In: *Les revenus et le patrimoine des ménages 2016*, pp. 9–21. Paris: Insee.
https://www.insee.fr/fr/statistiques/2017606?sommaire=2017614

Bouchet-Valat, M. (2014). Les évolutions de l'homogamie de diplôme, de classe et d'origine sociales en France (1969-2011) : ouverture d'ensemble, repli des élites. *Revue française de sociologie*, 55(3), 459–505. / Changes in Educational, Social Class and Social Class of Origin Homogamy in France (1969-2011): Greater Openness Overall but Increased Closure of Elites. *Revue française de sociologie*, 55(3), 324–364.

Breen, R. & Andersen, S. H. (2012). Educational Assortative Mating and Income Inequality in Denmark. *Demography*, 49(3), 867–887.

Breen, R. & Salazar, L. (2010). Has Increased Women's Educational Attainment Led to Greater Earnings Inequality in the United Kingdom? A Multivariate Decomposition Analysis. *European Sociological Review*, 26(2), 143–157.

Breen, R. & Salazar, L. (2011). Educational Assortative Mating and Earnings Inequality in the United States. *American Journal of Sociology*, 117(3), 808–843.

Burtless, G. (1999). Effects of growing wage disparities and changing family composition on the U.S. income distribution. *European Economic Review*, 43(4-6), 853–865.

Cabannes, P.-Y. (2014). Trois décennies d'évolutions du marché du travail. In: *Trente ans de vie économique et sociale*, pp. 55–67. Paris: Insee.
https://www.insee.fr/fr/statistiques/1374370?sommaire=1374377

Cancian, M. & Reed, D. (1998). Assessing the Effects of Wives' Earnings on Family Income Inequality. *Review of Economics and Statistics*, 80(1), 73–79.

Cancian, M. & Reed, D. (1999). The impact of wives' earnings on income inequality: Issues and estimates. *Demography*, 36(2), 173–184.

Chardon, O., Daguet, F. & Vivas, É. (2008). Les familles monoparentales : des difficultés à travailler et à se loger. *Insee Première* N°1195.
https://www.insee.fr/fr/statistiques/1281271

Charnoz, P., Coudin, É. & Gaini, M. (2013). Une diminution des disparités salariales en France entre 1967 et 2009. In: *Emploi et salaires 2013*, pp. 75–86. Paris : Insee.
https://www.insee.fr/fr/statistiques/1374085?sommaire=1374092

Courtioux, P. & Lignon, V. (2015a). Homogamie éducative et inégalités de revenu salarial : une perspective de cycle de vie. *Économie et Statistique*, 481, 149–183.
https://www.insee.fr/fr/statistiques/1305199?sommaire=1305205

Courtioux, P. & Lignon, V. (2015b). L'effet de l'homogamie éducative sur les inégalités : une note méthodologique avec une illustration pour le cas français. Centre d'économie de la Sorbonne, Document de travail N° 2015.91.
https://halshs.archives-ouvertes.fr/halshs-01254789/

Daly, M. C. & Valletta, R. G. (2006). Inequality and Poverty in United States: The Effects of Rising Dispersion of Men's Earnings and Changing Family Behaviour. *Economica*, 73(289), 75–98.

Devereux, P. J. (2004). Changes in relative wages and family labor supply. *Journal of Human Resources*, 39(3), 698–722.

Eika, L., Mogstad, M. & Zafar, B. (2014). Educational assortative mating and household income inequality. Federal Reserve Bank of New York, Staff Report 682.
http://www.nber.org/papers/w20271

Esping-Andersen, G. (2007). Sociological Explanations of Changing Income Distributions. *American Behavioral Scientist*, 50(5), 639–658.

Frémeaux, N. (2014). Le rôle de l'héritage et du revenu du travail dans les choix matrimoniaux. *Population*, 69(4), 551–587.

Frémeaux, N. & Lefranc, A. (2015). *Assortative mating and earnings inequality in France*. Université de Cergy-Pontoise – THEMA, Mimeo.
http://www.ecineq.org/ecineq_lux15/FILESx2015/CR2/p224.pdf

Funes Leal, V. (2015). *Efectos del Emparejamiento Selectivo en la Distribución del Ingreso en América Latina*. Universidad Nacional de La Plata, Thesis de Maestría en Economía.
http://www.depeco.econo.unlp.edu.ar/maestria/tesis/117-tesis-Funes-Leal.pdf

Godechot, O. (2012). Is finance responsible for the rise in wage inequality in France? *Socio-Economic Review*, 10(3), 447–470.

Godechot, O. (2013). Financiarisation et fractures socio-spatiales. *L'Année sociologique*, 63(1), 17–50. / [2014] Financialization and Sociospatial Divides

Greenwood, J., Guner, N., Kocharkov, G. & Santos, C. (2014). Corrigendum to Marry Your Like: Assortative Mating and Income Inequality. *American Economic Review: Papers & Proceedings*, 104(5), 1–2.

Harkness, S. (2013). Women's Employment and Household Inequality. In: J. Gornick & M. Jäntti (Ed.), *Income Inequality: Economic Disparities and the Middle Class in Affluent Countries* (p. 207–233). Stanford: Stanford University Press.

Hryshko, D., Juhn, C. & McCue, K. (2014). Trends in Earnings Inequality and Earnings Instability among US Couples: How Important Is Assortative Matching? Forschungsinstitut zur Zukunft der Arbeit (IZA), Discussion Paper 8729.
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2543924

Hyslop, D. R. (2001). Rising U.S. Earnings Inequality and Family Labor Supply: The Covariance Structure of Intrafamily Earnings. *The American Economic Review*, 91(4), 755–777.

Insee (2016). Marché du travail - Séries longues. *Insee Résultats* N° 175.
https://www.insee.fr/fr/statistiques/1406870

Karoly, L. A. & Burtless, G. (1995). Demographic change, rising earnings inequality, and the distribution of personal well-being, 1959–1989. *Demography*, 32(3), 379–405.

Landais, C. (2007). Les hauts revenus en France (1998-2006) : une explosion des inégalités ? École d'économie de Paris, Document de travail. / [2008] Top Incomes in France: booming inequalities? Paris School of Economics, Working paper.
http://piketty.pse.ens.fr/fichiers/Landais2007.pdf

Larrimore, J. (2014). Accounting for United States Household Income Inequality Trends: The Changing Importance of Household Structure and Male and Female Labor Earnings Inequality. *Review of Income and Wealth*, 60(4), 683–701.

Meron, M. (2008). Femmes et hommes dans l'emploi : permanences et évolutions. In: *L'emploi, nouveaux enjeux* (Édition 2008), pp. 85–93. Paris: Insee.
https://www.insee.fr/fr/statistiques/1374398?sommaire=1374407

Meron, M. & Maruani, M. (2012). *Un siècle de travail des femmes en France. 1901-2011*. Paris: La Découverte.

Minni, C. (2015). Femmes et hommes sur le marché du travail : les disparités se réduisent mais les emplois occupés restent très différents. *Dares analyses* N° 17.
http://travail-emploi.gouv.fr/IMG/pdf/2015-017.pdf

Morin, T. (2014). Écarts de revenus au sein des couples : Trois femmes sur quatre gagnent moins que leur conjoint. *Insee Première* N° 1492.
https://www.insee.fr/fr/statistiques/1281400

OCDE (2008). *Croissance et inégalités : Distribution des revenus et pauvreté dans les pays de l'OCDE*. Paris: Organisation de coopération et de développement économiques. / **OECD**. *Growing Unequal: Income Distribution and Poverty in OECD Countries.*

OCDE (2011). *Toujours plus d'inégalité : Pourquoi les écarts de revenus se creusent*. / **OECD**. *Divided We Stand: Why Inequality Keeps Rising.* Paris: Organisation de coopération et de développement économiques.

O'Prey, S. (2009). La non-réponse partielle aux variables financières de l'enquête Logement 2006 : mise en œuvre de nouvelles procédures de redressement et comparaison de méthodes d'imputation. Insee, Document de travail F0901.
http://www.insee.fr/fr/publications-et-services/docs_doc_travail/F0901.pdf

Pasqua, S. (2002). Wives' work and income distribution in European countries. *The European Journal of Comparative Economics*, 5(2), 197–226.

Pencavel, J. (2006). A life cycle perspective on changes in earnings inequality among married men and women. *The Review of Economics and Statistics*, 88(2), 232–242.

Piketty, T. (2013). *Le capital au XXI*[e] *siècle*. Paris: Seuil. / [2014] *Capital in the Twenty-First Century.* Cambridge: Harvard University Press.

Pujol, J. & Tomasini, M. (2009). Les inégalités de niveaux de vie entre 1996 et 2007. *Insee Première* N° 1266.
https://www.insee.fr/fr/statistiques/1286908

Ravel, C. (2007). La polarisation de l'emploi au sein des ménages de 1975 à 2002. *Économie et statistique*, 402(1), 3–23.
https://www.insee.fr/fr/statistiques/1376852?sommaire=1376860

Reed, D. & Cancian, M. (2001). Sources of inequality: measuring the contributions of income sources to rising family income inequality. *Review of Income and Wealth*, 47(3), 321–333.

Reed, D. & Cancian, M. (2012). Rising Family Income Inequality: The Importance of Sorting. *Journal of Income Distribution*, 21(2), 3–14.

Schwartz, C. R. (2010). Earnings Inequality and the Changing Association between Spouses' Earnings. *American Journal of Sociology*, 115(5), 1524–1557.

Solard, J. (2010). Les très hauts revenus : des différences de plus en plus marquées entre 2004 et 2007. *In*: *Les revenus et le patrimoine des ménages 2010*, pp. 45–64. Paris: Insee.
https://www.insee.fr/fr/statistiques/1373531?sommaire=1373536

Vanderschelden, M. (2006). Homogamie socioprofessionnelle et ressemblance en termes de niveau d'études : constat et évolution au fil des cohortes d'unions. *Économie et statistique*, 398-399, 33-58.
https://www.insee.fr/fr/statistiques/1376345?sommaire=1376355

Verbakel, E. (2008). *The Partner as Resource or Restriction? Labour market careers of husbands and wives and the consequences for inequality between couples*. Doctoral thesis, Radboud University, Nijmegen.

Verdugo, G. (2014). The great compression of the French wage structure, 1969–2008. *Labour Economics*, 28(4), 131–144.

Verdugo, G., Fraisse, H. & Horny, G. (2012). Évolution des inégalités salariales en France : le rôle des effets de composition. *Revue économique*, 63(6), 1081–1112.

Western, B., Bloome, D. & Percheski, C. (2008). Inequality among American families with children, 1975 to 2005. *American Sociological Review*, 73(6), 903–920.

Worner, S. M. (2006). The Effects of Assortative Mating on Income Inequality: A Decompositional Analysis. Centre for Economic Policy Research – Australian National University, Discussion Paper 538.
https://ideas.repec.org/p/auu/dpaper/538.html

APPENDIX

FIT STATISTICS FOR THE MODELS USED FOR THE DECOMPOSITION

	D. F.[1]	Deviance	Δ (%)[2]	BIC[3]	AIC[4]
M1 : Full independence	14,480	57,114	19.86	- 120,844	28,154
M2 : M1 + Employment rate of women	13,840	49,073	17.57	- 121,019	21,393
M3 : M2 + Employment rate of men	13,200	46,593	17.08	- 115,634	20,193
M4 : M3 + Stable employment-wage association	13,180	40,948	14.56	- 121,033	14,588
M5 : M4 + Full stable association	12,800	16,091	9.51	- 141,219	- 9,509
M6 : M5 + Varying employment-wage association	12,780	15,886	9.33	- 141,179	- 9,674

[1] Degrees of freedom. [2] Index of dissimilarity. [3] Bayesian Information Criterion. [4] Akaike Information Criterion. These last two indicators measure the quality of the description of the data performed by each model taking into account their parsimony (number of parameters to be estimated): a lower value (or a more negative one) indicates a statistically significant improvement compared to a model displaying a higher value.
Note: N = 217,489. Number of cells in the table: 14,553.
Coverage: men and women cohabiting, where the partners are aged between 30 and 59 and neither is self-employed or retired.
Source: Insee, Labour Force Surveys (*enquêtes Emploi*), 1982-2014.

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

* *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 observed[3], whereas 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

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

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* : $h(t) = P(T = t \mid T \geq t)$. 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 $E[T] = \sum_{t\geq 0} tf(t) = \sum_{t\geq 1} S(t)$.

The median is defined here as $Med(T) = \dfrac{u(1 - S(u) - S(u+1)) + 0,5 - S(u)}{S(u) - S(u+1)}$, 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/λ. 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.

lives[4]. 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.

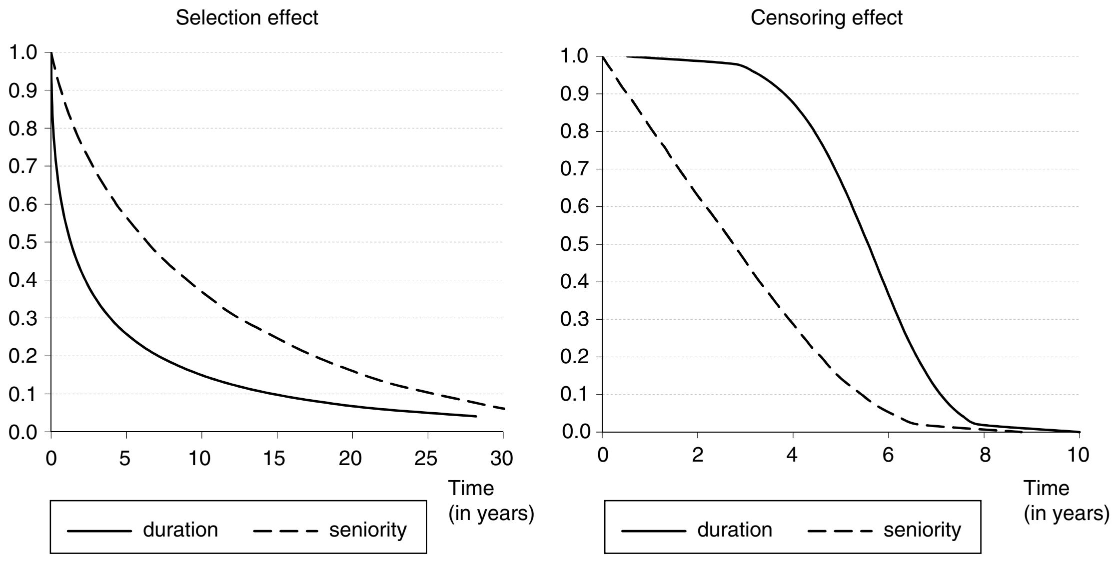
We illustrate these censoring and selection phenomena using simulations that randomly generate observations of durations and seniorities. Based on a uniform distribution, we randomly generate a date on which the situation of single parenthood started between 1950 and 2010 for 10,000 individuals. For 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.

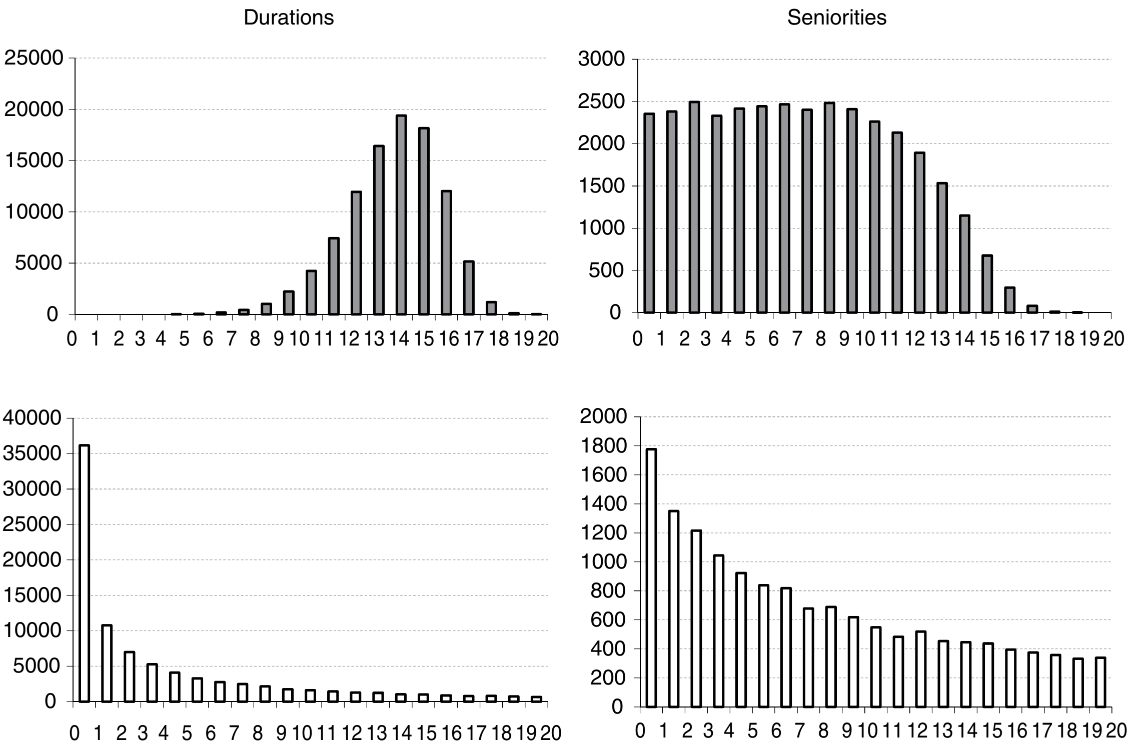
If the selection bias is stronger than the censoring bias, the seniorities observed will, on average, be longer than durations, whereas if the censoring effect is stronger, the seniorities will, on average, be shorter than the durations. If the comparison of two groups shows that, on average, the seniorities 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:

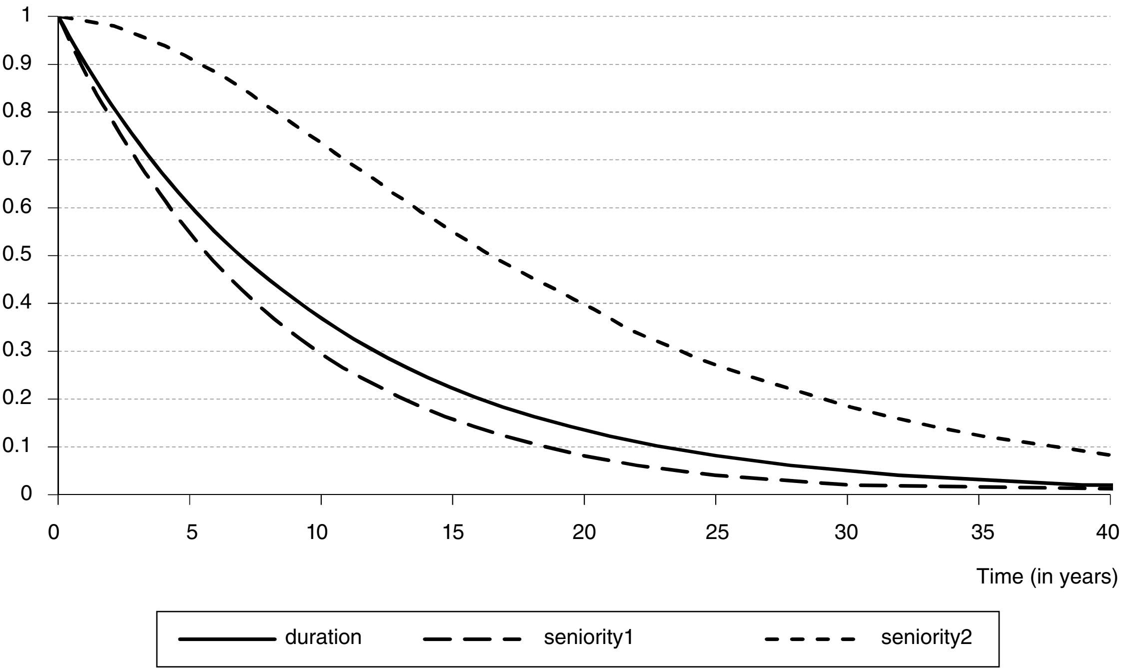
*Stock = Flows * E[T]*, where *E[T]* is the expected duration.

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 A1et 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 uniformely 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 $h(u,x) = P(T = u \mid T \geq u, x)$, 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 $d_1, \ldots, d_n$. 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:

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

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

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

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

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

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:

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

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

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.

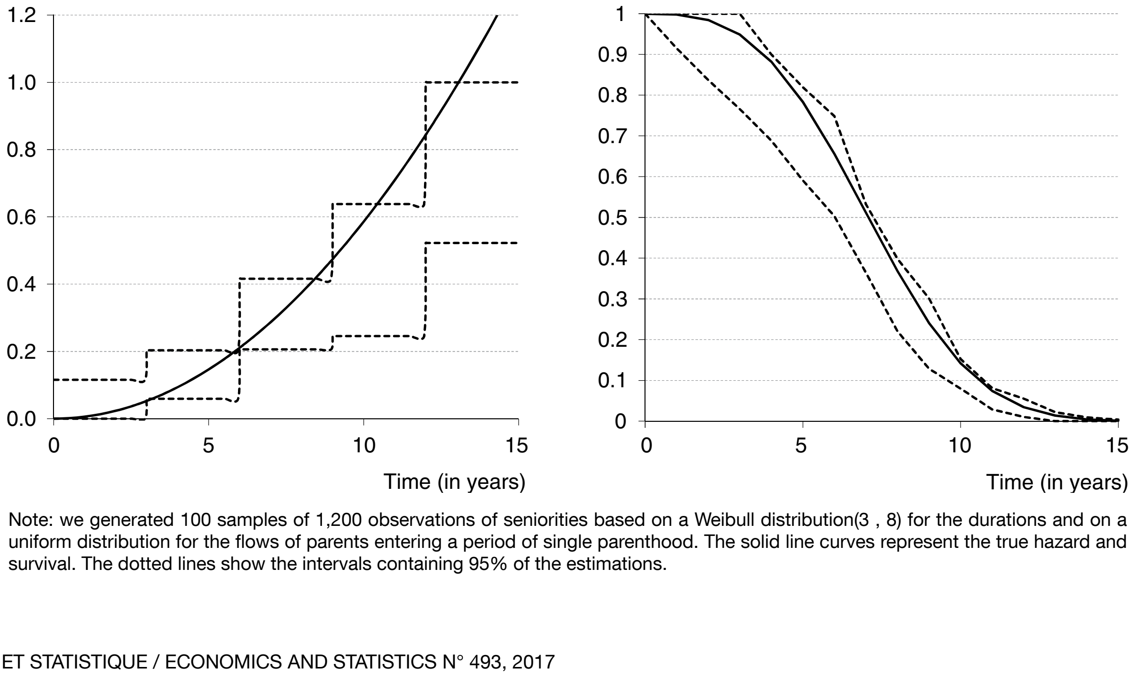
If a person experiences several periods of single parenthood during his/her life, only the last period is taken into account for the contribution to the flow.

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

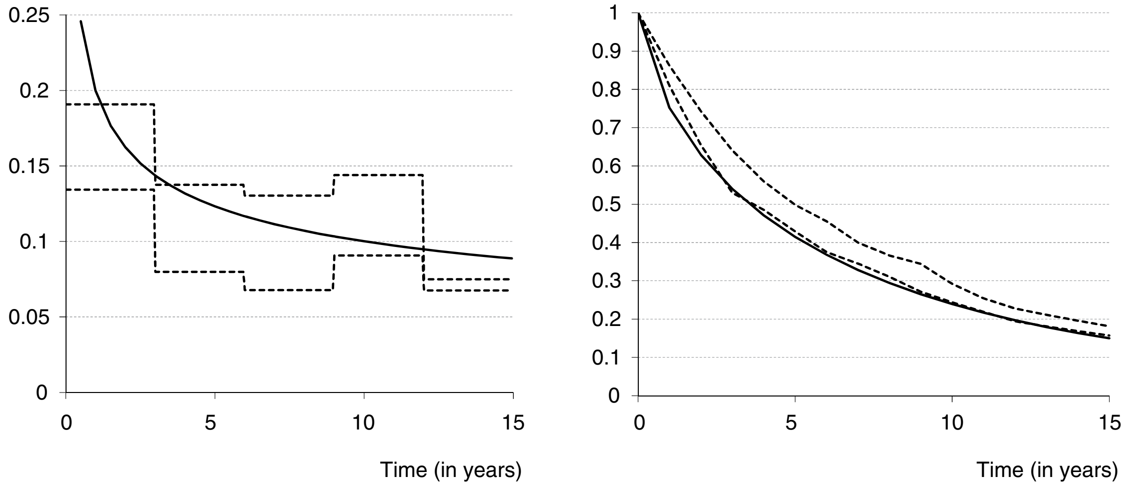
In order to fully understand how the estimations change depending on the sample size, we conduct a series of experiments, each of which consists of simulating flows of people entering the situation and durations for a given number of individuals, then deducing the distribution of discrete durations (based 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).

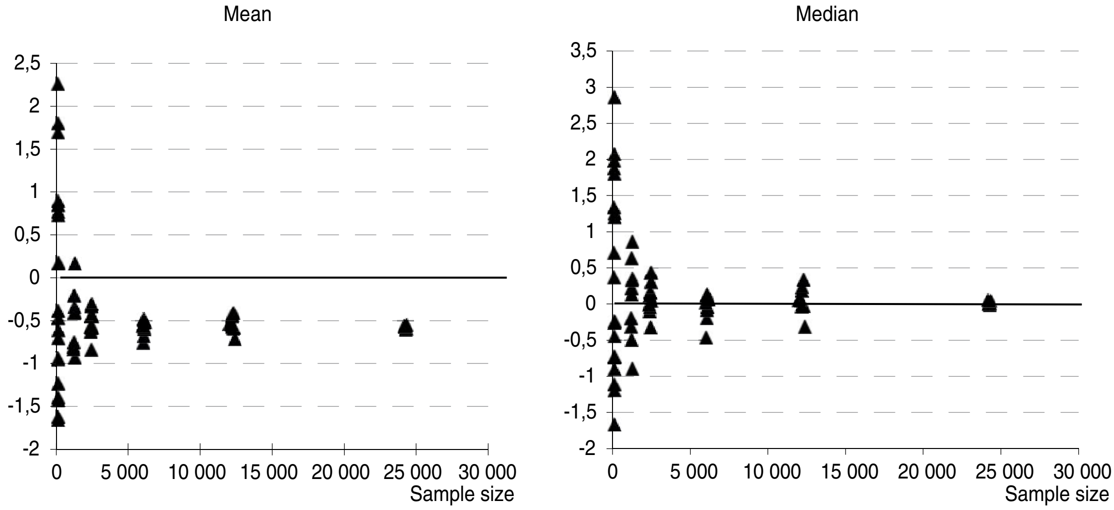
Estimating the risk factors

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):

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

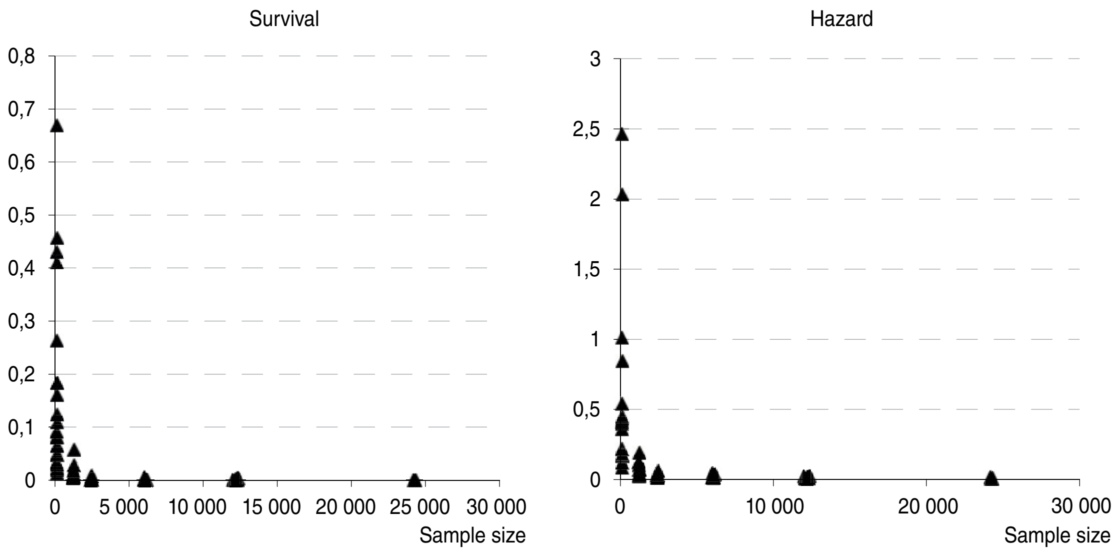
These probabilities are, as established in assumption 3 of our model, proportional to each other. It thus suffices to estimate only β_1 and β_2 to determine the differences in the instantaneous probabilities (and therefore in the durations) between the different groups. If the coefficient is positive, 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 hazarddepending 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 $\sum_{t\geq 0}\left(f(t)-g(t)\right)^2$.

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

		β_1	β_2
True value		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

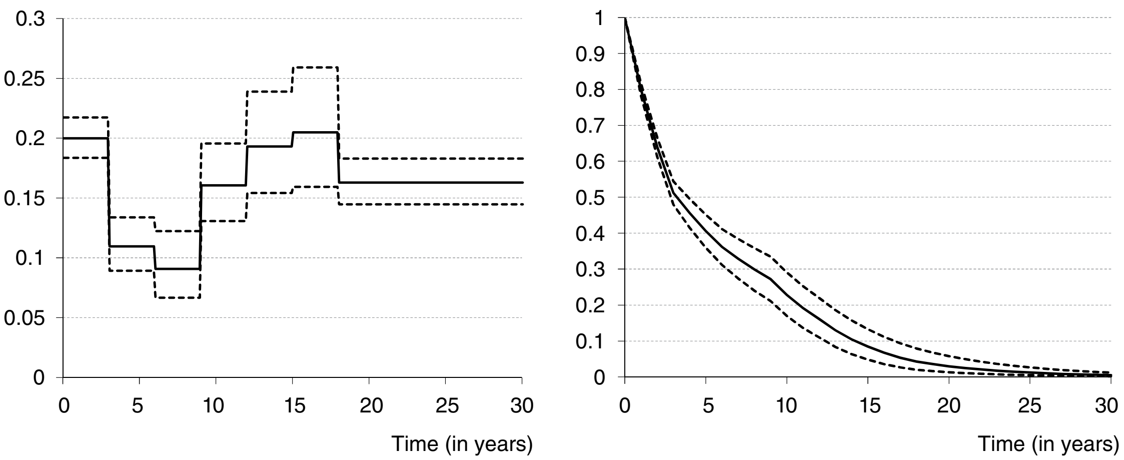
Of the 359,770 respondents to the *FHS*, 12,519 were in a situation of single parenthood at the time of the survey[5], 1,073 of whom were men and 11,446 women. In light of the previous results, we have a sufficiently large sample to be able to infer the distribution of durations for women, but the sample 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

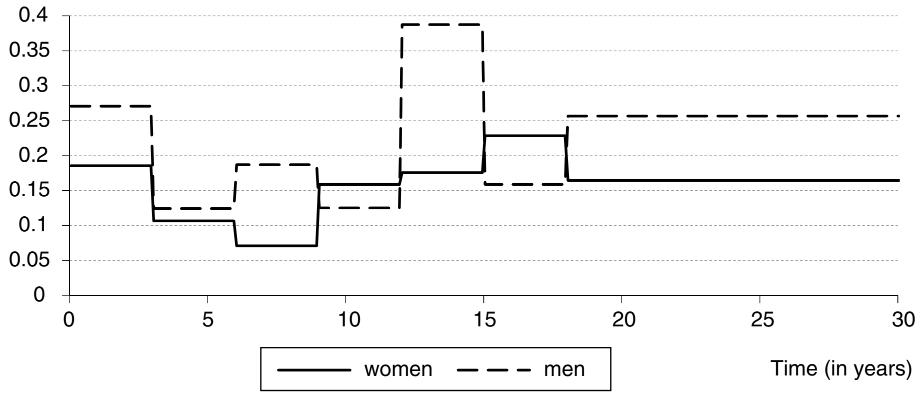
5. We eliminated the 205 respondents who had entered the situation of single parenthood in the year of the survey (in 2011), as they do not contribute any information to our model, as well as the LAT's (parents living apart together), which is a situation whose seniority cannot be determined.

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



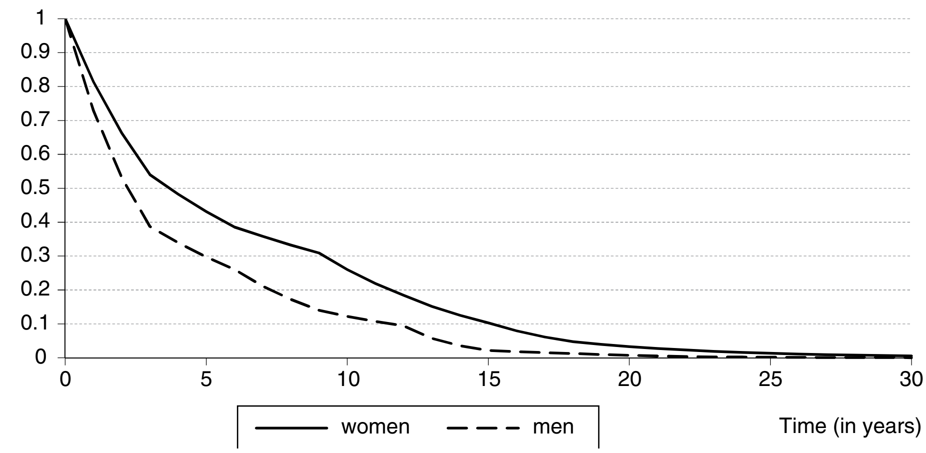
Note: the dotted lines represent the 95% confidence interval obtained with the maximum likelihood method.
Coverage: single parents with minor children, mainland France.
Source: Insee, *Family and Housing Survey* (*FHS*) 2011 and Ined-Insee, *Survey of Family and Intergenerational Relations* (*Erfi*), 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.
Source: 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.
Source: 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 boostrapping 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.
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.

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 categories[6]: 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.

* *
*

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.

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.

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*'e-ssentiel* 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/.

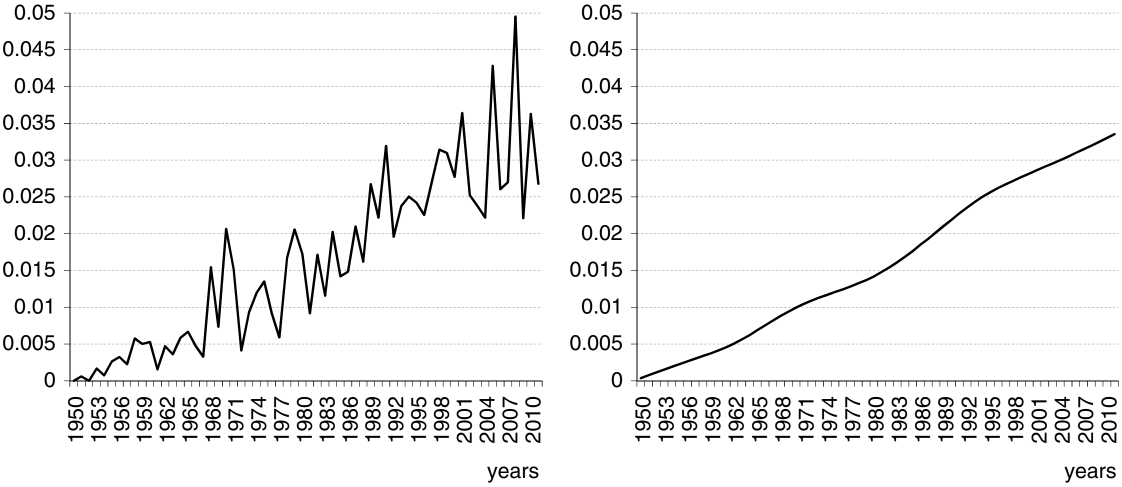
Toulemon, L. (2011). Individus, familles, ménages, logements : les compter, les décrire. *Travail, Genre et Société*, 26, 47−66.

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.

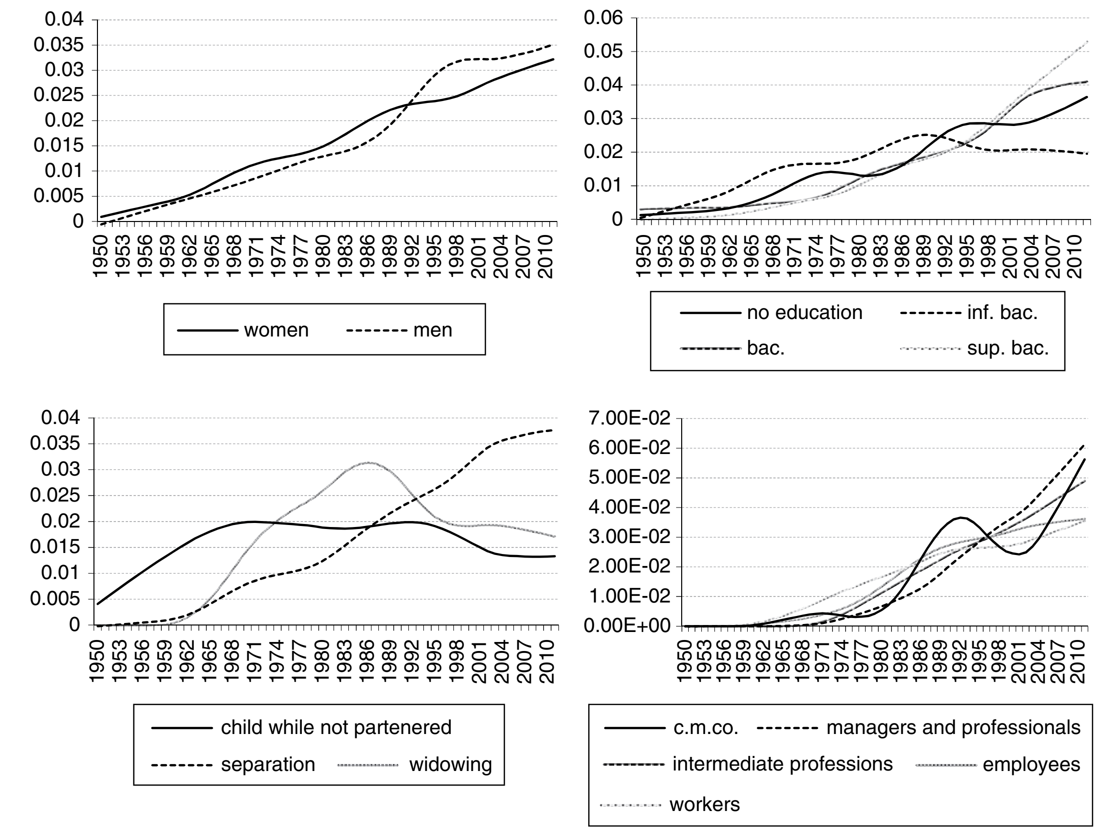
SMOOTHED ESTIMATIONS OF ENTRY FLOWS

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.

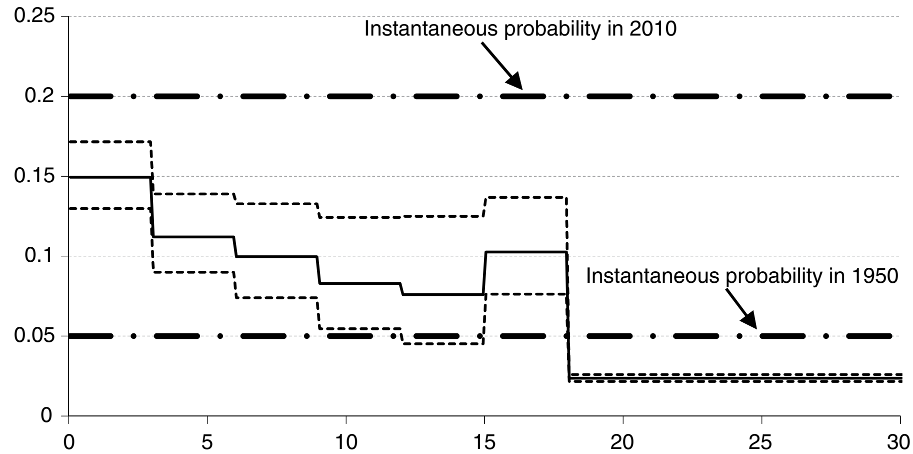
APPENDIX 3

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* . *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.

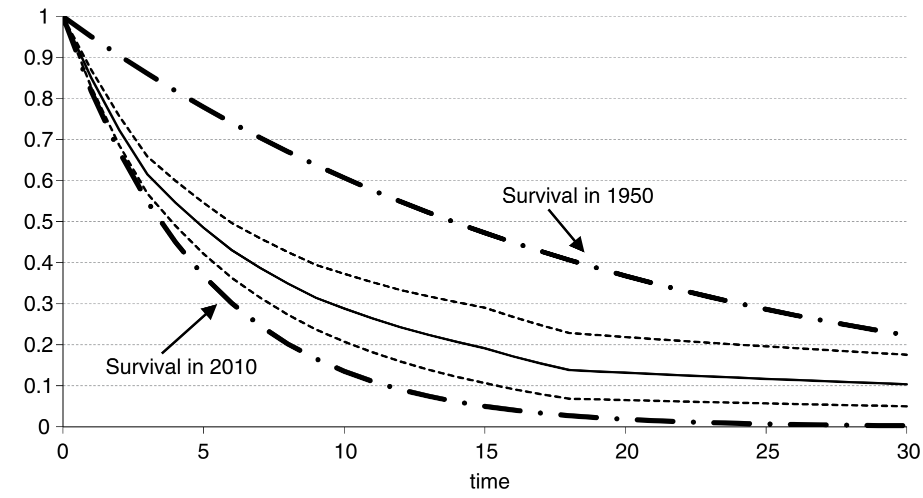
The estimated average duration is 8.2 years. We thus tend to underestimate the durations when there is a negative correlation between the duration *T* and the entry date in single-parenthood *D*. Nevertheless, the estimations actually provide estimates of an instantaneous probability and a survival which 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.*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.*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.

APPENDIX 4

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, $P(T_d \geq t) = P(T_c \geq t) \forall t \in N$. Then we show that $E[T_d] \approx E[T_c] - 0{,}5$.

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

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 $E[T_d] \approx E[T_c] - 0{,}5$.

N °488-489 - 2016

LES SANS-DOMICILE

- Avant-propos. L'enquête *Sans-Domicile* 2012 : histoire et place en Europe
- Introduction de cadrage. Les sans-domicile en France : caractéristiques et principales évolutions entre 2001 et 2012
- Placement dans l'enfance et précarité de la situation de logement
- Un « désordre » dans la catégorisation : le déclassement statutaire atypique de diplômés du supérieur sans domicile

LES ÉCARTS DE RÉMUNÉRATION ENTRE LES HOMMES ET LES FEMMES DANS LA FONCTION PUBLIQUE

- Introduction. Analyser les inégalités salariales entre les hommes et les femmes dans la fonction publique avec les données du *Système d'information sur les agents des services publics*
- Égalité professionnelle entre les hommes et les femmes : des plafonds de verre dans la fonction publique
- Les sources de l'écart de rémunération entre femmes et hommes dans la fonction publique
- La contribution des écarts de rémunération entre les femmes et les hommes à l'inégalité des rémunérations dans la fonction publique : une approche par la décomposition des inégalités

N °490 - 2016

COMPÉTENCES DES ADULTES À L'ÉCRIT ET EN CALCUL

- Introduction. L'enquête *Information et vie quotidienne* 2011 parmi les dispositifs nationaux et internationaux de mesure des compétences des adultes
- Compétences à l'écrit des adultes et événements marquants de l'enfance : le traitement de l'enquête *Information et vie quotidienne* à l'aide des méthodes de la statistique textuelle
- Les difficultés des adultes face à l'écrit : l'apport de l'épreuve de production écrite de l'enquête *Information et vie quotidienne*
- L'évolution des compétences des adultes : effet « génération » et effet « cycle de vie »
- Mesurer les compétences à l'écrit et à l'oral des nouveaux migrants

N °491-492 - 2017

ÂGES ET GÉNÉRATIONS / *AGE AND GENERATIONS*

- Éditorial : *Economie et Statistique* devient *Economie et Statistique / Economics and Statistics*
- Avant-propos / *Foreword : On the importance of taking a life-cycle view in understanding generational issues* / L'importance d'une perspective de cycle de vie pour la compréhension des questions générationnelles
- Âges et générations : une introduction générale / *Age and generations: a general introduction*
- Les différences de retraite entre secteur public et secteur privé : une analyse par simulations sur carrières types / *Differences between public and private sector pensions: an analysis on standard career simulations*
- Commentaire – Comparer les efforts contributifs pour comparer les retraites entre secteur public et secteur privé ?/ Comment – *Taking contributions into account in public-private comparison of pensions?*
- Le déficit de cycle de vie en France : une évaluation pour la période 1979-2001 / *Lifecycle deficit in France: an assessment for the period 1979-2011*
- Les inégalités de niveaux de vie entre les générations en France / *Intergenerational inequalities in standards of living in France*
- Calculer le niveau de vie d'un ménage : une ou plusieurs échelles d'équivalence ? / *Calculating the standard of living of a household: one or several equivalence scales?*
- Les méthodes de pseudo-panel et un exemple d'application aux données de patrimoine / *Pseudo-panel methods and an example of application to Household Wealth data*

Economie et Statistique / Economics and Statistics

Objectifs généraux de la revue

Economie et Statistique / Economics and Statistics publie des articles traitant de tous les phénomènes économiques et sociaux, au niveau micro ou macro, s'appuyant sur les données de la statistique publique ou d'autres origines. Une attention particulière est portée à la qualité de la démarche statistique et à la rigueur des concepts mobilisés dans l'analyse. Pour répondre aux objectifs de la revue, les principaux messages des articles et leurs limites éventuelles doivent être formulés dans des termes accessibles à un public qui n'est pas nécessairement spécialiste du sujet de l'article.

Soumettre un article

Les propositions d'articles, en français ou en anglais, doivent être adressées à la rédaction de la revue (redaction-ecostat@insee.fr), en Word et mis en forme suivant les consignes aux auteurs (accessibles sur https://www.insee.fr/fr/information/2410168). Il doit s'agir de travaux originaux, qui ne sont pas soumis en parallèle à une autre revue. Un article standard fait environ 11 000 mots (y compris encadrés, tableaux, figures, annexes et bibliographie, non compris éventuels compléments en ligne). Aucune proposition initiale de plus de 12 500 mots ne sera examinée.

La soumission doit comporter deux fichiers distincts :

- Un fichier d'une page indiquant : le titre de l'article ; le prénom et nom, les affiliations (maximum deux), l'adresse e-mail et postale de chaque auteur ; un résumé de 160 mots maximum (soit environ 1 050 signes espaces compris) qui doit présenter très brièvement la problématique, indiquer la source et donner les principaux axes et conclusions de la recherche ; les codes JEL et quelques mots-clés ; d'éventuels remerciements.
- Un fichier anonymisé de l'article complet (texte, illustrations, bibliographie, éventuelles annexes) indiquant en première page uniquement le titre, le résumé, les codes JEL et les mots clés.

Les propositions retenues sont évaluées par deux à trois rapporteurs (procédure en « double-aveugle »). Une fois acceptés, les articles peuvent faire l'objet d'un travail éditorial visant à améliorer leur lisibilité et leur présentation formelle.

Publication

Les articles sont publiés en français dans l'édition papier et simultanément en français et en anglais dans l'édition électronique. Celle-ci est disponible, en accès libre, sur le site de l'Insee, le jour même de la publication ; cette mise en ligne immédiate et gratuite donne aux articles une grande visibilité. La revue est par ailleurs accessible sur le portail francophone Persée, et référencée sur le site international Repec et dans la base EconLit.

Main objectives of the journal

Economie et Statistique / Economics and Statistics publishes articles covering any micro- or macro- economic or sociological topic, either using data from public statistics or other sources. Particular attention is paid to rigor in the statistical approach and clarity in the concepts and analyses. In order to meet the journal aims, the main conclusions of the articles, as well as possible limitations, should be written to be accessible to an audience not necessarily specialist of the topic.

Submissions

Manuscripts can be submitted either in French or in English; they should be sent to the editorial team (redaction-ecostat@insee.fr), in MS-Word and follow the guidelines for authors (available at https://www.insee.fr/en/information/2591257). The manuscript must be original work and not submitted at the same time to any other journal. The standard length of an article is of about 11,000 words (including boxes if needed, tables and figures, appendices, list of references, but not counting online complements if any). Manuscripts of more than 12,500 words will not be considered.

Submissions must include two separate files:

- A one-page file providing: the title of the article; the first name, name, affiliation-s (at most two), e-mail et postal addresses of each author; an abstract of maximum 160 words (about 1050 characters including spaces), briefly presenting the question(s), data and methodology, and the main conclusions; JEL codes and a few keywords; acknowledgements.
- An anonymised file of the article (including the main text, illustrations, bibliography and appendices if any), mentioning only the title, abstract, JEL codes and keywords on the front page.

Proposals that meet the journal objectives are reviewed by two to three ("double-blind" review referees). Once the article accepted, further editorial changes may be made in order to improve their presentation and readability.

Publication

The articles are published in French in the printed edition, and simultaneously in French and in English in the electronic edition. The online issue is available, in open access, on the Insee website the day of its publication; this immediate and free online availability gives the articles a high visibility. The journal is also available online on the French gateway Persée, and indexed in Repec and EconLit.

N° 493 - 2017

Economie et Statistique

Economics and Statistics

Au sommaire du prochain numéro :
La crise 10 ans après

Forthcoming:
The crisis 10 years after

ISBN 978-2-11-151320-4 - ISSN 0336-1454 - ECO 493
Parution juillet 2017 - PRIX : 9,10 €