

Economie Statistique **ET**

Economics **AND** Statistics

Mélanges

Varia

Economie Statistique ET

Economics AND Statistics

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.

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 PONTHEUX

Assistante de rédaction / Editorial Assistant: Véronique ALEXANDRE

Maquette PAO et impression / CAP and printing: JOUVE

1, rue du Docteur-Sauvé, BP3, 53101 Mayenne

Conseil scientifique / Scientific Committee

Alain CHENU, président (Observatoire sociologique du changement - Sciences Po/CNRS)

Laurence BLOCH (Insee)

Jérôme BOURDIEU (École d'économie de Paris)

Pierre CAHUC (École Polytechnique, Ensaie et Centre de recherche en économie et statistique)

Gilbert CETTE (Banque de France et École d'économie d'Aix-Marseille)

Jacques LE CACHEUX (Université de Pau et des pays de l'Adour)

Yannick L'HORTY (Université de Paris-Est - Marne la Vallée)

Joël MAURICE (École nationale des ponts et chaussées)

Katheline SCHUBERT (École d'économie de Paris- Université Paris 1)

Claudia SENIK (Université Paris 1 Panthéon-Sorbonne et École d'économie de Paris)

Louis-André VALLET (Observatoire sociologique du changement-Sciences Po/CNRS)

François-Charles WOLFF (Université de Nantes)

Comité éditorial / Editorial Advisory Board

Luc ARRONDEL (École d'économie de Paris)

Antoine BOZIO (Institut des politiques publiques/École d'économie de Paris)

Clément CARBONNIER (Théma/Université de Cergy-Pontoise)

Erwan GAUTIER (Banque de France et Université de Nantes)

Pauline GIVORD (Insee/Direction de la méthodologie et de la coordination statistique internationale et Crest)

Florence JUSOT (Université Paris-Dauphine, Leda-Legos et Irdes)

François LEGENDRE (Erudite/Université Paris-Est)

Claire LELARGE (Banque de France et Crest)

Laurent LESNARD (Observatoire sociologique du changement/ Sciences Po-CNRS)

Claire LOUPIAS (Direction générale du Trésor)

Sophie PONTHEUX (Insee et Crest)

Thepthida SOPRASEUTH (Théma/Université de Cergy-Pontoise)

Economie Statistique **ET**

Economics **AND** Statistics

Les jugements et opinions exprimés par les auteurs n'engagent qu'eux mêmes,
et non les institutions auxquelles ils appartiennent, ni *a fortiori* l'Insee.

Economie et Statistique / Economics and Statistics

Numéro 493 - 2017

ÉVALUATION DE POLITIQUES PUBLIQUES EN FAVEUR DES PME

5 L'effet des aides à la R&D sur l'emploi : une évaluation pour les petites entreprises en France

La forte augmentation des aides à la R&D perçues par les petites entreprises entre 2003 et 2010 a eu un effet positif et croissant sur l'emploi hautement qualifié. Cependant, elle s'est accompagnée d'un effet d'aubaine significatif, notamment depuis la réforme du CIR de 2008.

Vincent Dortet-Bernadet et Michaël Sicsic

23 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

Les données descriptives font ressortir un effet significatif de l'aide publique sur la survie à cinq ans des entreprises créées par les chômeurs. Mais lorsque l'on prend en compte un effet de sélection des chômeurs les plus aptes à la création, cet effet apparaît le plus souvent nul.

Dominique Redor

45 Commentaire : L'efficacité des aides publiques à la R&D et à l'entrepreneuriat

Pierre Möhnen

MÉLANGES

53 Turnover élevé du personnel soignant dans les EHPAD privés en France : impact de l'environnement local et du salaire

Les départs fréquents d'infirmiers et d'aides-soignants travaillant en Ehpad, souvent déterminés par l'environnement local, dégradent la qualité de la prise en charge des personnes âgées. Ils pourraient être réduits par une hausse de l'encadrement en personnel et des salaires des aides-soignants.

Cécile Martin et Mélina Ramos-Gorand

71 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

Contrairement à une idée reçue, l'accroissement du taux d'emploi des femmes n'a pas entraîné d'augmentation des inégalités de salaire entre couples en France : elles sont restées globalement stables entre 1982 et 2014 pour les couples de 30 à 59 ans.

Milan Bouchet-Valat

91 Combien de temps durent les situations de monoparentalité ? Une estimation sur données françaises

Un modèle original permet, de façon inédite en France, d'estimer combien de temps un parent avec enfant(s) mineur(s) à charge demeure isolé. Cette phase dure 3 ans ou moins pour la moitié des parents et est plus courte pour les personnes séparées que celles ayant eu un enfant sans être en couple.

Vianney Costemalle

L'effet des aides à la R&D sur l'emploi : une évaluation pour les petites entreprises en France

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

Vincent Dortet-Bernadet * et Michaël Sicsic **

Résumé – Entre 2003 et 2010, le montant des aides délivrées par les pouvoirs publics français pour financer les activités de R&D des PME a été multiplié par quatre. Cette très forte hausse est due aux réformes du crédit d'impôt recherche (CIR), notamment en 2008, à la mise en place d'un dispositif dédié aux jeunes entreprises innovantes en 2004 et à l'augmentation des subventions sur la période. Cet article présente, pour la première fois à partir de données exhaustives sur l'emploi en France, une évaluation de l'effet de l'augmentation de ces aides dans les petites entreprises. À partir d'une méthode combinant un modèle de demande de travail et une méthode d'appariement, nous mettons en évidence que l'effet du soutien public sur l'emploi consacré aux activités de R&D est positif et croissant entre 2004 et 2010. Néanmoins, l'augmentation des aides, particulièrement après l'importante réforme du CIR en 2008, s'accompagne d'un effet d'aubaine significatif : selon nos estimations, seulement entre 18 et 34 % du supplément d'aides obtenu par les entreprises entre 2008 et 2010 aurait servi à financer de nouveaux emplois hautement qualifiés.

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.

Codes JEL / JEL codes : O38, H25, C33, C36

Mots-clés : aides à la R&D, évaluation de politique publique, estimation par appariement, modèle de demande de travail

Keywords: R&D subsidies, R&D tax incentives, public policy evaluation, matching, labour demand model

Rappel :

Les jugements et opinions exprimés par les auteurs n'engagent qu'eux mêmes, et non les institutions auxquelles ils appartiennent, ni a fortiori l'Insee.

* Insee (vincent.dortet-bernadet@insee.fr).

** Insee et Cred, Université Panthéon-Assas Paris II (sicsic.michael@gmail.com).

Les auteurs remercient Louis de Gimel, Claire Lelarge, Corinne Prost, Loriane Py, Fabienne Rosenwald, Frédérique Sachwald, trois rapporteurs anonymes, ainsi que les participants des séminaires D2E de l'Insee, évaluation des politiques publiques de l'AFSE/DG Trésor, et du GT économie de la fiscalité du Cred. Les auteurs remercient la Direction générale pour la recherche et l'innovation du Ministère de l'Éducation nationale, de l'enseignement supérieur et de la recherche pour leur avoir permis d'utiliser la base de gestion du crédit d'impôt recherche. Les auteurs demeurent seuls responsables des erreurs et omissions restantes

Les activités de recherche et développement (R&D) visent l'apparition de nouveaux matériaux, produits et procédés de fabrication et leur amélioration. En stimulant l'innovation et le progrès technique, la R&D est une source importante de croissance économique (Griffith et al., 2003 et 2004). En France, le poids des dépenses intérieures de R&D des entreprises (DIRDE)¹ reste relativement faible par rapport aux autres grands pays : 1.45 % du PIB en 2014 contre 1.6 % pour la moyenne de l'OCDE, près de 2 % en Allemagne et 2.8 % au Japon (OCDE, 2017). Pourtant, au cours des années 2000, les pouvoirs publics ont fortement développé les dispositifs d'aides pour accroître les dépenses privées de R&D. L'État a ainsi alloué aux entreprises près de 8 milliards d'euros d'aides à la R&D en 2013, soit près de 0.4 point de PIB (contre moins de 0.2 en 2003). En 2013, la France est le 3^e pays au monde en termes de financement public de la R&D et le premier en termes d'incitation fiscale à la R&D (OCDE, 2016).

Cet article étudie l'impact sur l'emploi de la forte augmentation des aides publiques à la R&D reçues par les petites entreprises au cours de la période 2004-2010. Les petites entreprises ont en effet bénéficié de la mise en place en 2004 du statut de jeunes entreprises innovantes (JEI) pour les entreprises de moins de huit ans spécialisées dans les activités de R&D. De façon plus importante, les différentes réformes du crédit d'impôt recherche (CIR), à partir de 2004 et particulièrement en 2008, ont permis d'accroître fortement le nombre de petites entreprises utilisant ce dispositif. Enfin, les petites entreprises ont eu aussi la possibilité de recevoir des aides directes attribuées par Oséo² organisme créé en 2005, ayant vocation à accompagner les projets d'innovation des petites et moyennes entreprises (PME).

Peu d'études se sont attachées à évaluer l'effet des aides à la R&D reçues par les petites entreprises. Pourtant, ces entreprises connaissent le taux d'aide le plus élevé : en 2010, 50 % des dépenses de R&D déclarées par les très petites entreprises³ (TPE) bénéficiant du CIR ont été financées par des aides, contre 42 % pour les autres PME, 36 % pour les entreprises de taille intermédiaire (ETI) et 34 % pour les plus grandes entreprises (Dortet-Bernadet & Sicsic, 2015, p. 15).

Pourquoi aider les petites entreprises à faire de la R&D ?

Le soutien public aux entreprises pour leurs dépenses en R&D se justifie par le fait que,

en l'absence d'aide, les entreprises auraient tendance à réaliser moins de R&D que ce qui serait souhaitable pour l'ensemble de l'économie (Jones & Williams, 1998 ; Bloom et al., 2013). En diminuant le coût privé des activités de R&D, les aides publiques sont susceptibles d'augmenter les dépenses de R&D pour atteindre un niveau socialement optimal. Aider plus particulièrement les entreprises petites ou jeunes peut se justifier car elles sont soumises à des contraintes de financement plus fortes que les autres entreprises. Ces contraintes engendraient des investissements en R&D trop faibles et évoluant de façon pro-cyclique⁴ (Aghion et al., 2012). Les aides peuvent aussi bénéficier à des petites entreprises qui n'auraient sinon pas débuté des activités de R&D (González et al., 2005).

Aider les entreprises les plus jeunes peut aussi s'avérer efficace car ces entreprises seraient davantage à l'origine d'innovations radicales (Schneider & Veugelers, 2010 ; Cincera & Veugelers, 2012 ; Akcigit & Kerr, 2010). À partir de simulations issues d'un modèle théorique, Acemoglu et al. (2013) obtiennent qu'il serait plus efficace d'aider les entreprises entrantes (en particulier les jeunes et petites) à faire de la R&D que celles déjà installées. Mais il faut noter que dans une analyse empirique, Garcia-Macia et al. (2016) trouvent que l'essentiel de la croissance de la productivité ne vient pas des entreprises jeunes mais plutôt de l'amélioration de produits d'entreprises déjà établies.

Si le soutien public à la R&D peut avoir pour conséquence d'accroître le financement privé de la R&D (effet amplificateur des aides), il peut aussi simplement s'additionner sans le modifier (effet additif), ou même le réduire (effet d'aubaine ou d'éviction) (David et al., 2000). Dans ce dernier cas, les entreprises utiliseraient les fonds publics pour financer des projets qu'elles auraient réalisés sans aide.

1. La dépense intérieure de recherche et développement (DIRD) correspond aux travaux de recherche et développement (R&D) exécutés sur le territoire national quelle que soit l'origine des fonds. Une partie est exécutée par les administrations, l'autre par les entreprises (la DIRDE). Elle comprend les dépenses courantes (masse salariale des personnels de R&D et dépenses de fonctionnement) et les dépenses en capital (achats d'équipements nécessaires à la réalisation des travaux internes à la R&D et opérations immobilières réalisées dans l'année).

2. Oséo a été créé en 2005, à partir de l'Anvar, de la BDPME et de la Sofaris, puis intégré à BPIFrance en 2013.

3. Voir l'encadré 2 pour la définition des catégories d'entreprises.

4. La part de la R&D dans l'investissement baisserait dans les phases de récession et cette baisse ne serait pas parfaitement compensée lors des phases de reprise économique.

Des résultats variés sur l'effet des aides à la R&D selon la taille d'entreprise

L'évaluation de l'impact des aides à la R&D a fait l'objet d'une importante littérature (Ientile & Mairesse, 2009 ; Kohler et al., 2012 ; Zuñiga-Vicente et al., 2014, pour des revues de littérature). Elle décrit des résultats variés sur l'efficacité des aides à la R&D reçues par les petites entreprises. Une partie de cette variabilité s'explique par l'hétérogénéité des dispositifs d'aide mis en place dans les différents pays de l'OCDE. Ainsi, pour Busom et al. (2014), les jeunes entreprises espagnoles sans expérience de R&D auraient surtout recours aux aides directes et les mécanismes de crédit d'impôt seraient moins adaptés pour débiter une activité de R&D. Toujours en Espagne, Corchuelo et Martinez-Ros (2009) montrent que les incitations fiscales à la R&D sont plus efficaces pour les grandes entreprises que pour les PME. Toutefois, selon Lokshin et Mohnen (2012) qui travaillent sur un mécanisme de crédit d'impôt proportionnel aux volumes des dépenses aux Pays-Bas, les incitations fiscales sont surtout efficaces pour les petites entreprises, seule catégorie d'entreprise pour laquelle les effets d'aubaine peuvent être rejetés. Hægeland et Møen (2007) parviennent à une conclusion similaire pour un crédit d'impôt en Norvège. Cependant, comme Lokshin et Mohnen (2013), ils montrent aussi que les aides ont un effet significatif sur l'augmentation des salaires versés aux chercheurs. En Italie, Bronzini et Iachini (2014) mettent en évidence

un effet additif d'un programme de subvention à la R&D pour les petites entreprises, mais pas pour les grandes. Enfin, à partir d'un dispositif existant au Québec, Baghana et Mohnen (2009) mettent en évidence qu'un crédit d'impôt proportionnel aux volumes des dépenses de R&D ne serait pas efficace pour les grandes entreprises mais le serait pour les petites entreprises : pour ces dernières, l'augmentation des dépenses de R&D serait supérieure au niveau de l'aide.

Les études sur données françaises utilisent des données partielles sur les petites entreprises

En France, les évaluations sont la plupart du temps réalisées à partir des données de l'enquête R&D du Ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche (MENESR). Cette enquête fournit une information très détaillée sur les dépenses des grands producteurs de R&D. Toutefois, elle ne couvre que partiellement la population des entreprises jeunes ou petites et l'évolution de son champ rend difficile son utilisation exclusive pour une évaluation de l'impact des aides à la R&D sur les petites entreprises (encadré 1). Les travaux d'évaluation de l'efficacité des aides à la R&D réalisés à partir de cette enquête se sont ainsi concentrés sur les entreprises moyennes et grandes. C'est principalement le CIR qui a été évalué : parmi les travaux les plus récents, on peut citer Duguet (2012), Mulkay et Mairesse (2013), Bozio et al. (2015).

Encadré 1

L'ENQUÊTE R&D : UNE COUVERTURE PARTIELLE DES PETITES ENTREPRISES QUI A ÉVOLUÉ AU COURS DES ANNÉES 2000

Évaluer l'efficacité des aides à la R&D pour les petites entreprises uniquement à partir de l'enquête R&D du MENESR pose des difficultés statistiques : l'enquête ne fournit qu'une information partielle et évolutive sur les entreprises jeunes et petites.

D'une part, l'enquête ne couvre que partiellement la population des petites entreprises qui reçoivent des aides indirectes (crédit d'impôts et allègements fiscaux). En effet, si on utilise l'échantillon et les pondérations de l'enquête, on sous-estime systématiquement le montant des aides indirectes reçues par les petites entreprises. Par exemple, entre 2003 et 2010, on ne retrouve que 61 % du montant de CIR reçu par les TPE (complément en ligne C1).

D'autre part, l'enquête fournit des estimations peu réalistes de l'évolution des dépenses de R&D des petites entreprises car son champ évolue au cours

des années 2000. En effet, comme la base de sondage est actualisée à partir de listes d'entreprises demandant une aide, le nombre croissant de PME ayant recours au CIR a mécaniquement élargi le champ de l'enquête, ce qui a induit des estimations de l'évolution de l'emploi consacré à la R&D par les PME très volatiles : - 18 % en 2005, + 40 % en 2006, - 10 % en 2007, + 22 % en 2008.

L'enquête n'est pas non plus adaptée pour le suivi individuel des dépenses de R&D des plus petites entreprises dans un panel. En effet, les entreprises les plus jeunes (moins de deux ans) et les plus petites ont une très faible chance d'être interrogées deux années d'affilée (Bellégo & Dortet-Bernadet, 2014). Par ailleurs, l'enquête n'est pas adaptée pour observer les entreprises qui débiter une activité de R&D car l'actualisation de la base de sondage ne tient compte que des entreprises qui ont déjà réalisé de la R&D.

Ces évaluations concluent globalement à un effet additif du CIR ou à un faible effet d'amplification. Duguet (2004) obtient un résultat similaire pour les aides directes versées au cours de la période 1985-1997. Lhuillery et al. (2013) mettent globalement en évidence un effet additif des subventions et du CIR mais ils observent aussi des effets d'aubaine pour les entreprises bénéficiant de montants d'aides faibles ou modérés⁵.

Les études qui excluent de leur champ les plus grandes entreprises aboutissent à des résultats plus partagés. Selon Serrano-Velarde (2008), l'obtention de subventions de l'Anvar aurait été accompagnée d'une baisse de la dépense privée en R&D (dépense en R&D nette des aides) des PME et des ETI aidées. En revanche, sur un champ comparable, Bellégo et Dortet-Bernadet (2014) montrent que le supplément d'aide reçu à la suite d'une participation aux pôles de compétitivité n'aurait pas entraîné de baisse de la dépense privée. Néanmoins, ces deux études ont été réalisées à partir des données de l'enquête R&D, ce qui exclut les plus petites entreprises.

En ne travaillant pas exclusivement à partir de l'enquête R&D mais aussi sur des données similaires aux nôtres, Lelarge (2009) montre qu'au cours des premières années du dispositif JEI (2004-2005), les entreprises auraient augmenté les rémunérations versées ce qui permettrait de conserver les employés les plus qualifiés.

Étudier l'effet des aides à la R&D à partir de données exhaustives sur l'emploi

Pour éviter les difficultés liées à l'utilisation de l'enquête R&D pour les petites entreprises (encadré 1), nous proposons de n'étudier qu'une partie des dépenses de R&D : celles relatives à l'emploi consacré à la R&D. Nous évaluons l'effet des aides à la R&D sur l'emploi hautement qualifié⁶ qui est connu exhaustivement à partir des *Déclarations annuelles de données*

5. Ils montrent aussi que les aides les plus efficaces seraient celles à très petites et très fortes « doses ».

6. Défini comme le total de l'emploi des cadres, professions intellectuelles supérieures et dirigeants d'entreprises (catégories socioprofessionnelles 2 et 3 dans les DADS).

Encadré 2

DONNÉES UTILISÉES

Les données sur les aides et l'emploi liés à la R&D

Plusieurs bases de données sont utilisées dans cette étude pour mesurer les montants d'aides reçues par les entreprises et estimer le nombre d'emplois consacrés à la R&D :

- la base de gestion du CIR (GECIR, source MENESR). En plus du montant du crédit d'impôt, cette base donne des informations sur l'ensemble des subventions reçues par les entreprises pour financer leurs activités de R&D.
- le fichier des participations au dispositif JEI (source Acoess). Ce fichier fournit le montant global des exonérations de cotisations employeurs dont bénéficient les entreprises participant au dispositif. Les exonérations fiscales (type CIR) ne sont pas prises en compte, mais elles représentent seulement 10 % du montant global en 2010.
- la liste des agréments du MENESR (source MENESR). Une entreprise agréée réalise de la R&D pour d'autres entreprises qui peuvent ainsi bénéficier du CIR.
- la base de données de l'enquête R&D (source MENESR) : l'enquête est utilisée pour mesurer les montants d'aides directes et estimer le nombre d'emplois consacrés à la R&D entre 2008 et 2010.

Les autres sources d'informations

Pour reconstituer les catégories d'entreprises et estimer le nombre d'emplois consacrés à la R&D, différentes bases de données de l'Insee sont utilisées : les

données fiscales (Ficus/Esane), les déclarations de données sociales (DADS), la base de données sur les liaisons financières (Lifi) et de bases issues du Système informatique pour le répertoire des entreprises et des établissements (Sirene). Il faut noter que les montants de chiffre d'affaires, de valeur ajoutée, de salaires, et les aides sont déflatés à l'aide des indices de prix de la valeur ajoutée de chaque branche, avec pour date de référence juillet 2000).

Les DADS apportent une description exhaustive de l'emploi salarié par catégorie socioprofessionnelle (CS). Les effectifs par CS ont été recalculés en 2009 et 2010 pour contrôler l'influence du changement de méthode de codage des CS. Le coût du travail a été estimé à partir des salaires bruts auxquels ont été ajoutées des estimations de niveau des cotisations sociales employeurs proposées par Cottet et al. (2012).

Le champ de l'étude comprend les petites et moyennes entreprises (PME) et plus particulièrement les très petites entreprises (TPE) : entreprises de moins de 250 (respectivement de moins de 10 pour les TPE) personnes et dégageant un chiffre d'affaires annuel n'excédant pas 50 (respectivement 2 pour les TPE) millions d'euros ou présentant un total de bilan n'excédant pas 43 (respectivement 2 pour les TPE) millions d'euros (voir la définition dans Béguin et al., 2012). Les « entreprises » correspondent uniquement à des unités légales indépendantes ou à des groupes : les unités légales appartenant à des grands groupes ont été exclues du champ de l'étude.

sociales (DADS, encadré 2). L'effet des aides sur l'emploi R&D en est ensuite déduit.

Seul l'effet du montant total des aides est étudié : si ce choix ne permet pas de comparer les mérites respectifs de chaque mécanisme de soutien à la R&D, il permet de traiter simplement les nombreux cas où les entreprises ont recours simultanément à plusieurs dispositifs d'aide⁷. Pour estimer le montant des aides à la R&D reçues par les petites entreprises, nous utilisons des listes de participation aux dispositifs d'aide indirecte (CIR et JEI), ainsi que l'enquête R&D ce qui permet de prendre en compte les aides directes régionales, nationales et européennes (encadré 2).

L'article est organisé de la façon suivante. Dans une première partie, nous rappelons les grandes évolutions des dispositifs d'aide à la R&D pour les PME au cours de la période 2003-2010. Puis nous estimons l'évolution agrégée de l'emploi consacré à la R&D et montrons que pour les PME, l'emploi non financé par des aides publiques aurait fortement baissé. Dans une deuxième partie, nous restreignons à de petites entreprises qui appartiennent à des secteurs intenses en R&D, nous évaluons l'effet des aides à la R&D sur l'emploi pour les années 2004-2010 à partir d'un panel d'entreprises aidées et non aidées. Les résultats indiquent que les aides auraient des effets positifs sur l'emploi hautement qualifié (et R&D) mais qu'il existerait aussi des effets d'aubaine, notamment en fin de période.

Évolution agrégée des aides et de l'emploi R&D au cours de la période 2003-2010

Évolution des dispositifs d'aide à la R&D utilisés par les PME

Les dispositifs de soutien à la R&D se composent d'aides indirectes (crédits d'impôt et

allègements de charge) et d'aides directes. Les aides indirectes recouvrent principalement le CIR et les allègements de cotisations sociales accompagnant le statut de JEI, tandis que les aides directes sont attribuées par différents organismes en charge du développement de la R&D des entreprises.

Le CIR est une réduction d'impôt accordée aux entreprises de toutes tailles qui réalisent de la R&D. De 1983 à 2003, le dispositif a globalement conservé la même architecture : le montant du crédit d'impôt dépendait de l'accroissement des dépenses de R&D d'une année sur l'autre et était plafonné (MENESR, 2014). Une première grande réforme du CIR est intervenue en 2004 avec l'introduction d'une part supplémentaire de crédit d'impôt calculée sur le volume des dépenses de R&D. Celle-ci correspondait à 5 % des dépenses en 2004 et 2005, puis à 10 % à partir de 2006. Le crédit d'impôt calculé sur l'accroissement de ces dépenses était par contre progressivement réduit (tableau 1) et le plafond du crédit d'impôt relevé jusqu'à 16 millions d'euros en 2007.

La réforme du CIR engagée en 2008 supprime le crédit d'impôt calculé à partir de l'accroissement des dépenses de R&D, relève le taux applicable au volume des dépenses à 30 % jusqu'à 100 millions d'euros, puis à 5 % au-delà, et supprime le plafond sur le montant du crédit d'impôt. Des taux bonifiés ont également été appliqués aux entreprises demandant à bénéficier pour la première fois du CIR (un taux de 50 % pour la première année et de 40 % pour la deuxième).

Suite aux réformes successives du CIR, le montant de cette aide a été multiplié par 11 entre 2003 et 2010 pour atteindre à cette date près de 5 milliards d'euros.

7. En 2010, plus de 80 % des entreprises ayant reçu une aide directe ont aussi bénéficié d'une aide indirecte.

Tableau 1
Évolutions des paramètres du crédit d'impôt recherche (CIR) entre 2003 et 2010

	de 1991 à 2003	2004-2005	2006	2007	de 2008 à 2010
Taux (%) sur le volume des dépenses		5	10	10	30 % jusqu'à 100 millions d'euros 5 % au-delà taux bonifiés de 50 % pour la 1 ^{re} année et de 40 % pour la 2 nd e année (*)
Taux (%) sur l'accroissement des dépenses	50	45	40	40	///
Plafond (en millions d'euros)	6.1	8	10	16	Déplafonnement

(*) Pour les entreprises demandant à bénéficier du CIR pour la 1^{re} fois.
Source : réalisé à partir de documents du MENESR.

Le statut de JEI créé en 2004 accorde des allègements de cotisations patronales aux seules PME de moins de huit ans dont les activités de R&D représentent au moins 15 % de leurs charges. Le montant global des aides liées au statut de JEI est beaucoup moins important que celui du CIR (près de 140 millions d'euros en 2010), mais a doublé entre 2004 et 2010. De plus, pour les TPE, il correspond à près de 20 % de l'ensemble des aides indirectes.

Les aides directes correspondent à des subventions bénéficiant à des projets ou couvrant un type précis de dépenses. Ces subventions comprennent des avances remboursables (leur remboursement est lié au succès du projet soutenu), des primes, des prêts bonifiés, des garanties, et des commandes publiques. Elles sont délivrées par les collectivités locales, différents acteurs nationaux comme Oséo ou le Fonds unique ministériel (FUI)⁸, ou par l'Union européenne⁹. À partir de 2005, Oséo-Innovation a notamment été spécialement chargé du financement de la R&D des PME. Les aides directes ont au total augmenté de 64 % pour les PME entre 2003 et 2010 alors qu'elles sont restées stables pour les ETI et grandes entreprises (Dortet-Bernadet & Sicsic, 2015).

Finalement, le montant total des aides à la R&D (directes et indirectes) reçues par les PME a augmenté de 300 % entre 2003 et 2010 et atteint près de 2 milliards d'euros dont 26 %, soit près de 500 millions d'euros, ont été perçus par les TPE (Dortet-Bernadet & Sicsic, 2015).

Estimation agrégée de l'évolution de l'emploi R&D des PME non financé par les aides publiques

Dans cette partie, on présente une estimation en trois étapes de l'évolution entre 2003 et 2010 du total de l'emploi consacré à la R&D (appelé indifféremment par la suite emploi R&D) des PME et de la somme des aides reçues pour le financer. On estime d'abord l'évolution des dépenses en emploi liés à la R&D des PME (étape 1), puis celle des aides à la R&D consacrées à l'emploi (étape 2). Enfin dans une troisième étape, on compare ces deux résultats pour estimer l'évolution des emplois R&D qui n'ont pas été financés par des aides publiques.

Étape 1 : estimation de l'emploi consacré à la R&D

Pour estimer l'évolution de l'emploi R&D, on procède en deux temps. Dans un premier temps,

on le calcule dans les PME (y compris TPE) sur la période 2008-2010 à partir des données de l'enquête R&D, de la base Gecir¹⁰, de la liste des JEI, et de la liste des agréments du MENESR (encadré 2). On fait l'hypothèse que pour la période 2008-2010, le développement des aides permet d'avoir une liste quasiment exhaustive des PME qui mènent des activités de R&D. Puis dans un deuxième temps, on calcule l'évolution de l'emploi R&D sur la période 2003-2010 en supposant que pour chaque secteur (niveau 5 de la NAF) et catégorie d'entreprise le rapport¹¹ entre le nombre d'emplois R&D et le nombre d'emplois « hautement qualifiés » ($Emploi HQ_t^{secteur, catég.}$, source DADS) est stable dans le temps :

$$Estimation \quad emploi R \& D_t^{secteur, catég.} = \frac{Emploi R \& D_{2008-2010}^{secteur, catég.}}{Emploi HQ_{2008-2010}^{secteur, catég.}} \times Emploi HQ_t^{secteur, catég.}$$

Cette hypothèse de stabilité peut paraître forte, ce rapport ayant pu croître avec l'augmentation des aides à la R&D. Cependant, un calcul à partir des données de l'enquête communautaire sur l'innovation montre que le rapport entre DIRDE et dépense en emploi hautement qualifié serait resté globalement stable pour les PME hors TPE entre 2004 et 2008 (il serait passé de 11 % à 9 %).

Selon nos estimations, l'emploi R&D des PME (hors TPE) aurait plus augmenté que celui des TPE entre 2003 et 2008 et la crise de 2008/2009 aurait eu un impact plus fort pour les TPE (figure I). Les dépenses en emploi R&D suivent une évolution similaire. Néanmoins, contrairement à l'emploi R&D, la dépense en emploi R&D des TPE n'a pas diminué entre 2003 et 2010 : elle a augmenté de 5 % (16 % pour les autres PME). Ces évolutions sont très différentes de celles obtenues avec l'enquête R&D mais elles paraissent plus réalistes (voir la discussion dans le document complémentaire en ligne C1).

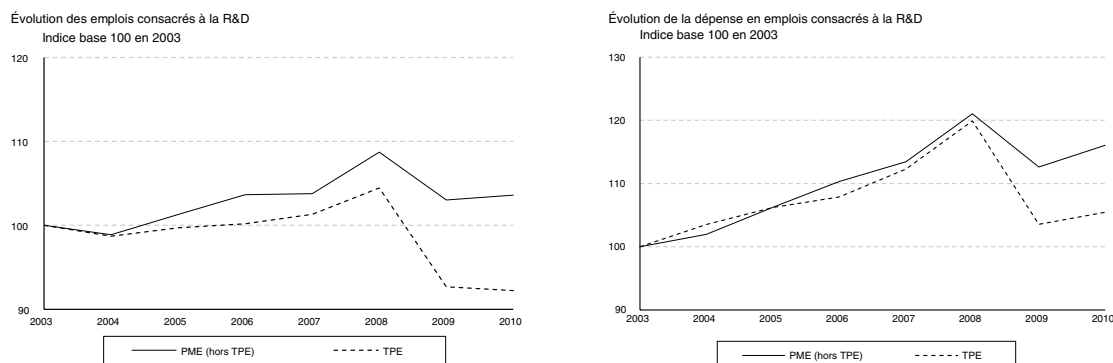
8. Par exemple pour les projets labellisés par les pôles de compétitivité (Dufau, 2017 ; Bellégo & Dortet-Bernadet, 2014).

9. Des financements peuvent être obtenus grâce au Programme-Cadre pour la recherche et le développement technologique ou aux Fonds européens de développement régional.

10. À partir du montant des dépenses de personnel consacrées à la R&D renseigné dans la base, divisé par le coût moyen des ingénieurs (source DADS).

11. Seulement une partie de l'emploi hautement qualifié est affectée à la R&D mais, pour les PME interrogées par l'enquête R&D, on obtient une corrélation de 62 % entre l'emploi R&D réel et l'estimation réalisée avec notre méthode et de 72 % pour les entreprises du panel utilisé dans la dernière partie de l'article.

Figure I

Évolution de l'emploi et de la dépense en emploi consacrés à la R&D

Lecture : entre 2003 et 2010, l'emploi consacré à la R&D a baissé de 8 % pour les TPE et augmenté de 4 % pour les autres PME. La dépense en emploi consacré à la R&D a augmenté de 5 % pour les TPE et de 16 % pour les autres PME.

Champ : France, sociétés commerciales, de l'industrie, du commerce et des services principalement marchands.

Source : MENESR, base Gecir et *enquête R&D* ; Acoess, base JEI ; Insee, Lifi, Ficus/Esane, *DADS*, calculs des auteurs.

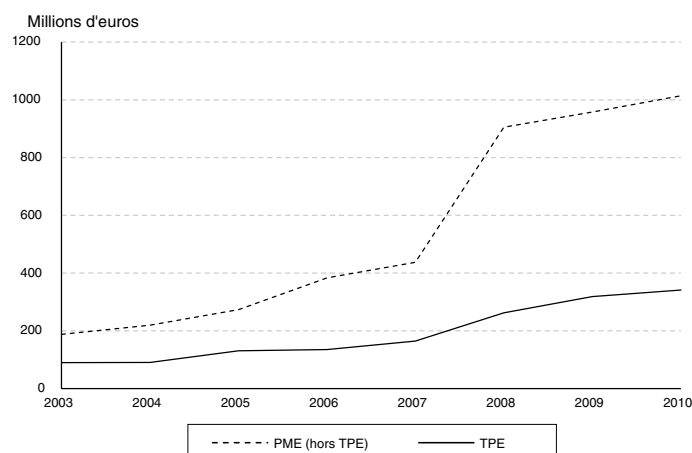
Étape 2 : estimation des aides qui servent à financer des emplois consacrés à la R&D

Pour estimer le montant des aides servant à financer l'emploi R&D, des règles différentes par dispositif sont adoptées. Pour les JEI, les exonérations de cotisations sociales employeurs sont intégralement considérées comme des aides à l'emploi R&D. La part du CIR qui sert à financer l'emploi correspond à la part des dépenses de personnel et des dépenses de fonctionnement dans l'assiette du CIR. En effet, pour calculer l'assiette du CIR, les dépenses de fonctionnement sont fixées à 75 % des dépenses de personnel : le crédit d'impôt qui en découle peut donc être interprété comme une aide à l'emploi.

Contrairement aux aides indirectes, aucune base ne répertorie les aides directes de façon exhaustive : leur montant doit être estimé. Pour chaque entreprise, l'estimation des aides directes qui servent à financer de l'emploi est réalisée à partir de l'information reportée dans la base du CIR, complétée et le cas échéant recoupée par l'information issue de l'enquête R&D. Ces sources étant parfois divergentes, c'est le montant de subventions le plus élevé reporté dans l'une de ces sources qui est retenu.

Selon nos estimations, les aides à la R&D consacrées à l'emploi représentent environ trois quarts des aides à la R&D perçues par les TPE et les autres PME. Entre 2003 et 2010, les aides à la R&D consacrées à l'emploi ont augmenté

Figure II

Évolution des aides à la R&D consacrées à l'emploi

Lecture : en 2010, les aides à la R&D consacrées à l'emploi s'élevaient à 341 millions d'euros pour les TPE et 1,01 milliard d'euros pour les autres PME.

Champ : France, sociétés commerciales, de l'industrie, du commerce et des services principalement marchands.

Source : MENESR, base Gecir et *enquête R&D* ; Acoess, base JEI ; Insee, Lifi, Ficus/Esane, calculs des auteurs.

de 280 % pour les TPE et 440 % pour les autres PME, avec une hausse particulièrement forte en 2008 liée à la réforme du CIR (figure II).

Étape 3 : évolution de l'emploi R&D non financé par les aides

En rapportant le niveau des aides (estimé dans la deuxième étape) au coût du travail moyen de l'emploi R&D (estimé à partir des résultats de la première étape), on peut estimer l'emploi consacré à la R&D « financé par les aides publiques ». On peut alors déduire, à partir de l'estimation de l'emploi R&D de la deuxième étape, l'emploi R&D « non financé par les aides publiques », c'est-à-dire la composante de l'emploi R&D qui n'aurait pas bénéficié des aides publiques. Selon nos estimations, cette dernière a diminué fortement (de - 46 %, entre 2003 et 2010 pour les TPE) et de façon moins importante (- 16 % sur la période) pour les autres PME (figure III). Pour les TPE, la baisse a été de 9 % entre 2004 et 2007, puis plus forte à partir de 2008, année de la réforme du CIR (baisse de 41 % entre 2007 et 2010). Globalement, pour l'ensemble des PME (y compris TPE), la baisse est d'environ 20 % sur la période.

Ces évolutions de l'emploi en R&D non financé par les aides suggèrent des effets d'aubaine, notamment à partir de 2008 avec la forte augmentation des aides à la R&D. Néanmoins, cette période est aussi marquée par la crise financière qui a pu peser sur l'emploi de chercheurs employés par les petites entreprises. En outre, les entreprises sont susceptibles d'avoir changé

de catégorie de taille au cours du temps, notamment grâce aux aides reçues. Par exemple, les TPE les plus dynamiques, ayant reçu des aides, ont pu devenir des PME (hors TPE), ce qui peut expliquer une partie de la baisse de l'emploi R&D non financé par les aides au sein des TPE (de même pour le passage PME/ETI).

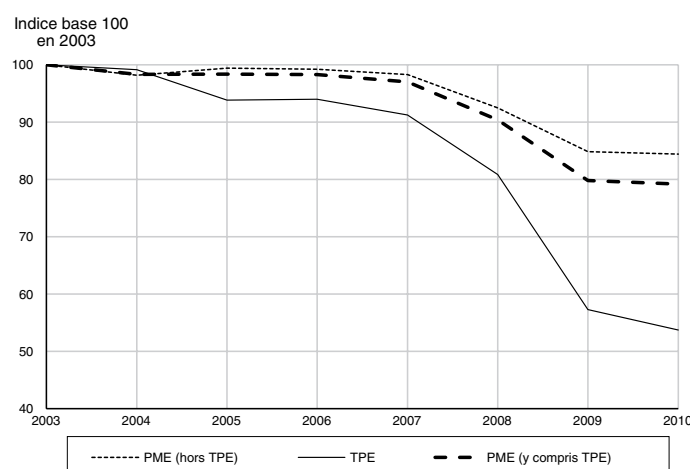
Évaluation de l'effet des aides à la R&D sur l'emploi

Dans cette partie, nous cherchons à évaluer l'effet des aides sur l'emploi consacré à la R&D par les petites entreprises. Cette évaluation est menée en suivant les petites entreprises, indépendamment de leurs changements de catégorie de taille et en les comparant à des entreprises initialement « semblables », soumises à la même conjoncture.

Plus précisément, l'effet des aides à la R&D sur l'emploi est estimé à partir d'un panel de petites entreprises suivies sur plusieurs années entre 2003 et 2010. Comme dans la partie précédente, les aides prises en compte sont celles qui peuvent être reliées à de l'emploi en R&D.

Les entreprises aidées une année donnée peuvent toutefois ne plus demander d'aide l'année suivante, l'effet des aides déjà reçues se prolongeant dans le temps. Pour prendre en compte cet effet différé, les entreprises considérées comme

Figure III
Évolution de l'emploi R&D non financé par les aides publiques



Lecture : l'emploi R&D « non financé par les aides publiques » estimé en 2003 sert de référence (niveau 100). En 2010, il atteint le niveau 54 pour les TPE, ce qui correspond à une baisse de 46 % depuis 2003.

Champ : France, sociétés commerciales, de l'industrie, du commerce et des services principalement marchands.

Source : MENESR, base Gecir et enquête R&D ; Acoess, base JEL ; Insee, Lifi, Ficus/Esane, DADS, calculs des auteurs.

« traitées » une année donnée comprennent les entreprises ayant reçu une aide cette année-là et toutes celles qui en ont déjà reçu une au cours des années précédentes.

Les estimations sont obtenues tout d'abord pour l'emploi hautement qualifié, puis transformées en emploi en R&D par la méthode présentée à la partie précédente.

Construction du panel servant aux estimations

L'évaluation est réalisée sur un panel de petites entreprises appartenant aux 75 secteurs les plus intenses en R&D (Dortet-Bernadet & Sicsic, 2015, p. 48). Ces entreprises sont qualifiées de petites car elles ont toutes été une TPE pendant au moins une année au cours de la période 2000-2010. Elles continuent à être suivies même si elles grandissent et deviennent une PME de 10 salariés ou plus ou une ETI¹² ; toutefois, la majorité des entreprises du champ restent des TPE. Par ailleurs, le champ inclut la très grande majorité des jeunes entreprises.

Le champ de l'évaluation est restreint aux entreprises qui comptent de l'emploi hautement qualifié. On calcule un effet des aides pour les entreprises présentes en 2003 (c'est-à-dire qui comptent de l'emploi hautement qualifié en 2003) et un effet pour les entreprises présentes en 2007¹³ : ces deux années de référence ont été choisies afin de pouvoir observer les entreprises avant les deux principales réformes du CIR, de 2004 et 2008. Le panel n'est pas cylindré : certaines entreprises disparaissent avant 2010 ou sont créées après 2003. Cependant, chaque entreprise doit compter de l'emploi hautement qualifié pendant au moins deux années (dont l'année de référence, 2003 ou 2007).

Le panel est composé d'entreprises aidées et d'autres qui ne l'ont jamais été mais qui ressemblent aux entreprises aidées. Les entreprises jamais aidées ont été sélectionnées en fonction de leur âge et d'un score de propension estimant la probabilité qu'une entreprise soit aidée au moins une fois entre 2004 et 2010 à partir de différentes variables (complément en ligne C2). Ce modèle indique que les dispositifs d'aide sont plus fréquemment utilisés par des entreprises jeunes, qui réalisent des investissements, sont exportatrices, et ont beaucoup d'emploi qualifié. Ces résultats semblent cohérents avec l'idée que les dispositifs d'aide sont plus utilisés par les jeunes entreprises qui sont en phase de développement et ne produisent pas encore beaucoup ou pas du tout (le niveau du chiffre d'affaires joue négativement et est peu significatif). Au total, le panel contient 15 128 entreprises, dont 4 597 ayant reçu au moins une fois une aide entre 2003 et 2010 (tableau 2).

Les secteurs sont regroupés en trois grandes catégories : le secteur de l'industrie, celui de l'information et de la communication (l'informatique, l'édition, les télécommunications, etc.) et une troisième catégorie regroupant les secteurs des activités spécialisées, scientifiques et techniques (R&D, ingénierie, etc.). Les entreprises des secteurs des services sont largement majoritaires (79 %). Presque toutes les entreprises aidées ont eu recours au moins une fois au CIR. La répartition sectorielle des entreprises subventionnées est proche de celle des entreprises ayant eu recours au CIR. Le dispositif JEI

12. En 2010, un tiers des entreprises du panel sont des PME et 3 % des ETI. Des entreprises ont été achetées par des grands groupes mais ces cas sont très peu nombreux. Comme ils peuvent donner lieu à des ambiguïtés sur la continuité de l'activité initiale, ils ont été retirés des bases servant aux estimations.

13. Ces deux groupes de traitement ne sont pas disjoints car des entreprises comptent à la fois de l'emploi hautement qualifié en 2003 et 2007.

Tableau 2

Aides à la R&D obtenues par les petites entreprises du panel entre 2003 et 2010

Entreprises qui ont...	... reçu au moins une fois une aide à la R&D	... eu recours au CIR	... reçu une subvention	... été JEI	qui n'ont jamais reçu d'aides
Nombre d'entreprises	4 597	4 064	2 334	1 348	10 531
Répartition (en %)					
Industrie	20	20	19	9	22
Services informatiques	42	42	41	55	33
Activités scientifiques et techniques	38	38	40	36	46

Lecture : sur les 15 128 entreprises du panel, 4 597 ont reçu au moins une fois une aide entre 2003 et 2010, 4 064 ont eu recours au CIR, 2 334 ont reçu une subvention, 1 348 ont été JEI, et 10 531 n'ont jamais reçu d'aides.

Champ : panel composé de petites entreprises existant en 2003 et/ou en 2007, appartenant à 75 secteurs intenses en R&D, qui ont reçu une aide à la R&D entre 2003 et 2010 ou qui sont « semblables » aux entreprises aidées (sélection par score de propension).

Source : MENESR, base GECIR et *Enquête R&D* ; Acoess, base JEI ; Insee, Lifi, Ficus/Ésane, DADS, calculs des auteurs.

se distingue par la très forte proportion d'entreprises des services informatiques.

Stratégie d'identification

Problèmes posés par l'utilisation de la méthode des différences de différences

L'effet des aides sur l'emploi correspond à la différence entre le nombre d'emplois hautement qualifiés observés dans les entreprises aidées et ce qu'il aurait été en l'absence d'aide. Pour calculer cet effet, on doit se placer dans la situation fictive où les entreprises aidées (traitées) ne reçoivent pas d'aide (ou de supplément d'aide). Cette estimation peut être réalisée par différences de différences en supposant qu'en l'absence d'aide les entreprises traitées se seraient comportées comme les entreprises jamais aidées (non traitées) qui ont un score de propension proche (voir les résultats en complément en ligne C3).

Cette méthode pose cependant plusieurs problèmes. Le premier concerne celui du choix du groupe de contrôle. En effet, l'hypothèse de tendance commune des traités et non-traités n'est pas respectée : avant même de recevoir une première aide, l'emploi des entreprises aidées a été plus dynamique que celui des entreprises jamais aidées. Les entreprises qui ne sont pas encore aidées constituent par contre un groupe de contrôle plus satisfaisant : les entreprises aidées dès l'année t et celles qui ne sont aidées qu'après t ont des évolutions de l'emploi hautement qualifié comparables jusqu'en $t-1$ ¹⁴.

Le deuxième concerne l'interprétation à donner au « traitement » reçu par chaque génération d'entreprises aidées. La méthode de différences de différences simple ne permet pas de prendre en compte l'hétérogénéité au sein de chaque génération d'entreprises aidées du niveau des aides reçues et de l'évolution du taux d'aide.

Enfin la méthode des différences de différences simple ne permet pas de prendre en compte les entreprises aidées dès l'année de référence (ou avant) en tenant compte de l'importance des aides déjà reçues à ce moment là.

Une méthode d'évaluation combinant modèle de demande de travail et appariement

Pour pallier ces différents problèmes, nous ajoutons au groupe de contrôle toutes les entreprises qui n'ont pas encore reçu d'aides mais

qui vont en recevoir une avant fin 2010 et nous estimons un modèle de demande de travail. Ce modèle permet de calculer un niveau d'emploi à partir du coût de l'emploi diminué des aides (assimilées à un allègement du coût du travail) et du chiffre d'affaires. Cela permet de porter la comparaison entre les entreprises traitées et non traitées non plus sur l'évolution de l'emploi mais sur l'évolution du coût de l'emploi et du chiffre d'affaires pour lesquels l'hypothèse de tendance commune est mieux vérifiée¹⁵.

Utiliser le modèle de demande de travail où la demande en emploi dépend de son coût nous permet aussi de contrôler le montant initial d'aide reçue et donc de prendre en compte de façon satisfaisante toutes les entreprises aidées et pas seulement celles qui obtiennent une première aide après l'année de référence (2003 ou 2007). On obtient alors des résultats sur la marge extensive (effet sur les nouvelles entreprises aidées) et sur la marge intensive (augmentation du taux d'aide).

Le modèle de demande en travail hautement qualifié

Le modèle de demande en emploi hautement qualifié (e_{it}^{16}) utilisé est dérivé d'un programme de minimisation des coûts d'une entreprise ayant une fonction de production de type Cobb-Douglas : il est proche de celui présenté dans l'article de Bresson et al. (1992). Il suppose que les entreprises choisissent leur niveau de l'emploi hautement qualifié en fonction de leur chiffre d'affaires (ca_{it}) et du coût relatif moyen de l'emploi hautement qualifié (c_{it}) par rapport aux autres formes d'emploi.

Le modèle n'est estimé que pour les entreprises aidées. Pour ces dernières, les aides s'apparentent à un allègement du coût du travail hautement qualifié (et pas seulement celui des emplois R&D¹⁷) : le coût de l'emploi considéré ici est un coût final après déduction du montant des aides à la R&D¹⁸. Le modèle prend en compte l'ajustement avec délai de la demande

14. En prenant comme groupe de contrôle la génération des entreprises qui ne reçoivent des aides qu'à partir de 2010, on obtient un effet négatif des aides sur l'emploi hautement qualifié non financé par les aides à partir de 2008 (complément en ligne C3, tableaux C3-4 et C3-5).

15. Avant de recevoir leur première aide en t , les entreprises ressemblent aux entreprises pas encore aidées : elles connaissent des évolutions du coût de l'emploi hautement qualifié et du chiffre d'affaires similaires. Pour les entreprises jamais aidées, le coût de l'emploi connaît des évolutions semblables mais le chiffre d'affaires reste moins dynamique (complément en ligne C4).

16. Toutes les variables non dichotomiques mentionnées pour définir le modèle sont prises en logarithme.

en emploi des entreprises et s'exprime sous la forme d'un modèle autorégressif :

$$e_{it} = \rho \cdot e_{it-1} + \alpha_1 \cdot ca_{it} + \alpha_2 \cdot ca_{it-1} + \beta \cdot c_{it} + \mu_i + \delta_t + \varepsilon_{it} \quad (1)$$

où le niveau de l'emploi hautement qualifié à la date t dépend du niveau atteint l'année précédente, du chiffre d'affaires en t et $t-1$ et du coût final relatif moyen de l'emploi en t .

L'emploi hautement qualifié est hétérogène : il comprend des postes consacrés à la R&D et des postes sans lien avec ce type d'activité. Bresson et al. (1992) recommandent en cas d'hétérogénéité de l'emploi de compléter le modèle en ajoutant la variable de coût mesurée en $t-1$ mais cette variable n'a pas été retenue car elle s'est avérée trop corrélée au coût en t et non significative¹⁹.

L'hétérogénéité inobservée des entreprises est prise en compte en introduisant un effet fixe (μ_i) propre à chaque entreprise : le modèle autorégressif permet de prendre ainsi en compte l'hétérogénéité des évolutions d'emploi (plutôt que les niveaux). Des effets temporels (δ_t) ont été ajoutés pour chaque année d'observation. De plus, des modèles différents ont été estimés selon le classement des entreprises dans l'un des trois grands types d'activité (industrie, services informatiques et activités scientifiques et techniques).

Estimation du modèle

Pour estimer le modèle nous prenons en compte l'endogénéité du coût relatif de l'emploi

hautement qualifié. Deux arguments au moins justifient cette hypothèse. D'une part, les organismes en charge des aides directes attribuent des subventions en fonction du dynamisme des entreprises ou de l'aspect novateur de leurs activités. Ces deux caractéristiques inobservées expliquent la croissance de l'emploi des entreprises mais comme elles conditionnent l'obtention d'aide elles sont aussi corrélées à la baisse du coût relatif de l'emploi. D'autre part, au cours de la période 2003-2007, le CIR est encore en partie calculé en fonction de l'accroissement des dépenses de R&D, ce qui implique des évolutions endogènes du coût de l'emploi.

Pour corriger de l'endogénéité du coût relatif de l'emploi (voir le test du complément en ligne C4, tableau C4-2), une variable instrumentale est construite à partir des différentes variations exogènes du dispositif du CIR lors des réformes de 2004, 2006 et 2008 (encadré 3).

17. Dans le cas des petites entreprises, cette hypothèse apparaît relativement réaliste car souvent, les cadres qui font de la R&D n'y consacrent qu'une partie de leur temps de travail. Le fait que de nombreuses petites entreprises utilisent les aides à la R&D de façon intermittente (entre 2004 et 2010, environ 25 % des TPE aidées une année ne le sont plus l'année suivante) rend ce point vraisemblable car cela semble indiquer que ces entreprises ne font pas de la R&D tous les ans.

18. Pour certaines entreprises (notamment celles qui ont obtenu des subventions pour un projet sur plusieurs années) le montant des aides reçues peut dépasser le coût total du travail. Dans ce cas, le coût du travail est annulé et le supplément d'aide est reporté pour diminuer le coût de l'année suivante.

19. Le modèle devrait aussi comporter un terme mesurant le coût relatif de l'emploi hautement qualifié par rapport au capital. Le coût du capital est difficile à évaluer : il peut être approximé en utilisant différents taux d'intérêt qui varient en fonction du niveau d'endettement des entreprises. Cependant les entreprises du champ se différencient peu par leur niveau d'endettement et les valeurs estimées pour le coût du capital sont trop homogènes pour être utilisées dans l'estimation.

Encadré 3

CALCUL DE L'INSTRUMENT POUR CORRIGER L'ENDOGENÉITÉ DU COÛT DE L'EMPLOI

L'évolution du coût relatif de l'emploi hautement qualifié diminué des aides (c_{it}) peut être décomposée en fonction de l'évolution du taux d'aide (τ) et de l'évolution du coût relatif de l'emploi non diminué des aides (c_{it}^*). Sous l'hypothèse que l'évolution du taux d'aide n'a pas d'influence (à court terme) sur la part de l'emploi hautement qualifié (d) consacré à la R&D, on a la décomposition :

$$\Delta c_{it} \approx d_i \cdot \Delta \log(1 - \tau_{it}^{aide}) + \Delta c_{it}^*$$

Cette décomposition permet de trouver un instrument corrélé à l'évolution du coût relatif de l'emploi en remplaçant chaque terme par une variable exogène :

- Pour le terme $\Delta \log(1 - \tau_{it}^{aide})$ on utilise les différentes réformes du CIR de 2004, 2006 et 2008 qui correspondent à des variations exogènes du taux d'aide

(16 variations de taux sont utilisées sur la période 2004-2010, voir le tableau C4-1 du complément en ligne C4).

- Pour la part de l'emploi R&D dans l'emploi hautement qualifié (d) on réalise des estimations en fonction des caractéristiques des entreprises avant les différentes réformes du CIR pour ne pas prendre en compte les éventuelles modifications (augmentation de la part de la R&D) dues à l'augmentation des taux d'aide.

- La variable Δc_{it}^* est simplement remplacée par la variable retardée Δc_{it-1}^* .

L'instrument ainsi obtenu est bien corrélé positivement avec l'évolution du coût de l'emploi et la régression de l'évolution du coût de l'emploi sur les variables exogènes et l'instrument donne un coefficient positif et fortement significatif pour l'instrument.

L'estimation des paramètres du modèle autorégressif à effets fixes peut être faite en travaillant sur l'évolution de la variable endogène entre deux dates et en prenant en guise d'instrument les valeurs retardées des variables explicatives : le but est de contrôler l'endogénéité liée au terme autorégressif (Arellano & Bond, 1991). Cependant, pour le modèle qui vient d'être présenté, la valeur du coefficient ρ est particulièrement élevée et la variation de l'emploi au cours d'une année est peu corrélée à la variation au cours de l'année suivante. Les instruments habituellement utilisés sont donc faibles, ce qui nuit à la qualité de l'estimation. Nous préférons retenir la solution proposée par Blundell et Bond (1998) : à partir d'une hypothèse de stationnarité de l'effectif initial des entreprises, ces auteurs estiment un modèle de demande de travail en utilisant l'évolution retardée de l'emploi comme un instrument du niveau de l'emploi retardé²⁰.

20. En moyenne sur l'ensemble des années et l'ensemble des secteurs, la corrélation entre la variation de l'emploi au cours d'une année et la variation au cours de l'année suivante est très faiblement négative (- 6 %). La corrélation entre l'évolution retardée de l'emploi et le niveau d'emploi est plus importante et positive (+ 27 %).

Enfin, pour estimer le modèle, les entreprises sont aussi supposées faire face à une demande contrainte : elles ne peuvent pas décider directement du niveau de leur chiffre d'affaires, considéré comme exogène à court terme. Cette hypothèse est en partie justifiée par la taille modeste des entreprises étudiées.

L'estimation est réalisée en utilisant la méthode des moments généralisées : les hypothèses d'orthogonalité s'appliquent aux termes résiduels $\mu_i + \varepsilon_{it}$ (pour l'évolution retardée de l'emploi, l'évolution du chiffre d'affaires et la variable instrumentale) et aux variations des résidus $\varepsilon_{it} - \varepsilon_{it-1}$ (pour les mêmes variables sauf l'évolution retardée de l'emploi).

Le calcul de l'effet des aides sur l'emploi hautement qualifié

L'estimation de l'effet des aides sur l'emploi hautement qualifié est réalisée en utilisant le modèle de demande de travail et en calculant des différences de différences. Pour chaque entreprise aidée, on estime l'évolution de l'emploi

Encadré 4

MÉTHODE D'ESTIMATION DE L'EFFET DES AIDES SUR L'EMPLOI

Le calcul de l'effet des aides sur l'emploi hautement qualifié se fait en deux étapes.

Étape 1 : Calcul de la variation de l'emploi à coût de l'emploi et chiffre d'affaires fixés

Le modèle de demande de travail (1) permet d'estimer par récurrence le logarithme du niveau de l'emploi à partir d'une valeur initiale de l'emploi (à la date t_0), et de l'évolution du chiffre d'affaires et du coût de l'emploi :

$$\hat{e}_{it} = \hat{\rho} \hat{e}_{it-1} + \hat{\alpha}_1 ca_{it} + \hat{\alpha}_2 ca_{it-1} + \hat{\beta} c_{it} + \hat{\mu}_i + \hat{\delta}_t$$

$$= f(e_{it_0}, ca_{it_0}, \dots, ca_{it_0+1}, \dots, c_{it}, \hat{\mu}_i, \hat{\delta}) \quad (2)$$

Dans la formule (2), il est possible d'isoler ce qui dépend de la valeur initiale de l'emploi et l'effet fixe, de ce qui dépend de l'évolution du coût de l'emploi et du chiffre d'affaires :

$$\hat{e}_{it} = g_t(e_{it_0}, \hat{\mu}_i, \hat{\delta}) + h_t(ca_{it_0}, \dots, ca_{it_0+1}, \dots, c_{it}, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta})$$

Pour une entreprise i aidée à partir de la date T (cette date n'est pas identique pour toutes les entreprises) l'obtention d'aide va modifier à la fois le coût du travail et le chiffre d'affaires. On peut estimer la variation de l'emploi (entre $T-1$ et t) due à cette évolution du coût et du chiffre d'affaires en fixant ces deux variables à

leur valeur en $T-1$. Cette variation de l'emploi est proportionnelle à :

$$\Delta_{it} = \exp\left(h_t(ca_{iT-1}, \dots, ca_{it}, c_{iT}, \dots, c_{it})\right) - \exp\left(h_t(ca_{iT-1}, \dots, ca_{iT-1}, c_{iT-1}, \dots, c_{iT-1})\right)$$

Étape 2 : Comparaison aux entreprises qui ne sont pas encore aidées

Le calcul précédent suppose qu'en l'absence d'aide le chiffre d'affaires n'aurait pas évolué, ce qui semble être une hypothèse particulièrement forte. Travailler en différence entre traitées et non traitées (et donc en différences de différences) permet de revenir à un cas plus réaliste en comparant l'évolution mesurée pour une entreprise aidée i à l'évolution moyenne obtenue pour des entreprises qui n'ont pas été aidées au cours de la période T à t et qui ont un score de propension proche de celui de i (cette évolution moyenne est signalée par l'exposant C). L'estimation de l'effet des aides sur l'entreprise i est finalement estimé par :

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

Il faut noter que pour contrôler le biais induit par le passage à l'exponentielle, on introduit une correction en utilisant une estimation de l'écart-type de la prévision à l'horizon t du modèle de demande de travail σ_t .

due à la variation du coût et du chiffre d'affaires en fixant ces deux variables au niveau atteint l'année précédant la première aide reçue²¹ (encadré 4). Le modèle de demande de travail permet de décomposer cette évolution en deux termes, l'un dépendant de l'effet fixe, des effets temporels et de la valeur initiale de l'emploi et l'autre dépendant du coût de l'emploi et du chiffre d'affaires. Le groupe des entreprises non traitées est alors utilisé pour contrôler uniquement l'évolution de ce second terme²².

Résultats

L'estimation du modèle de demande en emploi hautement qualifié des petites entreprises aidées au moins une fois et présentes en 2003 (tableau 3) montre que la demande d'une année dépend fortement de la demande de l'année précédente (les coefficients de l'emploi en $t-1$ sont relativement élevés, environ 0.8) et que les entreprises augmentent leurs effectifs si le chiffre d'affaires augmente ou le coût de l'emploi baisse. Les différences entre secteurs s'avèrent relativement réduites sauf pour l'effet du coût de l'emploi : son coefficient n'est pas significativement différent de 0 dans le secteur des services informatiques alors qu'il est négatif

et significatif entre - 0.16 et - 0.18 dans les deux autres secteurs.

L'effet des aides supplémentaires reçues par les entreprises du panel par rapport à l'année de référence est présenté ci-après.

Le supplément d'aide à la R&D par rapport à 2003 reçu par les petites entreprises présentes en 2003 est nettement croissant : il passe de 1 million d'euros 2000 en 2004 (dernière colonne du tableau 4) à 106 millions d'euros 2000 en 2010. Selon nos estimations, ce supplément d'aide à la R&D aurait entraîné en 2010 une augmentation du nombre d'emplois hautement qualifiés de 1 160 emplois équivalent temps plein (EQTP) (tableau 4, première colonne). L'effet du supplément d'aide sur l'emploi hautement qualifié est croissant chaque année : proche de 0 en 2004 et 2005, il progresse à partir de 2006. Pour chaque entreprise aidée, le nombre d'emplois susceptibles d'avoir été financés par les aides supplémentaires se calcule en rapportant le montant de ces aides au coût moyen d'un emploi R&D. Finalement, l'effet sur le nombre d'emplois hautement qualifiés non financés par les aides (ou financés en propre par les entreprises, troisième colonne) correspond à la différence entre l'effet sur l'emploi hautement qualifié et le nombre d'emplois hautement qualifiés susceptibles d'avoir été financés par le supplément d'aides : il est négatif et significatif sauf en 2004, 2007 et 2010. En 2010 l'estimation de l'effet sur l'emploi hautement qualifié non financé par des aides progresse mais reste négatif ; l'estimation devient aussi plus imprécise et finalement non significative.

21. Pour les entreprises déjà aidées l'année de référence (2003 ou 2007), on utilise les niveaux de l'année de référence ce qui revient à estimer l'effet du supplément d'aide obtenu depuis cette date.

22. Les entreprises qui n'ont pas encore été aidées (à une date donnée) sont réparties en fonction de leur score de propension en 10 groupes. Les résultats moyens obtenus sur chacun de ces groupes servent de point de comparaison à l'évolution des entreprises aidées.

Tableau 3

Modèle de demande en emploi hautement qualifié des petites entreprises aidées au moins une fois entre 2003 et 2010 (équation (1))

	Secteurs		
	Industrie	Services informatiques	Activités scientifiques et techniques
Emploi hautement qualifié en $t-1$ (log.)	0.76***	0.86***	0.77***
Chiffre d'affaires en t (log.)	0.08***	0.1***	0.07***
Chiffre d'affaires en $t-1$ (log.)	- 0.02*	- 0.05***	0
Coût relatif moyen de l'emploi hautement qualifié en t diminué des aides à la R&D (log.)	- 0.16***	- 0.04	- 0.18**

Note : modèle estimé par la méthode des moments généralisés (GMM) avec un instrument pour le coût de l'emploi. Emploi en équivalent temps plein (EQTP). Les coefficients sont significativement différents de zéro pour des tests de niveau 10 % (*), 5 % (**) et 1 % (***). Ces tests sont obtenus par *bootstrap*.

Lecture : pour les secteurs de l'industrie, le niveau de l'emploi hautement qualifié en t est expliqué par le niveau de l'emploi hautement qualifié en $t-1$ (coefficient estimé de 0.76), le niveau du chiffre d'affaires en t et $t-1$ et le coût relatif moyen de l'emploi hautement qualifié par rapport au coût des autres types d'emploi. Les modèles comprennent aussi des contrôles pour chaque année d'observation (coefficients estimés non présentés).

Champ : petites entreprises du panel ayant de l'emploi hautement qualifié en 2003, et aidées au moins une fois sur la période 2003-2010 (2 261 entreprises, panel non cylindré).

Source : MENESR, base GECIR et *Enquête R&D* ; Acoss, base JEI ; Insee, Lifi, Ficus/Ésane, DADS, calculs des auteurs.

Tableau 4

Estimation de l'effet du supplément d'aide à la R&D obtenu par les petites entreprises par rapport à 2003 sur l'emploi hautement qualifié total et comparaison au supplément d'aide reçu

	Effet sur l'emploi hautement qualifié	Effet sur l'emploi R&D	Effet sur l'emploi hautement qualifié non financé par des aides	Effet sur l'emploi R&D non financé par des aides	Effet sur l'emploi R&D non financé par des aides hors "CIR dépenses de fonctionnement"	Supplément d'aide à la R&D par rapport à 2003 (en millions d'euros 2000)
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 : emploi en équivalent temps plein (EQTP) ; supplément d'aide en millions d'euros constants de 2000. Les résultats sont significativement différents de zéro pour des tests de niveau 10 % (*), 5 % (**) et 1 % (***). Ces tests sont obtenus par bootstrap.

Lecture : par rapport à 2003, les aides supplémentaires reçues en 2005 s'élèvent à 18 millions d'euros (colonne 6). L'effet de ce supplément d'aide sur l'emploi hautement qualifié serait de + 40 EQTP en 2005 (colonne 1) et de + 10 EQTP pour l'emploi R&D (colonne 2). L'effet sur l'emploi non financé par les aides serait une baisse de 270 EQTP pour l'emploi hautement qualifié (colonne 3), de 290 EQTP pour l'emploi consacré à la R&D (colonne 4) et de 290 EQTP si le « CIR dépenses de fonctionnement » n'est pas compté dans les aides (colonne 5).

Champ : petites entreprises du panel ayant de l'emploi hautement qualifié en 2003, et aidées au moins une fois sur la période 2003-2010 (2 261 entreprises, panel non cylindré).

Source : MENESR, base GECIR et *Enquête R&D* ; Acoess, base JEL ; Insee, Lifi, Ficus/Ésane, DADS, calculs des auteurs.

Seule une partie des emplois hautement qualifiés sont réellement consacrés à la R&D : pour estimer leur nombre (deuxième colonne), on fait de nouveau l'hypothèse que pour chaque secteur et catégorie d'entreprise, le rapport entre le nombre d'emplois R&D et le nombre d'emplois hautement qualifiés est égal à celui estimé sur la période 2008-2010. Sous cette hypothèse, le déficit d'emplois R&D non financés par les aides (quatrième colonne) serait de 220 EQTP en 2006 (contre 180 EQTP pour l'emploi hautement qualifié). Comme pour l'emploi hautement qualifié, on observe des effets nettement plus négatifs à partir de la réforme du CIR de 2008 ; ils sont de plus toujours négatifs et significatifs sauf en 2004.

Finalement, par sommation des estimations obtenues aux différentes années, pour les entreprises présentes en 2003, seulement 63 %²³ du supplément d'aide versé entre 2004 et 2010 aurait servi à financer de nouveaux emplois hautement qualifiés (44 % si on ne considère que les emplois R&D).

Cette évaluation repose sur un majorant des aides à l'emploi, du fait notamment de la prise en compte de la partie du CIR liée aux dépenses de fonctionnement²⁴. Dans l'hypothèse où cette partie du CIR ne serait pas intégrée, les

emplois R&D financés en propre ne diminueraient plus que de 30 EQTP (évolution non significativement différente de zéro) en 2010 (cinquième colonne du tableau 4) au lieu de 480 EQTP (quatrième colonne) et 58 % du supplément d'aide versé entre 2004 et 2010 aurait servi à financer de nouveaux emplois R&D.

Les résultats précédents ont été établis pour des entreprises présentes en 2003. Pour mesurer l'effet des aides sur un plus grand nombre d'entreprises, on s'intéresse maintenant aux entreprises qui sont présentes en 2007. Cette nouvelle estimation permet de mieux prendre en compte la réforme du CIR de 2008. Le total des aides reçues par ces entreprises en 2010 est de 344 millions d'euros 2000, alors qu'il n'était que de 135 millions d'euros 2000 pour le groupe précédent (le supplément d'aides reçu en 2010 passe de 106 millions d'euros 2000 par rapport à l'année de référence 2003 (tableau 4, dernière colonne) à 171 millions d'euros 2000 par rapport à l'année 2007, dernière colonne du tableau 5). Pour ce groupe élargi, l'effet du supplément d'aide reçu sur l'emploi hautement qualifié serait maximum en 2010 avec 830 EQTP (tableau 4, colonne 1) ; parallèlement, le supplément d'aides reçu cette année-là correspondrait au financement de 2 140 EQTP ce qui correspond à une baisse de 1 310 EQTP de l'emploi hautement qualifié non financé par les aides (tableau 4, colonne 3). Pour l'emploi R&D la baisse serait plus importante avec une baisse de 1 520 EQTP de l'emploi financé en propre par les entreprises.

23. Ce résultat a pour intervalle de confiance à 95 % l'ensemble [42 % ; 84 %].

24. En effet elles sont fixées à 75 % des dépenses de personnel R&D (cf. supra). Ne pas le prendre en compte diminuerait de 43 % les montants de CIR servant à financer de l'emploi.

Ces baisses sont statistiquement significatives pour toutes les années. C'est aussi le cas si on ne prend pas en compte la partie du CIR liée aux dépenses de fonctionnement (tableau 4, colonne 5). Finalement, on obtient que pour les entreprises présentes en 2007, seulement 24 % (entre 18 et 34 % compte tenu de l'intervalle de confiance à 95 %) du supplément d'aide versé entre 2008 et 2010 par rapport à 2007 aurait servi à financer de nouveaux emplois hautement qualifiés (19 % si on ne considère que les emplois R&D et 29 % si on ne prend pas en compte le CIR liée aux dépenses de fonctionnement).

Les résultats détaillés par secteur d'activité montrent que la baisse de l'emploi hautement qualifié financé en propre par les entreprises concerne tous les secteurs mais qu'elle a été plus forte dans l'industrie et plus modérée dans les activités scientifiques et techniques (complément en ligne C5, tableaux C5-1 et C5-2). Pour les entreprises présentes en 2007, seulement 9 % du supplément d'aide reçu par les entreprises de l'industrie a servi à financer de nouveaux emplois hautement qualifiés contre 15 % pour les entreprises des services informatiques et 43 %²⁵ pour les entreprises des activités scientifiques et techniques.

Discussion des résultats et robustesse

Dans cette partie nous discutons les hypothèses utilisées pour construire le groupe de contrôle et

estimer le modèle et leur influence sur les résultats de l'évaluation.

Pour les deux populations d'entreprises étudiées (présentes en 2003 ou présentes en 2007) on constate une augmentation de l'effet des aides sur l'emploi hautement qualifié en fin de période. Cet effet est en partie lié à la moindre qualité du groupe de contrôle en 2010²⁶, ce qui a pu conduire à surestimer l'effet des aides pour les dernières années.

Le champ des tableaux 4 et 5 ne comprend que des entreprises appartenant aux 75 secteurs les plus intenses en R&D et comptant un ou des emplois hautement qualifiés en 2003 ou 2007. Ces deux restrictions permettent de trouver des contrefactuels de meilleure qualité. Toutefois ces restrictions sont plus techniques que justifiées par la réalité économique et il semble raisonnable d'étendre les estimations à un champ élargi à l'ensemble des TPE et des autres petites PME.

L'estimation de l'effet des aides est faite en comparant des entreprises aidées à des entreprises non aidées sans tenir compte d'un possible effet des aides sur la survie des entreprises. Si les aides ont permis aux entreprises de prolonger leur activité (ou de conserver leur emploi hautement qualifié), l'effet des aides aurait été sous-estimé. Toutefois, il ressort de la comparaison des entreprises aidées avant 2004 du panel et des entreprises jamais aidées qui ont un score de propension proche, que les entreprises

25. Les intervalles de confiance à 95 % sont respectivement pour chaque secteur de [6 % ; 21 %], [6 % ; 24 %], et [31 % ; 61 %].

26. Il ne comprend que des entreprises jamais aidées qui constituent un contrefactuel de moins bonne qualité que les entreprises non encore aidées (complément en ligne C4).

Tableau 5

Estimation de l'effet du supplément d'aide à la R&D obtenu par les petites entreprises par rapport à 2007 sur l'emploi hautement qualifié total et comparaison au supplément d'aide reçu

	Effet sur l'emploi hautement qualifié	Effet sur l'emploi R&D	Effet sur l'emploi hautement qualifié non financé par des aides	Effet sur l'emploi R&D non financé par des aides	Effet sur l'emploi R&D non financé par des aides hors "CIR dépenses de fonctionnement"	Supplément d'aide à la R&D par rapport à 2007 (en millions d'euros 2000)
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 : emploi en équivalent temps plein (EQTP) ; supplément d'aide en millions d'euros constants de 2000. Les résultats sont significativement différents de zéro pour des tests de niveau 10 % (*), 5 % (**) et 1 % (***). Ces tests sont calculés en utilisant une méthode de *bootstrap*.

Lecture : par rapport à 2007, les aides supplémentaires reçues en 2008 s'élèvent à 131 millions d'euros (colonne 6). L'effet de ce supplément d'aide sur l'emploi hautement qualifié serait de + 210 EQTP en 2008 (colonne 1) et de + 160 EQTP pour l'emploi R&D (colonne 2). L'effet sur l'emploi non financé par les aides serait une baisse de 1 710 EQTP pour l'emploi hautement qualifié (colonne 3), de 1 760 EQTP pour l'emploi consacré à la R&D (colonne 4) et de 1 150 EQTP si le « CIR dépenses de fonctionnement » n'est pas comptabilisé dans les aides (colonne 5).

Champ : petites entreprises du panel ayant de l'emploi hautement qualifié en 2007, et aidées au moins une fois sur la période 2007-2010 (4 117 entreprises, panel non cylindré).

Source : MENESR, base GECIR et *Enquête R&D* ; Acoss, base JEI ; Insee, Lifi, Ficus/Esane, DADS, calculs des auteurs.

aidées sont observées un peu plus longtemps mais que la différence de « survie » entre les deux groupes n'est pas statistiquement significative (complément en ligne C5, tableau C5-4).

Pour définir le groupe de contrôle d'entreprises, on ne dispose pas d'une information exhaustive sur les aides directes : le groupe de contrôle peut contenir quelques entreprises qui ont reçu des subventions. Ce problème peut entraîner une sous-estimation de l'effet des aides car l'évolution de l'emploi des entreprises qui ne sont pas aidées est peut-être surestimée. Ce biais devrait néanmoins se réduire avec l'augmentation des recours au CIR car les entreprises subventionnées sont de plus en plus répertoriées dans la base Gecir.

Les entreprises jamais aidées du groupe de contrôle sont sélectionnées à partir de caractéristiques observables qui déterminent le score de propension. Or, des variables inobservables ont pu influencer à la fois la participation des entreprises aux dispositifs d'aide et leur demande d'emploi hautement qualifié : ne pas les prendre en compte a pu biaiser nos estimations. Toutefois, parmi ces variables inobservées, le coût de recherche de subventions est nettement plus élevé pour les petites entreprises que pour les plus grandes (Arqué-Castells et Mohnen, 2015) et la connaissance des dispositifs d'aide dépend probablement de l'âge des entreprises : ces variables sont donc partiellement contrôlées grâce au score de propension qui intègre différentes caractéristiques liées à la taille et l'âge des entreprises.

Dans le modèle de demande d'emploi hautement qualifié les aides sont assimilées à un allègement immédiat du coût du travail. Si les subventions et les allègements de cotisations du dispositif JEI sont effectivement perçus par les entreprises dès que les dépenses de R&D sont réalisées ce n'est pas le cas du CIR en particulier pour les entreprises qui ne déclarent pas d'impôt sur les sociétés : ces entreprises détiennent une créance qui parfois n'est remboursée intégralement par l'État qu'au bout de trois ans. Cependant, depuis 2007 des régimes dérogatoires permettent aux petites et jeunes entreprises²⁷ d'obtenir un remboursement dès l'année suivante des créances de CIR, ce qui rend plus crédible l'hypothèse utilisée pour le modèle de demande d'emploi.

27. Plus précisément, des régimes dérogatoires ont été appliqués pour les JEI et PME dynamiques en 2007 et 2008, pour toutes les entreprises en 2009 et seulement pour les PME à partir de 2010.

Les résultats présentés peuvent être en partie biaisés car des entreprises agréées par le MENESR peuvent appartenir au groupe de contrôle. Il peut s'agir d'entreprises classées dans les secteurs intenses en R&D qui, bien que ne recevant pas toujours d'aide²⁸, bénéficient indirectement du CIR reçu par leurs clients. Toutefois, si on exclut toutes les entreprises agréées du champ de l'étude on obtient des résultats très proches de ceux présentés dans les tableaux 4 et 5 (complément en ligne C5, tableaux C5-5 et C5-6).

* *
*

Les analyses économétriques menées (par différences de différences et à l'aide d'un modèle de demande de travail combiné à une méthode d'appariement) confirment les effets d'aubaine suggérés par l'analyse agrégée : elles montrent que l'effet des aides à la R&D sur les dépenses de personnel hautement qualifié et en R&D a été positif mais nettement inférieur à l'augmentation des aides reçues, notamment à partir de 2008. La très forte augmentation des taux d'aide à la R&D durant les années 2000 n'aurait pas conduit à augmenter efficacement l'emploi consacré à la R&D des petites entreprises. Ce résultat se différencie nettement de ceux obtenus jusqu'à présent dans la plupart des travaux empiriques sur données françaises. Il est obtenu sur un champ comprenant l'ensemble des très petites entreprises, généralement ignoré par ces travaux réalisés à partir de l'enquête R&D.

Il faut cependant noter que nos résultats sont interprétables surtout à court terme et qu'ils ne présagent pas des éventuels effets de long terme des aides à la R&D. Enfin, cette étude ne prend pas en compte les évolutions récentes des règles de calcul du CIR. Ainsi, la diminution en 2011 puis la disparition des taux majorés à 50 % et 40 % en 2013²⁹, en réduisant sensiblement le taux d'aide des petites entreprises, peuvent avoir réduit l'effet d'aubaine mis en évidence dans cette étude. □

28. La très grande majorité des entreprises agréées du champ de l'étude ont cependant reçu au moins une fois une aide (75 % des TPE et 87 % des autres PME agréées en 2010 ont reçu une aide entre 2005 et 2010).

29. À noter aussi la diminution de l'assiette des dépenses de fonctionnement en 2011 (de 75 % à 50 % pour les dépenses de personnel).

BIBLIOGRAPHIE

- Acemoglu, D., Akcigit, U., Bloom, N. & Kerr, W. R. (2013).** Innovation, Reallocation and Growth. *NBER Working Paper* 18993.
- Aghion, P., Askenazy, P., Berman, N., Cetto, 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 additivity 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 Research 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.

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

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 *

Résumé – L'aide à la création d'entreprise est devenue un élément à part entière des politiques actives de l'emploi dans la plupart des pays de l'OCDE. En France, le dispositif de l'Accre (Aide aux chômeurs créateurs ou repreneurs d'entreprises) a connu une forte expansion depuis sa mise en place en 1979. Le nombre d'entrées dans le dispositif a atteint, après plus de 80 000 en 2006, un maximum de 220 000 en 2010. Nous étudions l'effet de l'Accre sur la survie à cinq ans de quatre cohortes d'entreprises créées par des chômeurs en 1994, 1998, 2002 et 2006, à partir des enquêtes *Sine* (Système d'information sur les nouvelles entreprises) de l'Insee. Selon les statistiques descriptives, les perspectives de survie des entreprises dont les créateurs ont bénéficié de l'Accre, sont supérieures à celles dont les créateurs ne l'ont pas reçue. Cependant, à l'aide d'une modélisation à équations simultanées de l'obtention de l'Accre et de la survie des entreprises créées, nous mettons en évidence une sélection des bénéficiaires liée au processus administratif d'attribution de l'aide, ainsi qu'une auto-sélection des créateurs eux-mêmes. L'effet de l'Accre sur la survie des entreprises aidées apparaît alors nul pour la plupart des catégories de chômeurs.

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.

Codes JEL / JEL codes : C26, D21, H25, J68, L38

Mots-clés : création d'entreprise, démographique d'entreprises, politique de l'emploi, aide aux chômeurs, évaluation des politiques publiques

Keywords : start-up, firm demographics, employment policy, support for the unemployed, public policy assessment

Rappel :

Les jugements et opinions exprimés par les auteurs n'engagent qu'eux-mêmes, et non les institutions auxquelles ils appartiennent, ni a fortiori l'Insee.

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

Je remercie Richard Duhautois pour les nombreux travaux et discussions menés en commun sur la création d'entreprises. Je remercie aussi mes collègues du CEE avec lesquels nous avons travaillé sur l'exploitation des enquêtes Sine de l'Insee : Lionel Désiage (décédé en mai 2012), Géraldine Rieucan, Guillaumette de Larquier, Bernard Gornel. Une version antérieure de ce texte a été présentée au séminaire interne du CEE le 16/12/2014. Elle a bénéficié des précieuses remarques, critiques et suggestions de F. Guillaumat-Taillet (Dares), qu'il en soit très sincèrement remercié. Je remercie également deux rapporteurs anonymes pour leurs remarques et suggestions fructueuses.

Au cours des dernières décennies, les économistes et les responsables des politiques de l'emploi ont accordé une attention grandissante aux créations / disparitions d'entreprises. La démographie des entreprises se caractérise en effet par des taux d'entrée et de sortie des marchés souvent élevés dans les économies avancées. Cette rotation rapide des firmes est généralement associée au processus de « destruction créatrice » mis en avant par l'économiste Schumpeter au début des années 1940. Il serait continuellement à l'œuvre dans les économies contemporaines, et se traduirait par un mouvement simultané de création d'activités nouvelles innovantes et de destructions d'activités obsolettes. Les start-ups seraient ainsi des éléments essentiels du processus de destruction créatrice, reconnu comme une des sources clés de la croissance économique (Aghion et al., 2014). Une large littérature empirique s'est développée sur ce thème depuis les années 1980-1990 : elle met en évidence l'impact important des nouvelles entreprises, que ce soit des start-ups (de moins de trois ans), ou plus largement des jeunes entreprises (de moins de cinq ans) sur la dynamique de l'emploi dans les pays avancés, aussi bien en Europe qu'aux États-Unis (voir, notamment, Audretsch & Mahmood, 1994 ; Davis et al. 2007 ; Haltiwanger, 2011 ; Haltiwanger et al. 2013 ; Mata & Portugal, 1994). À partir d'une base de données constituée spécialement à cet effet par l'OCDE (<http://www.oecd.org/fr/sti/dynemp.htm>), Criscuolo et al. 2014 montrent que, parmi les PME (qui emploient moins de 250 personnes), les firmes jeunes (créées depuis cinq ans au plus) « contribuent plus que proportionnellement à la création d'emplois » dans les dix-huit pays étudiés¹, confirmant ainsi des travaux récents sur données américaines (Haltiwanger et al., 2013). De plus, la plupart de ces créations d'emploi sont le fait des entrées de nouvelles firmes, et dans une moindre mesure, du développement des start-ups.

Les implications de ces recherches en termes de politique de l'emploi sont importantes. En effet, les aides à la création d'entreprises ont été progressivement intégrées aux politiques actives du marché du travail dans l'Union européenne² et dans la plupart des pays de l'OCDE, qui ont, de façon générale, pour objectif de réintégrer les

chômeurs au marché du travail. Pour leur part, les politiques d'aides à la création d'entreprises doivent permettre à des individus au chômage, souvent peu qualifiés et ayant parfois à faire face à des discriminations, de créer leur entreprise, et ainsi leur propre emploi, en les aidant à surmonter les difficultés premières liées à la création d'une activité. En les sortant du chômage, elles doivent leur permettre d'améliorer leur employabilité et leur capital humain. D'autres arguments sont en faveur de ces politiques : elles peuvent être porteuses d'un « double dividende », lorsque des emplois additionnels sont créés dans ces nouvelles entreprises. Enfin, elles auraient un effet positif sur la croissance économique, en relation avec la contribution de ces nouvelles entreprises à l'innovation et à la diffusion des technologies. Ces politiques d'aide à la création d'entreprise font néanmoins l'objet de sérieuses critiques. Tout d'abord, elles peuvent induire des effets d'aubaine : les créateurs d'entreprise auraient de toute façon créé leur entreprise, avec ou sans subvention. La survie et le succès de l'entreprise ne sont alors pas liés aux aides. Elles peuvent conduire à des effets d'éviction en désavantageant les entreprises existantes relativement aux start-ups subventionnées. Elles peuvent aussi avoir des effets de sélection adverse dans un contexte de forte incertitude, par exemple en facilitant l'accès à l'entreprenariat à des individus qui n'ont pas les capacités de gérer leur propre entreprise à court-moyen terme. Au contraire, en réduisant les barrières à l'entrée, ces politiques pourraient permettre de révéler à des individus qu'ils disposent des capacités nécessaires à l'entreprenariat. Enfin, ces subventions peuvent induire des effets d'aléa moral. Les créateurs aidés pourraient réduire leurs efforts, puisqu'ils n'auraient pas à subir de coûts ou de pertes de revenus en cas d'échec de leur entreprise.

Caliendo (2016) souligne que la plupart des études réalisées sur l'impact des programmes de soutien public à la création d'entreprises destinés aux chômeurs dans les pays de l'OCDE restent descriptives, et que rares sont celles susceptibles d'être qualifiées d'évaluation à moyen-long terme. Il souligne qu'elles ne sont en général pas convergentes, en raison de l'hétérogénéité des dispositifs institutionnels selon les pays et des méthodes statistiques et économétriques utilisées³. Pour l'Allemagne,

1. Autriche, Belgique, Canada, Costa Rica, Espagne, États-Unis, Finlande, France, Hongrie, Italie, Japon, Luxembourg, Norvège, Nouvelle Zélande, Pays-Bas, Portugal, Royaume-Uni, Suède, et Turquie sur la période 2001-2011.

2. Voir les dépenses par type d'action, notamment incitation à créer une nouvelle entreprise, sur le site d'Eurostat, tableau [imp_expsummm].

3. Sur les évaluations des différents programmes nationaux, on pourra consulter Caliendo et Künn (2011), Caliendo et al. (2015) et Pfeiffer et Reize (2000) pour l'Allemagne, Deidda et al. (2015) sur l'Italie, Gu et al. (2008) sur les États-Unis, ainsi que la synthèse de Caliendo (2016).

Caliendo et Künn (2011) comparent deux programmes d'aide à la création des entreprises (allocations chômage et subventions) sur la période 2003-2008. S'appuyant sur des méthodes d'appariement de propension de score, ils mettent en évidence qu'à l'horizon de cinq ans les deux programmes ont des effets significatifs et positifs sur l'emploi et le revenu des participants, particulièrement pour les chômeurs de longue durée. Enfin, alors que depuis août 2006, les deux programmes sont regroupés en un dispositif unique, une récente évaluation (Caliendo et al., 2015) montre, à partir d'une méthode d'appariement de propension de score, que les entreprises subventionnées ont un taux de survie, à 19 mois, supérieur aux autres entreprises. Mais leurs performances en terme de croissance, d'innovation et de revenu sont moindres en raison d'effets de sélection adverse et d'aléa moral négatifs. En France, suite à la mise en place en 1994 du dispositif d'enquête *Sine* (Système d'information sur les nouvelles entreprises) par l'Insee qui permet d'analyser le profil du créateur, les conditions de démarrage, et les conditions de développement des nouvelles entreprises, différentes évaluations des dispositifs d'aide à la création d'entreprises ont été menées. Elles s'intéressent principalement à la survie des entreprises et à leurs performances économiques. Crépon et Duguet (2003) étudient l'effet des aides publiques (toutes confondues) à la création sur les entreprises créées en 1994 sur une période de trois ans. En utilisant une méthode d'appariement sélectif reposant sur des variables observables (appariement avec propension de score), ils mettent en évidence un effet positif significatif des aides publiques sur la survie des entreprises créées par d'anciens chômeurs. De plus, l'obtention de prêts bancaires renforce fortement les chances de survie de ces firmes quand il est associé à une aide. Cabannes et Fougère (2012 et 2013) évaluent l'effet de l'Accre (Aide aux chômeurs pour la création et la reprise d'entreprises) sur la durée de vie des entreprises, sur une période de cinq ans, également à partir de l'enquête *Sine*, pour la cohorte née en 1998. Ils prennent en compte l'endogénéité de l'attribution de l'Accre et estiment un modèle à effets aléatoires constitué de deux équations simultanées, une équation d'attribution de l'Accre (*logit*) et une équation qui formalise la durée de vie des entreprises de la cohorte 1998. Ils mettent ainsi en évidence que l'effet causal de l'Accre sur la durée de vie à cinq ans des entreprises créées en 1998 par des personnes ayant passé au préalable moins d'un an au chômage, n'est pas significativement différent de zéro.

Désiage et al. (2010) et Duhautois et al. (2015) construisent une base en appariant l'enquête *Sine* de 1998 et des données d'entreprises issues de fichiers administratifs (Ficus : fichiers de comptabilité unifiés) et ils s'intéressent au devenir des entreprises créées au cours de leurs huit premières années. La population des créateurs bénéficiant de l'Accre est plus large que celle étudiée par Cabannes et Fougère, puisqu'elle inclut tous les bénéficiaires de l'Accre : les chômeurs quelle que soit la durée de leur chômage, et les inactifs au moment de la création. Utilisant une méthode d'appariement et de score de propension « à la Rubin », ils mettent en évidence un effet causal significatif de l'Accre sur la survie à 5 ans et à 8 ans des entreprises aidées.

Notre étude vise à évaluer l'effet de l'Accre sur la survie à 5 ans des cohortes d'entreprises créées par les bénéficiaires de cette aide, respectivement en 1994, 1998, 2002 et 2006. Ces années correspondent aux quatre premières enquêtes *Sine* menées par l'Insee. Notre démarche comporte un double intérêt par rapport aux études précédentes. Tout d'abord, comme chaque enquête est un sondage de l'ensemble des entreprises créées au cours du premier semestre de chacune des années considérées (encadré 1), les caractéristiques des créateurs bénéficiaires de l'Accre peuvent être comparées, à l'aide des mêmes variables, à celles des créateurs non bénéficiaires, ce qui n'est pas le cas de beaucoup d'études portant sur les pays étrangers, qui ne disposent pas d'enquêtes de ce type⁴. Ensuite, par rapport aux évaluations déjà menées pour la France, le fait de disposer de quatre cohortes différentes permet d'évaluer l'éventuelle variation au cours du temps de l'effet causal de cette aide sur la survie des entreprises, en prenant en compte les changements de la réglementation. Enfin, s'appuyant sur une méthodologie proche de Cabannes et Fougère (2012), nos résultats confortent et généralisent les leurs, en soulignant l'absence d'impact significatif de l'Accre sur la survie à trois et cinq ans des entreprises nouvellement créées.

Notre recherche est organisée de la façon suivante. Une première partie présente la réglementation de l'Accre en France et son évolution depuis sa création en 1979, ainsi que les effectifs des créateurs bénéficiaires, pour chaque cohorte de nouvelles entreprises. Une deuxième partie expose la constitution des quatre cohortes d'entreprises à partir des bases *Sine* et donne les indicateurs de survie des entreprises de ces quatre

4. Ce point est souligné par Caliendo (2016) p. 9.

cohortes, selon qu'elles ont été bénéficiaires ou non de cette aide. Dans un troisième temps, la stratégie d'estimation économétrique de l'effet causal de l'Accre sur la survie des entreprises bénéficiaires est exposée. La quatrième partie estime cet effet pour les créateurs qui étaient au chômage depuis moins d'un an au moment de la création. La partie suivante se concentre sur l'évaluation de cet effet pour d'autres catégories de créateurs (chômeurs depuis plus d'un an, et inactifs). Enfin la dernière partie teste la robustesse des estimations.

Évolution de la réglementation de l'Accre et de la population des bénéficiaires

Les conditions d'attribution de l'Accre

La réglementation de l'Accre a connu de nombreux changements depuis sa création (tableau 1). Ils concernent tant la population éligible que la nature de l'aide consentie aux créateurs, et ses conditions d'octroi. À l'origine, en 1979, c'était une « mesure guichet » octroyée automatiquement aux chômeurs à condition qu'ils soient indemnisés (Mouriaux, 1995). À partir de 1987, l'administration du travail a acquis le droit de refuser les projets qu'elle considérait comme non viables. Un comité départemental, placé auprès de la direction du travail, a été mis en place à la fin des années 1980 et jugeait de la réalité et de la consistance des projets.

La nature de l'aide et son importance ont aussi fluctué en fonction des aléas de la politique budgétaire (figure). La loi quinquennale de décembre 1993, qui ne fut appliquée effectivement qu'à partir du 5 avril 1994, introduisit une rupture par rapport au dispositif antérieur. Le bénéfice de l'Accre a été étendu à tous les chômeurs indemnisés ou non (avec cependant un délai de 6 mois d'ancienneté dans le chômage pour ces derniers). Surtout, la subvention forfaitaire devient égale pour tous, et a été portée à 32 000 francs (soit environ 4 880 euros en valeur 1994) à partir du second trimestre de 1994. La générosité du système et son caractère égalitaire ont induit une forte augmentation du nombre de nouveaux bénéficiaires de l'Accre dès le second trimestre de l'année 1994 (voir complément en ligne C1), ainsi qu'en 1995 et 1996 (figure). Mais le coût budgétaire élevé de l'Accre a conduit le gouvernement, dans sa loi de finances de 1997, à supprimer la prime de

32 000 francs (Daniel & Mandelblat, 2010). Le nouveau système – mis en place à partir de début 1998 – n'est réellement intéressant que pour les chômeurs indemnisés qui continuent à percevoir tout ou partie de leurs indemnités de chômage pendant une période maximale de 15 mois (à condition qu'ils ne soient pas rémunérés par leur nouvelle entreprise). De plus, si leur entreprise disparaît pendant cette période, les créateurs retrouvent ensuite leurs droits à indemnisation calculés à partir de la date de création de leur entreprise. Quant aux chômeurs non indemnisés, ils ne bénéficient que d'une exonération des charges sociales sur la rémunération qu'ils perçoivent pendant un an, et dans la limite de 1.2 fois le Smic. Cependant dès juillet 1998, la loi contre les exclusions a étendu le bénéfice de l'Accre à de nouveaux créateurs, sans changer la nature et le montant de l'aide.

À partir de 2007, plusieurs réformes très importantes de la réglementation de la création d'entreprise et des aides attachées à cette création sont intervenues. Tout d'abord, depuis janvier 2007, l'aide est accordée suivant des critères purement administratifs de conformité à la réglementation (Daniel & Mandelblat, 2010 ; Ould Younes, 2010) et, depuis le mois de septembre de la même année, le Centre de formalités des entreprises instruit les dossiers. Les indicateurs de viabilité économique des projets de créations ont été supprimés. Il en est résulté une forte hausse du nombre de nouveaux bénéficiaires de l'Accre (figure). À cela s'ajoute la création effective à partir de janvier 2009 du régime des autoentrepreneurs. Ce régime bénéficie de procédures de création très allégées et d'une réglementation fiscale particulière. Il a attiré de nombreux porteurs de petits projets et, de ce fait, a modifié les caractéristiques des créateurs d'entreprise recensées par l'enquête *Sine* de 2010, par rapport aux cohortes nées en 2002 et 2006 (Béziau & Bignon, 2017). En raison de ces profonds changements, nous avons pris le parti d'arrêter nos analyses à la cohorte née en 2006, c'est-à-dire avant l'introduction de la suppression des critères de viabilité des projets pour l'attribution de l'Accre (2007) et du régime des autoentrepreneurs (2009).

Les créateurs d'entreprises potentiellement concernés par l'Accre

Dans la suite de l'article nous qualifions de « chômeurs de CT », les créateurs qui étaient au chômage depuis moins d'un an au moment de la création, de « chômeurs de LT », les personnes

Tableau 1

Conditions d'attribution et formes d'aide apportées par l'Accre : évolution de la réglementation

Période	Population éligible	Nature de l'aide	Conditions et procédure d'attribution.
Avant avril 1994 (loi 80-1035 de décembre 1980 ; décret 87-202 du 28 mars 1987)	Les chômeurs indemnisés, les allocataires du RMI et de l'ASS.	- Un capital calculé en fonction du montant journalier de l'Allocation de solidarité spécifique (ASS) pour les chômeurs ayant épuisé leur droit à indemnisation. - Un capital calculé en fonction des droits à indemnisation restant à recevoir dans la limite de 6 mois pour les chômeurs indemnisés. - Exonération des charges sociales pendant 6 mois pour les chômeurs indemnisés. Concernant les bénéficiaires de l'ASS : possibilité de percevoir pendant un an leur allocation, et de plus le revenu de leur entreprise s'il est inférieur à un demi Smic. S'il est supérieur, l'ASS décroît. Concernant les bénéficiaires du RMI, les revenus de leur entreprise ne sont pris en compte qu'à 50 % dans le montant de leurs ressources qui sert de base au calcul de leur allocation d'insertion	Les bénéficiaires doivent créer ou reprendre une entreprise quels que soient le secteur d'activité et la forme juridique, et en exercer effectivement le contrôle. Demande introduite auprès de la direction départementale du travail, avant le début d'activité de l'entreprise. Un comité départemental est en charge de juger de la réalité et de la consistance du projet.
À partir d'avril 1994 (application de la loi quinquennale de décembre 1993)	<i>Idem</i> que précédemment plus les chômeurs non indemnisés inscrits depuis plus de 6 mois au chômage.	- Subvention égale pour tous les bénéficiaires : 32 000 francs (soit environ 4 880 euros en valeur 1994). - Exonération des charges sociales pendant 1 an pour les chômeurs indemnisés et les allocataires du RMI et de l'ASS, pas d'exonération pour les chômeurs non indemnisés.	<i>Idem</i> que précédemment
Premier semestre 1998 (loi de finances 1997).	<i>Idem</i> que précédemment plus les allocataires d'autres revenus sociaux : API, personnes reconnues handicapées.	- Suppression de la subvention de 32 000 francs. - Exonération des charges sociales sur la rémunération du créateur pendant 1 an pour les chômeurs indemnisés et non indemnisés dans la limite de 1,2 fois le Smic et pour les allocataires du RMI et de l'ASS. - Les chômeurs indemnisés et les allocataires de revenus sociaux continuent à percevoir leurs indemnités et allocations pendant 12 à 15 mois s'ils ne sont pas rémunérés par leur nouvelle entreprise. Si leur entreprise ferme, ils retrouvent leurs droits à indemnisation calculés à partir de la date de création de leur entreprise. - Chaque bénéficiaire de l'Accre reçoit des « chèques-conseils » utilisables auprès d'experts agréés.	<i>Idem</i> que précédemment
Du 2 ^e semestre 1998 à septembre 2007 (loi d'orientation de juillet 1998 : lutte contre les exclusions ; puis loi n° 2003-721 du 1 ^{er} août 2003 pour l'initiative économique).	À partir de juillet 1998, <i>Idem</i> que précédemment plus les titulaires d'un contrat d'appui au projet d'entreprise, les salariés repreneurs de leur entreprise en redressement ou liquidation judiciaire, les jeunes de moins de 26 ans éligibles aux « emplois-jeunes ».	Le mode de calcul de l'Accre n'a pas changé du premier semestre 1998 au premier semestre 2006. La loi du 1 ^{er} août 2003, d'initiative économique ne concerne pas l'Accre mais a facilité la création d'entreprise en offrant des garanties supplémentaires aux chefs d'entreprise (habitation principale insaisissable, dégrèvements fiscaux).	<i>Idem</i> que précédemment
Décret 2007-1396 de septembre 2007	Les bénéficiaires sont les mêmes que précédemment.	La nature de l'aide ne change pas par rapport aux textes précédents.	La procédure d'attribution de l'Accre change fondamentalement. Depuis janvier 2007, l'aide est accordée sur critères administratifs uniquement, et depuis septembre 2007 les dossiers sont traités par les centres de formalités des entreprises (CFE).

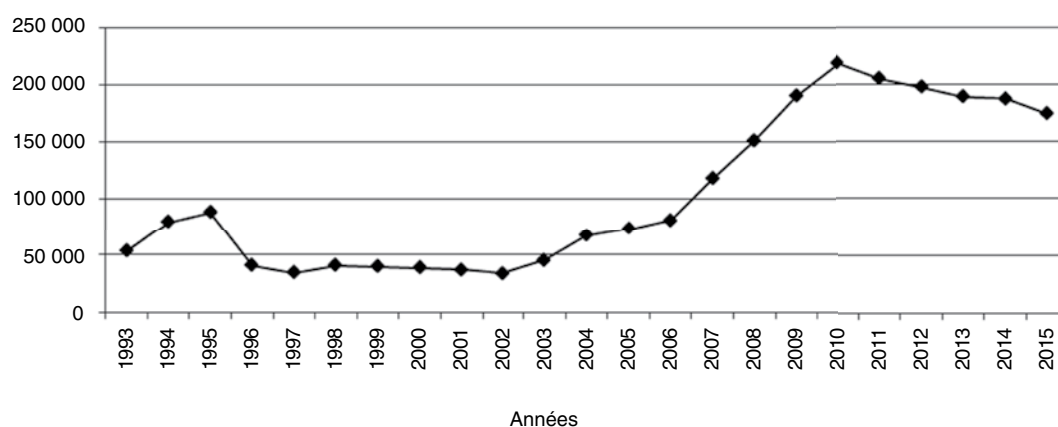
Source : Charpail (1995), Charpail (1996), Daniel & Mandelblat (2010), Guimiot & Mareau (2003), Mouriaux (1995), Ould Younes (2010).

qui étaient au chômage depuis au moins un an au moment de la création, et « d'inactifs » ceux qui ont déclaré ne pas avoir d'emploi, et ne pas en avoir recherché au même moment (il s'agit notamment des étudiants et des titulaires de minima sociaux qui ont déclaré ne pas rechercher d'emploi).

La part des bénéficiaires de l'Accre au sein des chômeurs et inactifs créateurs d'entreprises (tableau 2) apparaît fortement influencée par les changements de règles d'attribution de cette aide analysés *supra* (tableau 1). Le système mis en place en 1994, relativement généreux pour

Figure
Les entrées dans le dispositif d'aide à la création d'entreprise (Accre)

Nombre
d'entrées



Note : cumul annuel d'entrées dans le dispositif.
Champ : France métropolitaine.
Source : base statistique Poem de la Dares, disponible sur <http://poem.travail-emploi.gouv.fr>.

les créateurs, a été suivi par d'importantes restrictions à partir de 1996-1997. Les avantages financiers apportés par l'Accre ont été réduits et l'incitation à la demander a diminué. Cependant, pour les cohortes nées en 2002 et 2006, l'extension progressive par la législation de la population de créateurs éligibles a entraîné un accroissement significatif de la part des bénéficiaires parmi les chômeurs et inactifs.

Par ailleurs l'Accre constitue bien la plus grande part de l'aide publique aux créateurs, chômeurs et inactifs (tableau 2). Parmi les aides hors Accre, signalons les subventions locales et régionales, les exonérations de taxe professionnelle, et les allègements de charges sociales. L'ensemble de ces aides concerne aussi bien les créateurs salariés que les chômeurs et inactifs au moment de la création. Pour ces derniers, elles peuvent, dans certains cas, être cumulées avec l'Accre. Finalement, après la brève rupture de la période qui va du deuxième trimestre de l'année 1994, à la fin de l'année 1996, où l'Accre était d'un même montant pour chaque catégorie de bénéficiaire, l'aide a été calculée sur le montant des indemnités de chômage, et des minima sociaux versés aux créateurs, chômeurs ou inactifs. La législation a ainsi introduit de très grandes inégalités entre les avantages financiers consentis aux différents groupes de bénéficiaires.

Les données utilisées et les indicateurs de survie des entreprises des quatre cohortes

Les données utilisées pour cette étude sont issues du dispositif *Sine* (Système d'information sur les nouvelles entreprises) de l'Insee, et portent sur des cohortes d'entreprises créées en 1994, 1998, 2002 et 2006. Les méthodes statistiques et les concepts harmonisés permettent de comparer les quatre cohortes, moyennant certaines précautions qui sont détaillées dans l'encadré 1.

Survie des entreprises créées avec ou sans l'Accre

Nous avons retenu comme indicateur de performance, le taux de survie des entreprises cinq ans après leur création. Ce taux est défini comme le rapport entre le nombre d'entreprises créées au cours du 1^{er} semestre de la période considérée qui sont encore en vie au bout de cinq ans, et le nombre d'entreprises créées au début de la période considérée. Le tableau 3 fait apparaître que, lorsque les créateurs sont en activité juste avant la création (qu'ils soient salariés, artisans-commerçants, chefs d'entreprise, ou

Tableau 2

Bénéficiaires (chômeurs et inactifs au moment de la création) de l'aide à la création d'entreprise (Accre) dans les quatre enquêtes Sine

En %

	1994	1998	2002	2006
Chômeurs et inactifs par rapport à l'ensemble des créateurs	43.3	49.0	50.8	50.8
Parmi les chômeurs et inactifs :				
- Bénéficiaires d'une aide publique à la création (toutes aides confondues)	51.7	38.8	47.7	65.4
- Bénéficiaires de l'Accre	n.d.	30	40	59.0
Parmi les chômeurs de CT :				
- Bénéficiaires d'une aide publique à la création (toutes aides confondues)	69.2	49.5	58.5	76.2
- Bénéficiaires de l'Accre	n.d.	40.1	51.6	70.8
Parmi les chômeurs de LT :				
- Bénéficiaires d'une aide publique à la création (toutes aides confondues)	59.7	47.5	59.7	75.6
- Bénéficiaires de l'Accre	n.d.	39.4	52.8	69.8
Parmi les inactifs :				
- Bénéficiaires d'une aide publique à la création (toutes aides confondues)	6.7	13.4	18.2	27.9
- Bénéficiaires de l'Accre	n.d.	5.4	11.2	18.3

Note : chômeurs de CT : personnes au chômage depuis moins d'un an au moment de la création ; chômeurs de LT : personnes au chômage depuis un an et plus d'un an au moment de la création ; n.d. : non déterminé. Calculs sur la base de données des auteurs ; données pondérées à l'aide de la variable *poidsini* (encadré 1).

Champ : entreprises des secteurs marchands non agricoles créées au premier semestre de l'année de référence, France métropolitaine hors Corse.

Source : Insee, enquêtes Sine 1994, 1998, 2002, 2006.

Tableau 3

Taux de survie moyen des entreprises dans les quatre enquêtes suivant la situation des créateurs d'entreprise juste avant la création

En %

Année	Créateurs en emploi	Créateurs chômeurs et inactifs		
		Total	Bénéficiaires de l'Accre	Non bénéficiaires de l'Accre
Cohorte 1994				
Survie à 3 ans	57.6 [57.0 ; 58.2]	52.7 [52.0 ; 53.4]	54.1 [53.1 ; 55.2]	51.8 [51.0 ; 52.6]
Survie à 5 ans	44.3 [43.6 ; 45.0]	38.3 [37.8 ; 39.0]	42.0 [41.0 ; 43.0]	36.3 [35.5 ; 37.1]
Cohorte 1998				
Survie à 3 ans	68.8 [68.3 ; 69.3]	62.1 [61.6 ; 62.6]	70.2 [69.3 ; 71.0]	58.6 [58.0 ; 59.2]
Survie à 5 ans	55.0 [54.5 ; 55.5]	49.2 [48.7 ; 49.7]	59.2 [58.3 ; 60.0]	45.0 [44.4 ; 45.6]
Cohorte 2002				
Survie à 3 ans	72.3 [71.9 ; 77.7]	66.1 [65.7 ; 66.5]	67.2 [66.5 ; 67.9]	65.5 [64.9 ; 66.0]
Survie à 5 ans	58.4 [57.9 ; 58.9]	51.1 [50.6 ; 51.6]	53.0 [52.3 ; 53.7]	49.7 [49.1 ; 50.3]
Cohorte 2006				
Survie à 3 ans	70.8 [70.4 ; 71.2]	64.6 [64.3 ; 64.9]	65.9 [65.5 ; 66.3]	62.7 [62.2 ; 63.2]
Survie à 5 ans	n.d.	n.d.	n.d.	n.d.

Note : calculs de l'auteur ; données pondérées à l'aide de la variable *poidsini* (encadré 1). Entre crochets : intervalle de confiance au seuil de 10 % calculé à partir de l'écart-type du taux considéré ; n.d. : non déterminé. Pour la cohorte née en 1994, il s'agit de l'Accre plus éventuellement d'autres formes d'aides (encadré 1).

Champ : entreprises des secteurs marchands non agricoles créées au premier semestre de l'année de référence, France métropolitaine hors Corse.

Source : Insee, enquêtes Sine 1994, 1998, 2002, 2006.

LES DONNÉES ISSUES DU DISPOSITIF *SINE*

Depuis 1994, tous les quatre ans l'Insee réalise une enquête sur les créations d'entreprises selon le dispositif *Sine* (Système d'information sur les nouvelles entreprises). Ce dispositif permet notamment d'analyser le profil du créateur et les conditions de démarrage des nouvelles entreprises, les conditions de développement et l'évolution des effectifs des jeunes entreprises durant les cinq premières années de leur existence. Il s'agit d'une enquête par sondage au taux d'un tiers environ, portant sur les entreprises créées au cours du premier semestre de l'année. L'échantillon est tiré à partir du répertoire Sirene. Chaque cohorte est suivie pendant une période de cinq années. Elle est interrogée la première année de son existence, la troisième et la cinquième. Le champ de ces enquêtes concerne l'ensemble des secteurs d'activité marchands non agricoles. La grande majorité des nouvelles entreprises appartient aux secteurs du commerce, réparation et autres services et est constituée de microentreprises. Dans les cohortes que nous étudions 80 % des entreprises ne comptent qu'un seul emploi au démarrage.

Nous n'avons retenu dans notre étude que les « créations ex-nihilo » (créations pures) et avons éliminé les reprises et réactivations d'entreprises créées précédemment, ce qui rend les quatre enquêtes comparables de ce point de vue. La définition de la « création ex-nihilo » correspond à tout démarrage d'activité nouvelle faisant l'objet d'une immatriculation au fichier Sirene. Pour l'enquête 2006, c'est cette même définition que nous avons retenue, alors que cette enquête comprend aussi une variable qui inclut comme « créations » des réactivations d'entreprises préexistantes, mais dont l'activité est nouvelle par rapport à la précédente. Nous avons préféré ne pas tenir compte de

ces réactivations de façon à conserver une définition homogène pour les quatre cohortes. Nous avons aussi éliminé les entreprises qui avaient cessé leur activité avant la première interrogation (de la première année d'existence) par l'enquête *Sine*. Enfin, nous avons également éliminé les entreprises localisées dans les DOM et en Corse. La fiscalité particulière des entreprises dans ces deux régions peut en effet influencer le recours des créateurs à l'Accre.

Le plan de sondage a été construit pour assurer la représentativité de chaque enquête selon les critères de la région, du secteur d'activité (Nes16) et du critère « création pure ou reprise ». Le poids de chaque strate (région x secteur x critère de création) dépend de la dispersion des taux de survie à cinq ans au sein de la strate. Le poids d'une entreprise i (variable poidsini de l'enquête) au sein d'une strate donnée est égal à l'inverse de la probabilité de tirer une observation de cette même strate dans l'échantillon, rapportée à la probabilité de tirer dans la population une entreprise d'une strate donnée (Cabannes & Fougère, 2012).

Les quatre enquêtes utilisées contiennent sensiblement les mêmes variables. Une exception importante concerne l'enquête de 1994 pour laquelle la variable « aide publique », à la différence des enquêtes suivantes, ne distingue pas les différentes formes d'aide, et notamment l'Accre. Pour l'année 1994, nous avons retenu cette variable « aide publique » qui peut être considérée comme une proxy de l'Accre : notamment, elle capte bien la rupture dans la législation qui a incité un plus grand nombre de chômeurs et de bénéficiaires de revenus sociaux à demander l'Accre (tableaux 1 et 2).

exerçant une profession libérale), le taux de survie de leur entreprise est en moyenne supérieur de 5 à 7 points à celui des entreprises dont les créateurs sont initialement chômeurs ou « inactifs ». À échéance de trois ans et de cinq ans, l'écart est sensiblement le même.

Par ailleurs, la survie à trois ans et à cinq ans des entreprises dont les créateurs sont en activité au moment de la création, a augmenté fortement de 1994 à 2002, pour diminuer légèrement ensuite. Celle des entreprises créées par des personnes au chômage, ou inactives, au moment de la création a suivi la même évolution : une augmentation élevée pour les trois premières cohortes, puis une diminution légère. Enfin, selon ces données descriptives, les entreprises des créateurs ayant bénéficié de l'Accre ont un taux de survie nettement supérieur à celles dont les créateurs n'en ont pas bénéficié, particulièrement pour les deux

premières cohortes (à cinq ans, respectivement de 5.7 points en 1994 et de 14.2 points en 1998). Ce taux pour les entrepreneurs aidés a augmenté fortement en 1998, à tel point qu'il atteint plus de 59 % à cinq ans. Il est alors supérieur de plus de 4 points à celui des entreprises créées par des personnes en emploi avant la création. En revanche, pour les deux dernières cohortes, les différences de taux de survie entre les entreprises aidées et non aidées se réduisent à un peu plus de 3 points, tout en restant significatives.

Stratégie économétrique

On ne peut pas écarter l'hypothèse que les créateurs qui reçoivent l'Accre ne sont pas choisis de façon aléatoire. Tout d'abord, pour les

quatre cohortes étudiées, l'administration du travail est susceptible de cibler les projets dont les créateurs présentent les meilleures qualités pour assurer la survie et le développement de leur entreprise. Les entrepreneurs qui bénéficient de l'Accre sont issus d'un processus de sélection qui dépend de leurs caractéristiques personnelles et des caractéristiques de leur projet d'entreprise. De plus, des phénomènes d'auto-sélection peuvent jouer si certains candidats à la création sont mieux informés, ou (et) mieux à même de mener à bien les formalités administratives pour obtenir l'Accre. Ils ont alors une probabilité supérieure de l'obtenir, et leur niveau d'information et leur capacité à effectuer des démarches complexes, peuvent signifier qu'ils ont des caractéristiques personnelles qui les rendent plus aptes à gérer leur entreprise. Dans ce cas, ce sont ces caractéristiques individuelles qui peuvent expliquer la survie, ou la durée de vie de leur entreprise, et non le fait d'avoir bénéficié de l'Accre. Par ailleurs, la sélection inverse peut se produire en facilitant la création d'entreprise par des personnes qui font preuve de réelles capacités pour diriger une entreprise, lorsqu'on leur donne les moyens de tenter l'expérience (Jovanovic, 1982).

Traitement du biais de sélection, choix et validité des instruments

Pour prendre en compte ces phénomènes de sélection, suivant une méthodologie proche de celle de Cabannes & Fougère (2012, 2013), on estime conjointement une équation de durée de vie ou de survie des entreprises, et une équation d'attribution de l'Accre. Dans la mesure où l'éventuel processus de sélection repose en partie sur des variables inobservables (contenu précis du projet, personnalité du créateur, capacité à diriger une entreprise), on met en œuvre une méthode à variables instrumentales : on utilise au moins un instrument qui affecte le fait de recevoir l'Accre, mais qui n'a pas d'effet propre sur la survie des entreprises. Tout d'abord, nous faisons le choix d'un modèle de survie plutôt que de durée. En effet, les dates de cessation d'activité des entreprises ne sont pas mentionnées dans la base de l'année 1994, ou sont imprécises pour la cohorte de 1998, alors que les données annuelles sur la cessation d'activité au bout d'un an, deux ans, jusqu'à cinq ans, figurent dans les quatre bases (encadré 2). Un premier instrument envisageable est l'indicatrice du trimestre de création de l'entreprise (Cabannes & Fougère 2012). L'utilisation de

cet instrument est justifiée par les arguments suivants :

- La demande de l'Accre doit être faite préalablement à la création de l'entreprise ;
- L'aide est accordée si un refus n'est pas signifié au demandeur dans les trois mois suivant la demande ;
- En cas d'attribution de l'aide, le bénéficiaire doit créer son entreprise dans les trois mois de l'acceptation (l'entreprise est considérée comme créée le jour de son inscription au registre du commerce ou au registre des métiers).

Le créateur doit donc faire sa demande d'attribution de l'Accre trois mois au minimum avant le moment prévu de la création. Comme les crédits publics alloués à l'Accre concernent à la fois le budget de l'État (prise en charge des indemnités de chômage pendant un an, financement des chèques conseils pour la formation des chômeurs) et celui de la Sécurité sociale (exonération des cotisations sociales), dans un contexte de contraintes budgétaires, on peut penser que l'administration départementale du travail est plus restrictive au dernier trimestre de l'année $t-1$ qu'au premier trimestre de l'année t . On peut s'attendre à ce que, toutes choses égales par ailleurs, les entreprises créées au second trimestre de l'année t aient une probabilité plus forte de recevoir l'Accre que celles créées au premier trimestre. Comme les entreprises enquêtées, pour chaque cohorte, ont été créées au premier semestre de l'année (encadré 1), nous avons, dans un premier temps, testé dans l'équation d'attribution de l'Accre (équation 1 de l'encadré 2), l'effet de la création pour chacun des six mois correspondant. Le regroupement par trimestre s'est avéré pertinent (notamment la rupture entre les mois d'avril, mai, juin, et les mois précédents ressort clairement). Finalement, nous retenons comme variable instrumentale la variable indicatrice relative à la création de l'entreprise au deuxième trimestre, et nous anticipons une relation positive entre cette variable et l'attribution de l'Accre.

Le second instrument, utilisé dans plusieurs études de ce type (Pfeiffer & Reize, 2000 ; Cabannes & Fougère, 2012) est un indicateur de tension du marché local du travail défini comme le ratio du nombre de postes vacants (V) sur le nombre de chômeurs (U). L'échelon géographique retenu est départemental puisque c'est à ce niveau qu'est prise la décision d'attribution de l'Accre par l'administration du travail.

MÉTHODE ÉCONOMÉTRIQUE UTILISÉE

Nous estimons un modèle *biprobit* composé d'une équation d'attribution de l'Accre et d'une équation de survie à cinq ans.

Il peut être formalisé de la façon suivante pour un créateur d'entreprise i .

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

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

$ACCRE^*_i$ et $SURV^*_i$ sont des variables latentes qui correspondent respectivement aux scores de chaque entrepreneur i . Elles déterminent le fait qu'il obtient l'Accre ou non (1) et que l'entreprise qu'il a créée soit en activité, ou non, au bout de 5 ans (2). La règle de sélection est la suivante :

$$ACCRE_i = 1 \text{ si } ACCRE^*_i > 0 \text{ et } ACCRE_i = 0 \text{ si } ACCRE^*_i \leq 0$$

$$SURV_i = 1 \text{ si } SURV^*_i > 0 \text{ et } SURV_i = 0 \text{ si } SURV^*_i \leq 0$$

X_i est un vecteur des caractéristiques individuelles de l'entrepreneur i et de son projet (par exemple l'âge, la nationalité, le sexe, le niveau d'études, la catégorie socio-professionnelle avant la création, le secteur d'activité, le statut de l'entreprise créée...).

$instr_i$ est un vecteur constitué des variables instrumentales (deux dans notre modèle).

ε_{1i} et ε_{2i} sont les termes d'erreurs :

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \longrightarrow N(0, \Sigma)$$

Les termes d'erreur ont une distribution normale bivariable (avec une matrice variance-covariance formée de 1 sur la diagonale principale, les autres éléments de cette matrice étant formés par le coefficient de corrélation entre les termes d'erreurs). La corrélation éventuelle entre les termes d'erreur permet de tenir compte de l'hétérogénéité inobservée.

Les estimations ont été effectuées par la méthode du maximum de vraisemblance de la fonction cumulée de répartition normale bivariable, à l'aide du logiciel *Stata*. De plus, les régressions présentées dans le texte, les annexes et les compléments en ligne ont été réalisées en pondérant les observations i par la variable *poisini* (procédure *pweight* de *Stata*). Comme le souligne Cabannes & Fougère (2012), ce poids étant fonction de la dispersion de la survie à cinq ans au sein de chaque strate considérée (encadré 1), l'omission de ce poids peut introduire un biais d'endogénéité.

Si la tension sur le marché du travail départemental est faible (V/U petit), la probabilité pour un chômeur de trouver un emploi salarié est faible, alors les autorités administratives départementales vont chercher à faciliter la sortie du chômage en encourageant la création d'entreprises. Dans cette perspective, ils sont moins stricts dans l'attribution de l'Accre que dans les départements où la situation de l'emploi est plus favorable. On anticipe donc une relation décroissante entre l'indicateur V/U et l'attribution de l'Accre. Nous avons retenu le rapport de la moyenne mensuelle des flux de nouvelles offres et des nouvelles demandes d'emplois au cours de l'année $t-1$. Cet indicateur intègre donc la possibilité d'un retard dans la prise en compte de l'information sur le marché du travail par l'administration.

Cet instrument peut-il être endogène et avoir un effet propre sur la survie des entreprises, ou sur une variable omise ayant un effet sur cette survie ? L'hypothèse de causalité inverse, suivant laquelle les nouvelles entreprises qui survivraient, amélioreraient la situation du marché local peut être rejetée. En effet, l'indicateur de tension du marché du travail local qui a été choisi est décalé d'un an avant la création effective des entreprises. Par ailleurs, la possibilité

d'une corrélation entre cet instrument et le terme d'erreur de l'équation de survie nous paraît faible, même si les entrepreneurs persévèrent dans une activité peu rentable faute d'alternative, en raison de la persistance du chômage, au niveau du marché départemental de l'emploi. Une telle corrélation ne pourrait provenir que d'une situation du marché départemental durablement dégradée pendant cinq ans à partir de la date de création.

Prise en compte des caractéristiques individuelles des entrepreneurs et de leur entreprise

Nous utilisons le vaste ensemble de variables disponibles dans l'enquête *Sine* qui concernent les caractéristiques observables des entrepreneurs, ainsi que les caractéristiques économiques de l'entreprise qu'ils créent (encadré 1). Il s'agit du genre, de la nationalité, du niveau de formation, de la CSP avant la création d'entreprise, de l'âge, du nombre de créations précédentes par le créateur, du statut juridique de l'entreprise, de sa taille, et de son secteur d'activité. Nous introduisons également une variable dichotomique qui capte l'effet éventuel d'autres formes d'aide publique (tableau 2). Pour ce

qui est des moyens financiers mis en œuvre au moment de la création de l'entreprise, nous avons construit une variable dichotomique, petits moyens par rapport aux grands moyens, en répartissant également, pour chaque cohorte, les entreprises entre ces deux modalités. Il faut noter que cette variable ne prend pas en compte l'Accre qui correspond à un revenu (indemnités de chômage versées au créateur pendant un an) ou à une économie de charge (exonération de cotisations sociales) (tableau 1). Nous avons renoncé à intégrer l'éventuelle obtention de prêts bancaires par les entrepreneurs, en raison des risques de biais de sélection de cette obtention par rapport à la survie. Néanmoins, ci-après, nous étudions les effets de l'introduction de cette variable dans notre modèle sur les résultats (voir robustesse des estimations).

Effet de l'Accre sur la survie à cinq ans des entreprises créées par les chômeurs de court terme

Pour tenir compte de l'hétérogénéité de l'effet de l'Accre selon les différentes catégories de créateurs au moment de la création, chômeurs de CT, chômeurs de LT ou inactifs, nous nous sommes concentrés, dans un premier temps, sur les chômeurs de CT pour lesquels les droits à indemnisation du chômage – et ainsi l'aide apportée par l'Accre – sont relativement homogènes (tableau 1). Nous avons ensuite, par comparaison, pris en compte dans les estimations les autres catégories de bénéficiaires : les chômeurs de LT et les inactifs au moment de la création. Tout d'abord, chaque équation *probit* est estimée séparément, puis les deux sont estimées conjointement. La comparaison entre les deux ensembles de résultats permet de porter un jugement sur l'existence d'un biais de sélection dans l'attribution de l'Accre.

Estimation d'un modèle d'attribution de l'Accre (chômeurs de court terme)

Le tableau A (en annexe), relatif aux chômeurs de CT au moment de la création, met en évidence que le fait d'être de nationalité française favorise l'obtention de l'Accre. De plus, les femmes ne sont pas défavorisées par rapport aux hommes : en 2002 et 2006, cette aide est moins attribuée aux hommes qu'aux femmes, toutes choses égales par ailleurs. Pour ce qui concerne

le niveau d'éducation, les sans-diplômes ont clairement une probabilité inférieure aux titulaires de BEP et CAP de bénéficier de cette aide. En revanche, les titulaires du baccalauréat, de même que les titulaires d'un diplôme de l'enseignement supérieur (sauf en 1998), n'ont pas une probabilité supérieure de recevoir l'Accre que les précédents. Si l'on examine à présent la catégorie socio-professionnelle du créateur avant la création, les dispositions juridiques qui régissent l'attribution de l'Accre ressortent clairement. Les chefs d'entreprise et artisans-commerçants ont une probabilité inférieure aux cadres de recevoir cette aide (sauf en 2006, mais les résultats sont difficiles à interpréter car nous avons dû agréger dans une même catégorie tous les non-salariés). En effet, puisqu'ils ne bénéficient pas d'indemnités de chômage, l'intérêt pour eux de demander l'Accre est faible. Les étudiants qui ont terminé leurs études, et sont devenus créateurs d'entreprise, sont dans le même cas.

Par ailleurs, les petits projets, qui n'emploient aucun salarié, ont une probabilité supérieure aux autres de recevoir cette aide. De plus, l'Accre est souvent associée à d'autres subventions publiques (tableau 2). Dans nos estimations nous avons tenu compte de ce phénomène en introduisant une variable dichotomique (a obtenu / n'a pas obtenu d'autre aide publique), pour prendre en compte l'influence de ces subventions sur la survie des entreprises. Pour ce qui est du statut des entreprises, pour toutes les cohortes, les structures juridiques sous forme de professions libérales, SARL, SA, sociétés en nom collectif, prises globalement, ont une probabilité plus faible que les entreprises créées en nom personnel de recevoir l'Accre. Deux interprétations peuvent en être données. Tout d'abord, l'Accre serait attribuée par l'administration aux personnes les plus éloignées du marché du travail, disposant de moindres moyens juridiques et financiers que celles qui montent des sociétés pour réaliser leur projet, conformément aux objectifs fixés par la loi (voir tableau 1). Mais une toute autre hypothèse peut être avancée : les personnes qui utilisent des montages juridiques complexes n'ont pas le statut de salarié (gérant non salarié, rémunéré sur les profits par exemple), et de ce fait n'ont pas intérêt à demander l'Accre (Daniel et Mandelblat, 2010). Pour ce qui est des moyens financiers mis en œuvre au démarrage de l'entreprise, pour les quatre cohortes, les entreprises qui ont des moyens financiers relativement faibles ont une probabilité inférieure aux autres d'être aidées.

Enfin, les variables de tension du marché du travail au niveau départemental et d'indicatrice de l'octroi de l'Accre au deuxième trimestre de l'année considérée qui, dans l'estimation du modèle à deux équations, seront utilisées comme instruments, ont un effet du signe attendu (resp. < 0 pour la première et > 0 pour la seconde) et significatif (au seuil de 1 % pour la première et de 10 % pour la seconde) sur l'octroi de l'Accre. Ce résultat tend à valider le

raisonnement que nous avons avancé ci-dessus sur les déterminants de l'octroi de cette aide par l'Administration.

Estimation d'un modèle de survie à 5 ans (chômeurs de court terme)

L'estimation de l'équation *probit* de survie à 5 ans (tableau 4) fait apparaître que, pour les

Tableau 4
Estimation du modèle *probit* de variable dépendante : survie à 5 ans (oui/non) – créateurs chômeurs de court terme

Cohorte	1994	1998	2002	2006
Accre (oui/non)	0.30***[0.06]	0.50***[0.06]	0.16***[0.04]	0.14***[0.05]
Petits moyens financiers (ref. grands moyens)	- 0.32***[0.06]	- 0.25***[0.06]	- 0.17***[0.04]	- 0.25***[0.05]
Nationalité française (ref. étrangère)	0.26*[0.11]	0.47***[0.11]	0.11 [0.07]	0.27***[0.08]
Genre homme (ref. femme)	0.05 [0.07]	0.12*[0.07]	0.16***[0.050]	0.042 [0.050]
Classe d'âge > 50 ans (ref. 16 à 50 ans)	0.13 [0.16]	0.13 [0.12]	- 0.04 [0.07]	0.077 [0.067]
Niveau d'éducation, diplôme (ref. CAP, BEP)				
Sans diplôme	- 0.16*[0.08]	- 0.12 [0.09]	- 0.11 [0.07]	- 0.09 [0.08]
BEPC	intégré à BEP	- 0.13 [0.12]	- 0.09 [0.08]	- 0.20**[0.09]
Bac professionnel	- 0.01 [0.08]	- 0.01 [0.10]	0.06 [0.070]	0.02 [0.08]
Bac général	(Bac pro+gén)	- 0.16 [0.13]	- 0.04 [0.08]	- 0.18*[0.10]
Enseignement supérieur	0.01 [0.08]	- 0.12 [0.09]	0.10*[0.05]	0.05 [0.07]
Catégorie professionnelle antérieure (ref. cadre)				
Chef d'entreprise	0.11 [0.26]	- 0.18 [0.30]	0.30* [0.17]	- 0.27 [0.20]
Artisan-commerçant	0.13 [0.18]	- 0.19 [0.16]	avec chef entr	(ens. non salar.)
Agent de maîtrise	0.01 [0.13]	- 0.03 [0.15]	0.12 [0.11]	- 0.04 [0.08]
Profession intermédiaire	- 0.06 [0.13]	- 0.21*[0.13]	(ensemble des	- 0.11 [0.08]
Employé	- 0.06 [0.09]	- 0.23**[0.10]	salarisés	- 0.03 [0.07]
Ouvrier	0.12 [0.10]	- 0.10 [0.10]	sauf cadres)	- 0.02 [0.07]
Étudiant	- 0.17 [0.18]	- 0.26*[0.15]	0.07 [0.16]	(avec ens. non-
Autre : sans activité	- 0.03 [0.16]	- 0.51***[0.15]	- 0.02 [0.12]	salarisés)
Emploi d'aucun salarié (ref. un salarié ou plus)	0.18***[0.07]	- 0.07 [0.08]	0.05 [0.05]	- 0.02 [0.07]
Autres subventions publiques : en a reçu une autre (ref. n'en n'a pas reçu)	n.d.	0.05 [0.10]	- 0.06 [0.05]	0.03 [0.05]
Statut juridique : Sociétés et professions libérales (ref. entreprise personnelle)	0.40***[0.08]	0.16**[0.07]	0.21***[0.05]	0.34*** [0.05]
Constante	- 0.58***[0.20]	- 0.44***[0.18]	- 0.45***[0.15]	0.06 [0.15]
Nombre d'observations	4 230	3 355	5 588	7 300
Log vraisemblance	- 9 111	- 6 875	- 7 840	- 13 516
Pseudo R ²	0.061	0.068	0.031	0.036

Note : les observations ont été pondérées à l'aide de la variable *poidsini* de l'enquête *Sine* (encadré 1). Les écarts-types des coefficients de régression, qui figurent entre crochets, et leur significativité ont été calculés à l'aide de la procédure *robust* du logiciel Stata. Les étoiles indiquent les seuils de significativité de resp. 10 % (*), 5 % (**) et 1 % (***). Pour la cohorte 2006, la variable survie est calculée sur 3 ans, les données sur la survie à 5 ans n'étant pas disponibles. Certaines variables de contrôle ne figurent pas dans le tableau : les indicatrices relatives à la nomenclature d'activité en 6 secteurs (NES16), le nombre de créations d'entreprises précédant la création considérée, le fait de créer seul ou à plusieurs, le fait d'effectuer une étude de marché ou non avant de démarrer, le fait d'être localisé en région Île-de-France ou non. Lecture du tableau : le fait de recevoir l'Accre a un effet positif et significatif (au seuil de 1 %) sur la probabilité de survie à 5 ans des cohortes nées en 1994, 1998, 2002, et à 3 ans pour la cohorte née en 2006, à caractéristiques observables égales. À l'inverse, le fait de disposer de petits moyens pour démarrer l'entreprise a un effet négatif et significatif (au seuil de 1 %) sur la même probabilité de survie pour toutes les cohortes. Champ : entreprises des secteurs marchands non agricoles créées au premier semestre de l'année de référence, France métropolitaine hors Corse ; base de données des auteurs restreinte aux créateurs chômeurs de CT (moins d'un an au chômage au moment de la création de l'entreprise).

Source : Insee, enquêtes *Sine* 1994, 1998, 2002 et 2006.

quatre cohortes de nouvelles entreprises, l'effet de l'Accre sur la survie à 5 ans est significatif (au seuil de 1 %). Cet effet est le plus élevé en 1998, à tel point qu'il existe une différence significative (au seuil de 1 %) entre le coefficient estimé de la variable indicatrice Accre (oui/non) en 1998, d'une part, et ceux estimés pour les cohortes nées en 2002 et 2006, d'autre part. Ces résultats peuvent être rapprochés des données descriptives précédentes (tableau 3) qui montrent une réduction des écarts de taux de survie, entre les entreprises aidées et non aidées par l'Accre, en 2002 et 2006 par rapport à 1998.

Par ailleurs, la catégorie socio-professionnelle antérieure, de même que le niveau de diplôme, ont un effet faible – et le plus souvent non significatif – sur la survie à cinq ans des entreprises. En revanche, les entreprises créées sous la forme juridique de société, ou les professions libérales, ont une perspective de survie à cinq ans supérieure aux autres, créées en nom propre par l'entrepreneur. De plus, les autres formes d'aide publique à la création (régionales et locales), n'ont pas d'effet sur la survie à cinq ans des entreprises. Enfin, le fait de disposer de faibles moyens financiers diminue la probabilité de survie à cinq ans des entreprises.

Effet de l'Accre sur la survie des entreprises dans un modèle à équations simultanées (chômeurs de court terme)

L'estimation du modèle à deux équations simultanées tient compte des interdépendances entre les variables des deux équations et permet, si le modèle est correctement formalisé, d'éliminer l'effet sur la survie à 5 ans d'une éventuelle sélection dans l'attribution de l'Accre. Elle est réalisée par la méthode du maximum de vraisemblance à l'aide du logiciel *Stata* (encadré 2). Selon cette estimation, pour les chômeurs de CT (tableau 5), l'effet de l'Accre sur la survie à cinq ans n'est pas significativement différent de zéro (au seuil de 10 %) pour les quatre cohortes. Nous obtenons donc un résultat fondamentalement différent du précédent pour la simple équation de survie (tableau 4). Un effet de sélection est ainsi bien mis en évidence : il remet en question l'efficacité de l'Accre comme mesure de politique économique visant à augmenter la survie des entreprises à cinq ans pour les quatre cohortes.

Tableau 5

L'Accre comme déterminant de la survie des entreprises à 5 ans – Estimation du modèle à deux équations simultanées – créateurs, chômeurs de court terme

Cohorte	1994	1998	2002	2006
Équation de survie à 5 ans :				
Accre (oui/non)	0.19 [0.14]	0.18 [0.13]	- 0.07 [0.24]	0.03 [0.15]
Équation d'attribution de l'Accre :				
Tension sur le marché départemental du travail	- 0.65***[0.23]	- 0.93***[0.20]	- 0.26**[0.12]	- 0.30***[0.07]
Entreprise créée au 2 ^e trim. (ref. 1 ^{er} trim.)	0.10*[0.6]	0.08*[0.05]	0.13**[0.05]	0.07 [0.05]
Nombre d'observations	4 230	3 355	5 588	7 300
Corr. résidus des deux équations	- 0.23 [0.33]	- 0.13 [0.37]	0.23 [0.60]	0.01 [0.27]

Note : estimations du modèle par le maximum de vraisemblance à l'aide du logiciel *Stata* (encadré 2). Les observations ont été pondérées à l'aide de la variable *poidsini* de l'enquête *Sine* (encadré 1). Pour la cohorte 2006, la variable survie est calculée sur 3 ans, les données sur la survie à 5 ans n'étant pas disponibles. La base de données est restreinte aux créateurs chômeurs de CT (moins d'un an au chômage au moment de la création de l'entreprise). Les écarts-types des coefficients de régression, qui figurent entre crochets, et leur significativité ont été calculés à l'aide de la procédure *robust* du logiciel *Stata*. Les étoiles indiquent les seuils de significativité de resp. 10 % (*), 5 % (**) et 1 % (***). Certaines variables de contrôle ne figurent pas dans le tableau : les indicatrices relatives à la nomenclature d'activité en 6 secteurs (NES16), le nombre de créations d'entreprises précédant la création considérée, le fait de créer seul ou à plusieurs, le fait d'effectuer une étude de marché ou non avant de démarrer, le fait d'être localisé en région Île-de-France ou non.

Lecture du tableau : le fait de recevoir l'Accre n'a pas d'effet positif et significatif (au seuil de 10 %) sur la probabilité de survie à 5 ans des cohortes nées en 1994, 1998, 2002, et à 3 ans pour la cohorte née en 2006, à caractéristiques observables égales.

Champ : entreprises des secteurs marchands non agricoles créées au premier semestre de l'année de référence, France métropolitaine hors Corse ; base de données des auteurs restreinte aux créateurs chômeurs de CT (moins d'un an au chômage au moment de la création de l'entreprise).

Source : Insee, enquêtes *Sine* 1994, 1998, 2002 et 2006.

Les coefficients estimés des variables instrumentales sont du signe attendu et sont significatifs (au seuil de 1 % pour la variable de tension sur le marché départemental de l'emploi, et au seuil de 10 % pour la variable indicatrice du trimestre de création). Néanmoins, une rupture apparaît dans l'estimation de l'effet de la tension du marché du travail départemental sur l'octroi de l'Accre après 1998. En effet, le coefficient estimé est inférieur (au seuil de 1 %) en 2002 et 2006, à celui de 1998. Ceci peut être interprété comme une évolution dans le processus de sélection qui tendrait à diminuer après 1998, alors que dans le même temps la part des chômeurs qui bénéficient de l'Accre augmente sensiblement en 2002 et 2006 par rapport à 1998 (tableau 2). On observe une évolution semblable des coefficients estimés de la variable indicatrice Accre (oui/non), puisqu'ils sont significatifs au seuil de 20 % pour les cohortes nées en 1994 et 1998, puis ne le sont plus pour les cohortes suivantes. Dans cette perspective, la réduction de l'effet apparent de l'Accre dans les données descriptives (tableau 3) peut s'interpréter comme une diminution de la

sélection, elle-même liée à la plus large diffusion de l'Accre parmi les chômeurs.

Enfin, la corrélation entre les résidus des deux équations n'est jamais significative. Selon ce résultat, il n'existe pas de variable inobservée qui serait liée à la fois à l'attribution de l'Accre d'une part, et à la survie des entreprises d'autre part.

Effet de l'Accre sur la survie à cinq ans des entreprises créées par différentes catégories de bénéficiaires

Nous étendons ci-après l'analyse portant sur les chômeurs de CT à d'autres catégories de créateurs éligibles à l'obtention de l'Accre. Nous estimons le modèle à deux équations sur la population des créateurs, en incluant non seulement les chômeurs de CT, mais aussi les chômeurs de long terme (LT) et les inactifs au moment de la

Tableau 6
L'Accre comme déterminant de la survie des entreprises à 5 ans – Estimation du modèle à deux équations simultanées – créateurs chômeurs de court terme, de long terme et inactifs

Cohorte	1994	1998	2002	2006
Équation de survie à 5 ans :				
- Accre (oui/non)	0.11 [0.09]	0.23***[0.06]	0.06 [0.08]	- 0.09 [0.07]
- Catégorie de bénéficiaires (ref. chômeurs de CT)				
Chômeurs de LT	- 0.01 [0.05]	- 0.12 [0.08]	- 0.03 [0.03]	- 0.02 [0.03]
Inactifs	0.08 [0.20]	0.13 [0.09]	0.11 [0.11]	- 0.21* [0.12]
Équation d'attribution de l'Accre :				
- Catégorie de bénéficiaires (ref. chômeurs de CT)				
Chômeurs de LT	- 0.19***[0.05]	0.04 [0.04]	0.09***[0.03]	0.02 [0.03]
Inactifs	- 1.65***[0.07]	- 1.12***[0.06]	- 1.13***[0.04]	- 1.38***[0.05]
- Tension sur le marché départemental du travail	- 0.56***[0.19]	- 0.74***[0.13]	- 0.31***[0.07]	- 0.21***[0.04]
- Entreprise créée au 2 ^e trim. (ref. 1 ^{er} trim.)	0.15***[0.05]	0.09**[0.04]	0.091***[0.031]	0.06*[0.03]
Nombre d'observations	8 256	8 269	13 792	18 416
Corr. résidus des deux équations	- 0.14 [0.22]	- 0.33* [0.16]	- 0.07 [0.18]	0.29 [0.15]

Note : estimations du modèle par le maximum de vraisemblance à l'aide du logiciel Stata (encadré 2). Les observations ont été pondérées à l'aide de la variable *poidsini* de l'enquête *Sine* (encadré 1). Pour la cohorte 2006, la variable survie est calculée sur 3 ans, les données sur la survie à 5 ans n'étant pas disponibles. La base de données est restreinte aux créateurs chômeurs de CT (moins d'un an au chômage au moment de la création de l'entreprise). Les écarts-types des coefficients de régression, qui figurent entre crochets, et leur significativité ont été calculés à l'aide de la procédure *robust* du logiciel Stata. Les étoiles indiquent les seuils de significativité de resp. 10 % (*), 5 % (**) et 1 % (***). Certaines variables de contrôle ne figurent pas dans le tableau : les indicatrices relatives à la nomenclature d'activité en 6 secteurs (NES16), le nombre de créations d'entreprises précédant la création considérée, le fait de créer seul ou à plusieurs, le fait d'effectuer une étude de marché ou non avant de démarrer, le fait d'être localisé en région Île-de-France ou non. Seules les variables d'intérêt figurent dans le tableau.

Lecture du tableau : équation de survie à 5 ans, le fait de recevoir l'Accre n'a pas d'effet significatif (au seuil de 10 %) sur la probabilité de survie à 5 ans des cohortes nées en 1994, 2002, et à 3 ans pour la cohorte née en 2006, à caractéristiques observables égales. En revanche, cet effet est positif et significatif (au seuil de 1 %), pour la cohorte née en 1998.

Source : Insee, enquêtes *Sine* 1994, 1998, 2002 et 2006.

création. Nous introduisons dans les deux équations une variable dichotomique pour chacune de ces deux catégories de créateurs par rapport aux chômeurs de CT. L'objectif est d'évaluer si l'effet de l'Accre sur la survie est affecté par la prise en compte de ces deux catégories supplémentaires de créateurs.

Le fait de prendre en compte la totalité des créateurs éligibles à l'Accre multiplie par un facteur qui va de deux à trois la taille de l'échantillon de chaque cohorte. Cet accroissement peut permettre une estimation plus précise des coefficients estimés des deux équations. Si l'on considère les résultats de l'estimation de l'équation de l'attribution de l'Accre, il est confirmé que les inactifs en bénéficient beaucoup moins que les chômeurs de CT, ce qui correspond aux résultats obtenus en statistique descriptive (tableau 2). En revanche les chômeurs de LT, suivant les cohortes, ont une probabilité supérieure (2002), ou inférieure (1994), aux chômeurs de CT de la recevoir. Pour ce qui est de la survie, l'estimation de l'équation correspondante montre que la probabilité d'être encore en activité au bout de 5 ans, pour les entreprises créées par des chômeurs de LT et les inactifs, n'est pas significativement différente des chômeurs de CT (sauf en 2006 pour les inactifs, où elle est inférieure).

Finalement l'effet de l'Accre sur la survie des entreprises, toutes catégories de créateurs confondus, n'est pas différent de zéro, de même que pour les créateurs chômeurs de CT, à l'exception notable de la cohorte née en 1998. En effet, pour cette cohorte, son effet sur la survie est positif et significatif (au seuil de 1 %). Une première hypothèse serait que, pour la cohorte née en 1998, cet effet soit positif pour les chômeurs de LT et inactifs. Cependant nous pouvons la rejeter, car l'estimation de notre modèle, pour chacune de ces catégories de créateurs prise séparément, a montré que ce n'était pas le cas. C'est plutôt l'accroissement de la taille de l'échantillon qui accroît la précision de l'estimation. Pourtant, malgré des tailles d'échantillons nettement supérieures en 2002 et 2006, liées à l'accroissement du nombre de bénéficiaires (tableau 6 par rapport au tableau 5), pour ces deux cohortes, l'effet de l'Accre n'est pas significatif (et même négatif pour 2006). Il y a bien une rupture dans l'effet de l'Accre entre la cohorte née en 1998 et les suivantes.

À partir de 2002, le pourcentage de chômeurs et inactifs bénéficiaires de l'Accre a cru sensiblement (tableau 2), attirant de nouvelles

catégories de créateurs. La relation trouvée entre cette aide et la survie est, pour certaines catégories de créateurs, négative (mais non significative). C'est le cas pour les chômeurs de CT en 2002 (tableau 5) et l'ensemble des bénéficiaires de l'Accre en 2006. La population de ceux-ci a changé quantitativement et qualitativement. L'Accre est apparue de plus en plus, à partir du début des années 2000, comme une prestation sociale supplémentaire qui était donnée automatiquement, de façon administrative, aux créateurs. Ce fait a été consacré officiellement par le décret de 2007 (tableau 1). L'extension de la législation a pu inciter des personnes à créer une entreprise, alors qu'elles n'auraient pas tenté l'aventure, s'il n'y avait pas l'attrait de l'Accre. L'auto-sélection incite dans ce cas davantage de personnes à demander cette aide, alors que leurs caractéristiques individuelles font qu'elles sont moins qualifiées pour créer une entreprise que lorsque les contraintes dans l'attribution de l'Accre étaient plus dures. L'étude particulière des titulaires de minima sociaux, créateurs d'entreprises des cohortes 2002 et 2006, confirme ce jugement (voir le complément en ligne C2).

Robustesse des estimations

Prise en compte des variables financières

Nous avons pris en compte jusqu'à présent les moyens financiers utilisés au démarrage de l'entreprise (tableaux 4 et A1-1), sans distinguer les prêts bancaires dont elle peut disposer. En effet, l'on ne peut pas exclure que cette variable soit endogène par rapport à la survie. La comparaison de l'estimation des deux équations pour les chômeurs de CT, suivant que la variable « moyens financiers dont dispose l'entreprise au démarrage » est incluse ou non dans les deux équations du modèle, montre que les coefficients de régression sont très proches et, dans tous les cas, ne diffèrent pas significativement de zéro, pour ce qui est de l'effet de l'Accre sur la survie (tableau C4-1 dans le complément en ligne). Il en est de même des coefficients des variables instrumentales qui ne sont pas significativement modifiés. De plus, lorsque l'on remplace dans les deux mêmes équations, la variable « moyens financiers au démarrage » par la variable « obtention d'un prêt bancaire », les résultats des estimations obtenus⁵ ne

5. Non présentés ici et disponibles auprès de l'auteur.

diffèrent pas significativement de notre estimation initiale (tableau 5). Ces conclusions sont importantes car elles montrent que les moyens financiers mis en œuvre ont un effet significatif sur la survie à 5 ans (tableau 4) mais que, selon cette analyse, ils sont indépendants de l'effet de l'Accre sur cette survie.

Le cycle économique a-t-il un effet différencié sur la survie des entreprises suivant qu'elles sont aidées ou non aidées ?

La question de l'impact de la période d'analyse sur les résultats de la survie doit être posée. La survie des entreprises est en effet sensible à l'évolution du cycle économique. Jacobson et al. (2011), d'une part, et Fougère et al. (2013), d'autre part, ont mis en évidence un fort effet de la crise de 2008 sur les défaillances d'entreprises, avec cependant un décalage important, selon ces derniers, puisque l'effet de la crise sur les défaillances ne deviendrait significatif qu'à partir de 2009, pour s'accroître jusqu'à fin 2010,

date à laquelle se termine leur étude. Dans le cas présent, il s'agit de savoir s'il existe un effet différencié du cycle sur l'effet de l'Accre sur la survie des entreprises. Il faut considérer aussi que l'enquête *Sine* ne concerne, pour chaque cohorte, que les entreprises nées au premier semestre.

Le premier enseignement du tableau 7 est que l'effet comparé de l'Accre sur la survie des entreprises créées par des chômeurs de CT à trois et quatre ans (deux ans pour la cohorte née en 2006), n'est pas significativement différent de l'effet sur la survie à cinq ans (trois ans pour la cohorte 2006). Il généralise nos analyses à cinq ans précédentes à des échéances plus courtes. L'effet de l'Accre n'est pas différent de zéro pour les cohortes 1994, 2002 et 2006. La cohorte née en 1998 se distingue des autres, puisque cet effet, qui n'était pas significatif au seuil de 10 % sur la survie à 5 ans, le devient pour la survie à 4 ans et à 3 ans. Néanmoins, les coefficients et les écarts-types associés aux trois régressions correspondantes montrent que ces différences entre les coefficients ne

Tableau 7

Effet de l'Accre sur la survie des entreprises à 3, 4, et 5 ans – Estimation du modèle à deux équations simultanées – créateurs chômeurs de CT

	Équation de survie : effet de l'Accre sur la survie	Équation d'attribution de l'Accre : effet de la variable « Tension du marché du travail »	Équation d'attribution de l'Accre : effet de la variable « Attribution au 2 ^e trimestre »
Cohorte 1994			
3 ans	0.17 [0.18]	- 0.66***[0.23]	0.08*[0.05]
4 ans	0.17 [0.13]	- 0.64***[0.23]	0.09*[0.05]
5 ans	0.19 [0.14]	- 0.65***[0.24]	0.10*[0.06]
Cohorte 1998			
3 ans	0.20*[0.11]	- 0.94***[0.20]	0.08*[0.05]
4 ans	0.20*[0.12]	- 0.94***[0.20]	0.07*[0.04]
5 ans	0.18 [0.13]	- 0.93***[0.20]	0.08*[0.05]
Cohorte 2002			
3 ans	0.08 [0.32]	- 0.30**[0.14]	0.10*[0.06]
4 ans	0.10 [0.23]	- 0.31***[0.11]	0.09*[0.05]
5 ans	- 0.07 [0.24]	- 0.26**[0.12]	0.13*[0.05]
Cohorte 2006			
2 ans	0.03 [0.11]	- 0.30***[0.06]	0.09*[0.05]
3 ans	0.03 [0.15]	- 0.30***[0.07]	0.07*[0.05]

Note : estimations du modèle par le maximum de vraisemblance à l'aide du logiciel Stata (encadré 2). Les observations ont été pondérées à l'aide de la variable *poisini* de l'enquête *Sine* (encadré 1). Pour la cohorte 2006, la variable survie est calculée sur 3 ans, les données sur la survie à 5 ans n'étant pas disponibles. La base de données est restreinte aux créateurs chômeurs de CT (moins d'un an au chômage au moment de la création de l'entreprise). Les écarts-types des coefficients de régression, qui figurent entre crochets, et leur significativité ont été calculés à l'aide de la procédure *robust* du logiciel Stata. Les étoiles indiquent les seuils de significativité de resp. 10 % (*), 5 % (**) et 1 % (***). Certaines variables de contrôle ne figurent pas dans le tableau : les indicatrices relatives à la nomenclature d'activité en 6 secteurs (NES16), le nombre de créations d'entreprises précédant la création considérée, le fait de créer seul ou à plusieurs, le fait d'effectuer une étude de marché ou non avant de démarrer, le fait d'être localisé en région Île-de-France ou non. Seules les variables d'intérêt figurent dans le tableau.

Lecture du tableau : le fait de recevoir l'Accre n'a pas d'effet positif et significatif (au seuil de 10 %) sur la probabilité de survie à 3 ans, 4 ans, 5 ans pour les cohortes nées en 1994 et 2002. Il n'a pas d'effet non plus sur la probabilité de survie à 2 ans et à 3 ans des entreprises nées en 2006. Pour l'année 1998, ce même effet est positif sur la survie à 3 ans et à 4 ans (au seuil de 10 %).

Source : Insee, enquêtes *Sine* 1994, 1998, 2002 et 2006.

sont pas significatives. Nous nous en tenons à notre appréciation initiale d'un effet faible de l'Accre sur la cohorte née en 1998, quelle que soit l'échéance considérée, pour les chômeurs de CT.

Compte tenu de ces résultats, il n'est pas possible de mettre en évidence une influence significative du cycle du chômage sur l'effet de l'Accre sur la survie des entreprises. Par exemple, pour la cohorte née en 1998, le palier bas du taux de chômage du premier semestre 2001 jusqu'au premier semestre 2002, a été suivi par une remontée au début 2003 (figure 4-I dans le complément en ligne). Peut-on interpréter la légère baisse du coefficient estimé de l'effet de l'Accre à 5 ans par rapport aux années précédentes comme une conséquence de la dégradation du marché de l'emploi ? Nous ne le pensons pas, car cette baisse n'est pas significative par rapport aux années précédentes. De plus, les études précédemment citées font ressortir un décalage de plusieurs mois, voire de plusieurs semestres, entre l'évolution du cycle et les défaillances d'entreprises, donc si un tel effet existait, il le serait sur les années suivantes.

Pour la cohorte née au premier semestre 2002, un pic de chômage apparaît au dernier semestre 2005 et au premier semestre 2006, suivi par une légère amélioration en 2007. Le coefficient négatif de l'effet de l'Accre sur la survie à l'échéance de cinq ans (2007) peut-il être interprété comme un effet retardé du pic de chômage des deux années précédentes ? On ne peut répondre par l'affirmative étant donné la faiblesse de ce coefficient. Pour la cohorte née en 2006, le premier semestre 2008 correspond à un minimum de chômage suivi d'une brusque augmentation au premier semestre 2009. Un tel choc conjoncturel n'a pas eu, selon nos estimations, d'incidence sur l'effet de l'Accre, tout au moins à court terme.

* *

*

Notre recherche sur l'efficacité de l'Accre, comme mesure favorisant les chances de survie à cinq ans des entreprises créées par des chômeurs et des inactifs, met en évidence un biais de sélection. En effet, pour les cohortes nées en 1994, 2002 et 2006, la correction de ce biais fait ressortir un effet nul de l'Accre

sur la survie des entreprises aidées, quelle que soit la catégorie de créateur (chômeur de CT, chômeur de LT, inactif). Pour la cohorte née en 1998, cet effet est faible pour les chômeurs de CT (significatif au seuil de 10 % à 4 ans et de 20 % à 5 ans), mais significatif (au seuil de 1 %) lorsque l'échantillon est élargi à l'ensemble des personnes éligibles.

Cette différence peut s'expliquer par les modifications de la réglementation. L'année 1998 correspond à une période de forte restriction budgétaire qui a concerné toutes les catégories de bénéficiaires (tableau 2). Au contraire, lors du second trimestre de l'année 1994, le nombre de bénéficiaires a augmenté, en liaison avec la générosité de l'aide qui s'est accrue. Pour les années 2002 et 2006, ce sont les conditions d'éligibilité qui ont été élargies. On ne peut exclure que ces changements de la réglementation aient constitué des signaux qui auraient conditionné non seulement les effectifs, mais aussi les caractéristiques individuelles des chômeurs et inactifs, créateurs d'entreprise. En 1998, le caractère dissuasif du signal a pu limiter la population des candidats aux plus compétents et aux plus capables pour créer. Il y aurait alors eu un effet d'auto-sélection des créateurs qui peut expliquer nos résultats sur l'effet de l'Accre pour cette cohorte. En effet, les variables instrumentales de notre modèle sont conçues en premier lieu pour capter les évolutions économiques départementales prises en compte par les responsables de l'Administration dans leur décision d'attribution de l'aide. Il est possible que ces variables saisissent moins bien le comportement des créateurs potentiels, plus sensibles aux évolutions nationales (restrictions des crédits alloués à l'Accre au niveau national, ou encore situation macroéconomique générale).

En revanche, les législations plus accommodantes du second trimestre de l'année 1994, et à partir du début des années 2000, ont pu inciter de nouvelles catégories de personnes (notamment les allocataires des minima sociaux) à créer leur entreprise. La législation plus généreuse a entraîné un accroissement important du nombre de bénéficiaires, mais n'a pas eu d'impact sur la survie à 5 ans des entreprises aidées. L'auto-sélection peut alors avoir joué en sens inverse : des personnes moins compétentes et moins aptes à la création ont tenté leurs chances. Un effet d'aléa moral a pu également jouer. Les personnes qui ont été incitées à créer une entreprise ont pris un risque et un engagement réduit, dans la mesure

où elles avaient une forte probabilité d'obtenir l'aide, et pouvaient être moins préparées et motivées pour mener à bien leur projet. À nouveau, il est possible que ces effets de sélection ou d'aléa moral aient été imparfaitement contrôlés dans notre modèle. Ce résultat sur l'efficacité réduite, voire nulle, du dispositif

sur la survie des entreprises, ne doit cependant pas cacher un éventuel aspect qualitatif. L'élargissement des conditions d'éligibilité de l'Accre est susceptible d'avoir incité des populations éloignées du marché du travail à créer leur propre entreprise et pourraient ainsi avoir contribué à améliorer leur employabilité. □

BIBLIOGRAPHIE

- 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'ACCRES 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.org/10.1787/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 area: evidence from a matching approach. *IZA Journal of Labour Economics*, 4(7), 35–53.
- Duhautois, R., Désiège, 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.
- Désiège, 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 Institute for Entrepreneurship Public Policy N° WR-603-EMKF.

Guimirot, 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 aggregate 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.

ANNEXE

Tableau A

Estimation du modèle *probit* de variable dépendante : obtention de l'Accre (oui/non) - créateurs chômeurs de CT

Cohorte	1994	1998	2002	2006
Petits moyens financiers (ref. grands moyens)	- 0.44** [0.07]	- 0.18***[0.07]	- 0.16**[0.05]	- 0.26**[0.05]
Nationalité française (ref. étrangère)	0.47***[0.11]	0.26**[0.13]	0.28***[0.07]	0.19**[0.08]
Genre homme (ref. femme)	- 0.01 [0.08]	- 0.11 [0.07]	- 0.12**[0.05]	- 0.09* [0.05]
Classe d'âge > 50 ans (ref. 16 à 50 ans)	- 0.03 [0.17]	- 0.33**[0.13]	- 0.02 [0.07]	- 0.10 [0.07]
Niveau d'éducation, diplôme (ref. CAP, BEP)				
Sans diplôme	- 0.11 [0.08]	- 0.30***[0.09]	- 0.28***[0.07]	- 0.30***[0.09]
BEPC	intégré à BEP	- 0.10 [0.11]	- 0.02 [0.08]	- 0.22**[0.09]
Bac professionnel	- 0.02 [0.08]	- 0.07 [0.10]	- 0.03 [0.07]	- 0.10 [0.09]
Bac général	(Bac pro +gén)	- 0.02 [0.13]	0.06 [0.08]	- 0.15 [0.11]
Enseignement supérieur	0.06 [0.10]	0.17** [0.09]	0.08 [0.06]	- 0.10 [0.07]
Catégorie professionnelle antérieure (ref. cadre)				
Chef d'entreprise	- 0.97***[0.23]	- 0.50* [0.33]	- 0.16 [0.17]	- 0.24 [0.26]
Artisan-commerçant	- 0.91***[0.19]	- 0.51***[0.17]	avec ch. entr	(ens. non salariés)
Agent de maîtrise	0.16 [0.14]	0.25 [0.16]	0.21*[0.12]	- 0.03 [0.08]
Profession intermédiaire	0.15 [0.14]	- 0.01 [0.13]	(ensemble des salariés sauf cadres)	- 0.06 [0.09]
Employé	0.11 [0.10]	- 0.08 [0.10]		- 0.16** [0.07]
Ouvrier	- 0.06 [0.11]	- 0.04 [0.11]		- 0.18**[0.09]
Étudiant	- 0.92***[0.18]	- 0.51***[0.14]	0.09 [0.17]	(avec ens.non salariés)
Autres inactifs	- 0.73***[0.17]	- 0.41***[0.15]	0.01 [0.12]	
Emploi d'aucun salarié (ref. emploi d'un salarié ou plus)	0.046 [0.072]	0.49***[0.08]	0.17***[0.05]	0.19**[0.06]
Autres subventions publiques : en a reçu (ref. n'en n'a pas reçu)	d.m	0.62***[0.12]	0.38***[0.05]	0.28***[0.06]
Statut juridique : Stés et professions libérales (ref. entr. personnelle)	- 0.36*** [0.08]	- 0.55***[0.08]	- 0.36***[0.05]	- 0.20***[0.06]
Entreprise créée au 2 ^e trimestre (ref. 1 ^{er} trimestre)	0.10* [0.06]	0.09* [0.05]	0.11***[0.04]	0.05 [0.05]
Tension sur le marché départemental du travail	- 0.62*** [0.23]	- 0.94***[0.20]	- 0.28*** [0.11]	- 0.30***[0.07]
Constante	1.04***[0.23]	- 0.21 [0.21]	- 0.03 [0.17]	0.297*[0.160]
Nombre d'observations	4 230	3 355	5 588	7 300
Log vraisemblance	- 7 691	- 6 170	- 7 437	- 12 037
Pseudo R2	0.116	0.135	0.087	0.077

Note : les observations ont été pondérées à l'aide de la variable *poidsini* de l'enquête *Sine* (encadré 1). Écarts-types, champ, variables de contrôle ne figurant pas dans ce tableau : voir 4 dans le texte.

Pour la cohorte 2006, la variable survie est calculée sur 3 ans, les données sur la survie à 5 ans n'étant pas disponibles.

Lecture du tableau : les personnes de nationalité française ont une probabilité supérieure (significative au seuil de 1 % ou de 5 % suivant les cohortes de recevoir l'Accre.

Source : Insee, enquêtes *Sine* 1994, 1998, 2002 et 2006.

L'efficacité des aides publiques à la R&D et à l'entrepreneuriat

Effectiveness of public support for R&D and entrepreneurship

Commentaire sur les articles « L'effet des aides à la R&D sur l'emploi : une évaluation pour les petites entreprises en France » de Vincent Dortet-Bernadet et Michaël Sicsic, et « 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 » de Dominique Redor.

Pierre Mohnen *

Résumé – L'article de Dortet-Bernadet et Sicsic et l'article de Redor parus dans ce numéro examinent, pour le premier, l'efficacité des programmes d'aide financière à la R&D destinée à stimuler la R&D du secteur privé et, pour le second, la réussite des start-ups subventionnées pour les chômeurs créant des entreprises durables. Les deux articles se concentrent sur les petites entreprises françaises. L'impact de chacun des dispositifs étudiés apparaît limité par des effets d'aubaine. Ce commentaire discute des résultats obtenus et tire des enseignements pour les politiques publiques. Nous y verrons que l'effet d'aubaine est en partie inévitable, mais qu'il est possible d'en limiter l'ampleur, par exemple en utilisant un mélange d'incitations fiscales à la R&D et de subventions, privilégiant l'incitation fiscale pour les petites entreprises et les subventions pour les grandes entreprises. Nous rappelons aussi qu'une politique doit être évaluée selon plusieurs perspectives. Outre l'effet additif de la R&D et la survie des entreprises, une analyse coûts-bénéfices complète devrait également prendre en compte les externalités de la R&D, le maintien d'activités de R&D sur le territoire et la baisse induite du chômage.

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 additivity and firm survival a full cost benefit analysis would also consider R&D externalities, firm retention and decrease in unemployment.

Codes JEL / JEL codes : O31, J68

Mots clés : aides à la R&D, subventions à la R&D, subventions aux start-ups, évaluation de politiques publiques

Keywords: R&D tax incentives, R&D subsidies, start-up subsidies, policy evaluation

Rappel :

Les jugements et opinions exprimés par les auteurs n'engagent qu'eux mêmes, et non les institutions auxquelles ils appartiennent, ni a fortiori l'Insee.

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

Les deux excellents articles parus dans ce numéro d'*Economie et Statistique / Economics and Statistics* contribuent de manière intéressante et utile au débat sur l'efficacité du soutien financier à la R&D d'une part et aux start-ups créées par des chômeurs d'autre part. Vincent Dortet-Bernadet et Michaël Sicsic étudient l'efficacité du soutien public direct et indirect aux investissements en R&D des très petites entreprises en France. Dominique Redor évalue quant à lui la réussite d'une autre politique française à destination des petites entreprises : le soutien à la création de start-ups par les chômeurs.

Comme la productivité marginale du capital diminue, le seul moyen de faire croître le PIB/habitant est de mieux utiliser les ressources, d'introduire de nouvelles technologies plus efficaces ou de fabriquer des produits intermédiaires nouveaux ou des produits finaux permettant d'obtenir de meilleurs rendements (ou étant plus utiles au consommateur), avec moins de ressources. Dans ce processus de destruction créatrice schumpeterienne, de nouveaux produits remplacent les anciens, et de nouveaux acteurs supplantent ceux qui sont déjà là. Dans une certaine mesure, cela se produit de manière endogène, sous l'effet des seules forces du marché. Cependant, même sans tenir compte des difficultés sociales induites par ce processus de destruction créatrice, le marché laissé à lui-même n'atteint pas toujours une croissance ni un développement économique optimaux, en raison de ses défaillances. Les entrepreneurs qui suivent « la main invisible » ne prennent pas toujours en compte les externalités, par exemple en n'investissant pas suffisamment dans la R&D d'un point de vue sociétal, ou en n'innovant pas sur des technologies moins polluantes. En raison des défauts de coordination, les entrepreneurs privés peuvent imposer trop de restrictions sur certaines ressources, ce qui empêche la réalisation d'autres objectifs sociétaux, sans parler des aléas moraux ou des comportements anticoncurrentiels volontaires. Enfin, la connaissance étant un bien public, les innovateurs peuvent être réticents à fournir les informations requises par ceux qui les financent pour justifier leurs emprunts. C'est particulièrement vrai pour les petites entreprises et les start-ups, qui ne disposent pas de caution ni d'autres garanties pour appuyer leurs demandes de financement.

Une difficulté technique dans l'évaluation de l'efficacité de ces interventions publiques vient du caractère endogène des entreprises recevant l'aide, ainsi que de l'auto-sélection des

candidats au dispositif. En effet, il est possible que le groupe d'entreprises recevant les subventions ou les incitations fiscales à la R&D et celui des chômeurs bénéficiant de l'Accre soient plus travailleurs, productifs ou efficaces que ceux qui ne reçoivent pas d'aide. Il est aussi possible que cette première catégorie recoure plus facilement aux aides. Dès lors, la meilleure performance économique des acteurs recourant à l'aide n'est pas seulement (voire pas du tout) due à l'aide elle-même. La difficulté économétrique est de filtrer l'effet de ces deux types de biais. Les deux articles traitent de sujets similaires et de catégories d'entreprises proches ; ils traitent de la question de l'endogénéité du soutien public avec sérieux et intelligence, mais avec une approche légèrement différente.

Notre commentaire sera organisé de la façon suivante : nous allons d'abord résumer la méthode d'analyse et les résultats obtenus par les deux articles. Nous en ferons ensuite une analyse critique, en les comparant à d'autres études économiques. Pour conclure, nous proposerons des recommandations de politiques à la lumière des conclusions apportées par les deux études.

Résumé des deux articles

Vincent Dortet-Bernadet et Michaël Sicsic évaluent conjointement le financement direct et indirect apporté à l'emploi en R&D dans les petites et moyennes entreprises (PME) françaises. De nombreux articles ont étudié l'efficacité des incitations fiscales et des subventions directes à la R&D en France et dans d'autres pays (voir les travaux de Ientile & Mairesse, 2009 ; Köhler et al., 2012 ; Commission européenne, 2014 ; Zuñiga-Vicente et al., 2014). Cette étude se distingue par trois aspects. Tout d'abord, elle inclut les très petites entreprises, c'est-à-dire celles qui ont moins de 10 employés, ainsi qu'un chiffre d'affaires et des actifs inférieurs à 2 millions d'euros, quand la plupart des études à partir d'enquêtes sur la R&D portent d'abord sur les grandes entreprises. Ici, les très petites entreprises constituent les deux tiers du panel mobilisé. Deuxièmement, l'étude combine de nombreuses bases de données, essentiellement celles du Crédit d'impôt recherche (CIR), du programme Jeunes entreprises innovantes (JEI), l'enquête sur la R&D, la liste d'entreprises accréditées par le ministère de la Recherche, ainsi que diverses données fiscales, sociales, financières et administratives mises à disposition par l'Insee. Cet effort en matière de sources donne une vision unique,

presque exhaustive, des PME et des TPE françaises. Troisièmement, l'étude examine en même temps les mesures de financement direct et indirect à la R&D, alors que la plupart des études précédentes n'abordent qu'un seul type de mesures de soutien, omettant ainsi un facteur important de l'effet des mesures.

Les auteurs construisent soigneusement leur échantillon d'étude, en associant d'abord chaque entreprise ayant reçu un soutien financier de R&D au moins une fois entre 2003 et 2010 à trois entreprises de même ancienneté mais n'ayant pas reçu de soutien financier sur la même période, alors qu'elles avaient les mêmes possibilités de recevoir de l'aide en 2003 ou en 2007. Ils obtiennent ensuite, à l'aide de la méthode des moments généralisés, une équation dynamique de la demande de travail en R&D, fonction d'une part des niveaux de chiffre d'affaires présents et passés, et d'autre part du coût relatif de l'emploi hautement qualifié, comparé à d'autres types de main d'œuvre. Cette dernière variable est ensuite instrumentée en multipliant les modifications officielles du taux de crédit d'impôt recherche par la part de l'emploi R&D dans l'emploi juste avant la modification. Pour terminer, les différences premières en emploi R&D liées aux changements du coût relatif de la main d'œuvre R&D et aux fluctuations de chiffre d'affaires dues aux modifications du CIR sont calculées et soustraites de celles des entreprises similaires mais n'ayant pas bénéficié d'un soutien financier à la R&D. Comme les sommes investies pour financer la R&D et le coût moyen de la main-d'œuvre R&D sont connues, on peut calculer les différences sur l'emploi R&D induites par les mesures incitatives et, partant, l'effet sur l'emploi ne dépendant pas des subventions publiques.

Les auteurs en concluent que l'aide à la R&D a accru l'emploi de R&D, mais qu'elle a partiellement engendré un effet d'éviction. Autrement dit, le nombre de postes de R&D financés par le secteur privé a diminué. Les entreprises privées ont utilisé une partie de l'aide financière publique pour réduire leur propre investissement dans les emplois de R&D. Cette diminution statistiquement significative est particulièrement observable après la réforme du crédit d'impôt menée en 2008, qui a remplacé le système incrémental par un système fondé sur le volume. Pour les entreprises existant depuis 2007, seuls 24 % du supplément d'aides financières à la R&D avaient été consacrés à de nouvelles embauches sur des emplois de R&D.

Dominique Redor examine l'effet de l'Accre – un dispositif visant à aider les chômeurs à créer ou à reprendre une entreprise – sur la survie des entreprises à cinq ans. Son analyse économétrique se fonde sur quatre cohortes d'entreprises (1994, 1998, 2002 et 2006), obtenues à partir d'un échantillon stratifié des entreprises nouvellement créées. Le modèle sous-jacent se fonde sur un modèle *probit* bivarié à équations simultanées, estimé par le maximum de vraisemblance. La première équation calcule les déterminants de l'obtention d'un soutien financier pour lancer une nouvelle entreprise, et la seconde estime la survie à cinq ans, selon que l'entreprise a obtenu une aide ou non (parmi d'autres facteurs). Deux restrictions d'exclusion sont utilisées dans l'équation de sélection : le fait que l'entreprise a été créée au deuxième trimestre (ce qui implique que l'aide a été demandée au premier trimestre, quand davantage de financements sont disponibles par rapport à la fin de l'année), et la tension sur le marché de l'emploi (c'est-à-dire la proportion d'offres d'emploi rapportée au nombre de chômeurs). Les caractéristiques individuelles observables de l'entrepreneur et de l'entreprise créée sont contrôlées. Les deux restrictions d'exclusion sont significatives, même si le niveau de significativité du facteur « date de création de l'entreprise » ne joue qu'à hauteur de 10 %. Lorsque les deux probits sont estimés séparément, l'Accre a un effet positif sur la probabilité de survie à 5 ans. Lorsqu'elles sont estimées conjointement, l'effet de l'Accre disparaît. L'endogénéité ne provient pas de déterminants non observables communs pour les deux variables endogènes, car la corrélation entre les termes d'erreur de la distribution normale bivariable n'est pas significative.

La conclusion est que l'Accre n'est pas efficace pour créer des entreprises dont la durée de vie est d'au moins 5 ans. Le résultat est robuste à différentes définitions des moyens financiers, différentes durées de survie et différents types de bénéficiaires (inactifs, chômeurs depuis moins d'un an et chômeurs depuis plus d'un an). Le seul effet significatif de l'Accre concerne l'année 1998, quand toutes les catégories de chômeurs sont incluses dans l'échantillon. Il est suggéré que ce résultat exceptionnel peut être dû au plus petit nombre de bénéficiaires de l'Accre en 1998, en raison de critères de financement moins favorables cette année-là. Les autres cohortes ont bénéficié d'un soutien plus généreux. Cela explique peut-être la différence de résultat pour 1998, même si l'effet n'est pas

significatif pour chaque sous-catégorie de bénéficiaires, y compris cette année-là.

Analyse des résultats

Même si les résultats de Dortet-Bernadet et de Sicsic contredisent ceux d'évaluations précédentes du dispositif de crédit impôt recherche en France, ils ne sont pas complètement surprenants.

Tout d'abord, ces estimations considèrent à la fois la marge intensive (augmentation de l'intensité de la R&D pour les entreprises effectuant de la R&D) et la marge extensive (augmentation du nombre d'entreprises effectuant de la R&D). Concernant la marge extensive, on constate que la plupart des entreprises ont décidé de faire de la R&D précisément l'année de la réforme de 2008 du CIR (Bozio et al., 2014). Pour mettre en place une activité de R&D, les entreprises doivent engager des coûts irrécupérables, outre les coûts fixes et variables de la R&D proprement dite. Arqué-Castells & Mohnen (2015) estiment que ces coûts irrécupérables atteignent jusqu'à 1 % du chiffre d'affaires total, et qu'ils sont plus élevés pour les petites entreprises que pour les grandes. L'échantillon est ici principalement composé de très petites entreprises, non prises en compte dans les études précédentes fondées sur la base de données R&D. Les bénéfices des petites entreprises ne sont pas toujours assez élevés pour leur permettre de bénéficier du CIR. Or c'est seulement depuis 2010 que les PME peuvent prétendre à des remboursements immédiats pour les crédits inutilisés (Commission européenne, 2014, Fiches d'information par pays). Enfin, il ne faut pas oublier les coûts de mise en conformité liés aux demandes de CIR. La réforme de 2008 a facilité les demandes de CIR ; pourtant, pour les très petites entreprises, de tels coûts, estimés en moyenne à 7 % au Canada et aux Pays-Bas, peuvent être deux, voire trois fois, plus importants pour les très petites entreprises. Cela explique que, pour s'engager dans la R&D et faire une première demande de CIR, ce qui a été plus fréquent l'année de la réforme, il faut supporter des coûts additionnels qui réduisent d'autant le montant de l'enveloppe disponible pour embaucher sur des postes de R&D. La réduction du financement privé des emplois de R&D a atteint son maximum en 2008. Il aurait été intéressant de montrer la différence d'effet dans le soutien à la R&D sur les marges intensives et extensives.

Ensuite, l'effet d'éviction partiel peut également être dû à l'introduction progressive de crédits d'impôts R&D basés sur le volume après 2004 en France, ainsi qu'à la substitution complète du dispositif incrémental par celui fondé sur le volume après 2008. L'« effet d'aubaine », c'est-à-dire l'investissement en R&D qui aurait de toute façon été réalisé, est un phénomène typique des programmes fondés sur le volume. Les études ont montré que la R&D persistait (Peters, 2009, Arqué-Castells & Mohnen, 2015). Une fois entrées dans le circuit de la R&D, les entreprises tendent à y rester. Dans ce cas, une part importante du soutien financier à la R&D aurait lieu même sans aides publiques. Les entreprises poursuivraient de toute façon leurs dépenses en R&D. Dans les dispositifs incriminaux de crédit d'impôt recherche, seules les augmentations d'investissement en R&D sont éligibles au crédit d'impôt recherche, et seule une partie de l'augmentation est financée par le programme d'aide. Il n'est donc pas étonnant que Mairesse-Mulkay (2004) et Duguet (2012) aient identifié une forte additionnalité pour la période précédant 2003 pour la France, quand le pays ne disposait que d'un CIR incrémental, tandis que Mulkay & Mairesse (2013) évoquent un effet multiplicateur de 0,7 avec le régime du CIR basé sur les niveaux, après 2008. Dans la plupart des autres études, l'effet multiplicateur est inférieur à 1 quand dominant les crédits d'impôt fondés sur le volume (voir Ientile & Mairesse, 2009 ; Caiumi, 2011 ; Commission Européenne, 2014). L'effet d'aubaine est particulièrement important pour les grandes entreprises, mais le phénomène est observable même pour les petites entreprises, comme celles de l'échantillon étudié.

Troisièmement, comme l'admettent les auteurs, on peut constater un effet d'éviction à court terme, en raison des coûts irrécupérables et des coûts d'ajustement engagés par les entreprises. Cependant, à long terme, il peut y avoir additionnalité. En raison du coefficient autorégressif élevé dans l'équation de travail estimée, l'élasticité à long terme de la main-d'œuvre de R&D par rapport au salaire relatif de la main-d'œuvre hautement qualifiée pourrait être très élevée. Cette inversion pourrait aussi expliquer la non significativité de la baisse de la R&D financée par le secteur privé en 2010, contrairement aux effets négatifs obtenus en 2010 pour les suppléments d'aide par rapport à 2003 (voir le tableau 4 de leur article).

Dortet-Bernadet et Sicsic associent les approches de *matching*, de différences de différences et

de modélisation structurelle, mais à chaque étape, ils ne contrôlent que des caractéristiques observables. Il est possible que des caractéristiques non observables amènent les entreprises à demander le CIR, et que ces mêmes caractéristiques impactent la demande d'emplois de R&D. La crise financière d'après 2008 est l'une des variables qui a pu affecter à la fois les demandes d'aide publique et le niveau de dépense en R&D. Dans son article, Dominique Redor prend en compte la présence de ces caractéristiques non observables à travers la corrélation des termes d'erreur des équations de choix et de survie. Dans son cas, les corrélations ne sont significatives pour aucune des quatre cohortes d'entreprises, ce qui indique que la survie est conditionnellement indépendante au fait d'avoir obtenu l'Accre. En raison de l'évaluation positive du dispositif de l'Accre qui ressort de l'étude de Duhautois, Désiage et Redor (2015) pour l'année 1998, il serait intéressant de refaire l'analyse d'appariement et de score de propension sur les quatre cohortes, sans aucun présupposé quant à la forme fonctionnelle de la spécification ou de la distribution des termes d'erreur. Il est vrai qu'on ne pourrait tenir aucun compte de la présence de caractéristiques non observables, mais celles-ci ne semblent pas être déterminantes.

Comparaison internationale

Quasiment tous les pays ont mis en place des politiques de soutien à la R&D (voir le rapport de l'OCDE, 2017 pour un inventaire récent de ces mesures). Même si les pays mettent en place des incitations fiscales et des subventions de manière très hétérogène, quelques points communs émergent. La grande majorité des États ont trouvé des moyens de permettre aux entreprises d'utiliser leurs crédits d'impôts même quand elles ne sont pas assujetties à l'impôt. La plupart des pays proposent des crédits d'impôts supérieurs aux petites entreprises. De plus en plus de gouvernements mettent en place des crédits d'impôts recherche proportionnels au volume des dépenses, car ils sont plus faciles à gérer, n'encouragent pas des dépenses de R&D en dents de scie afin d'obtenir le crédit d'impôt maximum, et ils fournissent un soutien continu même si les investissements en R&D ne s'accroissent pas. En principe, les incitations fiscales à la R&D sont neutres même si, dans la plupart des cas, les États proposent des aides supplémentaires pour la R&D en collaboration avec les universités. Certains pays comme l'Allemagne, la Finlande et le Luxembourg n'ont pas de crédit d'impôt fondé sur les dépenses, leur

préférant les subventions. Les programmes de subvention à la R&D sont beaucoup plus diversifiés et peuvent, dans une certaine mesure, être orientés vers des projets aux retombées sociales plus larges.

Les études empiriques sur l'efficacité des incitations fiscales à la R&D (Ientile & Mairesse, 2009, Commission Européenne, 2014) indiquent toutes que le crédit d'impôt recherche est efficace, c'est-à-dire qu'il génère un surcroît de R&D. Cependant, les incitations fiscales à la R&D proportionnelles au volume des dépenses sont assez inefficaces en termes de coûts-bénéfices. La perte liée à l'effet d'aubaine peut être importante : partant du principe que les entreprises ne réduiront pas leurs dépenses en R&D à cause des coûts irrécupérables, Lokshin et Mohnen (2012) ont évalué la R&D supplémentaire générée par un euro de crédit d'impôt recherche à 0,42 euro aux Pays-Bas. Pour la France, Mulkay et Mairesse (2013) ont montré qu'à long terme, il existait un effet multiplicateur de budget de 0,72. Dans leur tour d'horizon de la recherche, Zuñiga-Vicente et al. (2014) ont trouvé diverses indications sur les effets amplificateurs et les effets d'éviction du soutien direct à la R&D, bien que les dernières études sembleraient pencher davantage vers un effet amplificateur.

La subvention des start-ups créées par les chômeurs est une politique qui a également été mise en place dans beaucoup d'autres pays. Comme Caliendo (2016) le montre dans son analyse des études sur le sujet, ces politiques sont généralement efficaces en termes de création d'emploi, mais elles n'ont pas tant de succès pour créer des entreprises durables.

Recommandations de politiques

Les mécanismes directs et indirects d'aide à la R&D ne génèrent pas toujours immédiatement de l'emploi supplémentaire en R&D ; d'une part, les entreprises doivent faire face à d'autres coûts à court terme et, d'autre part, elles auraient de toute façon investi dans la recherche. Un certain niveau d'effet d'aubaine est inévitable, sauf si les subventions sont limitées aux efforts de R&D additionnels. Les incitations fiscales liées à l'accroissement des dépenses de R&D sont cependant coûteuses à gérer, pour les entreprises comme pour l'État, et leur capacité à générer un volume important de R&D supplémentaire est limitée. La question est de savoir si ces inefficacités sont compensées par les externalités générées par la R&D supplémentaire. Il

existe aussi des manières de limiter l'inefficacité. Les entreprises investissant beaucoup dans la R&D ont besoin de moins d'aides, car elles ont d'autres moyens d'obtenir des fonds pour financer leurs projets : bénéfices non distribués, meilleur accès au financement externe et aux marchés de capital-risque. Un autre moyen consisterait à mettre en place un éventail de dispositifs, en accordant des crédits d'impôt facilement accessibles aux petites entreprises et aux start-ups, et des subventions et des prêts directs aux grands projets et éventuellement aux projets collaboratifs impliquant grands et petits acteurs ou entreprises et universités, lorsqu'une analyse coûts-bénéfices robuste met en avant des avantages à long terme pour la société. Cela s'appuie sur l'idée que les petites entreprises sont les plus affectées par le problème d'asymétrie de l'information et par le manque de capitaux financiers, tandis que les grandes entreprises sont plus susceptibles de générer des effets d'entraînement dans la R&D (Bloom et al., 2013). Pour terminer, ces politiques de soutien financier peuvent être renforcées par les commandes publiques, par la protection des droits de propriété intellectuelle, la création d'un marché de capital-risque et l'acceptation des risques et des échecs.

L'autre aspect à garder à l'esprit est qu'une politique peut avoir plusieurs effets et qu'elle

doit être évaluée selon plusieurs perspectives. La plus grande générosité de la politique du crédit d'impôt recherche mise en place après 2008 visait aussi à conserver des activités de R&D en France, plutôt que de voir des établissements partir à l'étranger. Ainsi, même si le dispositif n'a pas été très efficace pour stimuler la R&D privée, il l'a peut-être été pour retenir la R&D en France. De la même manière, la subvention de start-ups créées par les chômeurs n'a peut-être pas créé beaucoup d'entreprises durables, mais elle a pu donner à ses bénéficiaires l'occasion d'acquérir de l'expérience, afin d'être en meilleure position pour trouver un nouveau travail ou créer une nouvelle société. Si l'objectif est de créer de nouvelles entreprises susceptibles de survivre à long terme, il serait plus judicieux d'octroyer les subventions après une sélection plus rigoureuse, puis de les accompagner par de la formation et du mentorat. Cependant, les subventions pour les chômeurs visent aussi à lutter contre le chômage, sans doute encore davantage qu'à créer de nouvelles entreprises. Plutôt que d'étudier la survie à cinq ans des sociétés nouvellement créées, un autre indicateur de performance pertinent pourrait être la situation professionnelle des chômeurs cinq ans après qu'ils ont reçu l'Accre, ou le nombre de nouveaux emplois issus de la création d'entreprises. □

BIBLIOGRAPHIE

- Arqué-Castells, P. et 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. et Van Reenen, J. (2013).** Identifying technology spillovers and market rivalry. *Econometrica*, 81(4), p. 1347–1393.
- Bozio, A., Irac, D. et 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), p. 405–435.
- Duhautois, R., Désiège, L., et 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, p. 11–42.
- Commission Européenne (2014).** A study on R&D tax incentives: Final report. DG TAXUD Taxation Paper 52.
- Ientile, D. et Mairesse, J. (2009).** A policy to boost the R&D: Does the tax credit work? article de la Banque européenne d'investissement 14(1).
- Köhler, C., Laredo, P. et 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. et Mulkay, B. (2004).** Une évaluation du crédit d'impôt recherche en France, 1980-1997. *Revue d'économie politique*, 114 (6), p. 747–778.
- Mulkay, B. et Mairesse, J. (2013).** The R&D tax credit in France: assessment and ex-ante evaluation of the 2008 reform. *Oxford Economics Papers*, 65(3), p. 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, p. 226–342.
- Zuñiga-Vicente, J. A., Alonso-Borrego, C., Forcadell, F. J. et Galán, J. I. (2014).** Assessing the effect of public subsidies on firm R&D investment: A survey. *Journal of Economic Surveys*, 28, p. 36–67.

Turnover élevé du personnel soignant dans les Ehpad privés en France : impact de l'environnement local et du salaire

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 * et Mélina Ramos-Gorand **

Résumé – Le turnover élevé du personnel soignant travaillant en établissement d'hébergement pour personnes âgées dépendantes (Ehpad) en France a des conséquences négatives à la fois en termes de coûts et de qualité de prise en charge des résidents. Nous étudions les causes de ce turnover à partir de l'estimation d'un modèle *probit* sur deux échantillons de personnel travaillant en contrat à durée indéterminée dans des Ehpad privés, l'un de 5 478 infirmiers et l'autre de 13 444 aides-soignants. La probabilité de départ des infirmiers et des aides-soignants est significativement influencée par des facteurs liés à l'environnement local de leur lieu de domicile, calculés à une échelle géographique fine, tels que la proximité d'un hôpital, la concurrence entre établissements pour personnes âgées, la pénurie de personnel soignant et l'attractivité du secteur libéral pour les infirmiers. Le niveau de salaire, corrigé de l'endogénéité, a un effet positif sur la fidélisation des aides-soignants travaillant en Ehpad, mais ne semble pas avoir d'effet sur les infirmiers.

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.

Codes JEL : C25, I11, J63

Mots clés : turnover, Ehpad, qualité, infirmiers, aides-soignants

Keywords: staff turnover, nursing homes for dependent elderly people, care quality

Rappel :

Les jugements et opinions exprimés par les auteurs n'engagent qu'eux mêmes, et non les institutions auxquelles ils appartiennent, ni a fortiori l'Insee.

* Académie de Caen (cecile.martin@ac-caen.fr).

** Lab'Urba, Université Paris-Est Créteil et Caisse nationale d'assurance vieillesse (Cnav), Paris (melina.ramos-gorand@cnav.fr).

Remerciements : les auteurs remercient la Direction de la recherche, des études, de l'évaluation et des statistiques (Drees) et le Centre Maurice Halbwachs pour l'accès aux données utilisées dans cette étude. Elles remercient également les deux rapporteurs anonymes pour l'ensemble de leurs suggestions et remarques.

Compte tenu de la médicalisation croissante des établissements d'hébergement pour personnes âgées dépendantes (Ehpad), le recours au personnel soignant dans ces établissements devrait s'intensifier ces prochaines années. Les postes d'infirmiers et d'aides-soignants représentaient déjà 4 équivalents temps plein (ETP) sur 10 dans les Ehpad en 2007 et en 2011¹. Or le secteur souffre d'un déficit d'attractivité et de fidélisation de ces professions, qui touche plus fortement les établissements privés et a des effets délétères sur la qualité de la prise en charge. Le taux de départ moyen des infirmiers est en effet de 61 % et celui des aides-soignants s'élève à 68 % dans les Ehpad privés en 2008². Ces départs sont plus importants que dans le secteur hospitalier. Les taux de *turnover*³ des infirmiers ne sont ainsi que de 22 % dans les cliniques privées associatives et de 30 % dans les cliniques privées à but lucratif en 2011 (Loquet & Nagou, 2014). Les causes des départs prématurés des infirmiers et des aides-soignants travaillant en Ehpad peuvent être diverses. Ceux-ci peuvent préférer aller travailler dans un Ehpad concurrent ou situé dans une autre zone géographique plus attractive. Ils peuvent également changer de mode d'exercice et choisir de travailler à l'hôpital, dans un service de soins à domicile, ou d'exercer en libéral. Barlet et Cavillon (2011) montrent ainsi que les infirmiers travaillant en établissement pour personnes âgées sont les plus nombreux à changer de mode d'exercice (6 % en moyenne chaque année contre 3 % pour l'ensemble des infirmiers entre 2004 et 2008).

Ce *turnover* important du personnel soignant peut avoir pour effet de dégrader la qualité de la prise en charge en Ehpad. Des départs massifs d'infirmiers ou d'aides-soignants peuvent en effet impliquer un fonctionnement temporaire de l'Ehpad en sous-effectif, alors que le taux d'encadrement est positivement corrélé à la qualité (Martin, 2014). Un fort *turnover* peut également provoquer des interruptions dans la continuité des soins (Cohen-Mansfield, 1997), engendrer des accidents iatrogènes dus à des erreurs de prescriptions et dégrader l'état de santé des résidents (Lerner et al., 2014 ; Antwi & Bowblis, 2016). Il empêche enfin les résidents de nouer des relations de confiance avec leurs soignants (Wiener et al., 2009). Comme l'indique une ancienne professionnelle du secteur, « le turnover important du secteur entraîne des conséquences : les nouveaux personnels ne connaissent pas les résidents, leurs pathologies, leurs besoins d'aide, leurs

habitudes, n'ont pas les formations nécessaires » (Nénin & Lapart, 2011, p. 51). Outre les conséquences néfastes en termes de qualité qu'un fort *turnover* peut engendrer, il peut également être générateur de surcoûts pour l'établissement en raison du besoin de recrutement de personnel remplaçant, parfois intérimaire. Il peut également être à l'origine d'une réduction de la productivité, compte tenu du temps de formation nécessaire du nouveau personnel (Brannon et al., 2002).

Nous analysons dans cette étude les déterminants des départs du personnel soignant dans les Ehpad privés en France. En contrôlant pour les caractéristiques de l'emploi, de l'établissement et du personnel employé, nous étudions en particulier les effets du salaire et de l'environnement sous divers angles – pénurie de personnel soignant dans la zone géographique, présence d'un hôpital, concurrence entre établissements accueillant des personnes âgées, attractivité du secteur libéral – sur la probabilité de départ des infirmiers et des aides-soignants. Notre analyse est réalisée à partir de deux échantillons de personnel en contrat à durée indéterminée (CDI) dans des Ehpad privés en 2008 : un échantillon de 5 478 infirmiers et un échantillon de 13 444 aides-soignants. Cette étude présente une triple originalité. Premièrement, il n'existe pas d'autre étude économétrique qui examine les causes du *turnover* du personnel travaillant en Ehpad en France. Deuxièmement, nous prenons en compte le caractère endogène du salaire, ce qui a rarement été fait dans la littérature étrangère consacrée à ce sujet. Troisièmement, nous intégrons de nombreuses variables environnementales permettant d'analyser les motifs de départs du personnel. Nous utilisons pour cela des données géographiques fines, construites à partir des lieux de domicile de chaque salarié et calculées à l'échelle du bassin de vie⁴.

Nous présentons dans la section suivante les imperfections des études étrangères portant

1. Source : <http://www.data.drees.sante.gouv.fr> (30 mars 2017).
2. Le taux de départ des infirmiers (ou des aides-soignants) correspond au ratio du nombre d'infirmiers (respectivement d'aides-soignants) ayant quitté l'établissement dans l'année sur le nombre de postes infirmiers (resp. d'aides-soignants) dans l'établissement en fin d'année. Les calculs ont été effectués à partir de deux échantillons de 1 393 Ehpad (taux de départ des infirmiers) et 1 392 Ehpad (taux de départ des aides-soignants). Source : DADS 2008 (Insee) et enquête Ehpa 2007 (Drees), calculs des auteurs.

3. Les taux de turnover correspondent dans cette étude à la moyenne des taux d'entrée et des taux de départ. Ils peuvent donc être différents des taux de départ si le personnel sorti de l'établissement n'est pas remplacé dans l'année.

4. Voir infra.

sur l'analyse des causes du *turnover* du personnel soignant. Nous détaillerons ensuite les variables et les données utilisées, ainsi que notre démarche. Nous analyserons enfin les résultats avant d'apporter des éléments de conclusion.

L'analyse des départs du personnel soignant dans la littérature étrangère

Plusieurs études réalisées à partir de données américaines ont eu pour objectif d'identifier les causes du *turnover* du personnel dans les établissements accueillant des personnes âgées. La plupart d'entre elles présentent néanmoins deux limites.

Tout d'abord, beaucoup de ces études sont réalisées à l'échelle de l'établissement et les variables liées à l'environnement sont ainsi souvent mal intégrées. Castle et Engberg (2006), par exemple, ont cherché à mettre en évidence, par une régression binomiale négative, un nombre plus faible de départs du personnel soignant dans les maisons de retraite situées en milieu rural. Les infirmiers et les aides-soignants bénéficieraient en effet d'opportunités d'embauches moins importantes sur ces territoires. Ils observent néanmoins l'effet inverse, ce qui est probablement dû au fait qu'ils aient pris en compte la localisation de la maison de retraite et non celle de la commune ou de la ville de résidence des employés. Or, il est possible que les employés travaillant dans ces maisons de retraite rurales vivent en ville, ce qui leur impose de longs trajets domicile-travail qui sont sources d'insatisfaction. Brannon et al. (2002) ont également étudié, par un modèle logistique multinomial, les facteurs d'un faible taux de *turnover* des aides-soignants travaillant en maison de retraite (inférieur à 6,6 % en 6 mois) et ceux d'un taux élevé (supérieur à 64 %). Ils ont intégré quelques variables liées à l'environnement de marché, telle qu'une variable de concentration des établissements et le taux de chômage local. Ces variables ayant néanmoins été construites à partir de l'implantation de l'établissement et non du lieu de résidence des employés, aucun effet significatif n'a été observé.

Ensuite, le rôle du salaire est peu ou mal étudié. Celui-ci n'est pas toujours ajouté aux variables explicatives du *turnover*, à l'instar des travaux de Brannon et al. (2002) et de Castle et Engberg (2006). Or, cette omission biaise l'effet des variables environnementales si le salaire s'ajuste aux difficultés locales.

À partir d'une enquête qualitative réalisée auprès de 345 aides-soignants dans 18 maisons de retraite américaines, Dill et al. (2013) ont montré que lorsque les employés se sentent récompensés financièrement, cela impacte positivement leur intention de rester dans l'établissement. L'impact chiffré de ce salaire sur le *turnover* n'est en revanche pas aisé à identifier. Quelques auteurs ont intégré cette variable de salaire, comme Temple et al. (2009) qui ont étudié les facteurs liés à un faible et à un fort taux de *turnover* des aides-soignants (caractérisés par le premier et le dernier quartile de leur échantillon) par une régression logistique multinomiale. Wiener et al. (2009) ont également étudié l'effet du salaire et de l'environnement sur l'ancienneté de chaque aide-soignant travaillant en maison de retraite par une régression en moindres carrés ordinaires. Ces auteurs ont ainsi montré qu'une hausse du salaire diminuait la probabilité d'avoir un fort taux de *turnover* et augmentait l'ancienneté des aides-soignants dans les établissements. Ils n'ont néanmoins pas tenu compte de l'éventuelle endogénéité du salaire. Des facteurs non observés, tels que l'expérience ou le caractère plus ou moins revendicatif du personnel soignant, peuvent en effet être corrélés à la fois au salaire et au départ du personnel. En outre, un individu qui change régulièrement d'emploi peut avoir un salaire plus faible puisqu'il ne bénéficie pas de prime d'ancienneté.

Nous testons, dans cette étude, si les relations empiriquement observées dans ces travaux s'appliquent au cas français, en corrigeant ces deux écueils. L'endogénéité du salaire est corrigée par une instrumentation, à l'instar de Baughman et Smith (2012) qui ont analysé, par un modèle de durée, les déterminants de la probabilité que des aides-soignants aux États-Unis quittent l'organisation dans laquelle ils travaillent ; ils ont observé un effet négatif du salaire instrumenté sur la probabilité de départ, alors que celui-ci n'était pas significatif lorsqu'il n'était pas corrigé de l'endogénéité. Peu de variables liées à l'environnement sont néanmoins prises en compte dans leur étude, et celles qui le sont (taux de chômage, salaires moyens dans d'autres professions) sont calculées à l'échelle de l'État. Contrairement à ces auteurs, nous intégrons plusieurs variables environnementales construites avec une échelle géographique fine. Nous utilisons des données individuelles et construisons domicile par résidence ces variables d'environnement à partir du lieu de domicile des salariés et non du lieu d'implantation de l'établissement.

Variables et données mobilisées

Nous étudions la probabilité pour un infirmier et un aide-soignant de quitter leur établissement variable « *Départ* ». À l'instar de plusieurs modèles théoriques (Cohen-Mansfield, 1997), nous supposons que les départs du personnel soignant sont liés à deux types de facteurs : ceux susceptibles de générer une insatisfaction au travail, tels que les caractéristiques individuelles du salarié, les caractéristiques de son emploi et de son établissement, et ceux influençant la prise de décision de quitter l'établissement comme le marché local du travail. Les variables utilisées pour étudier empiriquement ces déterminants sont décrites ci-dessous et sont présentées dans le tableau 1.

Caractéristiques individuelles des infirmiers et des aides-soignants

Barlet et Cavillon (2011) ont observé une relation entre l'âge des infirmiers et leur probabilité de changer de mode d'exercice : cette probabilité semble plus forte avant 35 ans et plus faible après 45 ans. Des départs à la retraite, parfois anticipés, peuvent néanmoins avoir lieu après 50 ans. Pour tenir compte de ces motifs de départs, nous intégrons une variable d'âge (*Âge*) déclinée en plusieurs catégories : âge inférieur à 35 ans, compris entre 35 et 45 ans, entre 45 et 50 ans, entre 50 et 55 ans et supérieur à 55 ans. Nous ajoutons également une variable de genre (*Homme*). À l'instar des travaux de Wiener et al. (2009), nous attendons un effet positif de cette variable sur la probabilité de départ. Enfin, plus le trajet entre la commune de travail de l'employé et celle de son domicile est long (*Distance*), plus son insatisfaction au travail peut être grande. Puisque l'effet de cette variable n'est sans doute pas linéaire, nous la catégorisons en quatre modalités : distance inférieure à 5 km, comprise entre 5 et 10 km, entre 10 et 20 km et supérieure à 20 km. Nous ne disposons pas de données sur la situation maritale des employés ou leur nombre d'enfants. À partir de données recueillies auprès de 1 589 employés travaillant en maison de retraite aux États-Unis, Zhang et al. (2014) n'observent néanmoins pas de lien entre ces variables et l'intention du personnel de quitter leur établissement.

Caractéristiques de l'emploi et de l'établissement

Diverses sources d'insatisfaction sont liées aux conditions d'emploi et peuvent, de ce fait,

constituer des leviers d'action pour les directeurs d'Ehpad qui souhaiteraient réduire les taux de rotation de leur personnel (Anderson et al., 1997). Selon la théorie du salaire d'efficacité (Stiglitz, 1974 ; Salop, 1979), un salaire élevé peut être efficace dans la mesure où il incite le personnel à rester dans l'établissement et permet ainsi de réduire le *turnover*. Dans son rapport annuel en 2012, le groupe Le Noble Age indique ainsi avoir « *mis en place [...] une gestion favorable des rémunérations* », « *afin de limiter le risque de sous-effectif et d'augmentation du taux de rotation du personnel* » (p. 41). Nous intégrons donc le logarithme du salaire net annuel de l'employé (*Log_Salaire*). La possibilité de travailler la nuit (*Nuit*) dans l'établissement peut également avoir un effet sur l'insatisfaction, qui peut être positif lorsqu'il s'agit d'une obligation non souhaitée par le salarié, ou négatif s'il s'agit d'une possibilité offerte au salarié. La qualité de la prise en charge a probablement aussi un impact sur l'insatisfaction au travail et donc la probabilité de départ des salariés (Irvine & Evans, 1995). Nous choisissons d'inclure comme variable de qualité un ratio lié au taux d'encadrement en personnel de l'Ehpad dans lequel travaille l'employé. Cette variable est un bon *proxy* de la qualité puisqu'elle a un effet significatif sur le bien-être et l'état de santé des résidents (Spilsbury et al., 2011). En outre, un faible encadrement impacte les conditions de travail puisqu'il induit nécessairement une charge de travail plus lourde pour les employés de l'établissement. De manière à prendre en compte la diversité des besoins d'encadrement selon les Ehpad, nous calculons un encadrement théorique optimal (N^*) en fonction des catégories de dépendance (catégories liées aux groupes iso-ressources *GIR*) des résidents. Nous utilisons pour cela les recommandations du Plan Solidarité grand âge 2007-2012 : une personne en *GIR* 1 a ainsi besoin quotidiennement de 1 ETP, en *GIR* 2 de 0.84 ETP, en *GIR* 3 de 0.66 ETP, en *GIR* 4 de 0.42 ETP, en *GIR* 5 de 0.25 ETP et enfin en *GIR* 6 de 0.07 ETP (Ratte & Imbaud, 2011). Nous rapportons ensuite l'encadrement effectif de l'Ehpad (hors employés administratifs et services généraux) (N) à cet encadrement théorique calculé (N/N^*). Plusieurs auteurs ont montré que le taux d'encadrement pouvait avoir un impact important sur l'insatisfaction et le taux de *turnover* dans les maisons de retraite aux États-Unis (Temple et al., 2009). Certains auteurs mentionnent néanmoins le problème d'endogénéité que peut poser l'intégration de cette variable (Kash et al., 2006). Comme pour le salaire, la relation

Tableau 1
Définition des variables

Variable	Définition
Départ	Variable binaire égale à 1 si l'employé est parti au cours de l'année de l'établissement dans lequel il travaillait, 0 sinon.
Caractéristiques individuelles	
Âge	Variable catégorielle indiquant l'âge de l'employé : inférieur à 35 ans, compris entre 35 et 45 ans, entre 45 et 50 ans, entre 50 et 55 ans ou supérieur à 55 ans.
Homme	Variable binaire égale à 1 si l'employé est un homme, 0 s'il s'agit d'une femme.
Distance	Variable catégorielle indiquant si la distance parcourue par l'employé entre son domicile et l'Ehpad dans lequel il travaille est inférieure à 5 km, comprise entre 5 et 10 km, entre 10 et 20 km ou supérieure à 20 km.
Caractéristiques d'emploi et de l'établissement	
Log Salaire	Logarithme du salaire net annuel de l'employé dont l'endogénéité est corrigée par instrumentation.
Nuit	Variable binaire égale à 1 si l'employé peut travailler la nuit dans son établissement, 0 sinon.
N/N*	Ratio de l'encadrement effectif de l'Ehpad dans lequel travaille l'employé sur les besoins calculés en personnel (en fonction des degrés de dépendance des résidents accueillis).
Statut	Variable binaire indiquant si l'Ehpad dans lequel travaille l'employé est de statut privé associatif ou à but lucratif.
N_Lits	Nombre de lits installés dans l'Ehpad dans lequel travaille l'employé.
% GIR1	Proportion de résidents fortement dépendants de groupe iso-ressources (GIR) 1 dans l'Ehpad dans lequel travaille l'employé.
% GIR2	Proportion de résidents fortement dépendants de GIR 2 dans l'Ehpad dans lequel travaille l'employé.
% GIR3	Proportion de résidents moyennement dépendants de GIR 3 dans l'Ehpad dans lequel travaille l'employé.
% GIR4	Proportion de résidents moyennement dépendants de GIR 4 dans l'Ehpad dans lequel travaille l'employé.
% GIR5	Proportion de résidents faiblement dépendants de GIR 5 dans l'Ehpad dans lequel travaille l'employé.
% GIR6	Proportion de résidents non dépendants de GIR 6 dans l'Ehpad dans lequel travaille l'employé.
Directeur > 2 ans	Variable binaire égale à 1 si le directeur est présent depuis plus de deux ans dans l'Ehpad dans lequel travaille l'employé, 0 sinon.
Environnement	
Hôpital	Coefficient de Huff mesurant l'attraction exercée par les emplois hospitaliers sur l'employé.
%Inf_libéraux_bv/pop	Ratio du nombre d'infirmiers libéraux travaillant dans le bassin de vie où est domicilié l'employé rapporté à la population (en milliers d'habitants).
HFI _{Ehpa} _bv	Indice d'Herfindhal-Hirschmann mesurant la concentration des emplois en Ehpa au sein du bassin de vie où est domicilié l'employé.
%Inf_résid_bv/pop	Ratio du nombre d'infirmiers salariés résidant dans le bassin de vie où est domicilié l'employé rapporté à la population (en milliers d'habitants).
%AS_résid_bv/pop	Ratio du nombre d'aides-soignants salariés résidant dans le bassin de vie où est domicilié l'employé rapporté à la population (en milliers d'habitants).
Paris	Variable binaire égale à 1 si l'employé est domicilié à Paris, 0 sinon.
Île-de-France	Variable binaire égale à 1 si l'employé est domicilié en Ile-de-France (hors Paris), 0 sinon.
Instruments exclus	
Ehpad non régulé	Variable binaire égale à 1 si le tarif hébergement de l'établissement dans lequel travaille l'employé est fixé librement, 0 s'il est fixé administrativement.
Taux d'occupation moyen_bv	Taux d'occupation moyen des Ehpad dans le bassin de vie dans lequel travaille l'employé.
Tarif dépendance moyen_dpt	Tarif dépendance moyen pour les personnes de GIR 1 et 2 résidant en Ehpad dans le département dans lequel travaille l'employé.
Groupe	Variable catégorielle indiquant si l'établissement dans lequel travaille l'employé est indépendant ou appartient à un groupe d'Ehpad inférieur à 5 établissements, comprenant entre 5 et 20 établissements, ou plus de 20 établissements.

de causalité peut être à double sens : de nombreux départs d'employés peuvent engendrer des difficultés de remplacement et un taux d'encadrement temporairement plus faible. Le ratio d'encadrement (N/N^*) est néanmoins calculé à partir de l'ensemble du personnel au contact des résidents et non pas uniquement des infirmiers et des aides-soignants, ce qui réduit le risque d'endogénéité. Nous avons vérifié que l'estimation de notre modèle sans cette variable d'encadrement ne changeait pas les résultats obtenus (cf. complément en ligne C1).

Certains facteurs propres à la structure de l'établissement peuvent avoir des impacts sur les départs du personnel. Plusieurs études américaines ont ainsi montré que les établissements privés à but lucratif étaient confrontés à un *turnover* plus important (Banaszak-Holl & Hines, 1996 ; Anderson et al., 1997 ; Brannon et al., 2002 ; Castle & Engberg, 2006). Selon Wiener et al. (2009, p. 200), « *les établissements à but non lucratif sont plus focalisés sur leur mission à accomplir que les établissements à but lucratif et ont un taux d'équipement plus élevé ainsi que d'autres caractéristiques qui peuvent améliorer la fidélisation du personnel* ». Le personnel peut préférer travailler dans un établissement dont la finalité n'est pas lucrative mais uniquement sociétale, *i.e.* dans lequel les dirigeants ont pour unique objectif d'améliorer le bien-être des résidents. Nous intégrons par conséquent une variable binaire indiquant si l'établissement dans lequel travaille l'employé est de statut privé non lucratif ou lucratif (*Statut*). La taille de l'établissement, *i.e.* le nombre de lits installés (N_Lits), semble avoir un effet plus indéterminé sur les départs : si Castle observe en 2005 un effet positif significatif du nombre de lits sur la probabilité d'avoir un fort *turnover* des infirmiers et des aides-soignants, Wiener et al. (2009) n'obtiennent pas d'effet significatif de la taille des établissements sur l'ancienneté des aides-soignants. Nous ajoutons également les proportions de résidents de chaque catégorie GIR dans l'Ehpad dans lequel travaille l'employé afin de prendre en compte leur degré de dépendance (% *GIR1* à % *GIR6*). Une variable binaire indiquant si le directeur est présent depuis plus de deux ans dans l'établissement (*Directeur > 2 ans*) est également intégrée au modèle. À l'instar de Castle (2005), nous nous attendons à ce qu'une plus grande stabilité du personnel de direction réduise la probabilité de départ des salariés, et cela pour les trois raisons mentionnées par l'auteur : une forte rotation du personnel de direction pourrait avoir une influence déstabilisante sur l'organisation,

générer un comportement de désengagement du personnel soignant vis-à-vis de l'institution, et enfin conduire à une dégradation de la qualité qui augmenterait de ce fait l'insatisfaction du personnel en place.

Facteurs environnementaux

La présence d'opportunités d'embauches extérieures peut inciter un employé non pleinement satisfait de son emploi à le quitter. Pour des infirmiers ou des aides-soignants travaillant en Ehpad, ces opportunités peuvent être de diverses natures.

Ils peuvent tout d'abord changer de mode d'exercice en se dirigeant vers le secteur hospitalier. Les rémunérations des infirmiers et des aides-soignants travaillant à l'hôpital sont proches de ceux exerçant en Ehpad. Selon l'Institut national de la statistique et des études économiques (Insee, 2011), en 2008, les salaires nets annuels des professions intermédiaires soignantes et sociales (intégrant les infirmiers, mais aussi les sages-femmes, les éducateurs spécialisés, les techniciens médicaux et les assistantes sociales) s'élèvent en moyenne à 24 820 euros dans les hôpitaux privés lucratifs et à 25 220 euros dans les hôpitaux privés non lucratifs. Or, dans notre échantillon, les infirmiers travaillant en Ehpad privé ont en moyenne une rémunération nette annuelle de 25 205 euros (cf. tableau 2). Si le secteur hospitalier est attractif, ceci serait alors plutôt lié à la nature du travail. « *Le travail en hôpital est perçu comme plus qualifiant [...] et moins limité car il semble plus facile de changer de service au sein de l'hôpital qu'en venant d'un Ehpad* » (Josse, 2012, p. 16). L'opportunité de trouver un poste à l'hôpital près de son lieu de résidence pourrait ainsi être une incitation à la mobilité professionnelle pour le personnel soignant travaillant en Ehpad. Nous calculons des coefficients de Huff (cf. encadré 1) pour mesurer l'attractivité exercée par la présence proche d'un hôpital sur chaque infirmier et chaque aide-soignant (*Hôpital*). Nous attendons un effet positif de cette variable sur la probabilité de départ du personnel soignant. Les infirmiers peuvent aussi se diriger vers le secteur libéral. Pour tenir compte de ce motif de départ, nous intégrons la densité pour 1 000 habitants des infirmiers libéraux exerçant dans le bassin de vie du lieu de résidence de chaque infirmier (%*Inf_libéraux_bv/pop*). Un autre mode d'exercice possible est celui des services d'aide et de soins à domicile, mais nous ne disposons pas de données pour pouvoir étudier l'attractivité de ce secteur.

Les départs peuvent également être intra-sectoriels, les infirmiers et les aides-soignants pouvant choisir d'exercer dans un autre établissement d'hébergement pour personnes âgées (Ehpa) (cf. encadré 2). Les possibilités d'embauches dans un autre Ehpa du bassin de vie dans lequel réside l'employé peuvent être évaluées par l'indice de concentration d'Herfindhal-Hirschmann ($HHI_{EHPA-bv}$). Nous utilisons pour calculer cet indice les parts de marché en termes de personnel, *i.e.* les parts en besoins théoriques d'encadrement (N^*) de chaque établissement i par rapport aux besoins des autres Ehpa du bassin de vie. L'indice d'Herfindhal est alors calculé de la manière suivante : $HHI_{EHPA-bv} = \sum_{i=1}^n s_i^2$, avec $s_i^2 = N_i^* / \sum_j N_j^*$ et n le nombre d'Ehpa dans

le bassin de vie. Plus cet indice d'Herfindhal est élevé et proche de 1, plus les offres d'emploi sont concentrées dans quelques Ehpa, et plus il est alors difficile pour les employés de changer d'établissement. Nous ajoutons enfin une variable de présence du personnel soignant dans le bassin de vie où réside l'employé, caractérisée par le ratio du nombre d'infirmiers salariés (respectivement d'aides-soignants salariés) domiciliés dans le bassin de vie sur la population (en milliers d'habitants) ($\%Inf_résid_bv/pop$ et $\%AS_résid_bv/pop$). La répartition du personnel soignant sur le territoire semble en

effet inégale. Les densités d'infirmiers sont par exemple beaucoup plus fortes dans les régions du Sud que dans le Nord de la France (Barlet & Cavillon, 2011). Nous nous attendons à un effet négatif de cette variable sur la probabilité de départ, ou autrement dit à un effet positif de la pénurie de personnel soignant. Cette pénurie localisée de personnel peut en effet expliquer certaines difficultés de fidélisation du personnel travaillant en Ehpad à double titre. D'une part, une pénurie globale de personnel soignant peut révéler des caractéristiques non attrayantes du bassin de vie (caractère rural, coût de la vie élevé, etc.) qui peuvent inciter le personnel à aller résider et travailler dans un autre bassin de vie. D'autre part, dans un marché fortement contraint par l'offre de travail, les opportunités d'embauches immédiates d'un infirmier ou d'un aide-soignant sont élevées et peuvent l'inciter à quitter son poste s'il n'en est pas satisfait.

Deux variables binaires indiquant si l'employé est domicilié ou non à Paris (*Paris*) ou en Île-de-France hors Paris (*Île-de-France*) sont également intégrées afin de tenir compte des spécificités de ces territoires.

Nous avons choisi de construire plusieurs des variables environnementales à l'échelle du bassin de vie, qui est défini par l'Insee (2003) comme « le plus petit territoire sur lequel les

Encadré 1

LE COEFFICIENT DE HUFF

Le modèle de Huff est un modèle gravitaire couramment utilisé en géographie (Pumain & Saint-Julien, 2010). Nous l'employons dans cette étude pour mesurer l'attraction exercée par les hôpitaux (hôpitaux publics et cliniques privées) sur les infirmiers et les aides-soignants. Nous considérons que plus la distance entre le lieu de domicile de ces professionnels et l'hôpital augmente, moins l'hôpital est attractif. Par ailleurs, plus le nombre de postes d'infirmiers ou d'aides-soignants dans l'hôpital est élevé, plus l'hôpital est attractif. Formellement, le professionnel résidant dans une commune i sera attiré par les hôpitaux d'une autre commune j proportionnellement au nombre de postes hospitaliers lui correspondant dans la commune j , mais en proportion inverse de la distance au carré séparant i et j :

$$A_{ij} = \frac{\text{Postes}_j}{D_{ij}^2}$$

avec A_{ij} l'attraction des postes hospitaliers de la commune j sur les professionnels résidant dans

la commune i ; Postes_j le nombre de postes hospitaliers dans la commune j ; D_{ij} la distance séparant les communes i et j .

Par convention, lorsqu'une commune dispose d'un hôpital, la distance entre les professionnels habitant cette commune et l'hôpital est de 1 km. Quand la distance séparant les communes i et j est supérieure à 250 km, l'attraction est considérée comme nulle.

Pour chaque professionnel de la commune i , nous pouvons ainsi calculer le potentiel de relation (PR_i) défini par Huff (1964), en effectuant la somme de tous les indicateurs d'attraction A_{ij} que nous divisons par 1 000 pour en réduire l'ampleur. Nous obtenons ainsi un indicateur synthétique de l'attraction exercée par les hôpitaux des communes j environnantes sur les professionnels résidant dans la commune i :

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

habitants ont accès à la fois aux équipements et à l'emploi ». On dénombre, à partir de 1999 et jusqu'en 2012, 1 916 bassins de vie en France. Bien que l'échelle de la zone d'emploi soit habituellement utilisée pour ce type d'étude, un zonage plus fin par bassin de vie nous semble ici plus approprié. 67 % des infirmiers et 70 % des aides-soignants de notre échantillon travaillent en effet dans un Ehpad situé dans le bassin de vie de leur lieu de domicile. Ce zonage est en outre construit pour mieux qualifier l'espace à dominante rurale. Or, une de nos hypothèses est que les établissements en milieu rural rencontrent davantage de difficultés à recruter et fidéliser leur personnel, en raison d'une main d'œuvre peu nombreuse à proximité. Nous testons néanmoins la robustesse des résultats en les confrontant à ceux obtenus par des régressions effectuées à partir de données environnementales construites à l'échelle de la zone d'emploi (cf. complément en ligne C2).

Bases de données utilisées

Les données sur les départs d'infirmiers et d'aides-soignants travaillant en Ehpad, ainsi que leur salaire et leur âge, proviennent des *Déclarations annuelles de données sociales* (DADS). Tous les établissements doivent fournir annuellement à la Caisse nationale d'assurance vieillesse (Cnav) des informations relatives à chacun de leurs salariés : salaire net, ETP, âge, type de contrat, départ en cours d'année, etc. Ces données sont ensuite retraitées par l'Insee. Elles permettent également d'avoir des informations précises sur le lieu de résidence de chaque employé, ce qui constitue un véritable atout pour ce type d'étude. Nous avons ainsi pu construire des variables environnementales à partir du lieu de domicile de l'employé et non uniquement à partir de son lieu de travail. Nous avons utilisé les DADS de 2008 afin de pouvoir apparier ces données avec la base *Ehpa* 2007 de la Direction de la recherche, des études, de l'évaluation et des statistiques (Drees).

L'enquête *Ehpa* est réalisée tous les quatre ans auprès de l'ensemble des établissements hébergeant des personnes âgées (Ehpad, logements-foyers, unités de soins de longue durée, etc.). 79 % des *Ehpa* ont répondu à cette enquête en 2007 (Perrin-Haynes, 2010). Celle-ci comporte plusieurs questions sur le fonctionnement de l'établissement (tarifs, nombre de places, statut, etc.), sur les employés, les résidents et les locaux au 31 décembre 2007. Elle nous a ainsi fourni des renseignements sur la plupart des

variables liées aux caractéristiques de l'emploi et de l'établissement.

L'identifiant de l'Ehpad utilisé dans les données DADS est le numéro Siret, alors qu'il s'agit du numéro Finess dans l'enquête *EHPA*. Une extraction de la base Finess en 2007 nous a permis de disposer des correspondances entre ces identifiants et d'apparier les deux bases de données. Plusieurs de ces correspondances sont néanmoins manquantes, ce qui nous a conduit à exclure des établissements. L'extraction Finess 2007 a également permis d'obtenir les codes des communes des établissements nécessaires pour mesurer la distance entre le lieu de domicile de chaque salarié et son lieu de travail.

Les données sur les infirmiers libéraux travaillant dans le bassin de vie proviennent quant à elles de la base permanente des équipements (BPE) 2007 de l'Insee.

Modèle estimé et échantillon d'étude

Modèle estimé

Nous estimons la probabilité pour un infirmier ou un aide-soignant de quitter son établissement par un modèle *probit*, en corrigeant l'endogénéité de la variable de salaire. Le modèle estimé est le suivant :

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

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

avec $i \in [1 ; N]$, N correspondant au nombre d'infirmiers ou d'aides-soignants étudiés, ω_i le logarithme du salaire de l'employé i considéré comme endogène, z_i un vecteur de 1^*k_z variables instrumentales, x_i un vecteur de 1^*k_x variables exogènes du modèle liées aux caractéristiques individuelles de l'employé i , aux facteurs environnementaux définis à partir de son lieu de résidence, à ses conditions d'emploi et aux caractéristiques de l'établissement dans lequel il travaille. Les termes d'erreurs (μ_i, v_i) suivent une distribution normale multivariée d'espérance nulle.

Nous n'observons pas la variable latente y_i^* , mais nous observons une variable dépendante dichotomique :

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

Le modèle est estimé par maximisation d'une fonction de vraisemblance. η , α et β correspondent au terme constant et aux vecteurs des paramètres du modèle. δ , Π_x et Π_z caractérisent le terme constant et les vecteurs des paramètres de l'équation d'instrumentation du salaire. Ces derniers sont estimés conjointement avec η , α et β ; si les instruments sont correctement spécifiés, cette méthode de correction de l'endogénéité permet d'obtenir des estimateurs non biaisés (Cameron & Trivedi, 2009).

Plusieurs instruments exclus sont intégrés à l'estimation de première étape (cf. tableau 1). Ils sont supposés avoir un effet sur la variable de salaire mais n'ont pas d'effet direct sur la probabilité de départ du personnel soignant.

Nous utilisons tout d'abord une variable binaire indiquant si le tarif hébergement de l'établissement dans lequel travaille l'employé est fixé librement, *i.e.* n'est pas régulé (*Ehpad non régulé*). La liberté de fixation des prix dans les Ehpad privés, et par répercussion des salaires, est variable selon le mode de régulation de l'établissement⁵ (cf. encadré 2). Les établissements dont les tarifs hébergement ne sont pas régulés devraient pouvoir ajuster les salaires de

5. Puisque la régulation des tarifs hébergement concerne davantage les Ehpad non lucratifs (cf. encadré 2), il est probable que la variable Ehpad non régulé ait également un effet direct sur la probabilité de départ du personnel par le biais du statut. Cet effet est néanmoins contrôlé puisqu'une variable de statut est intégrée dans la régression principale. Cela permet de garantir la condition d'exclusion de la variable Ehpad non régulé.

Encadré 2

AJUSTEMENT DES RÉMUNÉRATIONS ET RÉGULATION TARIFAIRE DES EHPAD

Les établissements d'hébergement pour personnes âgées (EHPA) regroupent essentiellement des Ehpad (établissements d'hébergement pour personnes âgées dépendantes), mais également des unités de soins de longue durée (USLD) très médicalisées ou, à l'inverse, des établissements non médicalisés tels que des logements-foyers accueillant des personnes âgées non ou peu dépendantes et dans lesquels peu de personnel soignant exerce. En 2007, les Ehpad concernaient 67 % des EHPA et 75 % des places en EHPA (Prévoit, 2009).

Contrairement aux établissements publics pour lesquels les salaires sont fixés par une grille indiciaire nationale, les directeurs d'Ehpad privés ne sont contraints que par des conventions collectives spécifiques, et ne le sont souvent que pour un montant plancher et non plafond des rémunérations. Ils pourraient donc en théorie élever les niveaux des salaires de leurs employés afin de compenser les bénéfices recherchés par la mobilité et ainsi les fidéliser. Cette hausse peut néanmoins être limitée pour des raisons budgétaires lorsque les prix ne peuvent pas suivre l'évolution des salaires, notamment en raison de leur régulation par les pouvoirs publics.

La tarification des Ehpad est ternaire : trois tarifs journaliers sont fixés et correspondent aux trois activités principales des Ehpad que sont l'hébergement, la prise en charge de la dépendance et les soins. Les charges sont réparties entre ces trois catégories selon des clés de répartition légalement imposées. Les charges de personnel infirmier sont par exemple en théorie couvertes par le tarif soins, tandis que celles liées aux aides-soignants le sont à la fois par le tarif soins et par le tarif dépendance. L'étanchéité des sections tarifaires ne semble cependant pas toujours parfaite. Martin (2014) a observé que le niveau de rémunération des aides-soignants pouvait

avoir un effet positif sur les tarifs hébergement des Ehpad privés.

Les tarifs de la dépendance et des soins sont déterminés administrativement par les Conseils départementaux et les Agences régionales de santé (ARS). Plusieurs tarifs sont fixés et correspondent aux différentes catégories de dépendance (groupes iso-ressources, GIR) des résidents. La mesure de la dépendance en Ehpad est en effet définie par une grille nationale Autonomie, gérontologie, groupes iso-ressources (AGGIR), permettant de classer les individus dans six GIR selon les activités qu'ils parviennent ou non à faire seuls, des individus les plus dépendants (GIR 1) aux moins dépendants (GIR 6). Ces tarifs étaient jusqu'en 2016 fixés de manière rétrospective, en fonction des coûts annoncés par les établissements (voir Bozio et al., 2016). Des forfaits globaux relatifs aux soins et à la dépendance sont désormais fixés depuis le 1^{er} janvier 2017 en fonction des catégories GIR et des pathologies des résidents (décret n° 2016-1814 du 21 décembre 2016).

Certains Ehpad ont également des tarifs hébergement fixés par les Conseils départementaux. Il s'agit des établissements habilités à accueillir des résidents bénéficiaires de l'aide sociale départementale, cette aide financière permettant de couvrir tout ou partie des frais liés à l'hébergement. Les Ehpad non habilités à recevoir des bénéficiaires de l'aide sociale départementale peuvent quant à eux déterminer librement leurs tarifs hébergement à l'entrée des résidents, leur évolution est néanmoins ensuite encadrée par un pourcentage fixé par arrêté ministériel (article L 342-3 du CASF). Cela concerne 15 % des Ehpad privés associatifs et 78 % des Ehpad privés à but lucratif en 2007 (Perrin-Haynes, 2010) ; ces ratios sont respectivement de 10 % et 66 % en 2011 (Volant, 2014).

manière à recruter et retenir plus aisément leur personnel à un niveau d'encadrement souhaité. Ces Ehpad peuvent toutefois être également contraints dans la fixation de leurs tarifs hébergement par le degré concurrentiel du marché et la solvabilité de la demande dans leur secteur géographique. Lorsque la part de marché d'un Ehpad dans une zone géographique donnée est faible, celui-ci est preneur de prix : il ne peut pas faire varier son prix en réponse à un ajustement nécessaire des salaires. À l'inverse, un établissement seul ou qui détient une forte part de marché peut plus facilement ajuster son prix et par conséquent les rémunérations de son personnel. Une variable du taux d'occupation moyen des établissements dans le bassin de vie où est situé l'Ehpad dans lequel travaille l'employé est donc aussi intégrée (*Taux d'occupation moyen_bv*). Lorsque le nombre de places vacantes est important (*i.e.* lorsque le taux d'occupation est faible), la possibilité d'augmenter les prix afin d'ajuster les salaires est probablement plus limitée. Pour appréhender le degré concurrentiel des Ehpad sur le marché des résidents, le zonage par bassin de vie est mieux adapté car la plupart des personnes âgées hébergées dans un Ehpad résidaient précédemment à proximité. En 2011, 75 % des résidents étaient ainsi dans un Ehpad situé à moins de 15 kilomètres de leur domicile précédent (Martin, 2014). Puisque les tarifs soins et dépendance sont déterminés par les autorités de tutelle (cf. encadré 2), nous ajoutons une variable mesurant les choix des Conseils départementaux dans la fixation de ces tarifs ; nous prenons pour cela la moyenne des tarifs dépendance pour les résidents de GIR 1 et 2 dans le département dans lequel travaille l'employé (*Tarif dépendance moyen_dpt*). Plus ce tarif est élevé, plus les établissements peuvent proposer des salaires élevés à leur personnel soignant. Nous intégrons enfin une variable catégorielle indiquant si l'établissement appartient à un groupe d'Ehpad (*Groupe*). Elle est décomposée en trois modalités : inférieure à 5 établissements, entre 5 et 20 établissements, et supérieure à 20 établissements. Cette variable peut avoir deux effets en sens inverse sur les salaires. Les établissements appartenant à une chaîne peuvent d'une part profiter d'économies d'échelle, ce qui peut leur permettre de proposer des salaires plus élevés en cas de difficultés locales d'attractivité et de fidélisation du personnel. D'autre part, certains établissements peuvent subir une pression financière exercée par leur société-mère, qui peut les contraindre à limiter les niveaux de rémunérations de leur personnel ou à réduire les taux d'encadrement (que nous avons intégré comme variable de contrôle

dans notre modèle principal). L'appartenance à un groupe pourrait donc avoir un effet indirect sur le *turnover* via les salaires et les taux d'encadrement, mais il est peu probable que cet effet soit direct. À partir de données américaines, Castle et Engberg (2006) n'observent ainsi pas d'impact significatif de l'appartenance à une chaîne sur le *turnover* du personnel soignant en maison de retraite. Brannon et al. (2002) et Castle (2005) observent un effet positif de cette variable sur le *turnover* mais ils n'ont pas intégré le salaire comme variable explicative dans leur modèle ; l'effet observé est donc probablement un effet indirect lié à l'impact de l'appartenance à une chaîne sur le salaire.

Échantillon d'étude

Les départs des salariés peuvent être volontaires (démissions) ou subis (licenciements, fins de contrat, départs à la retraite, démissions pour des déménagements contraints par des raisons extra-professionnelles). Nous ne disposons malheureusement pas des motifs de départs dans les données DADS 2008. Puisque nous cherchons à étudier les causes des départs volontaires, nous n'avons pris en compte que les employés disposant d'un contrat de travail à durée indéterminée (CDI) ; les étudiants stagiaires, le personnel intérimaire, ainsi que le personnel en contrat à durée déterminée (CDD) ont donc été exclus. Seuls les salariés âgés de moins de 60 ans ont également été intégrés à nos échantillons afin de ne pas prendre en compte les départs à la retraite. Certains salariés peuvent néanmoins être partis à la retraite avant 60 ans ; nous avons intégré une catégorie d'âge de 55 ans ou plus comme variable explicative dans nos estimations pour isoler ces départs.

Nous nous intéressons dans cette étude uniquement aux départs des infirmiers et des aides-soignants travaillant dans les Ehpad privés. Ces derniers représentent 49 % des Ehpad en 2007 et 51 % en 2011 (Volant, 2014). Les taux de départ moyens sont beaucoup plus faibles dans les établissements publics, ils s'élèvent à 26 % pour les infirmiers et à 21 % pour les aides-soignants en 2008⁶. Puisqu'une grande majorité de ces départs sont liés à des sorties de personnel en CDD⁷ (59 % des départs d'infirmiers et 68 % des départs d'aides-soignants), le nombre de

6. Les calculs ont été effectués à partir de deux échantillons de 1 169 Ehpad (taux de départ des infirmiers) et 1 184 Ehpad (taux de départ des aides-soignants). Source : DADS 2008 (Insee) et enquête Ehpa 2007 (Drees), calculs des auteurs.

7. Ce phénomène est commun à l'ensemble des organisations dans le secteur tertiaire (Bourieau et al., 2014).

départs volontaires dans les Ehpads publics est trop faible pour pouvoir en étudier les causes. Deux raisons peuvent expliquer ce faible nombre de départs volontaires. D'une part, les Ehpads publics emploient principalement des fonctionnaires ; ils représentent en effet 71 % des infirmiers et 75 % des aides-soignants en 2008. Ces derniers ne vont quitter leur établissement que s'ils trouvent un autre emploi dans la fonction publique, et ont de ce fait peut-être moins d'opportunités. D'autre part, il est possible que les infirmiers et les aides-soignants qui ont fait le choix de travailler dans la fonction publique recherchent plus de stabilité dans leur emploi que les individus non titulaires.

Nous présentons, dans le tableau 2, les statistiques descriptives de chacune des variables utilisées pour les infirmiers et les aides-soignants. Nous observons peu de différences entre ces deux catégories de professionnels, excepté le fait que l'âge, le salaire et la distance parcourue entre leur domicile et leur lieu de travail sont un peu plus faibles pour les aides-soignants que pour les infirmiers.

Résultats : les motifs de départs des infirmiers et des aides-soignants

Nous présentons les effets sur les salaires des différentes variables du modèle et des instruments exclus dans le tableau 3.

Les directeurs d'Ehpads ne semblent pas adapter les niveaux de salaires aux difficultés territoriales. Seul le fait d'être domicilié à Paris ou en Île-de-France et la présence d'infirmiers libéraux ont un effet significatif sur les salaires. Les variables d'attraction des hôpitaux, de pénurie d'infirmiers et d'aides-soignants, et de concentration en termes de postes n'ont pas d'impact sur les niveaux des rémunérations. Les établissements sont peut-être contraints dans la fixation des salaires par des tarifs régulés, ou des tarifs librement fixés mais dont la hausse est limitée par la concurrence. Le degré concurrentiel du marché, mesuré par le taux d'occupation dans le bassin de vie, a effectivement un impact positif sur les salaires des aides-soignants : plus les taux d'occupation sont élevés, plus les directeurs d'Ehpads peuvent ajuster leurs prix pour permettre une hausse des rémunérations de leur personnel.

Plusieurs tests ont été menés pour valider les instruments. Des tests de Fisher de significativité globale, réalisés à partir de régressions en

moindres carrés ordinaires de la variable de salaire sur les instruments et les variables exogènes du modèle, ont permis d'écarter l'hypothèse nulle d'instruments faibles. La statistique de Fisher est néanmoins plus faible pour la régression portant sur les infirmiers, la corrélation entre les instruments et la variable explicative est ainsi moins forte, ce qui peut nuire à la justesse des résultats. Des tests de suridentification d'Amemiya-Lee-Newey ont permis de vérifier l'exogénéité des instruments. Nous avons également réalisé des tests de Wald pour tester l'exogénéité de la variable de salaire. L'hypothèse nulle d'exogénéité du salaire des aides-soignants est rejetée, mais elle ne l'est pas pour celui des infirmiers. Cette même régression sans instrumentation du salaire conduit à des estimations similaires des effets des facteurs environnementaux, mais indique un impact positif du salaire sur la probabilité de départ des infirmiers qui semble biaisé (cf. complément en ligne C3). Ce même impact est observé lorsque l'endogénéité de la variable de salaire des aides-soignants n'est pas corrigée. Il nous semble donc préférable de présenter et d'analyser les résultats des estimations obtenues avec instrumentation du salaire pour chaque catégorie de personnel.

Nous présentons dans le tableau 4 la moyenne des effets marginaux, c'est-à-dire l'impact moyen de chaque variable sur la probabilité de départ des aides-soignants et des infirmiers.

Concernant tout d'abord les effets des caractéristiques personnelles sur la probabilité de départ des employés, l'âge semble avoir peu d'impact sur la mobilité. Seuls les aides-soignants âgés de 45 à 50 ans ont une plus forte probabilité de départ (de 4 à 5 points de pourcentage) que les aides-soignants âgés de moins de 35 ans. Quant à la distance, l'impact est sans surprise positif : plus les aides-soignants habitent loin de leur établissement, plus leur probabilité de départ est importante. Cet effet est toutefois moindre pour les infirmiers : seuls ceux qui habitent à plus de 20 km ont une plus forte probabilité de départ. Les hommes sont également plus mobiles, la probabilité qu'ils quittent leur établissement est plus importante de 7 points de pourcentage.

Les impacts des conditions d'emploi et des caractéristiques de l'établissement varient selon la profession d'infirmier ou d'aide-soignant. Le niveau de rémunération a ainsi un effet fortement significatif sur la fidélisation des aides-soignants : une hausse de 1 % de leur salaire net diminue leur probabilité de départ

Tableau 2

Statistiques descriptives relatives aux infirmiers et aides-soignants travaillant en CDI dans un Ehpad privé en France

Variables	Infirmiers			Aides-soignants		
	Médiane	Moyenne	(σ)	Médiane	Moyenne	(σ)
Salaire (en euros)	24 449	25 205	(5 629)	17 379	17 848	(3 180)
N/N* (ratio encadrement)	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	%
Départ = non	4 371	79.8
Départ = oui	1 107	20.2
Âge < 35	1 515	27.6
Âge [35 ; 45[	1 313	24.0
Âge [45 ; 50[	881	16.1
Âge [50 ; 55[	1 035	18.9
Âge ≥ 55	734	13.4
Homme = non	5 024	91.7
Homme = oui	454	8.3
Distance < 5km	1 974	36.0
Distance [5km ; 10km[	1 160	21.2
Distance [10km ; 20km[	1 360	24.8
Distance ≥ 20km	984	18.0
Nuit = non	615	11.2
Nuit = oui	4 863	88.8
Statut = privé associatif	3 514	64.1
Statut = privé lucratif	1 964	35.9
Directeur ≤ 2 ans	1 116	20.4
Directeur > 2 ans	4 362	79.6
Paris = non	5 425	99.0
Paris = oui	53	1.0
Île-de-France = non	4 923	89.9
Île-de-France = oui	555	10.1
Ehpad non régulé = non	3 102	56.6
Ehpad non régulé = oui	2 376	43.4
Groupe < 5 Ehpad	4 093	74.7
Groupe [5 ; 20[Ehpad	882	16.1
Groupe ≥ 20 Ehpad	503	9.2

	N	%
	11 058	82.3
	2 386	17.7
	5 191	38.6
	4 075	30.3
	1 960	14.6
	1 425	10.6
	793	5.9
	12 464	92.7
	980	7.3
	5 375	40.0
	2 842	21.1
	3 207	23.9
	2 020	15.0
	1 457	10.8
	11 987	89.2
	9 368	69.7
	4 076	30.3
	2 606	19.4
	10 838	80.6
	13 335	99.2
	109	0.8
	12 387	92.1
	1 057	7.9
	8 254	61.4
	5 190	38.6
	9 891	73.6
	2 425	18.0
	1 128	8.4

Note : la partie supérieure du tableau présente les statistiques descriptives (médiane, moyenne et écart-type (σ)) des variables continues utilisées dans les estimations. La partie inférieure présente les proportions de chaque modalité des variables catégorielles.

Lecture : dans le cadre de l'estimation portant sur les départs des infirmiers, le pourcentage médian (resp. moyen) de résidents en GIR 1 dans les établissements étudiés s'élève à 17,8 % (resp. 18,3 %). 64,1 % de ces établissements sont des Ehpad privés associatifs.

Champ : 5 478 infirmiers et 13 444 aides-soignants travaillant en CDI dans un Ehpad privé en France.

Source : Insee, DADS 2008 ; Drees, enquête Ehpa 2007 ; Insee, base permanente des équipements (BPE) 2007 ; Drees, répertoire Finess 2007 ; calculs des auteurs.

Tableau 3

Effets des variables exogènes du modèle et des instruments exclus sur le salaire

	Infirmiers		Aides-soignants	
	Coeff.	(σ)	Coeff.	(σ)
Caractéristiques individuelles				
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	
Caractéristiques d'emploi et de l'établissement				
Nuit	0.017*	(0.010)	ns	
N/N*	- 0.118**	(0.060)	ns	
(N/N*)²	ns		- 0.014***	(0.003)
Statut = privé associatif	ref	ref	ref	ref
Statut = privé lucratif	ns	ns	- 0.056***	0.004
N_Lits	0.001***	(0.0002)	0.001***	(0.0001)
N_Lits² (coeff. par 10 ⁻⁶ et σ par 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)
Environnement				
Hôpital (coeff. de Huff)	ns		ns	
%Inf_libéraux_bv/pop	0.008*	(0.005)	-	-
HHL _{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 exclus				
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				
Test de Fisher				
H0 : instruments faibles				
Statistique de test	F(5.5446) = 3.25		F(5.13413) = 7.49	
p-value	0.0062		0.0000	
Test d'Amemiya-Lee-Newey				
H0 : instruments exogènes				
Statistique de test	Chi²(4) = 5.101		Chi²(4) = 0.907	
p-value	0.2771		0.9235	
Test de Wald				
H0 : Log_Salaire exogène				
Statistique de test	Chi²(1) = 1.18		Chi²(1) = 28.35	
p-value	0.2772		0.0000	

Note : le tableau donne les estimations des paramètres de l'équation d'instrumentation permettant de corriger l'endogénéité de la variable Log_Salaire. Les coefficients ont été estimés par maximisation de fonctions de vraisemblance (cf. supra). Les astérisques indiquent le degré de significativité des coefficients : *** : significatif à 1 % ; ** : significatif à 5 % ; * : significatif à 10 % ; ns signifie que les coefficients ne sont pas significatifs. Des tests d'exogénéité et de validité des instruments sont présentés dans la partie inférieure du tableau.

Champ : 5 478 infirmiers et 13 444 aides-soignants travaillant en CDI dans un Ehpad privé en France.

Source : Insee, DADS 2008 ; Drees, enquête Ehpa 2007 ; Insee, BPE 2007 ; Drees, répertoire Finess 2007 ; calculs des auteurs.

de 1.2 à 1.3 %. Celui-ci n'a en revanche pas d'impact significatif sur la probabilité de départ des infirmiers. Ces derniers ont un niveau de salaire plus important que les aides-soignants, ils gagnent en moyenne 40 % de plus (cf. tableau 2). Ils rencontrent donc probablement

moins de difficultés financières dans leur vie quotidienne et sont alors sensibles à d'autres caractéristiques de leurs conditions de travail comme la qualité, évaluée ici par le ratio lié à l'encadrement. Plus l'encadrement augmente, plus la probabilité de départ des infirmiers

Tableau 4

**Effets marginaux sur la probabilité de départ des infirmiers et des aides-soignants
(modèle *probit* avec correction de l'endogénéité du salaire)**

	Infirmiers		Aides-soignants	
	Coeff.	(σ)	Coeff.	(σ)
Caractéristiques individuelles				
Âge < 35	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
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	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
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)
Caractéristiques d'emploi et de l'établissement				
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	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Statut = privé lucratif	0.095***	(0.016)	ns	
N_Lits	ns		0.001**	(0.0003)
N_Lits ² (par 10 ⁻⁶)	ns		- 2.72**	(1.12)
% GIR1	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
% 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	
Environnement				
Hôpital (coef. Huff)	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)

Note : le tableau présente la moyenne des effets marginaux des différentes variables du modèle sur la probabilité de départ dans l'année des infirmiers et des aides-soignants. L'endogénéité de la variable Log_Salaire a été corrigée par instrumentation. Les astérisques indiquent le degré de significativité des coefficients : *** : significatif à 1 % ; ** : significatif à 5 % ; * : significatif à 10 % ; ns signifie que les coefficients ne sont pas significatifs.

Lecture : le fait pour un infirmier d'être domicilié à Paris réduit par exemple sa probabilité de départ de 15 à 16 % en moyenne, avec un degré de significativité de 1 %.

Champ : 5 478 infirmiers et 13 444 aides-soignants travaillant en CDI dans un Ehpad privé en France.

Source : Insee, DADS 2008 ; Drees, enquête Ehpa 2007 ; Insee, BPE 2007 ; Drees, répertoire Finess 2007 ; calculs des auteurs.

diminue, et cela d'autant plus pour les établissements ayant un faible ratio d'encadrement car le coefficient associé à $(N/N^*)^2$ est positif. Le travail de nuit n'a quant à lui pas d'effet direct sur la fidélisation du personnel, mais il peut en avoir via l'effet sur le salaire (cf. tableau 3). Une hausse de la proportion de résidents de GIR 2 et 3, par rapport à la catégorie de référence GIR 1, augmente la probabilité de départ des infirmiers. Ces derniers préfèrent probablement réaliser des soins techniques plutôt que des soins de *nursing* et restent donc plus longtemps dans les Ehpad dans lesquels ils peuvent mettre en pratique leur savoir-faire (*i.e.* lorsque l'état de santé des résidents est très dégradé). Cet effet est également positif mais moins fortement significatif pour les aides-soignants. Plus l'établissement est grand, plus la rotation des aides-soignants est également importante. La taille de l'Ehpad influe probablement sur les relations qu'entretient l'équipe de direction avec le personnel. Dans les petits établissements, les directeurs peuvent être plus proches des salariés, plus à l'écoute de leurs souhaits en termes d'organisation du travail, d'horaires. Les départs d'infirmiers sont aussi plus élevés dans les Ehpad privés à but lucratif, mais le statut n'a pas d'impact sur les départs des aides-soignants. Enfin, la présence dans l'établissement d'un directeur depuis plus de deux ans n'a pas d'impact direct sur la fidélisation du personnel, mais il peut en avoir de manière indirecte par le salaire (cf. tableau 3).

Concernant enfin l'impact de l'environnement local, la présence proche d'un hôpital a un effet positif et significatif sur les départs d'infirmiers et d'aides-soignants. Ceux-ci sont alors davantage incités à changer de mode d'exercice et à partir vers le secteur hospitalier, qui peut proposer des évolutions de carrières plus intéressantes et une plus grande diversité de postes et de secteurs. Pour les infirmiers, la présence d'infirmiers libéraux travaillant dans leur bassin de vie a également un effet positif et significatif sur leur probabilité de départ. Ceci constitue une autre opportunité de changement de mode d'exercice. Il existe une forte inégalité territoriale en termes de densité d'infirmiers libéraux. Dans les territoires où cette densité est élevée, les infirmiers libéraux réalisent en proportion moins d'actes médicaux infirmiers (AMI) et plus de soins de *nursing* ou actes infirmiers de soins (AIS) (Barlet & Cavillon, 2011). Dans les régions sous dotées en infirmiers libéraux, ces soins de *nursing* sont effectués par des services d'aides à domicile. Cette substitution d'actes de nature différente permet aux infirmiers libéraux de conserver

un niveau d'activité convenable même lorsque l'offre est importante, ce qui peut attirer les infirmiers salariés résidant dans ces territoires⁸. La variable de concentration a un effet négatif significatif : plus le marché est concentré en termes de postes de personnel soignant, moins le personnel est incité à quitter son établissement. Nous observons également que plus leur bassin de vie est fortement doté en aides-soignants relativement à la population locale, moins les aides-soignants quittent leur emploi. Les raisons peuvent être d'une double nature. Les aides-soignants conservent d'une part leur emploi car les opportunités d'embauches sont rares compte tenu du nombre élevé d'aides-soignants sur le marché du travail local. D'autre part, ils travaillent probablement dans un bassin de vie relativement attractif et ne cherchent donc pas à travailler dans une autre localité. Enfin, le personnel est plus mobile en Île-de-France, en raison peut-être d'un meilleur réseau de transport en commun que dans les autres régions françaises. Les infirmiers domiciliés à Paris ont quant à eux une propension plus faible à quitter leur emploi. Ces infirmiers ont probablement choisi de payer plus cher leur logement pour réduire leur temps de transport et sont par conséquent moins disposés à changer d'établissement si cela joue sur leur trajet domicile-travail.

* *
*

Nous avons mis en évidence l'existence de difficultés locales de fidélisation du personnel soignant au sein des Ehpad privés : la présence proche d'un hôpital, la densité d'infirmiers libéraux, une pénurie globale d'aides-soignants et une pression concurrentielle accrue entre Ehpa peuvent inciter les infirmiers et les aides-soignants à quitter les établissements dans lesquels ils travaillent.

Les niveaux de salaires ont également un effet sur le comportement des aides-soignants. Plus ils sont élevés, plus leur probabilité de départ est faible. Une compensation salariale pourrait être un moyen pour réduire l'effet des disparités territoriales sur le *turnover* de cette catégorie de personnel. Cependant, leurs rémunérations

8. Depuis 2011, le conventionnement des infirmiers libéraux dans les zones « sur dotées » ne peut plus intervenir qu'en remplacement d'un départ (avenant n°3 à la convention nationale des infirmières et des infirmiers libéraux). Cet effet est donc probablement fortement atténué.

ne sont actuellement pas fixées en fonction de ces difficultés locales car elles sont plafonnées par les tarifs des Ehpad. Les tarifs soins et dépendance, qui couvrent en théorie les charges de personnel soignant, sont fixés administrativement par les Conseils départementaux et les Agences régionales de santé (ARS). Les tarifs hébergement peuvent parfois compenser l'insuffisance des rémunérations ; néanmoins, lorsqu'ils ne sont pas eux-mêmes également régulés, ils peuvent être contraints par une concurrence en prix, ce que nous observons dans certains bassins de vie. La probabilité pour les infirmiers de quitter leur emploi n'est quant à elle pas impactée par le salaire. Leur offre de travail semble ainsi inélastique au prix et déterminée par d'autres facteurs tels que la qualité mesurée par l'encadrement en personnel, ou les degrés de dépendance des résidents qui influent sur la nature du travail à effectuer.

Les directeurs d'établissements ont ainsi peu de leviers à disposition pour faire face au *turnover* de leur personnel soignant. La présence d'un directeur de plus de deux ans et le travail de nuit n'ont pas d'impact sur leur fidélisation. Seules une augmentation des salaires et un renforcement de l'encadrement en personnel semblent être en mesure de réduire la probabilité de départ des salariés. Or, ces mesures nécessitent une hausse de la masse salariale et, par répercussion, des tarifs soins et dépendance des Ehpad. Puisque ces tarifs sont respectivement couverts par l'Assurance maladie et en grande partie par les Conseils départementaux via le versement de l'Allocation personnalisée d'autonomie (APA), ces mesures induiraient

un surcoût pour les finances publiques. Elles semblent néanmoins indispensables compte tenu de l'impact que peut avoir une réduction du *turnover* du personnel soignant sur la qualité de la prise en charge des résidents.

Cette étude économétrique est la première à analyser les motifs de départ du personnel soignant en Ehpad en France. Il pourrait être intéressant, dans des études complémentaires, d'analyser plus en détail l'impact de certaines variables, notamment de qualité, sur les décisions de démission du personnel soignant. Nous ne disposons malheureusement que du taux d'encadrement qui ne permet bien évidemment pas d'approcher toute la multi-dimensionnalité de la qualité de la prise en charge. Nous manquons également dans cette étude de données sur les caractéristiques sociodémographiques des infirmiers et des aides-soignants, ainsi que sur la culture organisationnelle et la politique managériale du directeur d'établissement. Or, plusieurs auteurs ont montré que l'implication du personnel soignant dans la gestion des emplois du temps ou dans les décisions administratives peut avoir un impact non négligeable sur leur satisfaction au travail, et donc sur leur choix de rester travailler dans l'établissement (Donoghue & Castle, 2007). Enfin, nous ne disposons ici que de données en coupe qui, contrairement aux données de panel, ne permettent pas de corriger l'endogénéité liée à l'hétérogénéité non observée. Or, il se peut que les employés qui décident de quitter leur établissement aient des caractéristiques particulières qui ne soient pas toutes prises en compte par les variables explicatives du modèle. □

BIBLIOGRAPHIE

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.

Bouricau, 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/series-tat142-2.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* (2^e 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.

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

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

Milan Bouchet-Valat*

Résumé – Le développement de l'emploi des femmes a souvent été présenté comme un facteur d'accroissement des inégalités entre ménages, du fait notamment d'un renforcement supposé de l'homogamie sociale. À rebours de cette idée, une approche comptable de décomposition des inégalités montre, à partir des enquêtes Emploi de l'Insee, que les inégalités de salaire entre couples de 30 à 59 ans sont restées globalement stables entre 1982 et 2014 en France, alors qu'elles auraient augmenté si le taux d'emploi des femmes n'avait pas progressé. Cette stabilité globale recouvre deux évolutions inverses, liées au fort développement de l'emploi des femmes sur cette période : une diminution de l'inégalité de salaire entre femmes et une augmentation de la corrélation entre les salaires des conjoints au sein des couples. Toutefois, la progression quasi uniforme du taux d'emploi des femmes, quel que soit le niveau de salaire de leur conjoint, a limité l'augmentation de la corrélation entre les salaires des conjoints et évité un renforcement des inégalités de salaires entre couples.

Abstract – *It has often been argued that women's employment growth is a factor of increasing inequalities between households, in particular because it is assumed to reinforce social homogamy. In contrast to this idea, an accounting approach of decomposition based on Insee's Labor Force surveys (enquêtes Emploi) shows that wage inequalities between couples aged 30 to 59 remained stable between 1982 and 2014 in France, while they would have increased if women's employment rate had not risen. This overall stability is resulting from two converse developments, 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 spouses' wages within couples. However, the almost uniform increase in women's employment rate, regardless of the wage level of their spouse, has limited the increase in the correlation of spouses' wages and prevented an increase in wage inequalities between couples.*

Codes JEL / JEL codes : D10, D31, D63, J12, J22

Mots-clés : inégalités salariales, emploi des femmes, couple, homogamie

Keywords : wage inequality, women's employment, couples, homogamy

Rappel :

Les jugements et opinions exprimés par les auteurs n'engagent qu'eux-mêmes, et non les institutions auxquelles ils appartiennent, ni a fortiori l'Insee.

* Institut national d'études démographiques (Ined) et Crest, Laboratoire de sociologie quantitative (milan.bouchet-valat@ined.fr).
Les tables et le code permettant de reproduire les analyses sont disponibles sur la page personnelle de l'auteur à l'adresse <http://bouchet-valat.site.ined.fr>.

Le taux d'emploi des femmes a très nettement augmenté en France depuis les années 1960. Ce mouvement, qui se prolonge encore aujourd'hui à un rythme régulier, a eu des conséquences importantes sur les sources de revenus des ménages. Ainsi, entre 1982 et 2014, la contribution du total des salaires des femmes à la masse des salaires des couples entre 30 et 59 ans (hors indépendants et retraités) est passée de 27 % à 38 % (voir figure I plus bas).

Une certaine augmentation des inégalités de salaires ou de revenus, entre individus et entre ménages, a par ailleurs été mise en évidence en France, due principalement à la forte croissance des revenus dans le haut de la structure sociale (Landais, 2007 ; Amar, 2010 ; Solard, 2010 ; Godechot, 2012, 2013 ; Piketty, 2013). On peut se demander, comme l'a fait récemment l'Organisation pour la coopération et le développement économiques (OCDE, 2008, 2011), si l'augmentation du taux d'emploi des femmes, en induisant une plus grande proximité entre les salaires des conjoints (conséquence de l'homogamie sociale), a contribué à renforcer cette tendance.

De manière plutôt contre-intuitive, cet article apporte une réponse négative à cette question. Les inégalités de salaire entre couples de 30 à 59 ans (hors indépendants et retraités) sont restées assez stables en France entre 1982 et 2014, et ont même légèrement diminué au-dessus de la médiane. Cette évolution résulte principalement de l'effet égalisateur du développement de l'emploi des femmes, alors que la légère diminution du taux d'emploi des hommes a eu tendance à renforcer les inégalités entre couples.

Cette évolution d'ensemble cache des tendances contrastées selon la partie considérée de la distribution des salaires des couples. Ainsi, si l'augmentation du taux d'emploi des femmes a eu des effets égalisateurs sur l'ensemble de la distribution, la diminution du taux d'emploi des hommes a eu des effets inégalitaires en dessous de la médiane.

En France, l'augmentation des salaires des femmes ne s'est pas accompagnée d'un renforcement de l'association entre les salaires des conjoints au sein des couples suffisant pour augmenter les inégalités entre couples, contrairement à ce qui a été observé notamment aux États-Unis (Cancian & Reed, 1998 ; Hyslop, 2001 ; Schwartz, 2010). Ce phénomène s'explique notamment par le fait que l'augmentation du taux d'emploi des femmes

s'est faite quasiment au même rythme quel que soit le décile de salaire du conjoint (à l'exception notable du premier décile et des inactifs et chômeurs). L'association entre les salaires des conjoints s'est même affaiblie dans le haut de la distribution.

Ces tendances sont cohérentes avec celles constatées par ailleurs en France concernant les inégalités salariales et le choix du conjoint. Ainsi, contrairement à une croyance répandue, si l'on excepte le sommet de la distribution, c'est plutôt une légère baisse des inégalités de salaire qui prend place depuis les années 1960 (Verdugo et al., 2012 ; Charnoz et al., 2013 pour les salaires des hommes à temps complet dans le secteur privé ; Verdugo, 2014), alors que les inégalités de niveau de vie sont de leur côté relativement stables depuis les années 1990 (Pujol et Tomasini, 2009 ; Boiron et al., 2016). D'autre part, la ressemblance entre conjoints (homogamie sociale) tend à s'affaiblir aussi bien en termes de diplôme que de profession depuis plusieurs décennies (Vanderschelden, 2006 ; Bouchet-Valat, 2014).

L'étude s'appuie sur les données annuelles des enquêtes *Emploi* de l'Insee depuis 1982, et porte uniquement sur les salaires (encadré 1). Les couples dans lesquels l'un des conjoints est indépendant ou retraité sont exclus du champ. Pour ne pas alourdir les formulations, nous assimilerons dans la suite l'« emploi » au fait de percevoir un salaire ; le « taux d'emploi » désignera donc la proportion des individus de ce champ qui perçoivent un salaire. Aucune distinction n'est faite entre salariés à temps complet et à temps partiel, et l'absence de salaire est considérée comme un salaire nul. Les analyses portent donc sur le salaire effectivement perçu, résultat aussi bien de décisions d'activité plus ou moins choisies ou subies (et fréquemment liées à des situations familiales, voir Meron & Maruani, 2012, chapitre 2), que d'une position plus ou moins favorisée en termes de salaire horaire. Si l'absence de prise en compte des autres types de revenus, et notamment de ceux du patrimoine, revient à laisser de côté une partie importante de l'augmentation des inégalités de revenus (Landais, 2007 ; Piketty, 2013), cette approche garde néanmoins tout son intérêt pour étudier en particulier les effets du développement de l'emploi des femmes sur le lien entre les revenus des conjoints et donc sur les inégalités entre couples.

Nous présentons d'abord une revue de littérature internationale et française. Nous montrons

ensuite, grâce à une décomposition du coefficient de variation, que la croissance du taux d'emploi des femmes n'a pas entraîné de renforcement de l'inégalité de salaire entre couples. Ce résultat découle notamment du fait que la distribution sociale de l'emploi des femmes est demeurée relativement stable, limitant ainsi l'augmentation de la corrélation entre salaires des conjoints. Finalement, grâce à une méthode de décomposition fondée sur différents scénarios contrefactuels mis en œuvre par des modèles log-linéaires, nous mettons en évidence des effets différents des évolutions de l'emploi des femmes et des hommes sur les inégalités selon la partie de la distribution des salaires considérée, ainsi que le rôle assez limité de l'évolution de l'association entre salaires des conjoints.

Emploi des femmes, homogamie et inégalités de revenus entre couples dans la littérature

La question de l'évolution de l'association entre les salaires des conjoints au sein d'un couple et de ses effets sur les inégalités a été soulevée principalement dans des travaux portant sur les États-Unis : il s'agissait au départ d'évaluer l'effet de l'augmentation de l'activité professionnelle des femmes sur les inégalités entre ménages depuis les années 1960. Le contexte américain diffère assez nettement de celui de la France puisqu'on y observe une forte augmentation de l'homogamie d'éducation, qui a fréquemment été perçue comme un risque pour la cohésion de

Encadré 1

SOURCE, CHAMP, ET DÉFINITION DU SALAIRE

Cette étude se fonde sur la série des enquêtes *Emploi* de l'Insee de 1982 à 2014, et concerne l'ensemble des couples cohabitants (mariés ou non) dans lesquels les deux conjoints ont entre 30 et 59 ans et aucun n'est indépendant ou retraité. La restriction du champ aux individus de 30 ans au moins permet de limiter autant que possible les effets du recul de l'âge de première entrée en cohabitation et sur le marché du travail, qui exigeraient des analyses séparées. En effet, sur toute la période étudiée, environ un tiers des hommes et la moitié des femmes entre 20 et 30 ans vivent en couple cohabitant une année donnée, alors que ce taux dépasse les trois quarts pour les deux sexes parmi les trentenaires ; cette proportion a diminué de 6 à 8 points de pourcentage au fil des années (Bouchet-Valat, 2014, p. 467-468). Par ailleurs, si le taux d'emploi entre 20 et 30 ans a nettement baissé depuis 1982, aussi bien chez les hommes que chez les femmes, il est resté plus stable chez les hommes trentenaires et a augmenté (comme pour toutes les classes d'âge) chez les femmes trentenaires (Insee, 2016).

Les enquêtes *Emploi* offrent l'intérêt de fournir des données annuelles fondées sur des échantillons de taille importante, pour lesquels on dispose d'informations homogènes sur les salaires mensuels. En revanche, ceux-ci sont connus sur une base purement déclarative. Si cela peut entraîner une différence de niveau par rapport aux inégalités mesurées à partir de sources administratives une comparaison avec les salaires annuels des *Déclarations annuelles des données sociales (DADS)*, déclarés par les employeurs, montre que les évolutions sont largement similaires (Verdugo, 2014).

Le questionnaire des enquêtes *Emploi* interroge les individus sur le montant de leur salaire mensuel le

mois précédant l'enquête, ainsi que sur les éventuels compléments perçus annuellement (primes, 13^e mois...). Les enquêtés se voient proposer la possibilité de ne pas répondre, ou bien d'indiquer une tranche de salaire plutôt qu'un montant exact, auxquels cas l'Insee impute un montant à partir des autres variables disponibles.

Le salaire étant déclaré par tranches entre 1982 et 1989, nous avons recouru pour ces années à un modèle d'imputation par la méthode des résidus simulés (O'Prey, 2009, p. 17). Le modèle, appliqué séparément aux hommes et aux femmes (vivant en couple ou non), prend en compte la censure par intervalles liée aux tranches, et suppose une distribution log-normale des salaires. Les variables prises en compte sont le temps de travail habituel, l'ancienneté combinée au type de contrat, l'âge (et son carré), la catégorie socio-professionnelle (PCS niveau 3), le diplôme, la tranche d'unité urbaine et la région de résidence.

Du fait des importantes fluctuations liées à l'échantillonnage au sommet de la distribution des salaires, les salaires qui dépassent pour chaque année le 995^e millime (soit 0,5 % des cas) ont été ramenés à ce niveau. Les effectifs des enquêtes ne permettent pas d'étudier précisément ce groupe (représentant environ 40 individus par an) d'une année sur l'autre.

Seuls les actifs occupés au moment de l'enquête sont interrogés sur leur salaire. Nous avons attribué un salaire nul aux chômeurs et inactifs (parmi lesquels moins de 3 % déclarent un salaire). L'échantillon, limité aux individus interrogés pour la première fois dans l'enquête, comprend entre 5 300 et 7 000 couples par an avant 2009, et entre 7 700 et 8 700 depuis, pour un total de 217 000 couples.

la société américaine (Breen & Salazar, 2011). Ce phénomène a été considéré comme l'un des facteurs de l'augmentation des inégalités observée dans nombre de pays du monde (Blossfeld & Buchholz, 2009 ; Esping-Andersen, 2007). À rebours de cette lecture, les travaux concordent pour conclure que la croissance de l'emploi des femmes a au contraire eu tendance à limiter l'augmentation des inégalités de revenus aux États-Unis (Cancian & Reed, 1998, 1999 ; Reed & Cancian, 2001 ; Devereux, 2004 ; Pencavel, 2006 ; Western et al., 2008 ; Daly & Valletta, 2006 ; Hryshko et al., 2014) ainsi que dans tous les pays de l'OCDE, et particulièrement en France (OCDE, 2011, p. 226 ; Harkness, 2013 ; ces deux études comparatives se sont fondées sur la Luxembourg Income Study, constituée pour la France par l'enquête *Budget de famille*).

Le rôle propre de l'évolution de l'homogamie est plus discuté. En France, une analyse reconstituant l'ensemble des salaires perçus par les couples au cours de leur vie (Courtioux & Lignon, 2015a) a récemment montré que l'homogamie d'éducation n'atténue que très partiellement l'effet égalisateur de la mise en couple. En effet, le coefficient de Gini concernant les salaires perçus au cours de l'ensemble de leur vie par les individus d'une même génération diminue de 12 % si l'on considère les couples plutôt que les individus pris isolément. Cette diminution serait plus marquée de 3 points de pourcentage pour les femmes et de 7 points pour les hommes si les couples se formaient au hasard (absence d'homogamie). En recourant à une méthode différente, une autre étude (Frémeaux & Lefranc, 2015) a estimé que l'homogamie d'éducation augmente de 3 % à 10 % les inégalités entre salaires annuels des couples. Ces études laissent cependant de côté la question des évolutions temporelles. Étant donné l'affaiblissement de l'homogamie de diplôme et de classe sociale mis en évidence par des travaux portant sur la France (Vanderschelden, 2006 ; Bouchet-Valat, 2014), il semble peu probable que l'évolution de ce facteur ait pu contribuer de façon significative à une augmentation des inégalités entre ménages au cours des dernières décennies.

La plupart des travaux disponibles sur d'autres pays confirment que, contrairement à une idée commune, l'homogamie d'éducation n'entretient qu'un lien relativement lâche, bien que réel, avec l'association entre les revenus des conjoints¹, et que son évolution n'a eu que des effets négligeables – voire négatifs – sur l'augmentation des inégalités de revenus dans de nombreux pays (Worner, 2006 pour l'Australie ; Western

et al., 2008 ; Breen & Salazar, 2010, 2011 pour les États-Unis et le Royaume-Uni ; Breen & Andersen, 2012 pour le Danemark ; Eika et al., 2014 pour les États-Unis et la Norvège).

Nous nous concentrons donc ici sur la mesure de l'effet de l'évolution de l'association entre salaires des conjoints sur les inégalités entre couples : cette question doit à notre sens être traitée avant celle de l'influence des variations de l'homogamie d'éducation, qui n'entretient qu'un rapport indirect avec les salaires des conjoints. Les études existantes sur cette question ont observé soit que l'évolution de l'association entre revenus des conjoints explique entre 15 % et 30 % de l'augmentation totale des inégalités aux États-Unis selon les périodes étudiées et les méthodes retenues (Karoly & Burtless, 1995 ; Burtless, 1999 ; Cancian & Reed, 1999 ; Hyslop, 2001 ; Schwartz, 2010), soit qu'elle y prend une part encore plus modeste (Cancian & Reed, 1998 ; Hryshko et al., 2014 pour les États-Unis ; OCDE, 2011, p. 226 pour les pays membres de l'organisation ; Funes Leal, 2015 pour l'Argentine). Plus précisément, aux États-Unis, ce facteur aurait principalement contribué à la hausse des inégalités de revenus dans les années 1980, mais de manière assez négligeable depuis (Larrimore, 2014).

Dans le cas de la France, il apparaît d'ores et déjà que l'effet égalisateur de la croissance de l'emploi des femmes l'emportera sur l'effet contraire d'une éventuelle augmentation de l'association entre salaires des conjoints. En effet, c'est bien le premier phénomène qui constitue l'évolution majeure sur la période considérée.

L'apport de la présente étude est avant tout d'analyser les évolutions au cours du temps, alors que les travaux existants sur la France (Frémeaux & Lefranc, 2015 ; Courtioux & Lignon, 2015a) concernent un seul point temporel (que ce soit une cohorte ou une année d'enquête). Ce recul historique est nécessaire pour identifier le rôle du développement de l'emploi des femmes sur les inégalités. Comme Frémeaux et Lefranc, nous étudions ici directement les inégalités entre les salaires perçus par les couples, alors que l'étude de Courtioux et Lignon s'intéresse aux inégalités qui s'expliquent par l'existence d'une tendance à l'homogamie d'éducation (voir Courtioux & Lignon, 2015b, pour une

1. Si une étude affirmant que l'augmentation de l'homogamie d'éducation explique une part importante de la croissance des inégalités entre ménages aux États-Unis a connu un certain écho, les résultats ont fait l'objet d'un corrigendum qui atténue très fortement cette affirmation (Greenwood et al., 2014).

présentation des différentes méthodes utilisées par cette dernière approche).

Courtioux et Lignon (2015a) ont cherché à reconstituer les revenus perçus par les membres d'une même cohorte tout au long de leur vie à l'aide de modèles de microsimulation dynamique. Plus simple, l'analyse des salaires au moment de l'enquête retenue ici présente l'avantage de décrire l'association entre salaires des conjoints réellement observée chaque année, sans avoir à faire d'hypothèses visant à reconstituer la composition des couples en tenant compte notamment de l'homogamie d'éducation. Cependant, elle ne tient pas compte de l'ensemble des revenus des individus au cours de leur vie, ni du célibat, ni encore de la taille des ménages (qui permet de calculer des niveaux de vie).

De son côté, l'étude de Frémeaux et Lefranc (2015) s'est intéressée aux salaires annuels – réellement perçus et en équivalent temps plein – là où le présent article s'en tient aux salaires mensuels des conjoints. Cette restriction nous permet de remonter plus loin dans le temps grâce aux enquêtes *Emploi*. En revanche, elle introduit des variations des salaires sur courte période que le recours à des moyennes annuelles ou pluriannuelles permettrait de lisser. Cette limite conduit probablement à sous-estimer l'association entre salaires des conjoints ; cependant, l'estimation des contributions des différents facteurs aux inégalités et celle de leurs évolutions temporelles sont peu affectées par ce biais (Frémeaux & Lefranc, 2015, p. 11 ; Hryshko et al., 2014, p. 771).

Pour finir, soulignons que l'approche comptable (Courtioux & Lignon, 2015b) que nous adoptons vise à décomposer les effets des différents facteurs sur l'homogamie dans l'hypothèse où les comportements des individus resteraient inchangés. Il s'agit donc d'un exercice descriptif et illustratif mais qui ne vise pas à identifier des liens causaux.

Le développement de l'emploi des femmes n'a pas renforcé les inégalités de salaire entre couples

La croissance de l'emploi et des salaires des femmes en France depuis 1982

Le développement de l'emploi des femmes, à l'œuvre depuis le début des années 60, a été très

marqué en France depuis 1982. La forte croissance de la contribution des femmes au total des salaires des couples mentionnée en introduction a résulté de la conjonction de deux tendances : d'un côté, l'augmentation du taux d'emploi, c'est-à-dire du point de vue retenu ici la croissance de la proportion de femmes qui perçoivent un salaire ; de l'autre, la croissance des salaires parmi les femmes en emploi.

La croissance tendancielle du taux d'emploi des femmes est bien connue : la proportion des femmes de 30 à 59 ans en couple (hors indépendants et retraités) percevant un salaire est ainsi passée de 51 % à 78 % entre 1982 et 2014 (figure I). Une autre évolution est cependant passée plus inaperçue : parallèlement à l'augmentation de la contribution des femmes aux salaires des couples, la proportion des femmes percevant un salaire supérieur à celui de leur conjoint a doublé, passant de 12 % à 24 % sur la même période (voir aussi Morin, 2014).

Cette dernière tendance s'explique bien sûr par l'augmentation du taux d'emploi des femmes, mais aussi par une très légère augmentation des salaires des femmes en emploi par rapport à ceux des hommes (Minni, 2015) : les salaires médians des femmes en couple étaient inférieurs de 36 % à ceux des hommes en couple en début de période, et de 28 % en fin de période. Ce mouvement, assez faible, a été plus marqué en bas de la distribution (on passe d'un écart de 67 % à un écart de 49 % pour le premier décile).

Notons que cette évolution s'est faite en dépit d'une forte croissance de la part des emplois à temps partiel chez les femmes (Afsa Essafi & Buffeteau, 2006), qui passe de 19 % en 1982 à 32 % en 1999, et s'est stabilisée autour de ce niveau depuis (Insee, 2016). Ce taux a aussi progressé chez les hommes, passant de 3 % à 8 %, tout en demeurant à un niveau assez bas. À l'inverse, le taux de chômage a plus augmenté chez les hommes (chez lesquels il double, de 5 % en 1982 à 10 % en 2014) que chez les femmes (Cabannes, 2014 ; Insee, 2016). Il était en effet déjà relativement élevé chez ces dernières (à 8 % en 1982), et s'il a frôlé les 12 % au milieu des années 1990, il est revenu depuis au même niveau que chez les hommes (à 10 % en 2014). Les différences entre hommes et femmes concernant le temps partiel et le chômage ont donc suivi des tendances contraires, alors que l'augmentation du taux d'activité et du taux d'emploi ont de leur côté nettement favorisé la croissance des salaires des femmes par rapport à ceux des hommes.

L'évolution des salaires des femmes au cours des trente dernières années a été très marquée. Cependant, même si le rythme d'augmentation très stable observé depuis 1982 se maintenait, les salaires des femmes représenteraient la moitié du total des salaires des couples aux alentours de 2045. Il est en outre peu probable que ce calendrier soit tenu étant donné que le taux d'emploi des femmes en couple, qui est le facteur principal de cette évolution, atteindrait (sous la même hypothèse) 100 % une décennie plus tôt, et qu'il se stabiliserait probablement bien avant. Pour les mêmes raisons, il est loin d'être certain que la moitié des femmes perçoivent un salaire supérieur à celui de leur conjoint au cours du XXI^e siècle : même d'après la lecture très optimiste consistant à prolonger le rythme passé, cet événement aurait lieu autour de l'année 2080.

Par ailleurs, le développement de l'emploi des femmes n'a pas empêché la persistance d'une situation plus défavorable pour elles sur le marché du travail (Meron, 2008 ; Albouy et al. 2012) : plus grande précarité des emplois, temps partiel, salaires inférieurs du fait de la structure de ces emplois (ségrégation entre secteurs et professions), mais aussi à poste égal. Ces réserves et l'incertitude portant sur les perspectives futures ne doivent cependant pas

occulter l'importance des évolutions qui ont eu lieu depuis les années 1980.

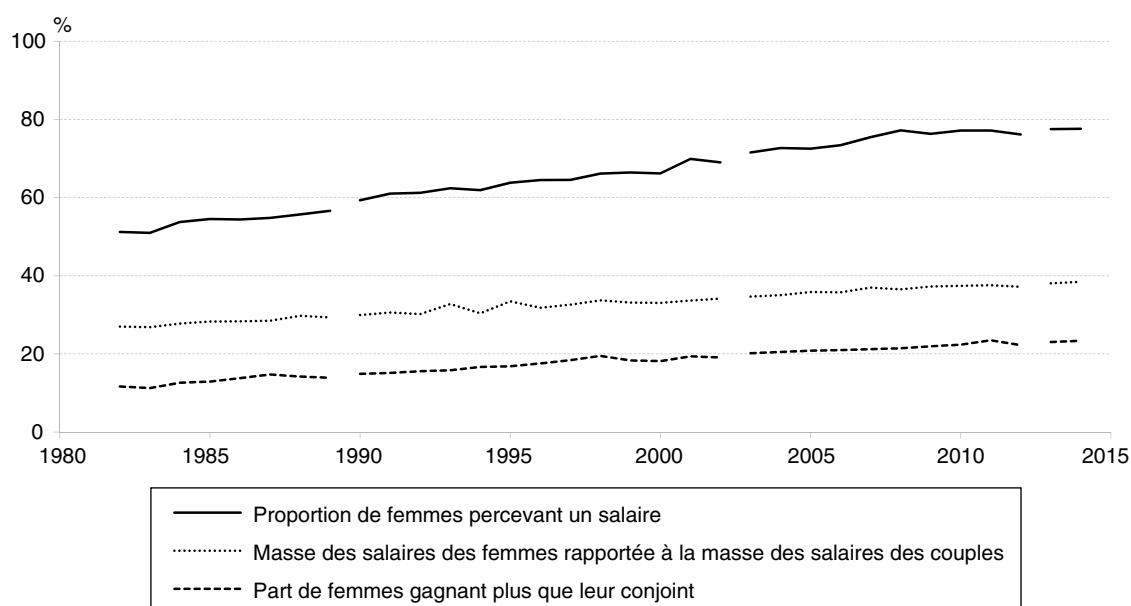
Une baisse des inégalités de salaire entre femmes et une stabilité des inégalités entre couples

L'augmentation de plus de dix points de pourcentage de la part des salaires des femmes dans le total des salaires des couples a-t-elle eu pour effet de réduire les inégalités entre couples ? Afin de répondre à cette question, nous procédons dans cette section à une première décomposition de l'inégalité entre salaires des couples, et de son évolution temporelle, en trois sources distinctes. Nous mesurons pour cela le niveau d'ensemble des inégalités à l'aide du coefficient de variation, qui présente une définition aisément interprétable en termes de dispersion par rapport à la moyenne. Il est défini comme le rapport entre l'écart-type σ et la moyenne μ d'une distribution, soit :

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

L'évolution de l'inégalité entre les salaires totaux des couples (en attribuant un salaire nul

Figure I
La croissance des salaires des femmes en couple depuis 1982 à travers trois indicateurs



Note : les chômeurs et inactifs sont inclus dans le champ et se voient attribuer un salaire nul.

Lecture : en 1982, 51 % des femmes en couple de 30 à 59 ans touchaient un salaire, mais ceux-ci ne représentaient que 27 % du total des salaires des couples, et seules 12 % d'entre elles gagnaient plus que leur conjoint.

Champ : hommes et femmes vivant en couple cohabitant dans lequel les conjoints ont entre 30 et 59 ans et aucun n'est indépendant ou retraité.

Source : Insee, enquêtes Emploi, 1982-2014.

aux chômeurs et inactifs) est présentée dans la figure II. On observe une fluctuation sans tendance claire, qui aboutit en 2014 à un niveau d'inégalité très proche de celui de 1982. La décomposition du coefficient de variation va nous permettre de comprendre ce résultat.

En effet, le carré du coefficient de variation peut être écrit comme la somme de trois termes correspondant chacun à une source de revenu bien identifiée :

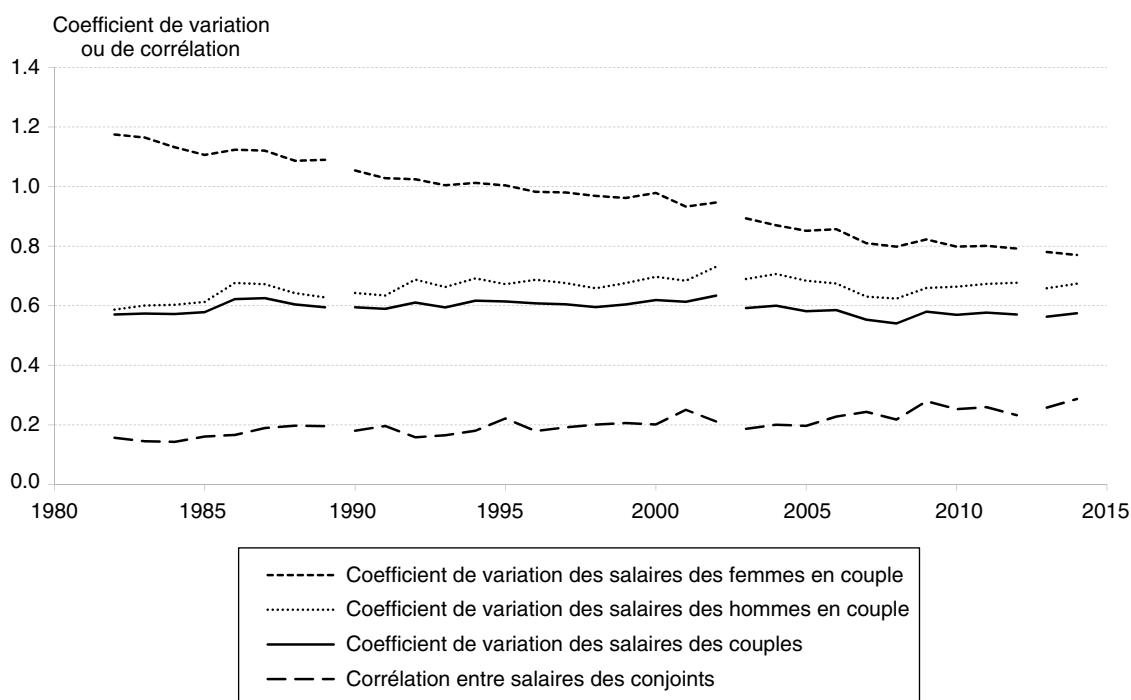
$$CV^2 = (1 - S_f)^2 CV_h^2 + S_f^2 CV_f^2 + 2\rho_{hf} S_f (1 - S_f) CV_h CV_f$$

S_f mesure la part des salaires des femmes dans le total des salaires des couples, CV_h et CV_f sont respectivement les coefficients de variation des salaires des hommes et des femmes, et ρ_{hf} correspond à la corrélation entre les salaires des conjoints au sein des couples. L'inégalité entre couples est donc d'autant plus prononcée que l'inégalité entre les individus de même sexe est

élevée, que le sexe pour lequel l'inégalité est la plus marquée contribue fortement au total des salaires des couples, et que la corrélation entre salaires des conjoints est forte.

D'après cette équation, la contribution de chaque sexe à l'inégalité entre couples peut être évaluée par rapport à (au moins) trois situations de référence dites contrefactuelles, qui permettent d'identifier trois sources différentes de l'évolution des inégalités (Cancian & Reed, 1998, p. 74). Première situation : si seuls les individus d'un sexe contribuaient au total des salaires des couples, les membres de l'autre sexe étant tous inactifs ou chômeurs, le niveau des inégalités entre couples serait respectivement de CV_h ou de CV_f selon que les actifs occupés sont les hommes ou les femmes. Deuxième situation : si les salaires des individus d'un sexe donné étaient tous égaux, et que par conséquent l'inégalité trouvait exclusivement son origine dans les salaires de l'autre sexe, alors le niveau d'inégalité entre les couples serait égal à $(1 - S_f) CV_h$ ou $S_f CV_f$, selon que l'inégalité proviendrait des hommes ou des femmes. Troisième

Figure II
Évolution des inégalités de salaire et de la corrélation entre salaires des conjoints depuis 1982 (ensemble des couples)



Note : les chômeurs et inactifs sont inclus dans le champ et se voient attribuer un salaire nul.

Lecture : le coefficient de variation des salaires des femmes en couple (y compris celles dont le salaire est nul) est passé de 1.17 à 0.77 entre 1982 et 2014.

Champ : hommes et femmes vivant en couple cohabitant dans lequel les conjoints ont entre 30 et 59 ans et aucun n'est indépendant ou retraité.

Source : Insee, enquêtes Emploi, 1982-2014.

situation : s'il n'existait aucune corrélation entre les salaires des conjoints, le coefficient de variation serait égal à la racine carrée de la somme des termes $(1 - S_f)^2 CV_h^2$ et $S_f^2 CV_f^2$.

Ainsi, l'association entre les salaires des conjoints joue potentiellement un rôle important dans la détermination des inégalités entre couples. Dans une situation hypothétique, irréaliste en 1982 mais de plus en plus crédible au fil du temps, où les inégalités de salaire seraient les mêmes parmi les hommes et parmi les femmes et où les deux sexes contribueraient à parts égales au volume total des salaires, passer d'une absence de corrélation à une corrélation parfaite entre salaires des conjoints équivaldrait exactement à doubler le niveau des inégalités entre couples. Une évolution aussi radicale de la corrélation est assez improbable : au contraire, la corrélation entre salaires des conjoints est généralement trop faible pour avoir des conséquences aussi fortes sur les inégalités².

De fait, la stabilité de l'inégalité d'ensemble résulte du fait que les différentes composantes de l'équation sont soit restées relativement stables, soit ont suivi des tendances opposées qui se sont largement compensées (cf. figure II). Tout d'abord, malgré une certaine tendance à l'augmentation jusqu'aux années 2000, l'inégalité entre les salaires des hommes (CV_h) a peu évolué, fluctuant autour d'un coefficient de variation de 0.7³. En revanche, l'inégalité entre les salaires des femmes (CV_f) a très nettement diminué, avec un coefficient de variation passant de 1.17 à 0.77 : elle s'est clairement rapprochée du niveau qui prévaut chez les hommes. Ce mouvement s'explique par le développement de l'emploi des femmes, qui a fait passer le salaire d'une partie importante de l'échantillon d'une valeur nulle (inactivité) à un niveau qui a toutes les chances d'être plus proche de la moyenne (Pasqua, 2002). Cette diminution de l'inégalité entre femmes associée à la nette augmentation de la part des salaires des femmes dans les salaires des couples (S_f) a eu un effet égalisateur très clair (cf. figure I).

Si cette évolution est ainsi une conséquence mécanique de l'augmentation du taux d'emploi des femmes, il est particulièrement intéressant de remarquer qu'elle ne s'est accompagnée d'une augmentation du troisième terme de l'équation – la corrélation entre salaires des conjoints – suffisante pour inverser la tendance égalisatrice : la croissance de cette corrélation (actifs et inactifs confondus) de 0.16 en 1982 à 0.26 en

2013 a tout juste suffi à compenser les effets de l'affaiblissement de l'inégalité entre femmes⁴. Ce niveau de corrélation, légèrement supérieur à celui déjà observé par les études existantes sur la France (Frémeaux & Lefranc, 2015, p. 10), est nettement plus élevé que celui rapporté par plusieurs auteurs concernant les États-Unis. Dans ce pays, la corrélation était légèrement négative⁵ avant 1980, et s'établit autour de 0.1 dans les années récentes (Schwartz, 2010, p. 1540 ; Cancian & Reed, 1998, p. 76 ; Reed & Cancian, 2012, p. 10). On peut y voir un reflet du modèle français d'activité des femmes, dans lequel l'emploi à temps plein occupe une place plus importante que dans d'autres pays (Meron & Maruani, 2012).

De son côté, la corrélation entre salaires des conjoints au sein des couples composés de deux salariés (qui représentaient 48 % des couples hors indépendants et retraités en 1982 et 68 % en 2014) est plus élevée que dans l'ensemble des couples : elle se situe sur la période autour de 0.35-0.40 (figure III). Cet écart a aussi été noté aux États-Unis (Schwartz, 2010), avec ici encore une corrélation moins forte qu'en France. Mais on voit que cette corrélation reste d'ampleur assez modeste. Ainsi, la croissance de la corrélation entre salaires des conjoints ne se retrouve pas parmi les couples de deux salariés, pour lesquels elle a fluctué sans tendance nette depuis 1982 (figure III). Du fait de cette relative stabilité, l'inégalité de salaire entre couples de deux salariés a suivi assez fidèlement les évolutions des inégalités de salaire entre hommes et entre femmes vivant dans ces couples : on relève ainsi une augmentation de 19 % jusqu'en 2002, puis une diminution ramenant l'inégalité entre couples à son niveau de départ. Des mécanismes très différents ont donc abouti à des évolutions assez semblables que l'on considère l'ensemble des couples ou seulement ceux percevant deux salaires.

2. Ce mécanisme correspond en fait au phénomène classique de régression vers la moyenne (Verbakel, 2008, p. 132), suivant lequel un individu percevant un salaire très éloigné de la moyenne (très élevé ou très bas, voire nul) a peu de chances d'être en couple avec une personne qui touche un salaire aussi extrême.

3. L'augmentation des inégalités salariales entre hommes de 1982 à 1986, qui s'est produite uniquement au-dessus de la médiane, a déjà été relevée par d'autres travaux fondés sur des données administratives (Charnoz et al., 2013, p. 73 ; Verdugo, 2014, p. 135). En revanche, les études existantes relèvent plutôt une baisse des inégalités de salaire entre hommes depuis les années 1960 hors chômage et inactivité (Verdugo, 2014 ; Verdugo et al., 2012).

4. La réalité du passage soudain à une corrélation de 0.29 en 2014 demande à être confirmée par de futures enquêtes.

5. Du fait notamment du faible taux d'emploi des conjointes des hommes appartenant aux derniers déciles.

Un développement quasi uniforme de l'emploi des femmes à tous les niveaux de salaire des conjoints

L'effet de l'augmentation de la proportion de couples de deux salariés sur la corrélation entre salaires des conjoints parmi l'ensemble des couples dépend en grande partie du lien entre emploi de la femme et niveau de salaire de son conjoint (Pasqua, 2002). Ainsi, aux États-Unis, l'augmentation de la corrélation entre salaires des conjoints s'explique largement par le fait que la relation négative entre salaire de l'homme et emploi de sa conjointe a progressivement disparu, laissant place à une courbe en U inversé suivant laquelle les classes moyennes présentent les plus forts taux d'emploi des femmes (Schwartz, 2010, p. 1541).

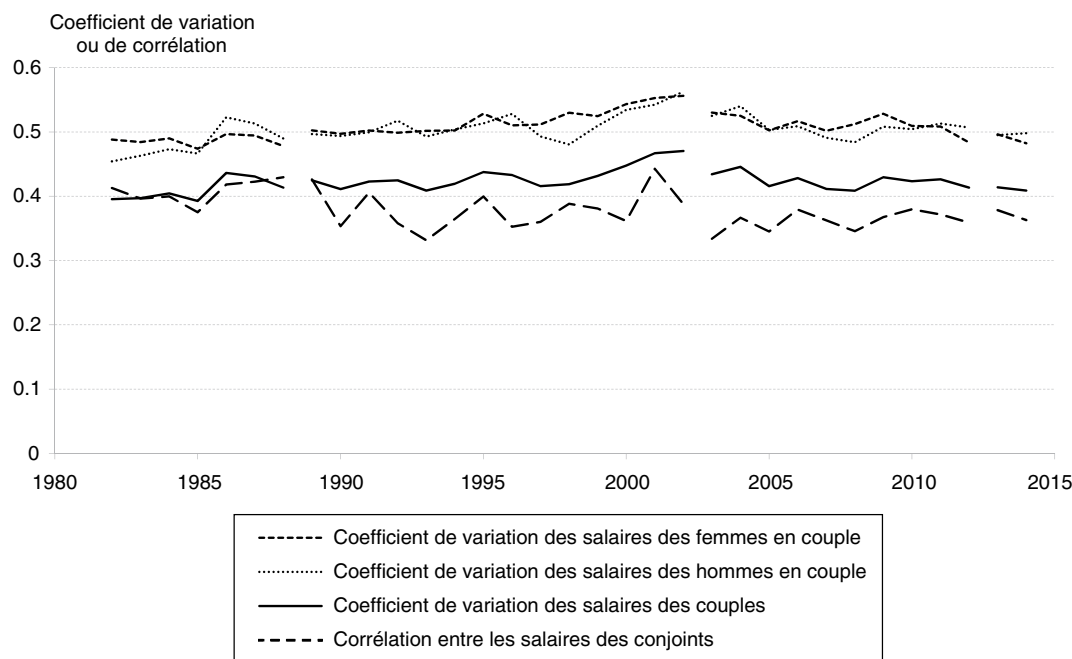
Mais le cas français apparaît relativement différent (figure IV, graphique de gauche). Dès 1982, le taux d'emploi des femmes le plus élevé (à près de 60 %) se rencontrait parmi les conjointes des hommes appartenant au septième décile de

salaire (confirmant les résultats de Frémeaux & Lefranc, 2015, p. 15). À l'opposé, son niveau était presque aussi faible pour le dernier décile (11 points de pourcentage en dessous du maximum) que pour le premier (13 points en dessous). Néanmoins, les conjointes des hommes ne percevant aucun salaire se distinguaient nettement par un taux d'emploi inférieur au maximum de 18 points.

Aux États-Unis à la même période, seuls les déciles supérieurs se distinguaient par un taux d'emploi des femmes clairement plus faible que les autres, les déciles inférieurs les ayant progressivement rejoints seulement après cette date. La France semble donc ici avoir connu plus précocement l'évolution observée aux États-Unis : en France, l'emploi des femmes avait déjà pour effet d'augmenter les salaires des couples du milieu de la distribution plus que ceux des couples situés à ses extrêmes.

En France, entre 1982 et 2014, le taux d'emploi des femmes a augmenté au même rythme quel que soit le décile du conjoint, à l'exception notable du décile inférieur et des inactifs et chômeurs, pour lesquels l'écart avec le taux d'emploi le plus élevé

Figure III
Évolution des inégalités de salaire et de la corrélation entre salaires des conjoints depuis 1982 au sein des couples composés de deux salariés



Lecture : le coefficient de variation des salaires des femmes en couple de deux salariés est resté relativement stable, passant de 0.49 à 0.48 entre 1982 et 2014, avec un pic à 0.56 en 2002.

Champ : hommes et femmes vivant en couple cohabitant dans lequel les conjoints ont entre 30 et 59 ans, touchent tous deux un salaire et aucun n'est indépendant ou retraité.

Source : Insee, enquêtes Emploi, 1982-2014.

s'est accru jusqu'à atteindre respectivement 18 et 28 points. Ainsi, les effets du développement de l'emploi des femmes sur la corrélation entre salaires des conjoints (en attribuant un salaire nul aux inactifs et chômeurs) sont restés limités et n'ont pas suffi à augmenter l'inégalité entre couples.

Le tableau est assez différent en ce qui concerne les hommes (figure IV, graphique de droite). Dans l'ensemble, on relève une légère baisse du taux d'emploi au cours du temps qui contraste avec la hausse observée pour les femmes. En outre, le taux d'emploi des hommes croît avec le salaire de leur conjointe jusqu'à la médiane, et se stabilise ensuite. Cependant, comme parmi les femmes, le groupe dont les conjointes ne touchent aucun salaire se détache assez nettement des autres. L'écart entre le taux d'emploi de ce groupe et le taux maximum s'est nettement renforcé au cours du temps (passant de 4 à 16 points), reflétant une certaine tendance à la polarisation de l'emploi entre couples (Ravel, 2007).

Ainsi, contrairement aux États-Unis, l'évolution du taux d'emploi des femmes n'a pas représenté en France une source importante d'évolution de la corrélation entre salaires des conjoints. Le caractère quasi général de l'accroissement

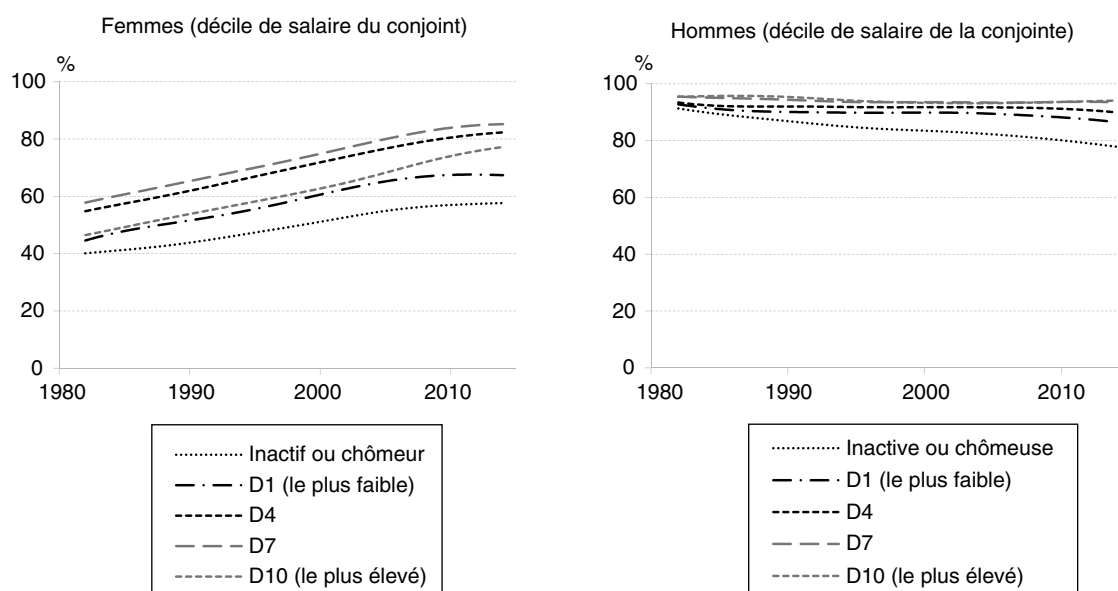
du taux d'emploi des femmes, quel que soit le salaire du conjoint, a permis de limiter l'augmentation de la corrélation entre salaires des conjoints, et ainsi d'éviter une augmentation de l'inégalité entre couples. En revanche, qu'il s'agisse des femmes ou des hommes, on relève un décrochage du taux d'emploi des conjoints des individus touchant les salaires les plus faibles, voire aucun salaire, dont l'effet est nécessairement inégalitaire. Dans la section suivante, une décomposition plus détaillée de l'inégalité nous permettra d'évaluer les effets de ces tendances en différents points de la distribution des salaires des couples.

Une diminution des inégalités qui n'a pas profité aux couples situés juste sous la médiane

Des évolutions divergentes selon la partie de la distribution considérée

La décomposition du coefficient de variation mise en œuvre dans la section précédente présente l'avantage d'une grande simplicité. En revanche, elle ne permet pas d'examiner si les

Figure IV
Évolution du taux d'emploi des hommes et des femmes selon le salaire de leur conjoint



Note : les chômeurs et inactifs sont inclus dans le champ et se voient attribuer un salaire nul. Les courbes sont lissées à l'aide d'une régression locale de degré 1 (LOESS).

Lecture : en 1985, 58 % des conjointes des hommes appartenant au 7^e décile de salaire de leur sexe étaient en emploi au sein des couples âgés de 30 à 59 ans.

Champ : hommes et femmes vivant en couple cohabitant dans lequel les conjoints ont entre 30 et 59 ans et aucun n'est indépendant ou retraité.

Source : Insee, enquêtes Emploi, 1982-2014.

effets mis en évidence ont été uniformes sur toute la distribution des salaires, ni d'évaluer l'évolution des inégalités qui aurait été observée dans d'autres situations contrefactuelles que celles déjà évoquées. Dans cette section, nous nous inspirons de la démarche élaborée par Schwartz (2010), qui consiste à modéliser la distribution jointe des salaires des deux conjoints à l'aide de modèles log-linéaires (Agresti, 2002). Ceux-ci permettent, en imposant des contraintes aux paramètres, de simuler plusieurs situations contrefactuelles, et ainsi d'évaluer la contribution de chacune des différentes tendances à l'évolution de l'inégalité entre couples.

Beaucoup plus souple que la précédente, cette décomposition ne se limite plus nécessairement au coefficient de variation. Afin d'étudier si les évolutions ont varié selon la partie de la distribution des salaires des couples considérée, nous mobilisons trois nouvelles mesures : le rapport entre le décile supérieur et la médiane ($D9/D5$), celui entre la médiane et le deuxième décile ($D5/D2$)⁶, ainsi que la proportion de couples ne percevant aucun salaire.

L'intérêt de cette approche plus détaillée apparaît dans la figure V à travers les évolutions très différentes des trois indicateurs. D'un côté,

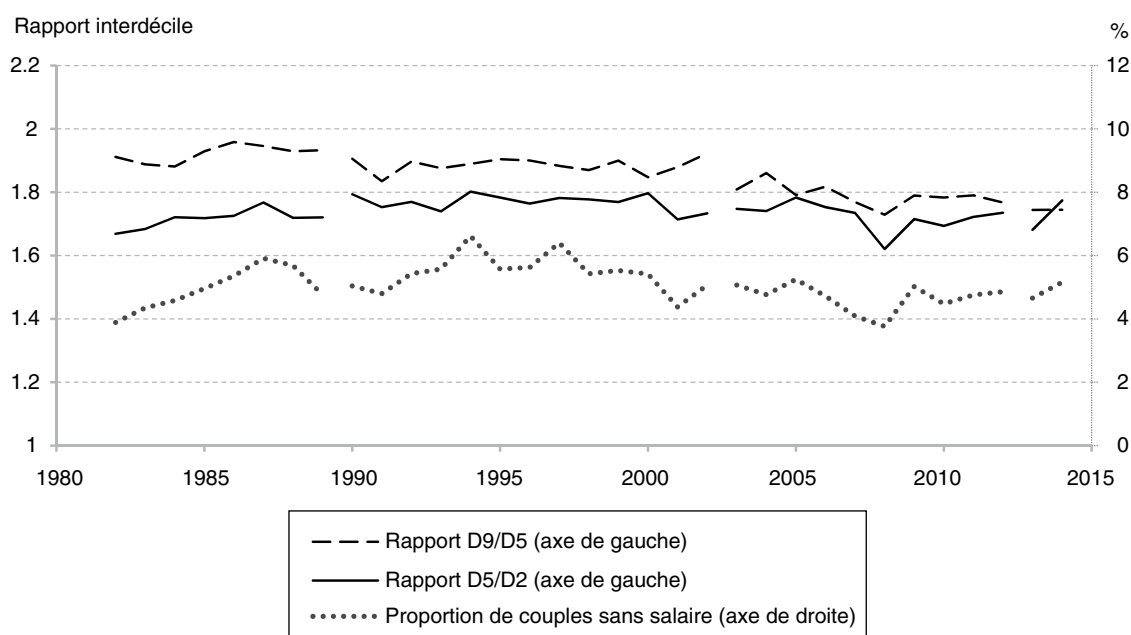
l'inégalité au-dessus de la médiane ($D9/D5$) a quelque peu augmenté dans les années 1980, avant de décroître plus nettement de 11 % entre 1986 et 2014. De l'autre, l'inégalité en dessous de la médiane ($D5/D2$) a augmenté de 8 % jusqu'en 1994 pour fluctuer ou diminuer légèrement depuis. Enfin, la proportion de couples ne percevant aucun salaire est passée de 3.9 % à 5.9 % entre 1982 et 1987, a connu un nouveau pic au milieu des années 1990 avant de diminuer quelque peu dans les années 2000 et de remonter après 2008.

Ainsi, la stabilité de l'inégalité entre couples relevée plus haut (figure II) masque des tendances divergentes. Si l'on relève d'un côté une diminution des écarts entre la médiane et les déciles supérieurs⁷, de l'autre, les couples situés sous la médiane ont vu leur situation relative

6. C'est souvent le premier décile (D1) qui est retenu plutôt que le deuxième. Ce choix poserait ici problème dans la mesure où, pour certaines années, près de 10 % des couples ne perçoivent aucun salaire : le premier décile suit des évolutions irrégulières liées aux variations des très bas salaires en fonction de la conjoncture économique et du taux chômage. Puisque l'inégalité dans le bas de la distribution est déjà mesurée par la proportion de couples ne percevant aucun salaire, nous avons jugé préférable de recourir au deuxième décile, dont les évolutions sont plus régulières.

7. Cette baisse s'observe pour tous les déciles au-dessus de la médiane.

Figure V
Évolution de trois indicateurs d'inégalité entre couples depuis 1982



Note : les chômeurs et inactifs sont inclus dans le champ et se voient attribuer un salaire nul.

Lecture : en 1982, le salaire séparant les 10 % des couples les plus riches des 90 % les plus pauvres était égal à 1.9 fois le salaire médian des couples.

Champ : hommes et femmes vivant en couple cohabitant dans lequel les conjoints ont entre 30 et 59 ans et aucun n'est indépendant ou retraité.

Source : Insee, enquêtes Emploi, 1982-2014.

se dégrader quelque peu, avant de finalement s'améliorer, mais de manière peu marquée. Nous chercherons dans ce qui suit à appréhender les facteurs à l'origine de ce phénomène.

La méthode de décomposition par la modélisation log-linéaire : six scénarios contrefactuels

La méthode de décomposition proposée par Schwartz (2010) consiste à découper la distribution des salaires de chaque sexe observée chaque année en quantiles relativement fins, plus un groupe pour les inactifs et les chômeurs, et à construire une table d'homogamie à trois dimensions croisant chaque année les quantiles d'appartenance de chacun des conjoints. Le coefficient

de variation et les rapports interdéciles peuvent être recalculés à partir de cette table, en attribuant à chaque individu le salaire médian de son quantile d'appartenance. Pour reproduire de manière assez fidèle les valeurs des indices mesurées à partir des données brutes, nous découpons ici la distribution en 20 quantiles représentant chacun 5 % des individus de chaque sexe ayant perçu un salaire, plus une catégorie comprenant ceux n'ayant perçu aucun salaire.

L'objectif de cette méthode est de modéliser, à partir de la table d'homogamie réellement observée, plusieurs tables, correspondant chacune à une situation contrefactuelle. Il s'agit de partir d'un modèle imposant des hypothèses très fortes sur l'association entre salaires des conjoints et son évolution, et de les relâcher peu à peu par l'introduction de termes supplémentaires jusqu'à arriver aux évolutions réellement observées : chaque modèle est une

Encadré 2

MODÈLES LOG-LINÉAIRES UTILISÉS POUR LA DÉCOMPOSITION

Le modèle de base (modèle 1), celui d'indépendance, inclut uniquement les paramètres nécessaires pour reconstituer correctement la proportion de couples dans chaque quantile de salaire des hommes et des femmes (dont les salaires nuls) et dans chaque année, mais sans interaction entre ces trois dimensions. En notant m_{hft} les effectifs prédits par le modèle pour la cellule à l'intersection de la ligne h (quantile de l'homme), de la colonne f (quantile de la femme) et de la couche t (année d'enquête), dans une table de dimensions $H \times F \times T$, ce modèle s'écrit :

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

Le modèle 2 tient compte en plus de l'évolution de la distribution marginale des femmes au cours du temps.

Le modèle 3 y ajoute celle des hommes. Ils s'écrivent :

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

Etant donné que les populations des deux sexes sont découpées chaque année en quantiles, les paramètres marginaux sont à ce stade de peu d'effet (à l'approximation liée au découpage près), à l'exception de ceux reflétant les proportions marginales d'inactifs et de chômeurs : c'est pourquoi ce modèle revient principalement à autoriser ces proportions à varier au cours du temps.

Le modèle 4 ajoute à la spécification du précédent l'association entre l'inactivité ou le chômage de la femme et le quantile de salaire de l'homme, en supposant qu'elle est restée stable dans le temps. Il s'écrit :

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

avec $f = 0$ indiquant l'absence de salaire de la femme (inactive ou chômeuse), et $1_{f=0}$ l'indicatrice correspondante.

Le modèle 5, celui de stabilité de l'association, inclut l'association complète mais stable entre salaires de l'homme et de la femme. Il intègre donc l'association entre inactivité ou chômage et salaire du conjoint du modèle précédent (qui n'apparaît cependant plus comme un terme spécifique). Il s'écrit :

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

Bien que ces cinq modèles supposent que l'association entre les salaires des conjoints est restée stable en termes d'odds ratios, les indices d'inégalité peuvent varier au cours du temps, puisqu'ils ne sont pas indépendants des marges de la table.

Le modèle 6 autorise en plus l'association entre l'inactivité ou le chômage de la femme et le quantile de salaire de l'homme à varier de manière linéaire dans le temps. Il s'écrit :

$$M6 : M5 + \left(\lambda^{F_0} + t \lambda^{F_0 T} \right) \lambda_h^{HF_0 T} 1_{f=0}$$

avec $f = 0$ indiquant l'absence de salaire de la femme.

Enfin, le dernier modèle correspond aux données réellement observées (modèle saturé). La comparaison avec le précédent permet de mesurer les effets de l'évolution de l'association entre salaires des conjoints dans les couples dans lesquels la femme est salariée sur les inégalités entre couples.

extension du précédent (voir encadré 2). En comparant les valeurs des indices d'inégalité obtenus pour chaque situation, on obtient une estimation de la contribution de la levée de chacune des hypothèses aux inégalités et à leur évolution temporelle.

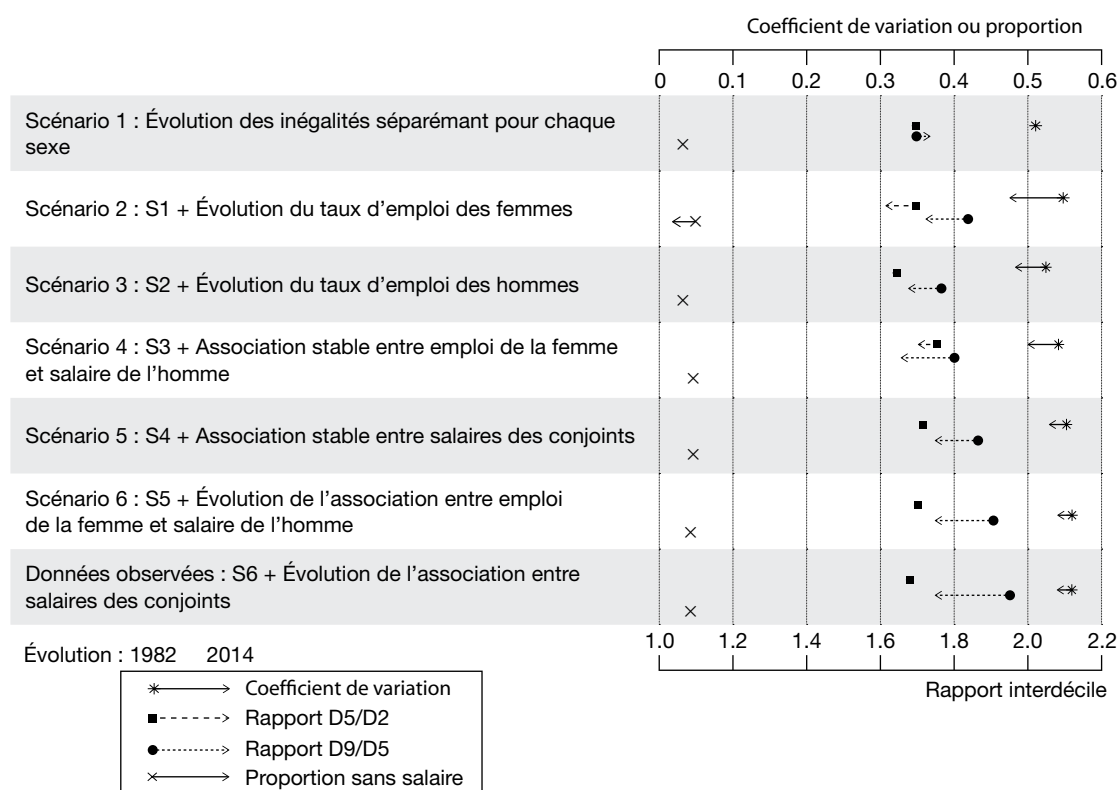
L'évolution des indicateurs entre 1982 et 2014 dans les différents scénarios contrefactuels est présentée dans la figure VI. On pourra retrouver les statistiques d'ajustement des modèles log-linéaires dans l'annexe. Soulignons simplement que chaque modèle plus complexe constitue une amélioration substantivement et statistiquement significative par rapport au précédent comme l'indique le critère d'Akaike (*Akaike Criterion Information, AIC*)⁸.

Le premier scénario (cf. modèle 1 dans l'encadré 2) suppose qu'il n'existe aucune

association entre les salaires des conjoints, ni entre le salaire d'un des conjoints et le fait que l'autre soit en emploi, et aussi que la proportion de personnes en emploi n'évolue pas dans le temps. De ce fait, le seul facteur d'inégalité et d'évolution des inégalités pris en compte ici correspond aux inégalités de salaire entre hommes en emploi d'un côté, et entre femmes en emploi de l'autre, dans l'hypothèse où les couples seraient constitués au hasard. On voit que ce facteur a très peu contribué aux évolutions temporelles des inégalités : les indicateurs sont stables (figure VI, première ligne). Le développement de l'emploi des femmes et l'évolution de l'association entre salaires des

8. En revanche, l'indicateur BIC (Bayesian Information Criterion), plus parcimonieux, n'incite pas à considérer comme dignes d'attention les apports respectifs des modèles 3 et 6, pour lesquels la proportion de couples classés dans la mauvaise cellule (indice de dissimilarité) diminue d'ailleurs très peu.

Figure VI
Évolution du coefficient de variation, des rapports D9/D5 et D5/D2 et de la proportion de couples ne percevant aucun salaire entre 1982 et 2014 dans les différentes situations contrefactuelles



Notes : les chômeurs et inactifs sont inclus dans le champ et se voient attribuer un salaire nul. Afin de ne pas tenir compte des variations liées à l'échantillonnage aléatoire ainsi qu'à la conjoncture économique, les flèches ne représentent pas directement les indices calculés à partir des effectifs prédits par les modèles, mais leur valeur lissée à l'aide d'une régression locale de degré 2 (LOESS).

Lecture : si seules les inégalités de salaire entre hommes et entre femmes avaient varié (première ligne), le niveau global des inégalités (coefficient de variation) serait resté stable (à 0.50) entre 1982 et 2014.

Champ : hommes et femmes vivant en couple cohabitant dans lequel les conjoints ont entre 30 et 59 ans et aucun n'est indépendant ou retraité.

Source : Insee, enquêtes Emploi, 1982-2014.

conjointes expliquent donc la plus grande part des évolutions.

Dans le second scénario, on suppose toujours qu'il n'existe aucun lien entre le niveau de salaire d'un conjoint et le statut de l'autre, mais on laisse la proportion de personnes en emploi évoluer dans le temps. Cette spécification permet d'évaluer la contribution de l'augmentation du taux d'emploi des femmes à l'évolution des inégalités entre couples. L'évolution de ce seul facteur entraîne une diminution très marquée des inégalités entre 1982 et 2014, qui fait apparaître très clairement le rôle égalisateur de l'emploi des femmes.

Ainsi, les inégalités (coefficient de variation) auraient baissé de 14 % dans leur ensemble, de 5 % en dessous de la médiane (rapport D5/D2) et de 8 % au-dessus de la médiane (rapport D9/D5), alors que la proportion de couples ne percevant aucun salaire serait passée de 4.9 % à 2.3 %.

Le troisième scénario ajoute au précédent l'évolution temporelle du taux d'emploi des hommes. La légère diminution du taux d'emploi des hommes a atténué l'affaiblissement des inégalités prédit par le scénario précédent (cf. figure IV). En particulier, l'inégalité d'ensemble (coefficient de variation) aurait baissé de 7 % seulement (contre 14 % dans le deuxième scénario), et l'inégalité en dessous de la médiane et la proportion de couples ne percevant aucun salaire seraient restées stables. On voit donc que le développement de l'emploi des femmes a permis de compenser les effets de la tendance inégalitaire à l'œuvre parmi les hommes.

Le quatrième scénario repose sur l'hypothèse que la seule association entre salaires des conjoints vient du lien entre emploi de la femme et salaire de son conjoint, tout en supposant que ce lien est resté stable au cours du temps. Il permet de mesurer l'effet des différences de taux d'emploi des femmes selon le salaire de leur conjoint sur les inégalités et leur évolution. Ce scénario ne modifie pas de manière majeure les évolutions temporelles prédites par le précédent. Cependant, les inégalités en dessous de la médiane (rapport D5/D2) et d'ensemble (coefficient de variation) y sont plus élevées, en 1982 comme en 2014. Ceci tient à la surreprésentation des femmes ne percevant aucun salaire parmi les couples dans lesquels les hommes sont situés dans le bas de la distribution des salaires.

Le cinquième scénario étend le précédent en supposant qu'il existe une association entre salaires des conjoints (y compris l'absence de salaire), mais que celle-ci est restée stable au cours du temps. Contrairement à ce que l'on pourrait attendre, ce modèle (ainsi que le précédent) prédit une évolution des inégalités potentiellement différente du modèle 3. En effet, même si l'association elle-même reste stable dans le temps, l'évolution de la répartition des hommes et des femmes entre emploi et non-emploi déplace les effectifs entre des zones de la table d'homogamie qui diffèrent en termes d'intensité de l'association entre salaires des conjoints. Ainsi, la comparaison de l'inégalité prédite par ce modèle et par le précédent permet d'évaluer l'effet de l'existence d'une association entre salaires des conjoints sur l'évolution des inégalités, même en supposant que cette association est restée stable. Ce scénario prédit des inégalités plus fortes que le précédent au-dessus de la médiane (rapport D9/D5) ainsi que dans l'ensemble (coefficient de variation), du fait de la forte tendance des hommes et des femmes percevant les plus hauts salaires à former des couples ensemble. L'écart est cependant modeste : par rapport au quatrième scénario, le coefficient de variation est supérieur de 5 % en 1982, et de 9 % en 2014.

De leur côté, les tendances temporelles demeurent essentiellement les mêmes que dans le scénario précédent. L'affaiblissement des inégalités d'ensemble (coefficient de variation) et en dessous de la médiane (rapport D5/D2) est légèrement moins marqué, puisque les femmes qui entrent sur le marché du travail perçoivent un salaire plus proche de celui de leur conjoint que dans l'hypothèse d'une absence totale d'association. Mais finalement, l'existence d'une tendance à l'homogamie n'a en soi qu'un faible effet sur les évolutions temporelles des inégalités.

Le sixième scénario introduit une première source d'évolution temporelle de l'association entre salaires des conjoints : celle du lien entre emploi de la femme et salaire de l'homme, l'association entre salaires des conjoints restant supposée stable. Comme on pouvait s'y attendre du fait de la relative stabilité du lien entre salaire de l'homme et emploi de la femme déjà relevée plus haut, ce scénario n'introduit aucune différence notable par rapport au précédent.

Enfin, le dernier scénario correspond à la table réellement observée (modèle saturé). La différence par rapport au scénario précédent

provient donc exclusivement de l'évolution au cours du temps de l'association entre salaires des conjoints parmi les couples dans lesquels la femme est salariée. Cette comparaison fait apparaître un résultat négatif qui mérite d'être souligné : contrairement à ce qui a été relevé aux États-Unis, l'évolution de l'association entre les salaires des conjoints en emploi (conséquence de l'homogamie sociale) n'a pas eu d'effet notable sur l'inégalité entre couples. On peut cependant relever que la diminution de l'inégalité au-dessus de la médiane (rapport D9/D5) est légèrement plus marquée, signe d'un affaiblissement de l'association dans le haut de la distribution.

La décomposition plus fine réalisée dans cette section a montré que l'augmentation du taux d'emploi des femmes est de loin le facteur principal de l'évolution de l'inégalité de salaire entre couples (dont aucun des conjoints n'est indépendant ni retraité) depuis 1982. La légère diminution du taux d'emploi des hommes a en revanche renforcé les inégalités, mais de manière peu marquée. Enfin, les évolutions des différentes composantes du lien entre salaires des conjoints ont joué un rôle relativement mineur.

Si le développement de l'emploi des femmes a eu un effet égalisateur sur l'ensemble de la distribution, le léger recul de l'emploi des hommes dont les conjoints perçoivent un salaire faible ou nul a eu un effet inégalitaire particulièrement marqué en dessous de la médiane. En outre, l'association entre les salaires des conjoints semble s'être quelque peu affaiblie dans le haut de la distribution. La combinaison de ces trois phénomènes explique les évolutions contrastées des différents indicateurs d'inégalités relevées plus haut : diminution des inégalités entre couples au-dessus de la médiane, stabilité ailleurs.

* *
*

Les résultats présentés ici, qui vont à l'encontre de l'idée selon laquelle on assisterait à une

augmentation généralisée à la fois de l'homogamie et des inégalités salariales, suggèrent des prolongements dans trois directions.

Premièrement, les inégalités sont ici mesurées uniquement à partir des salaires. Or l'augmentation des inégalités économiques provient en grande partie de l'évolution de la distribution du patrimoine et des revenus qui en découlent (Landais, 2007 ; Piketty, 2013). L'extension de l'étude à tous les types de revenus ainsi qu'aux héritages semble une voie prometteuse, même si les données permettant de les étudier sont plus rares (Frémeaux, 2014). Par ailleurs, la prise en compte des revenus de transfert – et notamment des allocations chômage – apparaît essentielle pour affiner l'analyse des bas revenus, appréhendés ici uniquement à travers la proportion de couples ne percevant aucun salaire.

Deuxièmement, nous avons retenu une perspective purement transversale, négligeant de ce fait les variations importantes des revenus des conjoints au cours de leur vie, notamment du fait des périodes de chômage ou d'inactivité (Courtioux & Lignon, 2015a). Les recherches futures devront s'attacher à combiner une étude des évolutions temporelles des inégalités avec une perspective longitudinale au niveau individuel de manière à mieux tenir compte des possibles compensations entre conjoints des variations du niveau d'activité (inactivité, chômage, temps de travail).

Troisièmement, notre analyse a porté exclusivement sur les individus en couple une année donnée. Or, suivant ce qui a été relevé aux États-Unis (Karoly & Burtless, 1995 ; Burtless, 1999 ; Cancian & Reed, 1999 ; Reed & Cancian, 2001 ; Daly & Valletta, 2006 ; Western et al., 2008 ; Larrimore, 2014), on peut penser que la baisse de la proportion de personnes vivant en couple a renforcé les inégalités parmi l'ensemble des ménages. En particulier, les familles monoparentales, qui ont plus fréquemment à leur tête des mères peu diplômées et éloignées de l'emploi (Chardon et al., 2008), ont pu voir leur situation relative se détériorer par rapport au salaire médian des couples. □

BIBLIOGRAPHIE

- 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* (2^e éd.). New York: Wiley.
- Albouy, V., Djider, Z. & Mainguéné, 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.
- 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.
- 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* (pp. 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.
<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.
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.
- OCDE (2011).** *Toujours plus d'inégalité : Pourquoi les écarts de revenus se creusent*. 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.
- 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 socio-professionnelle 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*. Radboud University, Nijmegen.
- Verdugo, G. (2014).** The great compression of the French wage structure, 1969–2008. *Labour Economics*, 28(4), 131–144.
<http://hdl.handle.net/2066/73433>
- Verdugo, G., Fraise, 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>

ANNEXE

STATISTIQUES D'AJUSTEMENT DES MODÈLES UTILISÉS POUR LA DÉCOMPOSITION

	D. L. ⁽¹⁾	Déviance	Δ (%) ⁽²⁾	BIC ⁽³⁾	AIC ⁽⁴⁾
M1 : Indépendance complète	14 480	57 114	19.86	- 120 844	28 154
M2 : M1 + Taux d'emploi des femmes	13 840	49 073	17.57	- 121 019	21 393
M3 : M2 + Taux d'emploi des hommes	13 200	46 593	17.08	- 115 634	20 193
M4 : M3 + Association emploi-salaire stable	13 180	40 948	14.56	- 121 033	14 588
M5 : M4 + Association complète stable	12 800	16 091	9.51	- 141 219	- 9 509
M6 : M5 + Association emploi-salaire variable	12 780	15 886	9.33	- 141 179	- 9 674

⁽¹⁾ Degrés de liberté. ⁽²⁾ Indice de dissimilarité. ⁽³⁾ Critère d'information bayésien. ⁽⁴⁾ Critère d'information d'Akaike. Ces deux derniers indicateurs mesurent la qualité de la description des données réalisée par chaque modèle en tenant compte de leur parcimonie (nombre de paramètres à estimer) : une valeur inférieure (ou plus fortement négative) signale une amélioration statistiquement significative par rapport à un modèle présentant une valeur plus élevée.

Note : N = 217 489. Nombre de cellules de la table : 14 553.

Champ : hommes et femmes vivant en couple cohabitant dans lequel les conjoints ont entre 30 et 59 ans et aucun n'est indépendant ou retraité.

Source : Insee, *enquêtes Emploi*, 1982-2014.

Combien de temps durent les situations de monoparentalité ?

Une estimation sur données françaises

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

Vianney Costemalle *

Résumé – Les familles monoparentales représentent aujourd’hui plus de 20 % des familles avec enfants mineurs en France comme en moyenne en Europe. La monoparentalité est généralement associée à de plus grands risques de précarité et d’exclusion auxquels répondent différentes politiques sociales. Il est ainsi important de savoir combien de temps dure cette situation. On présente dans cet article une méthode originale d’estimation de cette durée à partir d’un échantillon de familles monoparentales pour lesquelles seules les anciennetés dans la situation à la date de l’enquête sont observées (échantillonnage dans le stock). Elle combine un calcul de la fonction de vraisemblance des observations selon la méthodologie proposée par Nickell avec l’introduction de risques instantanés de sortie de la situation proportionnels comme dans le modèle de Cox. Plusieurs simulations reproduisant des cas réels et variés confirment la fiabilité de cette méthode. Appliquée aux données de l’*Enquête famille et logements* de 2011, elle permet d’estimer que la moitié des parents de famille monoparentale sortent de cette situation au bout de trois ans.

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

Mots-clés : monoparentalité, modèle de durée, estimation durées

Keywords : single parenthood, survival models, duration models

Rappel :

Les jugements et opinions exprimés par les auteurs n’engagent qu’eux mêmes, et non les institutions auxquelles ils appartiennent, ni a fortiori l’Insee.

* Insee, Division Enquêtes et études démographiques (vianney.costemalle@insee.fr).

La monoparentalité, c'est-à-dire la situation d'un parent – le plus souvent une mère – seul avec un ou des enfants à charge, est une situation familiale dont la fréquence s'est élevée depuis plusieurs décennies¹, en France comme à l'étranger. En France, selon le recensement de la population de l'Insee², il y avait 953 000 familles monoparentales en 1990 et 1 687 000 en 2010, ce qui correspond à une progression de 77 % en 20 ans. Parallèlement, le nombre de familles avec enfants mineurs n'a que peu évolué (7 944 000 en 2010, soit 3.8 % de plus qu'en 1990) ; la part des familles monoparentales parmi l'ensemble des familles avec enfants mineurs a ainsi fortement progressé, passant de 12 % en 1990 à 21 % en 2010, dans un mouvement de hausse constante depuis plusieurs dizaines d'années. En Europe, cette tendance a été observée en Grande-Bretagne à partir des années 1970 (David et al., 2004) avant de s'étendre à l'ensemble des pays dans les années 1980. En 2012, la part des familles monoparentales en Europe atteignait 19 %. Les contrastes sont marqués entre pays d'Europe du Nord et pays d'Europe de l'Est et du Sud ; les formes traditionnelles de la famille sont encore bien ancrées dans ces derniers même si la monoparentalité y a fortement progressé (Le Pape et al., 2015).

Plus souvent que d'autres situations familiales, la monoparentalité est associée à des situations difficiles : les familles monoparentales représentent la composition familiale dans laquelle le taux de pauvreté est le plus élevé, atteignant 33 % en France (Boiron et al., 2016), et elles connaissent plus souvent des conditions de logement défavorisées (Chardon et al., 2008) ; les mères de familles monoparentales, bien que leur taux d'activité soit un peu plus élevé que celui des mères en couple, ont un taux de chômage deux fois plus élevé (Rabier, 2014) et font face à des difficultés particulières pour articuler travail et contrainte familiale (Algava et al., 2005). Pour répondre à ces risques d'exclusion professionnelle et sociale, différentes politiques sociales ont été mises en œuvre en Europe dès les années 1950 (Eydoux & Letablier, 2009). En France, l'allocation de soutien familial (ASF) et le complément familial (CF) s'ajoutent ainsi aux divers dispositifs dont bénéficient les familles. Il est donc important de savoir combien de temps durent les épisodes de monoparentalité.

La monoparentalité est une situation forcément transitoire, car elle prend fin soit avec la (re)mise en couple du parent seul, soit lorsque les enfants atteignent un âge auquel ils ne sont plus considérés comme dépendants ou quittent

le foyer. Mais la durée des situations de monoparentalité reste à ce jour assez mal connue. L'objectif de cet article est de fournir une estimation de la durée de la monoparentalité, ce qui, à notre connaissance, n'a jamais été fait auparavant pour la France. Pour cela on propose une méthode originale d'estimation de durées à partir des anciennetés, basée sur des travaux anciens de modélisation de durées, notamment les travaux de Nickell (1979) et de Cox (1972).

Avant d'aller plus loin, il faut noter qu'il n'existe pas de définition unique de la monoparentalité. Toutes les approches considèrent néanmoins qu'il s'agit d'un parent isolé élevant des enfants à charge, mais plusieurs critères peuvent être retenus pour chacune de ces deux dimensions. Longtemps, le critère d'isolement fut basé sur le statut matrimonial légal du parent : étaient considérés comme parents isolés les femmes (ou les hommes) non mariés avec des enfants. Cette approche pose maintenant problème puisque de plus en plus de couples ne sont pas mariés. En outre, dans certains couples, les partenaires ne vivent pas dans un même logement, faisant alors partie de ce qu'on dénomme les LAT (*living apart together*). Par exemple, 10 % des parents de familles monoparentales en France déclaraient, en 1999, vivre en couple avec quelqu'un qui n'habite pas dans le même logement (Algava, 2002). Le statut marital n'est donc plus un bon indicateur d'isolement et la notion de « vie en couple » est devenue plus difficile à saisir (Toulemon, 2011). Par ailleurs, il y a aussi plusieurs possibilités pour définir un enfant à charge : le plus souvent, on retient les enfants de moins de 25 ans (par exemple Algava, 2002) ou de moins de 18 ans (Buisson et al., 2005) mais certains travaux ne retiennent pas de limite d'âge (David et al., 2004). L'Insee analyse en général les types de familles en considérant les enfants mineurs ; le Haut Conseil de la famille s'intéresse aussi aux familles avec enfants mineurs (Haut Conseil de la famille, 2014). Nous retiendrons également cette limite d'âge de 18 ans, et nous définissons pour cet article la monoparentalité comme la situation d'un parent ne vivant pas en couple ayant au moins un enfant célibataire âgé de moins de 18 ans vivant dans le même logement, qui n'a pas lui-même d'enfants vivant dans le logement.

La première partie de l'article présente la méthode d'estimation. Dans un premier temps on explique en quoi les anciennetés observées sur

1. Les familles monoparentales apparaissent comme une catégorie statistique en France avec le recensement de 1982.

2. Il s'agit des familles monoparentales avec enfant(s) de moins de 18 ans.

les personnes présentes dans le stock de familles monoparentales sont différentes des durées : les anciennetés peuvent être vues comme des durées doublement biaisées. On détaille ensuite comment il est possible d'inférer la distribution des durées à partir de celle des anciennetés à partir de la connaissance des flux d'entrée. On s'appuie ici sur des simulations numériques, générant des variables aléatoires selon le modèle présenté, pour illustrer différents cas particuliers et pour tester la robustesse de la méthode proposée. Dans la seconde partie, cette méthode est appliquée aux données de l'*Enquête famille et logements (EFL)* en utilisant de plus les données de l'enquête *Étude des relations familiales et intergénérationnelles (Erfi)* (voir l'encadré 1 pour une présentation de ces sources) ; on expose alors l'estimation de la distribution des durées passées en situation de monoparentalité.

Comment déterminer les durées à partir des anciennetés ?

La durée des situations de monoparentalité est mal connue car difficile à mesurer. Les périodes

de monoparentalité se déduisent en effet de la confrontation entre les périodes où les personnes vivent sans conjoint et les périodes où elles ont des enfants à charge. Connaître les dates de début et de fin de ces périodes requiert de disposer d'informations longitudinales (ou rétrospectives) et de telles données sont assez rarement disponibles. Des durées moyennes ont pu être ainsi mesurées aux États-Unis sur la base de données rétrospectives à partir de l'enquête *National Survey of Families and Households* de 1987 : en moyenne, les mères entrées en famille monoparentale entre 1970 et 1974 sont restées 4.5 ans dans cette situation, contre 3.4 ans pour celles entrées 10 ans plus tard (Bumpass & Raley, 1995). Au Royaume-Uni, l'enquête rétrospective *Survey of Family and Working Lives* de 1994 a permis d'estimer des durées médianes de monoparentalité : 5.8 ans pour l'ensemble des mères, 4.6 ans pour les mères célibataires, 4.7 ans pour les mères divorcées, 6.8 ans pour les mères séparées et 10.5 ans pour les mères veuves (McKay, 2002).

Les autres enquêtes plus classiques qui apportent des informations sur les familles monoparentales sont les enquêtes transversales,

Encadré 1

L'ENQUÊTE FAMILLE ET LOGEMENTS ET L'ENQUÊTE ÉTUDE DES RELATIONS FAMILIALES ET INTERGÉNÉRATIONNELLES

L'*Enquête famille et logements (EFL)* a été réalisée par l'Insee au début de l'année 2011 conjointement avec le recensement de la population. Dans chaque ménage sélectionné pour l'enquête, tous les hommes ou toutes les femmes de 18 ans ou plus ont été interrogés. Au total, 359 770 personnes de 18 ans ou plus, vivant en ménage ordinaire en France métropolitaine ont été interrogées sur leur vie familiale et résidentielle (Breuil et al., 2016). Cette enquête permet d'identifier les adultes à la tête d'une famille monoparentale au 1^{er} janvier 2011. Il s'agit donc d'un échantillonnage dans le stock. Les personnes qui déclaraient ne pas vivre en couple au moment de l'enquête, et qui avaient déjà vécu en couple auparavant, devaient indiquer l'année de la fin de l'union ainsi que la cause (séparation ou décès du conjoint). Cela permet notamment de déterminer l'ancienneté de la monoparentalité au moment de l'enquête ainsi que la cause (rupture d'union, enfant hors vie en couple, décès du conjoint) de la monoparentalité.

L'enquête *Étude des relations familiales et intergénérationnelles (Erfi)* est le volet français du programme international *Generations and Gender Survey (GGG)* visant à établir des statistiques comparatives par pays au niveau mondial, et principalement au niveau européen (Régnier-Loilier, 2012). Elle a été menée

conjointement par l'Ined et l'Insee en trois vagues successives, en 2005, 2008 et 2011. Lors de la première vague il y avait 10 079 répondants, puis 6 534 pour la deuxième vague et enfin 5 781 pour la dernière vague (certains ont répondu à la première et troisième vague sans répondre à la deuxième). Les répondants sont âgés de 18 à 79 ans au 31 décembre 2005, une seule personne par ménage étant tirée au sort. L'intérêt de cette enquête est qu'elle pose des questions à caractère rétrospectif, ce qui permet d'avoir des informations longitudinales sur les parcours familiaux et conjugaux. Les personnes interrogées décrivent l'ensemble des relations cohabitantes (dont la durée a été d'au moins trois mois) qu'elles ont vécues au cours de leur vie. Elles donnent aussi des informations sur tous les enfants qu'elles ont eus, avec notamment une information sur la date de départ du foyer de chaque enfant. On a pu, à partir de ces éléments, déterminer les périodes pendant lesquelles les personnes interrogées étaient à la tête d'une famille monoparentale avec enfants mineurs. Les inconvénients de cette enquête sont : une taille de l'échantillon des personnes ayant vécu au moins une fois en famille monoparentale assez faible et de possibles imprécisions sur les dates données, – les questions portant sur un passé parfois lointain – ce qui conduirait à une imprécision sur les épisodes de monoparentalité.

qui permettent de connaître le stock de ces familles à un instant donné. On peut sur cette base mesurer l'ancienneté – mais pas la durée – des situations de monoparentalité. En France, l'enquête *Étude de l'histoire familiale* avait ainsi permis de mesurer pour 1999 une ancienneté moyenne de la situation de 6 ans et 3 mois pour les femmes et 5 ans et 9 mois pour les hommes (Algava, 2002). L'*Enquête famille et logements* de 2011 a permis de mesurer que les 1.5 million de familles monoparentales de France métropolitaine s'étaient formées 5.5 années plus tôt en moyenne : 4.5 ans pour un parents séparé, 5.5 ans pour un parent dont le conjoint est décédé et 10 ans pour un parent n'ayant jamais vécu en couple (Buisson et al., 2015).

Les deux types d'approche, longitudinale et transversale, renvoient respectivement à ce qu'on appelle dans les modèles de durée l'échantillonnage dans le flux (*flow sampling*) et l'échantillonnage dans le stock (*stock sampling*).

Des anciennetés aux durées : le biais de censure et le biais de sélection

L'ancienneté d'une situation est le temps qui s'est écoulé entre le début de cette situation et le moment où la situation est observée³, tandis que la durée est le temps total écoulé entre le début et la fin de la situation (voir l'encadré 2 pour la modélisation des durées). Les anciennetés et les durées sont donc *a priori* deux concepts différents qui répondent respectivement aux questions « Depuis combien de temps est-on dans cette situation, à la date t ? » et « Combien de temps cette situation dure-t-elle ? ». L'ancienneté peut être vue comme une durée qui est censurée sur la droite, puisqu'on ne connaît pas la date de fin de la situation, mais seulement son commencement. Les anciennetés sont par conséquent plus courtes que les durées, ce

3. Le moment de l'observation sera, dans notre cas d'application, identique pour tous les individus : il s'agit de la date de l'enquête.

Encadré 2

MODÉLISATION DES DURÉES

Dans le cas où la variable aléatoire de durée T est discrète et prend des valeurs entières, la loi de T est donnée par sa densité $f(t) = P(T = t)$.

La survie au temps $t \in \mathbb{N}$, notée $S(t)$, correspond à la proportion de la population dont la situation d'intérêt a duré t unité de temps ou plus : $S(t) = P(T \geq t)$. Il en résulte que S est une fonction décroissante et que $S(0) = 1$.

Le risque instantané au temps $t \in \mathbb{N}$, noté $h(t)$, correspond à la part de personnes qui quittent la situation au temps t parmi les personnes qui étaient toujours dans la situation au temps t : $h(t) = P(T = t | T \geq t)$. Soit $f(t) = h(t)S(t)$.

On a alors une relation entre la survie et le risque instantané : pour tout $t > 0$, $S(t) = \prod_{u=0}^{t-1} (1 - h(u))$.

En remarquant que $f(t) = S(t) - S(t+1)$, on a déduit que l'espérance de la variable de durée est donnée par $E[T] = \sum_{t \geq 0} t f(t) = \sum_{t \geq 1} S(t)$.

La médiane est ici définie comme $Med(T) = \frac{u(1 - S(u) - S(u+1)) + 0,5 - S(u)}{S(u) - S(u+1)}$, où u est le plus grand entier tel que $S(u) \geq 0,5$. Ce qui donne bien $Med(T) = u$ lorsque $S(u) = 0,5$.

Connaître la loi de probabilité, la fonction de survie ou la fonction de risque instantané revient au même et donne toute l'information sur la distribution des durées. Il suffit donc de connaître une de ces trois fonctions pour calculer n'importe quel indicateur, comme la moyenne ou la médiane.

Dans le cas continu, on utilise la loi de Weibull pour simuler des variables de durées. Dans ce cas, on peut définir de la même façon que dans le cas discret la fonction de survie et la fonction de risque instantané.

On a alors pour tout $t \in \mathbb{R}$, $S(t) = \exp\left(-\int_0^t h(u) du\right)$.

La loi de Weibull est paramétrée par deux réels positifs : un paramètre d'échelle λ et un paramètre de forme k .

La survie s'écrit $S(t) = \exp\left(-\left(\frac{t}{\lambda}\right)^k\right)$ et le risque instantané $h(t) = \frac{k}{\lambda} \left(\frac{t}{\lambda}\right)^{k-1}$.

Lorsque le paramètre de forme k vaut 1, alors on obtient une loi exponentielle : la survie est exponentiellement décroissante et le risque instantané est constant et vaut $1/\lambda$. Si k est plus petit que 1, le risque instantané est décroissant tandis que s'il est plus grand que 1 le risque instantané augmente avec le temps t . De plus, la variance augmente lorsque k diminue.

qu'on nomme ici le biais de censure. Un autre problème apparaît dans le cas d'un échantillonnage dans le stock : la probabilité d'être enquêté alors qu'on est en situation monoparentale augmente avec la durée de la monoparentalité. Ceci implique que les personnes en situation monoparentale au moment de l'enquête ont des durées en moyenne plus grandes que l'ensemble des personnes ayant connu un épisode de monoparentalité au cours de leur vie⁴. On nomme cela le biais de sélection. Ce biais joue en sens contraire du biais de censure, ce qui fait qu'il est *a priori* impossible de dire si les anciennetés sont en moyenne plus longues ou plus courtes que les durées. Quel est le lien entre la distribution des durées et celle des anciennetés ? Afin de bien comprendre comment peuvent jouer ces effets de censure et de sélection, on va par la suite générer de façon aléatoire plusieurs jeux de données mimant différentes situations réelles. On pourra alors ajuster les paramètres de ces simulations à volonté pour mettre en évidence les deux biais dans des cas choisis.

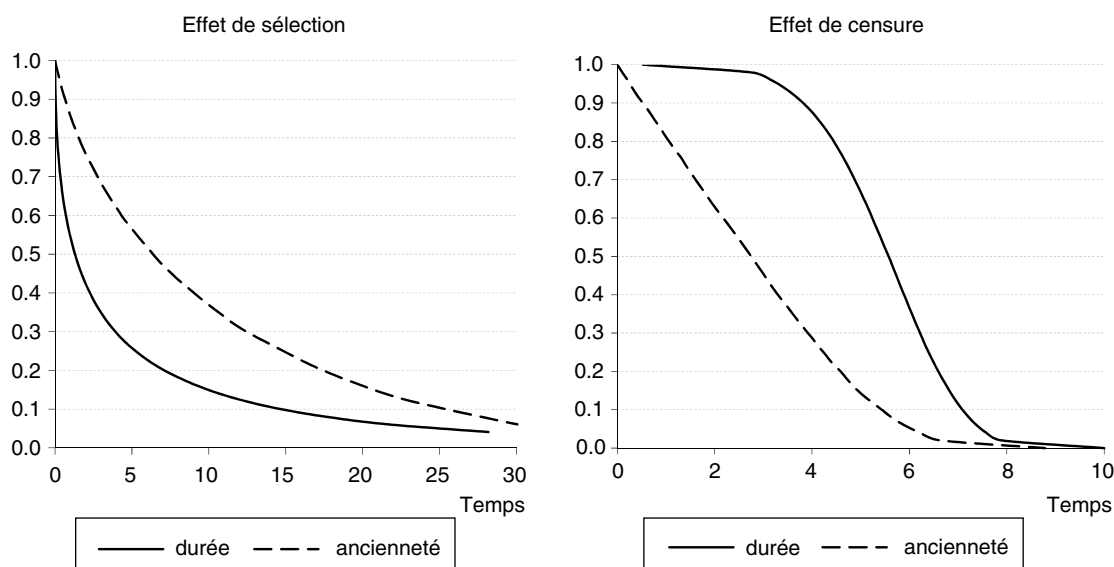
On illustre ces phénomènes de censure et de sélection à l'aide de simulations qui génèrent de manière aléatoire des observations de durées et d'anciennetés. On génère aléatoirement, selon une loi uniforme, une date d'entrée en monoparentalité, entre 1950 et 2010, pour 100 000 personnes. On génère de plus pour chacune de ces personnes la variable aléatoire de durée selon

une loi de Weibull (cf. encadré 2). On en déduit alors un stock de personnes qui sont toujours en situation de monoparentalité en 2011. Parmi les personnes de ce stock, on peut calculer l'ancienneté de la monoparentalité en 2011. Le paramètre de forme de la loi de Weibull est le paramètre clé : s'il est plus petit que 1, alors c'est l'effet de sélection qui l'emporte tandis que s'il est plus grand que 1 c'est l'effet de censure qui est le plus fort (figure I).

Ce résultat peut se comprendre intuitivement. Dans le cas où le paramètre de forme est inférieur à 1, les durées de la plupart des personnes sont très courtes et seulement quelques-unes d'entre elles ont des durées très longues (il y a une très grande hétérogénéité des durées). Il en résulte qu'au moment de l'enquête, toutes les personnes avec des durées très longues sont en famille monoparentale, tandis que seulement une fraction de celles qui ont des durées très courtes le sont (les personnes qui sont entrées en monoparentalité juste avant l'enquête). Les personnes avec des durées longues sont donc surreprésentées parmi les familles monoparentales au moment de l'enquête. En conséquence, les anciennetés observées sont globalement plus longues que les durées réelles.

4. Pour bien comprendre cela on peut aussi penser au jeu de la bataille navale : en choisissant une case au hasard, les gros bateaux ont plus de chance d'être touchés que les petits.

Figure I
Courbes de survie associées à la distribution des durées et à la distribution des anciennetés illustrant l'effet de sélection et l'effet de censure à partir de simulations



Note : données simulées. Pour la figure de gauche, la durée est générée selon une loi de Weibull(0.5 , 2.75) et pour la figure de droite selon une loi de Weibull(5 , 6). La date d'entrée en monoparentalité est générée selon une loi uniforme.

Dans le cas contraire, si le paramètre de forme est supérieur à 1, il y a peu de variabilité des durées : la plupart des personnes ont la même durée à quelques variations près. Celles qui sont en situation monoparentale au moment de l'enquête sont donc à peu près représentatives de l'ensemble des personnes (car il y a très peu d'hétérogénéité en termes de durées). Les anciennetés étant toujours plus courtes que les durées, il en résulte que globalement, les anciennetés des personnes en situation monoparentale au moment de l'enquête seront plus courtes que l'ensemble des durées.

Si l'effet de sélection est plus fort que l'effet de censure, les anciennetés observées seront en moyenne plus grandes que les durées, tandis que si l'effet de censure est plus fort, les anciennetés seront en moyenne plus petites que les durées. Si donc en comparant deux groupes, on trouve qu'en moyenne les anciennetés du premier groupe sont plus faibles que celles du deuxième, cela n'implique pas que les durées sous-jacentes du premier groupe sont en moyenne plus petites que les durées du deuxième groupe. L'ordre des durées ne respectera

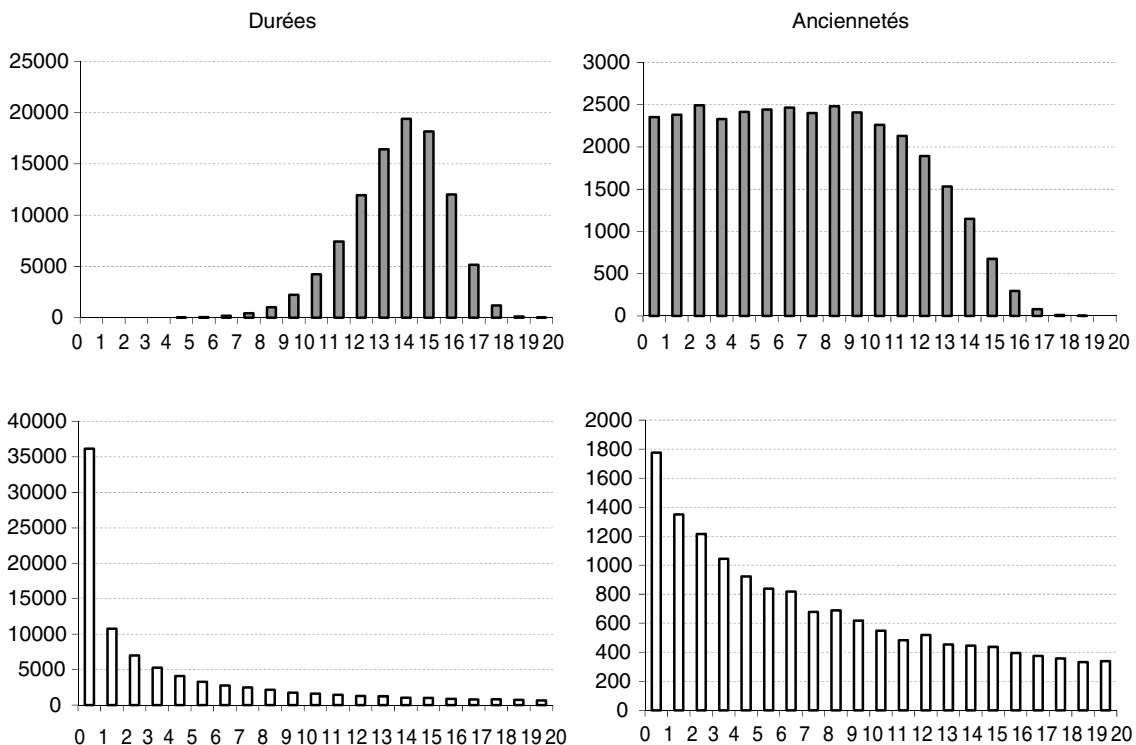
pas l'ordre des anciennetés si dans le premier groupe l'effet de censure est très fort et dans le deuxième l'effet de sélection est très fort, comme l'illustre la figure II.

L'influence des flux d'entrée sur les anciennetés

La distribution des durées dans la population a donc un impact direct sur les anciennetés. Un autre facteur influence également la distribution des anciennetés observées à un instant donné : il s'agit des flux d'entrée dans la situation de monoparentalité. Si, par exemple, de plus en plus de familles monoparentales se constituent chaque année (flux croissant), les anciennetés observées auront mécaniquement tendance à être courtes, la plupart des familles monoparentales observées dans l'enquête auront été formées peu de temps avant l'enquête. Un flux d'entrée croissant joue donc en faveur du biais de censure (un flux décroissant augmente le biais de sélection). Les simulations présentées dans la figure III illustrent bien ce phénomène.

Figure II

Comparaison de la distribution des durées et des anciennetés entre deux groupes



Note : les durées du groupe 1 (gris foncé), T_1 , ont été simulées selon une loi de Weibull(8, 15) et celles du groupe 2 (blanc) selon une loi de Weibull(0.5, 5). La distribution des durées est représentée à gauche et celle des anciennetés à droite. Cette figure illustre le fait que les durées peuvent être en moyenne plus grandes dans le groupe 1 que dans le groupe 2 (14 ans contre 10 ans), alors que les anciennetés sont en moyenne plus petites dans le groupe 1 (7 ans) que dans le groupe 2 (12 ans).

Pour résumer, la distribution des anciennetés observées au moment de l'enquête dépend à la fois des flux d'entrée et de la distribution des durées réelles. On a donc l'intuition que si la distribution des anciennetés et les flux d'entrée sont connus, on peut en déduire la distribution des durées. C'est l'objet de la section suivante.

Inférer les durées à partir des anciennetés et des flux d'entrée

Relation entre les flux, les durées et les stocks

On considère dans un premier temps la situation hypothétique dans laquelle les flux d'entrée en monoparentalité sont constants au cours du temps, et où la durée moyenne de la monoparentalité n'évolue pas au cours du temps non plus. Alors les flux de sortie compensent exactement les flux d'entrée et le stock n'évolue pas. Il y a dans ce cas précis, une relation simple entre le stock, les flux et les durées :

$Stock = Flux * E[T]$, où $E[T]$ est l'espérance de la variable aléatoire de durée.

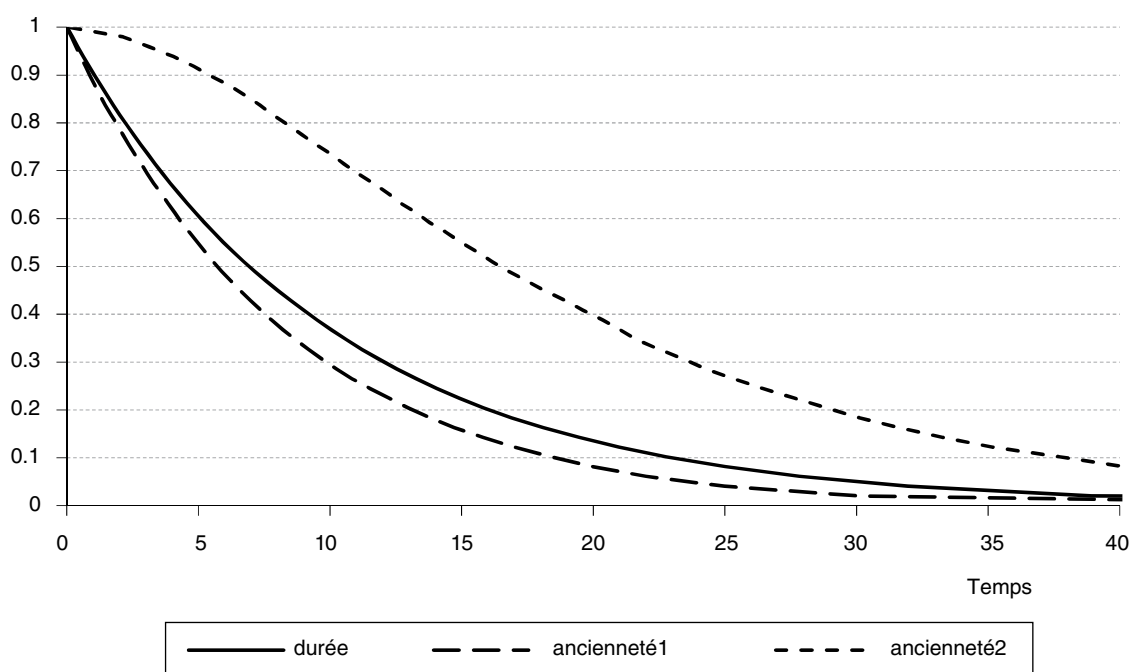
Cette relation se comprend facilement : plus les flux sont importants, plus les stocks observés à

un instant donné le sont également ; plus les personnes restent longtemps en situation de monoparentalité, plus elles ont de chance de faire partie du stock, et plus le stock est important. Connaissant les flux et le stock, cette relation nous donne donc directement accès à la durée moyenne passée en situation monoparentale $E[T]$. Dans le cas du régime stationnaire (flux constants) on peut donc directement déduire, à partir des flux d'entrée et de la taille du stock de familles monoparentales, la durée moyenne de la monoparentalité.

En dehors du régime stationnaire, l'égalité n'est plus vérifiée. En fait, si les flux augmentent régulièrement le stock une année donnée est inférieur au dernier flux multiplié par la durée moyenne : $Stock < Dernier Flux * E[T]$, ce qui donne une borne inférieure pour la moyenne des durées. On a la relation inverse dans le cas d'un flux décroissant.

La relation entre les flux, les durées et les stocks ne nous permet toutefois pas de déduire la distribution des durées. Pour cela, il faut en plus connaître la distribution des anciennetés. Comme nous souhaitons connaître la distribution des durées (et pas seulement la durée

Figure III
Influence des flux d'entrées sur les anciennetés observées



Note : on a simulé des dates d'entrées en monoparentalité et des durées de monoparentalité pour deux groupes. Dans les deux cas, la variable de durée est générée selon une loi exponentielle de moyenne 10. La différence entre les deux groupes vient de la distribution des dates d'entrées. Dans le premier groupe, le flux d'entrées en famille monoparentale est croissant tandis que dans le deuxième groupe il est décroissant. Il en résulte que les anciennetés observées A_1 et A_2 ont une distribution différente.

moyenne) et parce que nous ne sommes pas, avec la monoparentalité, dans une situation où les flux sont constants, nous présentons par la suite une méthode plus sophistiquée tenant compte de toute l'information disponible afin de déterminer la distribution des durées quels que soient les flux d'entrée.

Échantillonnage dans le flux et échantillonnage dans le stock

Les modèles de durée classiques, comme le modèle de Cox (1972), se servent la plupart du temps d'échantillonnages dans le flux : on observe les personnes qui entrent dans la situation d'intérêt à une date donnée et on suit ces personnes au cours du temps jusqu'à une date finale. Cela nécessite un dispositif particulier de suivi des personnes et c'est pourquoi ce genre d'études est en général mené sur de petits échantillons en sciences sociales qui nécessitent des enquêtes auprès des personnes, bien qu'on puisse appliquer ces modèles sur des données plus volumineuses si elles sont disponibles. Certaines durées sont alors complètement observées et d'autres sont censurées sur la droite, dans le cas des personnes qui ne sont pas sorties de la situation d'intérêt avant la date finale.

Au contraire, l'échantillonnage dans le stock permet d'étudier un échantillon de personnes qui sont dans la situation d'intérêt à un instant donné, quelle que soit leur date d'entrée dans cette situation. Même si ce type d'échantillonnage est plus rare lorsqu'on s'intéresse à des durées, plusieurs auteurs ont néanmoins développé des modèles de durée sur cette base. Wooldridge (2002) évoque un modèle dans lequel les individus échantillonnés dans le stock sont ensuite suivis pendant une certaine période de temps ; ainsi toutes les durées ne sont pas censurées sur la droite. Kiefer (1988) parle de la difficulté à estimer des durées de chômage à partir du *Current Population Survey* aux États-Unis, en raison du fait que les durées sont à la fois censurées à droite et à gauche. C'est bien souvent pour estimer des probabilités de sortie du chômage, au sein de la théorie de la recherche d'emploi (Atkinson et al., 1984) ou bien en épidémiologie (Keiding, 2006), que des modèles, spécifiques à l'échantillonnage dans le stock, ont été développés. Lorsque toutes les données sont censurées sur la droite, il est toujours possible d'estimer des probabilités de sortie de la situation d'intérêt, comme l'a montré Nickell (1979) dans un article auquel on se référera souvent par la

suite, pourvu que l'on connaisse les probabilités d'entrée aux dates précédant l'enquête. Pour cela il a développé un modèle entièrement paramétrique dont l'estimation repose sur le principe du maximum de vraisemblance. Lancaster remplace le problème de l'échantillonnage dans le stock dans le contexte plus général des processus de renouvellement (Lancaster, 1990).

La méthode utilisée ici reprend en très grande partie la solution proposée par Nickell (1979), mais plusieurs modifications ont néanmoins été apportées afin de mieux répondre aux questions que l'on se pose sur les durées de la monoparentalité. On reprend l'idée des risques instantanés proportionnels du modèle de Cox que l'on introduit dans le calcul de la vraisemblance donné par Nickell, ce qui rend le modèle proposé ici unique à notre connaissance. De plus, on aura besoin pour estimer les paramètres de ce modèle de disposer d'éléments permettant d'avoir une estimation des flux d'entrée dans cette situation, ce qui se révélera essentiel comme on le verra par la suite. On considère ici les données de l'EFL (*Enquête famille et logements*) pour le stock et les données d'Erfi (*Étude des relations familiales et intergénérationnelles*) pour les flux.

Principe général de la méthode

Dans le cas de l'EFL, on ne connaît que l'année de début de la situation de monoparentalité pour les personnes en famille monoparentale au moment de l'enquête. On ne connaît donc les anciennetés qu'à un an près. Cette incertitude va également se traduire par une incertitude sur les durées inférées. En fait, on choisit ici de modéliser les durées de façon discrète (cf. encadré 2) : on considère la variable aléatoire de durée correspondant à un nombre entier d'années (0, 1, 2, etc.). Par exemple, une personne entrée dans la situation en 2010 et sortie en 2012 aura une durée discrète de 2 ans, alors que sa durée réelle peut valoir entre 1 an (entrée fin 2010 et sortie début 2012) et 3 ans (entrée début 2010 et sortie fin 2012).

La méthode d'inférence consiste dans un premier temps à calculer la fonction de vraisemblance des observations (on observe les anciennetés). On montre que cette fonction de vraisemblance ne dépend que des anciennetés, des flux d'entrée et du risque instantané (voir encadré 3). On choisit ici de modéliser le risque instantané par une fonction constante par morceaux. Bien que le risque instantané

soit une fonction discrète, ce choix permet de réduire le nombre de paramètres à estimer et donc d'augmenter la précision des estimations. Il s'agit alors de trouver la fonction constante par morceaux qui maximise la vraisemblance. On peut rendre le modèle plus sophistiqué

en faisant dépendre le risque instantané de variables individuelles, comme dans un modèle de Cox (cf. encadré 3). L'inférence consiste alors à trouver les paramètres de régression et la fonction constante par morceaux qui maximisent la vraisemblance.

Encadré 3

MÉTHODE D'INFÉRENCE DES DURÉES À PARTIR DES ANCIENNETÉS ET DES FLUX D'ENTRÉES

On considère m individus, nés entre l'année a_0 et l'année a_1 , qui connaissent au cours de leur vie un ou plusieurs épisodes de monoparentalité. Pour chaque épisode, on note D l'année de début, F l'année de fin de la situation et $T = F - D$ la durée (mesurée de façon discrète) de la situation. Si l'entrée en monoparentalité et la sortie de la monoparentalité se font de manière uniforme sur l'année, l'espérance de T est alors égale à l'espérance de la véritable durée. D , F et T sont des variables aléatoires discrètes à valeurs entières.

Hypothèse 1 : on suppose que la variable aléatoire de durée suit la même loi quel que soit le rang de l'épisode de monoparentalité (il peut en effet y avoir plusieurs épisodes de monoparentalité au cours d'une vie).

L'objectif est d'estimer la loi de la variable de durée T conditionnellement à des caractéristiques individuelles x . Pour cela on estimera le risque instantané de cette variable, noté $h(u, x) = P(T = u | T \geq u, x)$, pour $u \in \mathbb{N}$.

Au début de l'année e , on réalise une enquête et on observe que n de ces individus ($n \leq m$) sont en situation de monoparentalité. On observe de plus leurs dates d'entrées dans la situation qu'on note $d_1, \dots, d_n$. C'est à partir de ces dates d'entrées qu'on pourra inférer la loi de T . Pour cela, on calcule dans un premier temps la vraisemblance des observations en faisant l'hypothèse suivante :

Hypothèse 2 : les variables aléatoires T et D sont indépendantes. Cette hypothèse s'interprète comme le fait que la loi de durée de la monoparentalité n'évolue pas au cours du temps : toutes les générations font donc face aux mêmes durées. La contribution à la vraisemblance d'un individu i ayant les caractéristiques x_i et étant entré en monoparentalité l'année d_i s'écrit :

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

où $A(d, x) = P(D = d | x)$ est la loi conditionnelle de D sachant x .

D'après l'hypothèse 2, on a :

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

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

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

On propose ici de modéliser le risque instantané comme dans le cas du modèle de Cox à risques instantanés proportionnels :

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

Prendre le minimum avec 1 permet de s'assurer que le risque instantané prend bien une valeur entre 0 et 1, comme il se doit dans le cas discret. Enfin, on modélise le risque instantané de base h_0 par une fonction constante par morceaux qu'il faudra estimer. On choisit un pas constant de 3 ans : h_0 est donc constante de 0 à 2 ans, de 3 à 5 ans, etc. jusqu'à 18 ans, puis elle est constante au-delà de 19 ans. Ceci revient donc à estimer 7 paramètres (un pour chaque période de temps sur laquelle la fonction est constante). Le choix d'un pas de 3 ans a été fait pour réduire le nombre de paramètres et augmenter la précision des estimations. L'inconvénient est que cela contraint un peu la forme de la fonction h_0 .

La vraisemblance du modèle s'écrit finalement :

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

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

On cherche alors la fonction constante par morceaux h_0 et le paramètre β qui maximise cette fonction de vraisemblance. On remarque qu'il suffit pour cela de connaître, x étant fixé, la fonction $A(d, x)$ à une constante multiplicative près, ce qui revient à connaître les flux d'entrées en monoparentalité chaque année pour les personnes nées entre l'année a_0 et l'année a_1 et ayant les caractéristiques x . Contrairement au modèle de Cox, on est ici obligé d'estimer conjointement le risque instantané de base h_0 et les paramètres β , bien qu'on ne précise pas de forme paramétrique pour h_0 .

D'autres modélisations seraient possibles afin d'établir un lien entre des covariables et le risque instantané. Par exemple, le modèle du temps de défaillance accéléré (*accelerated failure time*) est une alternative intéressante au modèle de Cox. Dans ce modèle, les risques instantanés ne sont plus proportionnels entre eux. L'idée est que chaque personne fait face au même risque de base, mais que le temps ne s'écoule pas à la même vitesse pour chacun : il est soit accéléré, soit ralenti selon que le coefficient associé est plus grand ou plus petit que un. Au final, cela a un effet multiplicatif sur l'espérance de la variable de durée. L'avantage de ce modèle est donc son interprétation très intuitive. L'inconvénient est qu'il faut spécifier de manière paramétrique le risque instantané de base. Or c'est justement ce que l'on évite de faire en ne modélisant pas le risque instantané de base par une fonction déterminée et en laissant les données « librement » estimer ce risque sans contraintes particulières. Tout autre modèle entièrement paramétrique, comme la modélisation *logit* de Nickell, conviendrait également pour modéliser le risque instantané en fonction de covariables. Ainsi, si on a une idée plus ou moins précise de la façon dont le risque instantané évolue au cours du temps et de la façon dont il est modifié selon certaines covariables, il est plus judicieux de construire un modèle entièrement paramétrique qui reflète ces connaissances *a priori*. Si nous avons ici retenu le modèle de Cox, c'est parce que, d'une part, il n'impose pas de donner une forme *a priori* au risque instantané de base et que, d'autre part, l'interprétation des coefficients estimés est simple.

Trois hypothèses ont été faites dans le cadre de cette modélisation. La première est que la durée passée dans la situation de monoparentalité ne dépend pas du rang de l'épisode de monoparentalité au cours de la vie, c'est-à-dire en supposant que, à caractéristiques égales, le fait de vivre un premier ou un deuxième épisode monoparental n'influence pas la durée de la monoparentalité. Il peut en effet arriver de vivre une deuxième fois cette situation, par exemple à la suite d'une remise en couple mettant fin à un épisode de monoparentalité, elle-même suivie d'une séparation ramenant ainsi à la situation de parent isolé. La seconde hypothèse stipule que la durée de la situation est indépendante de la date d'entrée en monoparentalité, ce qui revient à dire qu'il n'y a pas d'effet de génération. Cette hypothèse est forte et n'est sans doute pas vérifiée dans la réalité. Par exemple, le temps pour reformer une union après une séparation se réduit au fil des générations (Costemalle, 2015). Néanmoins nous n'avons pas pu créer un modèle permettant

d'estimer de façon robuste cette dépendance. Enfin, la troisième hypothèse est la même que dans un modèle de Cox, à savoir que les risques instantanés sont proportionnels. Cette hypothèse permet de calculer des effets d'une covariable sur le risque instantané indépendamment du temps. On pourra avoir une idée de sa validité en comparant le risque instantané de façon indépendante sur les sous-populations étudiées.

La méthode développée ici est donc originale puisqu'on a combiné deux modélisations connues (celle de Nickell et celle de Cox) pour en faire une troisième. La modélisation de Cox permet d'estimer un risque instantané de base et des paramètres qui indiquent comment évolue le risque instantané par rapport à ce risque instantané de base. Il ne peut malheureusement pas être utilisé dans le cas d'un échantillonnage dans le stock. La modélisation de Nickell, quant à elle, ne permet pas d'estimer un risque instantané de base, ni *a posteriori* une distribution des durées. Son but est de déterminer si une variable externe donnée a une influence positive ou négative sur la probabilité instantanée de sortie d'une situation (initialement chez Nickell, le chômage). Sa force réside dans la simplicité de l'expression du risque instantané (une fonction logistique). L'inconvénient d'une telle modélisation est que l'on force le risque instantané à avoir une certaine forme, ce qui contraint fortement le modèle. Une autre différence entre la modélisation proposée ici et celle de Nickell est que dans le premier cas, les durées suivent des lois discrètes alors que dans le second, elles suivent des lois continues. L'aspect discret du problème ajoute une difficulté supplémentaire.

Résultats de l'inférence sur des simulations

Si on connaît parfaitement les flux d'entrée en monoparentalité, la méthode présentée fonctionne correctement et permet de retrouver la distribution des durées. On peut alors en déduire plusieurs grandeurs associées aux durées comme la moyenne, la médiane ou d'autres quantiles (cf. encadré 2). Les figures IV et V montrent sur des exemples simulés que, quel que soit le biais dominant (biais de censure ou biais de sélection), la méthode d'inférence permet de retrouver les bonnes distributions des durées. Il reste néanmoins trois sources d'incertitude. La première vient du fait qu'on estime des durées discrètes et non continues. La deuxième provient de la taille de l'échantillon : plus celui-ci est petit, plus l'incertitude est grande. Enfin, la troisième concerne le fait qu'on ne connaît pas parfaitement les flux d'entrée mais que ceux-ci sont estimés à partir de l'enquête *Erfi*.

Encadré 4

DÉTERMINATION DES FLUX D'ENTRÉES À PARTIR D'ERFI

Pour des caractéristiques individuelles x fixées, il s'agit d'estimer $A(d, x)$ pour les personnes qui sont nées entre a_0 et a_1 , à une constante près (cf. encadré 3). Ceci revient alors à estimer le nombre de personnes ayant les caractéristiques x étant entrées dans la situation de monoparentalité l'année d . Pour cela, on utilise l'enquête *Erfi* dont les répondants sont nés entre 1926 et 1987. Les répondants à l'*EFL* ont quant à eux pu naître jusqu'en 1992. Il nous manque donc dans *Erfi* le flux d'entrée en monoparentalité des personnes nées entre 1988 et 1992. Pour tenir compte de cela, on estime, dans un premier temps, le flux d'entrée en monoparentalité pour les années inférieures ou égales à 2005 directement à partir de la troisième vague d'*Erfi*. Les personnes nées entre 1988 et 1992 ayant moins de 18 ans sur ces années ne contribuent alors presque pas au flux. Dans un deuxième temps, pour la période 2006-2010 on corrige chaque année le flux par un coefficient multiplicatif afin d'estimer les flux correspondants aux personnes de 18 ans ou plus. Ces coefficients, présentés dans le tableau ci-dessous, sont calculés à l'aide de la répartition des âges d'entrée en monoparentalité (cette répartition est elle-même estimée, à l'aide d'*Erfi*, à partir du flux de personnes entrées en monoparentalité entre 2000 et 2005).

Coefficients correcteurs pour estimer les flux sur la période 2006-2010

	2006	2007	2008	2009	2010
Femmes	1.026	1.026	1.056	1.1	1.125
Hommes	1	1	1	1	1

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

Si une personne connaît plusieurs épisodes de monoparentalité au cours de sa vie, seul le dernier est pris en compte pour la contribution au flux.

Enfin, on lisse les flux ainsi estimés pour atténuer le bruit inhérent aux faibles effectifs présents dans l'enquête *Erfi*. Pour cela on effectue, dans un premier temps, une moyenne mobile sur cinq années qui permet localement d'atténuer les variations aléatoires dues aux petits effectifs. Dans un deuxième temps, on fait une régression polynomiale sur les données ainsi lissées. Les flux d'entrées estimés selon cette méthode sont présentés dans les graphiques en annexe 1.

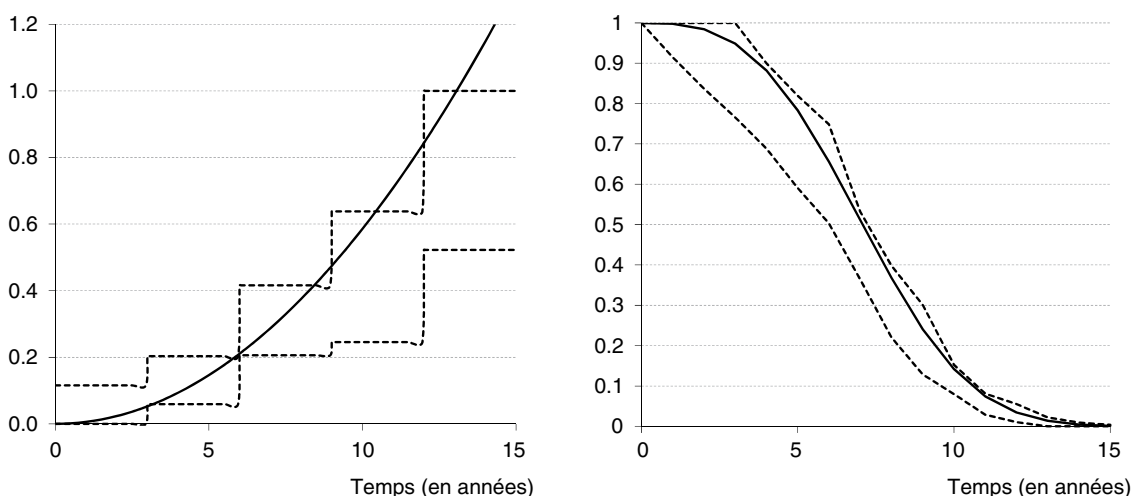
Influence de la taille de l'échantillon

Pour bien comprendre comment évoluent les estimations en fonction de la taille de l'échantillon, on a réalisé une série d'expériences qui consiste pour chacune d'elle à simuler des flux d'entrée et des durées pour un nombre donné

d'individus, puis d'en déduire la distribution des durées discrètes (selon la méthode exposée) et enfin d'estimer la moyenne et la médiane des durées. On teste quatre mesures de précision. La première est l'écart entre la moyenne estimée des durées et l'espérance réelle de la variable aléatoire continue de durée ; la deuxième est

Figure IV

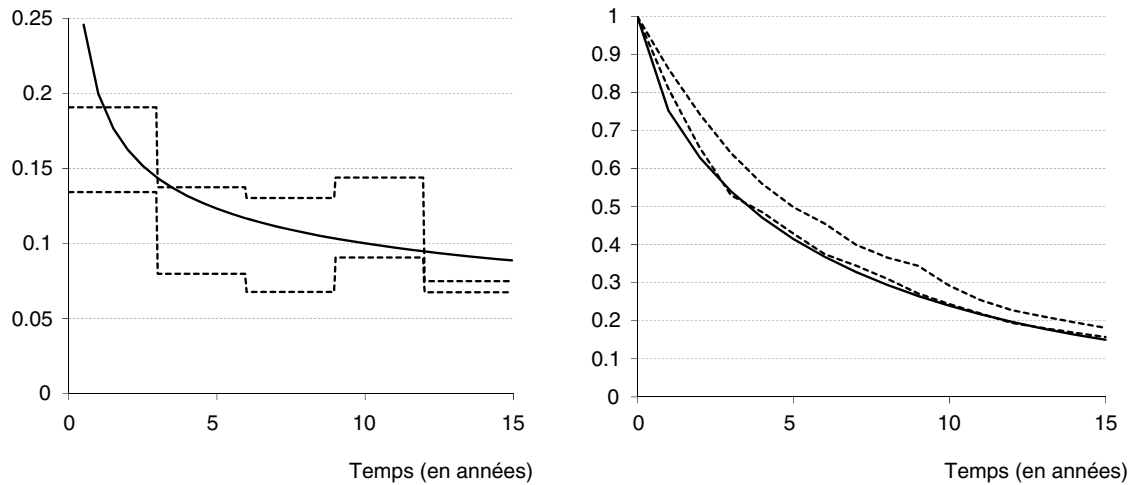
Comparaison entre le risque instantané estimé et réel (gauche) et entre la survie estimée et réelle (droite), lorsque le biais de censure est le plus fort



Note : on a généré 100 échantillons de 1 200 observations d'anciennetés à partir d'une loi de Weibull(3, 8) pour les durées et d'une loi uniforme pour les flux d'entrées. Les courbes en traits pleins représentent le risque instantané et la survie réels. Les courbes en pointillés donnent les intervalles contenant 95 % des estimations.

Figure V

Comparaison entre le risque instantané estimé et réel (gauche) et entre la survie estimée et réelle (droite), lorsque le biais de sélection est le plus fort



Note : on a généré 100 échantillons de 12 300 observations d'anciennetés à partir d'une loi de Weibull(0.7 , 6) pour les durées et d'une loi uniforme pour les flux d'entrées. Les courbes en traits pleins représentent le risque instantané et la survie réels. Les courbes en pointillés donnent les intervalles contenant 95 % des estimations.

l'écart entre la médiane estimée et la médiane réelle ; la troisième est l'écart entre la courbe de survie estimée et la courbe de survie réelle ; la quatrième est l'écart entre le risque instantané estimé et le risque instantané réel. Lorsque la taille de l'échantillon est grande les différentes mesures de précisions ne fluctuent presque pas d'une expérience à l'autre, alors que ce n'est pas le cas lorsque la taille de l'échantillon est petite. Le premier résultat est que la moyenne estimée est inférieure d'environ 0.5 an à la moyenne réelle (quelle que soit la loi de durée sous-jacente). Ainsi, l'inférence des durées conduit à une sous-estimation de la moyenne de 0.5 an, même si la taille de l'échantillon est très importante. Au contraire, la médiane est bien estimée et, lorsque l'échantillon est suffisamment grand, il n'y a presque pas d'écart entre la médiane estimée et la médiane réelle (figure VI). L'écart entre la fonction de survie estimée et la survie réelle et l'écart entre le risque instantané estimé et le risque réel tendent vers 0 lorsque la taille de l'échantillon augmente (figure VII). Ces résultats obtenus à partir de simulations numériques indiquent que l'inférence de la distribution de durées par la méthode de maximisation de la vraisemblance fonctionne correctement et que les résultats deviennent très précis si la taille de l'échantillon est suffisamment importante (de l'ordre de 5 000 personnes).

Par la suite, on ajoutera 0.5 an aux estimations de moyennes afin de tenir compte du biais de l'estimateur de la moyenne. Cette correction s'appuie sur une constatation empirique sur

des données simulées, mais n'est pas prouvée théoriquement. On peut néanmoins montrer mathématiquement que si la fonction de survie est bien estimée, comme c'est le cas ici, alors l'erreur sur la moyenne sera d'environ 0.5 an (voir la démonstration en annexe 4).

Estimer les facteurs de risque

On teste maintenant si le modèle estime correctement les coefficients de régression du risque instantané. On suppose que l'on dispose d'une population formée de trois groupes distincts et faisant chacun face à un risque instantané différent (h_0 pour le premier groupe qu'on désigne comme le groupe de référence, h_1 pour le deuxième et h_2 pour le troisième) :

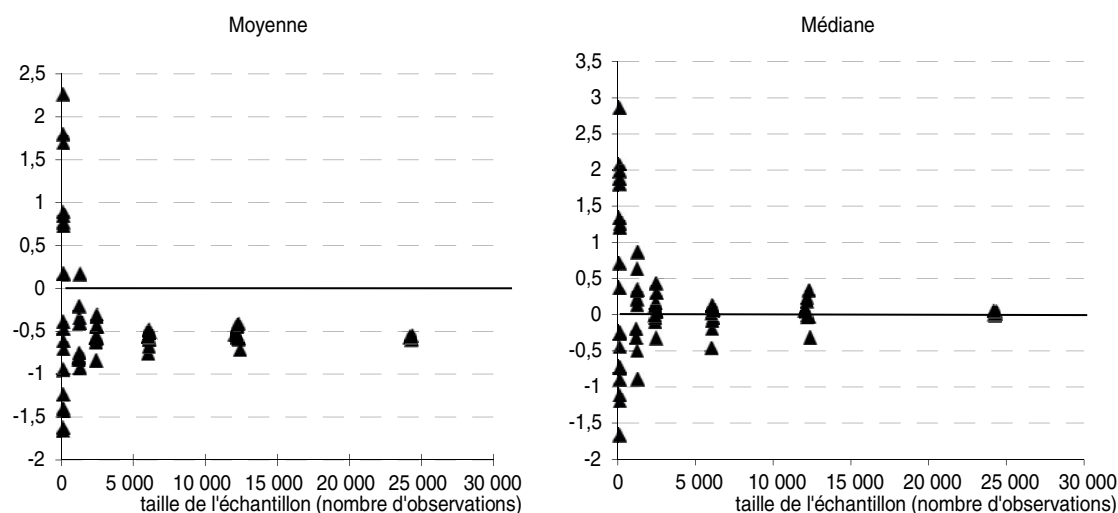
$$h_1(t) = h_0(t)\exp(\beta_1) ; h_2(t) = h_0(t)\exp(\beta_2).$$

Ces risques sont, comme l'hypothèse 3 de notre modélisation le suppose, proportionnels entre eux. Il suffit donc d'estimer seulement β_1 et β_2 pour connaître les différences de risques instantanés (et donc de durée) entre les groupes. Si le coefficient est positif, cela signifie que le risque instantané est plus important et que la durée moyenne est donc plus faible. C'est l'inverse si le coefficient est négatif.

Afin vérifier si notre méthode d'estimation permet de retrouver ces deux paramètres, on simule 200 fois une population totale de 50 000 individus équitablement répartis entre

Figure VI

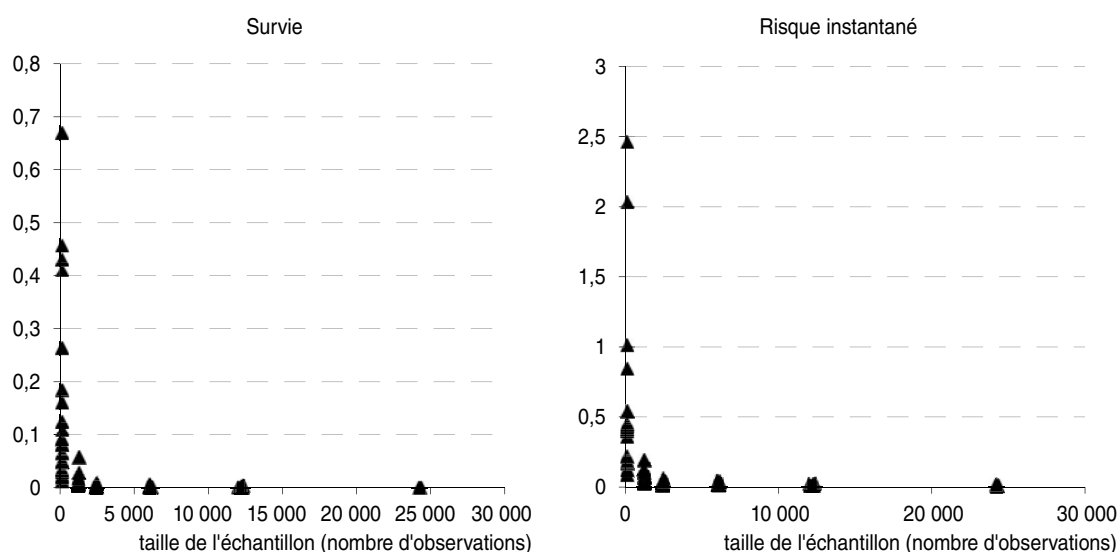
Écarts entre la moyenne estimée et réelle et entre la médiane estimée et réelle, selon la taille de l'échantillon



Note : chaque point correspond à une expérience, qui consiste à générer des durées de façon aléatoire, à estimer le risque instantané selon la méthode présentée et à en déduire la moyenne ou la médiane. La variable de flux a été générée selon une loi uniforme et la variable de durée selon une loi de Weibull(5, 7). Ce qui change d'une expérience à l'autre, c'est la taille de l'échantillon qui correspond aux nombre de personnes en famille monoparentale au moment de l'enquête.

Figure VII

Écarts entre la survie estimée et réelle et entre le risque instantané estimé et réel, selon la taille de l'échantillon



Note : chaque point correspond à une expérience. La variable de flux a été générée selon une loi uniforme et la variable de durée selon une loi de Weibull(5, 7). Si f est la fonction estimée (survie ou risque instantané) et si g est la fonction réelle, on définit l'écart entre ces deux fonctions comme $\sum_{t \geq 0} (f(t) - g(t))^2$.

les trois groupes, avec à chaque fois $\beta_1 = 0.5$ et $\beta_2 = -0.5$. Pour chacune de ces 200 simulations, on estime β_1 et β_2 par la méthode du maximum de vraisemblance. On fait ensuite la moyenne de ces estimations sur les 200 expériences. On peut également calculer le taux de couverture, c'est-à-dire la proportion d'estimations telles que la vraie valeur du coefficient se trouve dans l'intervalle de confiance estimé (intervalle de

confiance à 95 % de l'estimateur de maximum de vraisemblance, basée sur le comportement asymptotique de l'estimateur qui théoriquement doit suivre une loi normale).

Les résultats de ces simulations sont dans le tableau 1. Elles montrent que le modèle retrouve bien les valeurs des coefficients β_1 et β_2 , même si les estimations semblent un peu sous-estimer les

Tableau 1

Moyenne et taux de couverture des estimations des coefficients β_1 et β_2

		β_1	β_2
Vraie valeur		0.5	- 0.5
Estimation	Moyenne	0.47	- 0.48
	Taux de couverture	82.5	85.0

vraies valeurs. Les taux de couverture sont inférieurs à 95 % ce qui indique que les intervalles de confiances estimés sont légèrement trop étroits. Ces taux de couverture ont néanmoins une valeur assez élevée, de l'ordre de 85 %.

Une estimation des durées de la monoparentalité en France

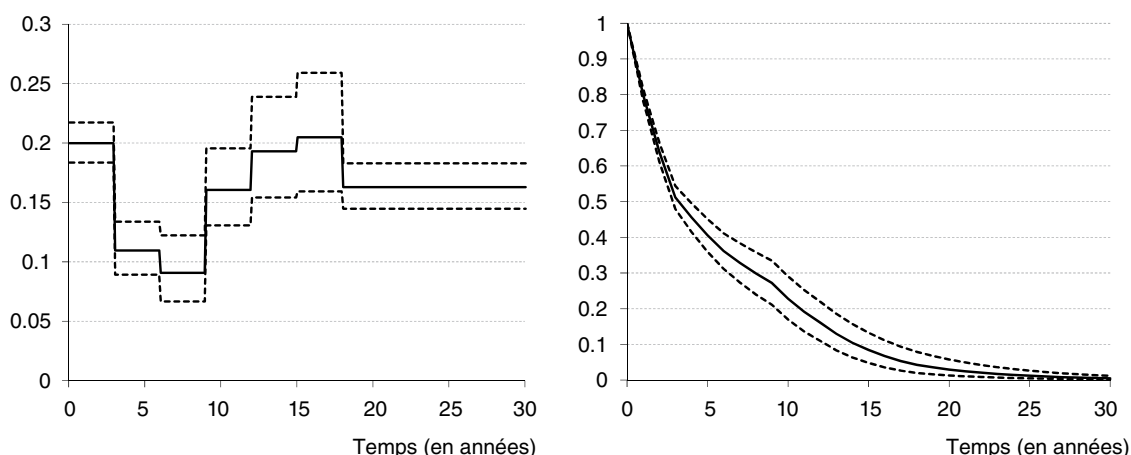
Parmi les 359 770 répondants à l'EFL, 12 519 sont en situation de monoparentalité au moment de l'enquête⁵, dont 1 073 hommes et 11 446 femmes. Au vu des résultats précédents, on dispose donc d'un échantillon suffisamment grand pour pouvoir inférer la distribution des durées pour les femmes, mais l'échantillon des hommes semble trop petit pour avoir des résultats robustes.

5. On a retiré les 205 personnes entrées en monoparentalité l'année de l'enquête (en 2011), car elles n'apportent pas d'information à notre modèle, ainsi que les parents vivant en couple non-cohabitant, situation dont on ne peut déterminer l'ancienneté.

Les durées estimées : une forme « en U »

On présente ici les résultats des estimations du risque instantané de sortir de la monoparentalité sans prendre en compte de covariables, c'est-à-dire en ne tenant compte que du temps, sans introduire d'autres facteurs qui pourraient influencer la probabilité de sortir de la monoparentalité. La figure VIII montre que le risque instantané global n'est pas monotone : initialement il diminue, puis se stabilise avant d'augmenter. Cette forme en « U » suggère que soit on sort rapidement de la monoparentalité, soit on y reste longtemps. En effet, la probabilité de sortir de cette situation est la plus faible entre 3 et 8 ans. La courbe de survie obtenue à partir du risque instantané montre qu'au bout de 3 ans la moitié des parents de famille monoparentale sont sortis de cette situation ; au bout de 8 ans il en reste encore 30 %, au bout de 12 ans 16 %, et seulement 4 % y restent plus de 18 ans. On peut aussi déterminer la moyenne de la durée passée dans cette situation (cf. encadré 2) : on trouve ici que la durée moyenne est de 5.7 ans. La moyenne des anciennetés quant à elle est de

Figure VIII

Estimation du risque instantané et de la survie associés aux durées de monoparentalité

Note : les traits en pointillés représentent l'intervalle de confiance à 95 % obtenus par la méthode du maximum de vraisemblance.

Champ : parents isolés avec enfants mineurs, France métropolitaine.

Source : Insee, *Enquête famille et logements (EFL)* 2011 et Ined-Insee, *Étude des relations familiales et intergénérationnelles (Erfi)*, vagues 1 et 3, 2005 et 2011.

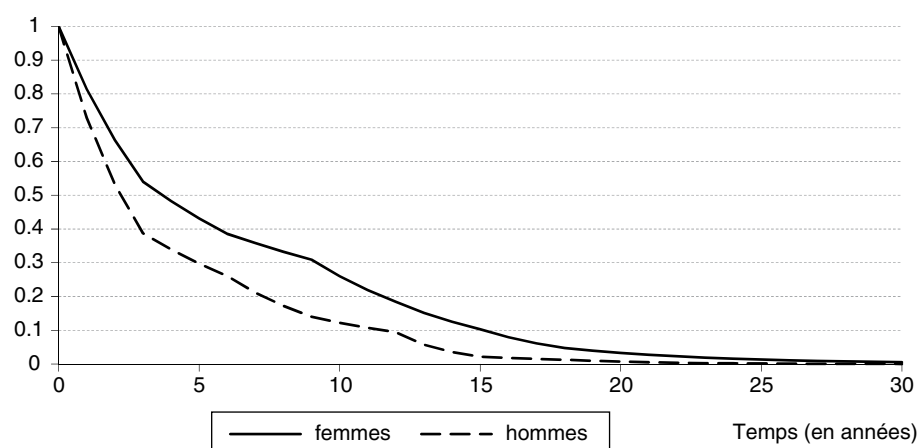
Figure IX

Risque instantané de sortie de la monoparentalité des hommes et des femmes

Champ : parents isolés avec enfants mineurs, France métropolitaine.

Source : Insee, *Enquête famille et logements (EFL)* 2011 et Ined-Insee, *Étude des relations familiales et intergénérationnelles (Erfi)*, vagues 1 et 3, 2005 et 2011.

Figure X

Fonctions de survie associées aux durées passées en famille monoparentale, pour les hommes et les femmes

Champ : Parents isolés avec enfants mineurs, France métropolitaine.

Source : Insee, *Enquête famille et logements (EFL)* 2011 et Ined-Insee, *Étude des relations familiales et intergénérationnelles (Erfi)*, vagues 1 et 3, 2005 et 2011.

Tableau 2

Estimations de la survie, de la moyenne et de la médiane des durées passées en famille monoparentale

	Moyennes			Médianes		
	Valeur estimée	Intervalles de confiance		Valeur estimée	Intervalles de confiance	
		Maximum de vraisemblance	Bootstrap		Maximum de vraisemblance	Bootstrap
Ensemble	5.7	[4.9 ; 6.6]	[5.5 ; 6.1]	3.2	[2.8 ; 3.9]	[2.8 ; 3.9]
Hommes	4.1	[2.8 ; 6.0]	[3.7 ; 4.7]	2.2	[1.8 ; 2.8]	[1.8 ; 3.0]
Femmes	6.1	[5.2 ; 7.2]	[5.9 ; 6.5]	3.7	[3.1 ; 4.6]	[3.0 ; 4.3]

Note : la moyenne est estimée ici selon l'estimateur présentée dans l'encadré 2 corrigée du biais de 0.5 an. Les intervalles de confiance obtenus par *bootstrap* reposent sur 320 échantillons tirés aléatoirement selon la distribution des anciennetés observées (tirage aléatoire avec remise). Pour chacun de ces échantillons on a estimé le risque instantané dont on a déduit la moyenne et la médiane des durées. Champ : France métropolitaine.

Source : Insee, *Enquête famille et logements (EFL)* 2011 et Ined-Insee, *Étude des relations familiales et intergénérationnelles (Erfi)*, vagues 1 et 3, 2005 et 2011.

5.5 ans. Elle est donc proche de la moyenne des durées estimées. On est ainsi dans une situation particulière où le biais de censure et le biais de sélection se neutralisent presque.

En estimant séparément le risque instantané pour les femmes et les hommes, on constate qu'ils n'ont pas la même forme ; par conséquent, l'hypothèse de proportionnalité des risques instantanés n'est pas appropriée pour la comparaison entre sexes (figures IX et X). Si le risque instantané des femmes a la forme en « U » observée pour l'ensemble, celui des hommes est plus fluctuant et plus élevé en moyenne.

Cette fluctuation laisse penser qu'il n'y a pas assez d'observations pour les hommes dans l'EFL : autour de 1 000. D'autre part, le fait que dans l'ensemble, il y a beaucoup plus de femmes que d'hommes (à peu près dix fois plus) explique de plus qu'on retrouve la forme en « U » à la fois pour l'ensemble des parents en famille monoparentale et pour les femmes seulement.

La durée moyenne de la monoparentalité estimée ici est de 6.1 ans pour les femmes et 4.1 ans pour les hommes (tableau 2). Des intervalles de confiance à 95 % ont été estimés successivement par la méthode du maximum de vraisemblance, puis par *bootstrap*. En ce qui concerne la moyenne, ces derniers sont plus étroits que ceux donnés par la méthode du maximum de vraisemblance. En revanche, pour la médiane, on obtient presque les mêmes intervalles de confiance. Ceci constitue une indication supplémentaire de la fiabilité des estimations.

Déterminer la durée moyenne à partir des flux et des stocks

On a pu déterminer des moyennes de durées de monoparentalité à partir de la distribution des durées estimée. Comme on l'a expliqué auparavant, il est néanmoins possible d'avoir une

estimation de la borne inférieure de la moyenne des durées à l'aide de la relation entre les stocks, les flux et les durées.

On peut appliquer ce principe à l'EFL car on connaît la taille du stock de familles monoparentales en 2011 (1 449 000) ainsi que la grandeur du dernier flux (254 000 entrées en 2010). On en déduit alors que la durée moyenne de la monoparentalité doit être supérieure à 6 ans pour les femmes et 4.4 ans pour les hommes (tableau 3), or nos estimations précédentes sont de 6.1 ans pour les femmes et 4.1 ans pour les hommes.

Le dernier flux est sans doute sous-estimé car on ne prend pas en compte les personnes qui sont entrées dans la situation en 2010 et qui en sont sorties la même année. Il en résulte que l'estimation de la borne inférieure de la moyenne de la durée passée en situation monoparentale est légèrement sous-estimée également. Au final, ces différentes estimations semblent tout à fait cohérentes entre elles.

Les durées et les anciennetés correspondent à deux concepts différents. Mais dans le cas présent, la répartition des anciennetés est assez proche de celles des durées. On est en effet dans une situation où le biais de censure et le biais de sélection se compensent approximativement, les anciennetés peuvent donc fournir une première approximation des durées.

Des différences sensibles des durées estimées selon la cause d'entrée dans la monoparentalité

On s'intéresse maintenant aux durées obtenues en introduisant des covariables indépendantes du temps. Compte tenu de la faible taille de l'échantillon d'hommes, on se restreint seulement aux résultats concernant les femmes. Les durées de monoparentalité sont estimées successivement selon la cause d'entrée dans la situation, selon le niveau de diplôme atteint

Tableau 3
Stock et flux de familles monoparentales

	Stock en 2011	Flux en 2010	Stock/Flux
Hommes	208 904	47 977	4.4
Femmes	1 239 843	206 067	6.0

Champ : France métropolitaine.

Source : Insee, *Enquête famille et logements* 2011.

au moment de l'enquête et selon la catégorie sociale au moment de l'enquête. À chaque fois, on présente les estimations des coefficients β du modèle à risques instantanés proportionnels, par rapport à un groupe de référence ; si β est positif, cela signifie que le risque instantané de sortie de la situation monoparentale est plus élevé que celui du groupe de référence, donc que sa survie dans la situation (la durée) est plus courte que celle du groupe de référence. Les durées moyennes sont déduites des risques instantanés estimés pour chaque groupe, eux-mêmes déduits du risque instantané de base (voir annexe 2) et des risques relatifs.

Les causes d'entrée en monoparentalité sont au nombre de trois : la séparation d'avec le conjoint (78 %) un enfant né hors de la vie en couple (16 %), le décès du conjoint (6 %). Des différences de durée notables sont observées entre ces causes d'entrée. Par rapport à une entrée associée à un enfant né hors de la vie de couple, la probabilité instantanée de sortie de la monoparentalité est 1.8 fois plus élevée lorsque la cause d'entrée est une séparation, et 1.7 fois plus élevée lorsque que c'est à la suite du décès du conjoint (tableau 4.A). Les femmes qui ont

vécu une situation de monoparentalité à l'issue d'une séparation sont donc celles qui y passent le moins de temps (5.4 ans), suivies des femmes qui se trouvent à la tête d'une famille monoparentale suite à un veuvage (5.7 ans) et celles qui ont eu un enfant hors couple restent nettement plus longtemps dans cette situation (9.1 ans).

Des différences apparaissent également selon le niveau de diplôme atteint. On distingue les personnes n'ayant aucun diplôme, les personnes ayant un diplôme de niveau inférieur au baccalauréat, celles ayant un diplôme de niveau équivalent au baccalauréat et celles ayant un diplôme de niveau strictement supérieur au baccalauréat. La probabilité instantanée de sortie de la monoparentalité apparaît toujours plus élevée pour les femmes ayant un diplôme, ce qui signifie que celles qui n'en ont aucun restent en moyenne plus longtemps en situation monoparentale que les autres (tableau 4.B). La probabilité de sortie ne s'élève cependant pas régulièrement selon le niveau de diplôme atteint ; ainsi, d'après ces estimations, ce sont les femmes qui ont un diplôme inférieur au baccalauréat et celles qui ont seulement le baccalauréat pour lesquelles les durées sont les moins

Tableau 4
Estimation des risques relatifs de sortie de la monoparentalité et des durées moyennes dans la situation

A. Selon la cause d'entrée

Cause d'entrée	Risque relatif	valeur-p	Durée moyenne
Enfant hors couple (16 %)	1	-	9.1
Séparation (78 %)	1.84	7.90E-77	5.4
Veuvage (6 %)	1.73	6.00E-31	5.7

B. Selon le niveau de diplôme

Diplôme	Risque relatif	valeur-p	Durée moyenne
Aucun diplôme (21 %)	1	-	7.6
Inférieur au bac (34 %)	1.40	2.10E-28	5.5
Baccalauréat (19 %)	1.38	6.70E-18	5.6
Supérieur au baccalauréat (27%)	1.20	4.70E-07	6.4

C. Selon la catégorie sociale

Catégorie sociale	Risque relatif	valeur-p	Durée moyenne
Ouvriers (11 %)	1	-	6.9
Artisans, commerçants, chefs d'entreprises (3 %)	1.02	7.70E-01	6.8
Cadres et profession intellectuelles supérieures (9 %)	0.90	1.00E-01	7.5
Professions intermédiaires (23%)	1.05	2.90E-01	6.6
Employés (54%)	1.11	7.60E-03	6.3

Note : les pourcentages entre parenthèses donnent la proportion de chaque catégorie.

Champ : femmes, France métropolitaine.

Source : Insee, *Enquête famille et logements (EFL)* 2011 ; Ined-Insee, *Étude des relations familiales et intergénérationnelles (Erfl)*, vagues 1 et 3, 2005 et 2011.

élevées (respectivement 5.5 et 5.6 ans) contre 7.6 ans pour celles n'ayant aucun diplôme et 6.4 ans pour celles qui ont un diplôme supérieur au baccalauréat.

On distingue enfin cinq catégories sociales⁶ : les artisans, commerçants ou chefs d'entreprise, les cadres, les professions intermédiaires, les employés et les ouvriers. La catégorie sociale semble avoir très peu d'effet sur la durée passée en situation de monoparentalité (tableau 4.C). Le seul écart de probabilité statistiquement significatif (au seuil de 5 %) s'observe pour les femmes employées, avec une probabilité instantanée de sortie de la monoparentalité légèrement plus élevée que les ouvrières. En moyenne, les femmes artisanes, commerçantes ou chefs d'entreprises restent 6.8 ans en famille monoparentale, les cadres 7.5 ans, les professions intermédiaires 6.6 ans, les employées 6.3 ans et les ouvrières 6.9 ans.

Les différences de durée les plus importantes sont ainsi observées entre les causes d'entrée dans la monoparentalité. Ces différences de durée selon la cause d'entrée sont associées en premier lieu à l'âge des enfants au début de la situation ; en effet, la durée maximale de la monoparentalité est limitée par l'âge du plus jeune des enfants au moment où la famille devient monoparentale. Les femmes qui ont eu un enfant sans avoir vécu en couple sont en situation monoparentale d'emblée, dès la naissance de l'enfant. Si elles ne forment pas de couple par la suite, elles peuvent ainsi rester dans cette situation 18 ans, voire plus longtemps, dans le cas (rare) où elles auraient plusieurs enfants sans jamais vivre en couple. Les femmes qui deviennent mères de famille monoparentale à la suite d'une séparation ou d'un décès ont, à l'entrée dans cette situation, des enfants plus âgés.

La cause d'entrée pourrait contribuer par ailleurs aux écarts de durée estimés selon les niveaux de diplôme : en effet, pour les femmes observées en situation de monoparentalité en 2011, l'entrée dans la situation est associée au fait d'avoir eu un enfant hors couple plus souvent lorsqu'elles n'ont pas de diplôme que lorsqu'elles ont un diplôme supérieur au baccalauréat (respectivement 20 % et 11 % des cas).

6. Les retraités et les chômeurs ayant déjà travaillé sont reclassés dans leur ancienne catégorie sociale. Les chômeurs n'ayant jamais travaillé, les inactifs de moins de 60 ans, les militaires, les étudiants les personnes sans activité de plus de 60 ans ainsi que les agriculteurs ne sont pas pris en compte.

Si l'âge des enfants à l'entrée dans la monoparentalité influence mécaniquement la durée dans la situation, un autre facteur peut aussi contribuer à cette durée : la mise ou remise en couple du parent, qui interrompt instantanément la monoparentalité. On sait par exemple qu'après une séparation, les femmes plus diplômées ne reforment pas plus rapidement un couple que les moins diplômées (Costemalle, 2015). Mais l'influence des mises ou remises en couple sur les écarts de durée de monoparentalité selon la cause d'entrée ne peut pas être évaluée ici car les données ne permettent pas ici de connaître la cause de sortie de la monoparentalité.

Restent quelques sources d'incertitude. L'une porte sur les durées de monoparentalité concernant les hommes ; on a vu que les effectifs d'hommes dans cette situation sont trop faibles pour pouvoir mener des inférences précises (les simulations ont montré que c'est à partir d'un échantillon d'environ 5 000 individus que les estimations deviennent précises), les résultats sont donc fragiles. Les autres sources d'incertitude concernent à la fois les estimations pour les hommes et pour les femmes : l'une est que l'on ne connaît les anciennetés qu'à un an près. D'après les simulations présentées, le fait de ne connaître les anciennetés qu'à un an près semble résulter en un biais de 0.5 an environ pour l'estimation de la moyenne des durées, mais cela n'affecte pas l'estimation de la médiane. Si l'on veut tenir compte de ce résultat observé à partir de plusieurs cas simulés, il faut donc rajouter 0.5 an aux estimations des durées moyennes. La seconde concerne l'estimation des flux d'entrée dans la situation : ils ont été estimés à partir d'une autre source, *Erffi*, dont les effectifs sont beaucoup plus faibles que ceux de l'*EFL* et l'impact d'une erreur d'estimation de ces flux sur les résultats finaux n'est pas mesuré ici.

* *
*

Revenons pour finir sur la méthode proposée pour l'estimation des durées passées en famille monoparentale, et d'abord sur les trois hypothèses formulées dans le cadre de la modélisation.

La première est que la durée de monoparentalité ne dépend pas du rang de l'épisode de

monoparentalité. Cette hypothèse semble *a priori* peu vraisemblable, car on peut penser que lors d'une deuxième expérience de monoparentalité, les enfants sont plus âgés, ce qui limite mécaniquement la durée passée en famille monoparentale pour les personnes qui ne reforment pas un couple cohabitant (d'après la définition retenue ici, les enfants doivent être âgés de moins de 18 ans). Néanmoins, la majorité des adultes étant à la tête d'une famille monoparentale ne vivent qu'une fois cette situation. Selon l'enquête *Erfi*, seules 16 % des personnes entre 18 et 72 ans en 2005 ayant connu au moins un épisode de monoparentalité au cours de leur vie, ont vécu deux fois cette situation. L'hypothèse formulée apparaît donc raisonnable.

La deuxième hypothèse suppose que la loi de la durée passée en famille monoparentale n'évolue pas avec le temps. Ceci semble peu probable, car les raisons de la formation de familles monoparentales évoluent avec le temps : il y a de moins en moins de veuves et de femmes ayant eu un enfant hors couple et de plus en plus de personnes séparées (selon l'enquête *Erfi*). On a vu que ces différentes causes d'entrée présentent des durées hétérogènes, ce qui fait que les durées passées en situation monoparentale évoluent avec la structure de la population des familles monoparentales, donc changent avec le temps. Il y a alors une corrélation entre la date d'entrée en famille monoparentale et la durée de la monoparentalité. On a tenté de comprendre l'effet de cette corrélation sur les estimations en s'appuyant sur des simulations (voir annexe 3), qui génèrent de façon aléatoire des durées corrélées soit positivement, soit négativement, avec la date d'entrée en famille monoparentale. Ainsi, si la durée moyenne de monoparentalité diminue avec la date d'entrée dans la situation, alors les estimations du modèle sous-estiment la moyenne des durées et au contraire, si la corrélation est positive, on surestime la moyenne.

Enfin la troisième hypothèse porte sur la proportionnalité des risques instantanés. Contrairement au modèle de Cox, nous ne disposons pas ici de test permettant de confirmer ou d'infirmer cette hypothèse. Il est néanmoins possible d'avoir une idée de sa validité en estimant séparément les courbes de risque instantané pour chaque sous-population. On a vu par exemple que cette hypothèse ne semble pas être vérifiée lorsque l'on veut distinguer les femmes des hommes.

Plus généralement, il convient de mentionner certaines limites de la méthode d'estimation développée dans cet article. Premièrement, il faut disposer d'échantillons de grande taille pour pouvoir obtenir des résultats fiables. C'est une limite importante, car il n'est alors pas possible d'appliquer cette méthode sur des enquêtes ou sur des sous-populations trop petites. Au niveau de la modélisation ensuite, il apparaît difficile de prendre en compte conjointement plusieurs variables explicatives, comme c'est normalement le cas dans un modèle de Cox à risques instantanés proportionnels. Il faudrait en effet pour cela pouvoir estimer des flux d'entrée selon le croisement de plusieurs variables, or les effectifs de l'enquête *Erfi* ne permettent pas d'avoir une si grande précision. Il en résulte qu'on n'a pas pris ici en compte l'interaction des différentes variables explicatives sur la probabilité instantanée de sortir de la monoparentalité. Il n'est donc pas possible d'estimer des effets « toutes choses égales par ailleurs ». De plus, même si la modélisation du risque instantané par une fonction constante par morceaux laisse beaucoup de liberté, on a contraint ce risque instantané à être constant sur des périodes de 3 ans, pour les besoins de l'estimation. En effet, s'il y a trop de paramètres à estimer pour le risque instantané, cela peut nuire à la précision des estimations.

La limite principale de la méthode présentée reste néanmoins qu'il faut connaître les flux d'entrée dans la situation que l'on étudie. Or ces flux ne peuvent pas être déduits d'une enquête par échantillonnage dans le stock. Il faut donc disposer d'une autre source d'information pour établir ces flux. La nécessité de connaître les flux associés à la situation d'intérêt est ainsi le point faible de la démarche. Toutefois, si l'on ne connaît pas ces flux, il est toujours possible soit d'en avoir une idée *a priori*, soit d'élaborer plusieurs scénarios. On pourrait même imaginer un modèle bayésien qui, partant d'une distribution *a priori* sur les flux d'entrée, estimerait une distribution *a posteriori* des durées.

Dans de nombreux domaines, les durées sont très difficiles à mesurer car la plupart des données résultent d'échantillonnages dans le stock. Malgré ces limites, la méthode développée dans cet article reste intéressante, offrant un moyen simple d'obtenir des estimations sur les durées à partir des anciennetés observées. □

BIBLIOGRAPHIE

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

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

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

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

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

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

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

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

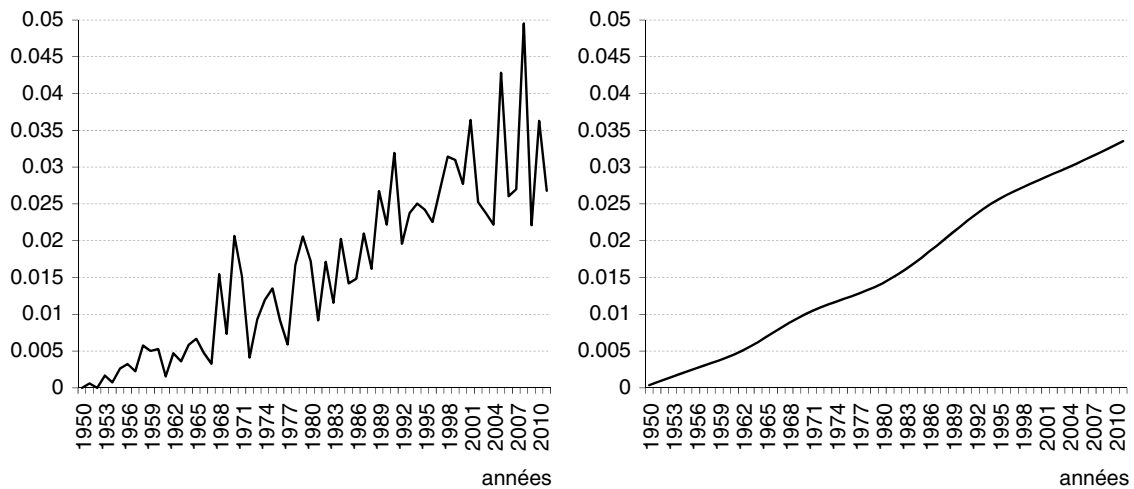
Toulemon, L. (2011). Individus, familles, ménages, logements : les compter, les décrire. *Travail, Genre et Société*, 26, 47–66.

Vespa, J., Lewis, J. M. & Kreider, R. M. (2013). America's families and living arrangements: 2012. *Current Population Reports*, P20-570. Washington, DC: U.S. Census Bureau.
<https://www.census.gov/library/publications/2013/demo/p20-570.html>.

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

LES ESTIMATIONS LISSÉES DES FLUX D'ENTRÉE

Figure A1-I

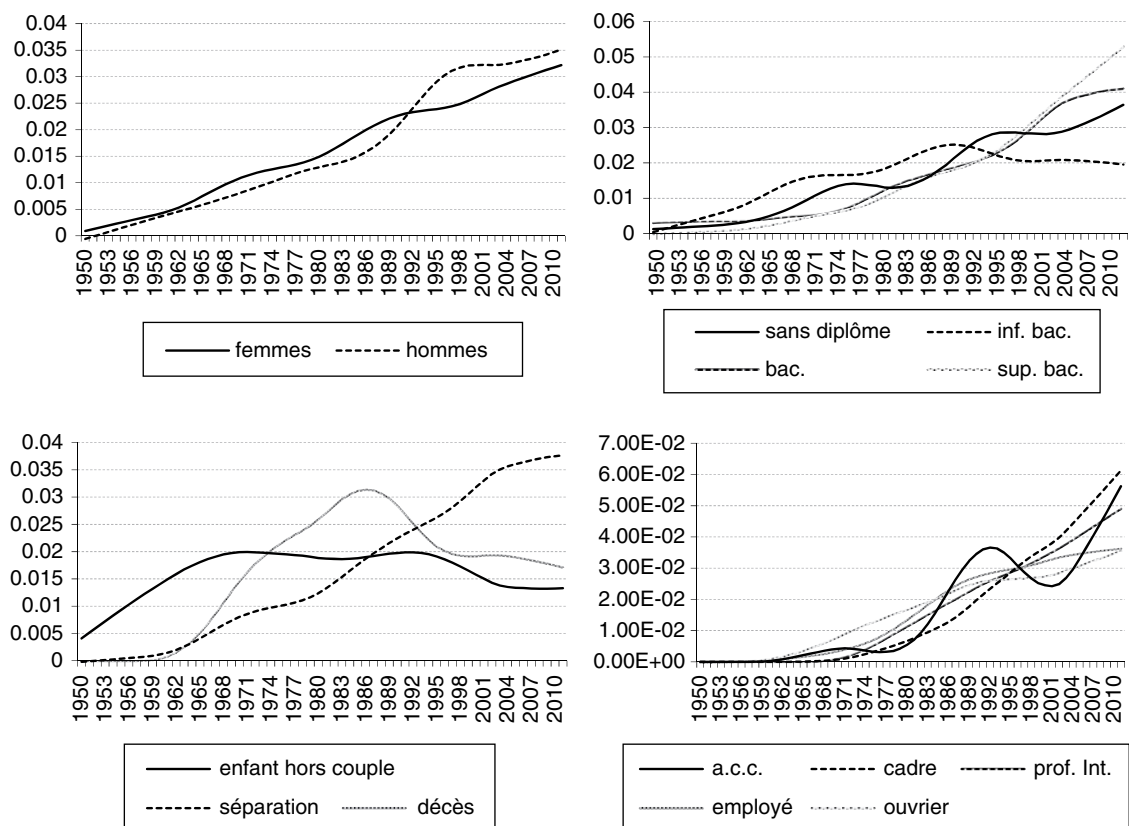
Estimations des flux annuels d'entrée en famille monoparentale, avant et après lissage

Note : le lissage est obtenu par une moyenne mobile suivie d'une approximation polynomiale.

Champ : France métropolitaine.

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

Figure A1-II

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

Note : le lissage est obtenu par une moyenne mobile suivie d'une approximation polynomiale. La catégorie sociale a.c.c. correspond aux artisans, commerçants et chefs d'entreprises.

Champ : France métropolitaine.

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

ANNEXE 2

ESTIMATIONS DES FONCTIONS DE SURVIE ASSOCIÉES AUX DURÉES DE MONOPARENTALITÉ

Tableau A2-1

Fonctions de survie estimées dans le modèle sans covariables

	Ensemble			Hommes			Femmes		
	estimation	borne inférieure	borne supérieure	estimation	borne inférieure	borne supérieure	estimation	borne inférieure	borne supérieure
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

Champ : France métropolitaine.

Source : Insee, *Enquête famille et logements (EFL)* 2011 ; Ined-Insee, *Étude des relations familiales et intergénérationnelles (Erfi)*, vagues 1 et 3, 2005 et 2011.

Tableau A2-2

Fonctions de survie estimées pour les femmes, associées aux risques instantanés de base lorsque l'on considère un ensemble de covariables (cause d'entrée, diplôme ou catégorie sociale)

	Cause d'entrée			Diplôme			Catégorie sociale		
	estimation	borne inférieure	borne supérieure	estimation	borne inférieure	borne supérieure	estimation	borne inférieure	borne supérieure
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

Champ : femmes, France métropolitaine.

Source : Insee, *Enquête famille et logements (EFL)* 2011 ; Ined-Insee, *Étude des relations familiales et intergénérationnelles (Erfi)*, vagues 1 et 3, 2005 et 2011.

ANNEXE 3

SIMULATION DES CORRÉLATIONS ENTRE D ET T

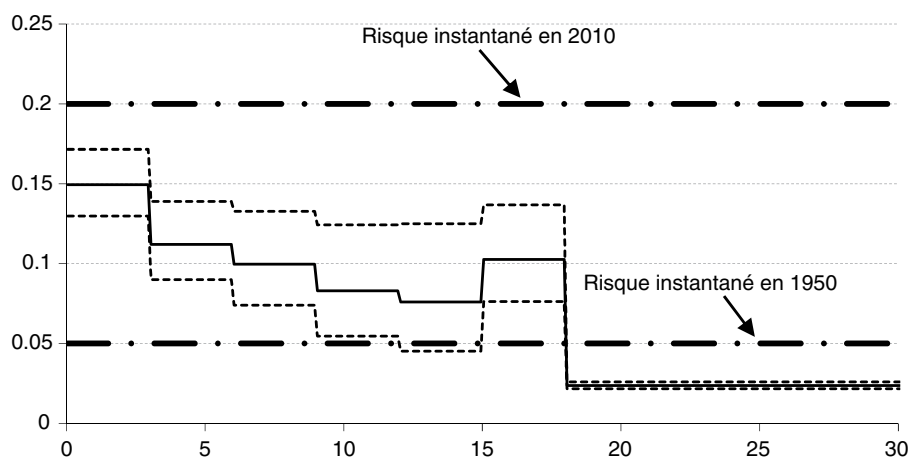
Pour tenter de saisir ce qui se passe lorsque l'on relâche l'hypothèse d'indépendance entre la variable de flux D et la variable de durée T , on simule une population de 100 000 personnes entrant en famille monoparentale de manière uniforme entre 1950 et 2010, mais dont la durée est négativement corrélée à la date d'entrée. Pour cela, on simule T selon une loi de Weibull de paramètre de forme constant qui vaut 1 (cas particulier pour lequel la loi est exponentielle) et un paramètre d'échelle qui vaut $a-b.D$, où $a=507.5$ et $b=0.25$. Ainsi, pour les personnes entrées en monoparentalité en 1950, la durée moyenne est de 20 ans tandis que pour celles entrées en 2010, la

durée moyenne est de 5 ans (bien sûr cette situation est exagérée et ne correspond pas à la réalité). Sur la population simulée, la durée moyenne est de 12.4 ans.

La durée moyenne estimée est de 8.2 ans. On a donc tendance à sous-estimer les durées lorsqu'il y a une corrélation négative entre la durée T et la date d'entrée en famille monoparentale D . Néanmoins, les estimations donnent bien un risque instantané et une survie qui se situent entre les risques instantanés et survies extrêmes, à savoir ceux de 1950 et de 2010 (figures A3-I et A3-II).

Figure A3-I

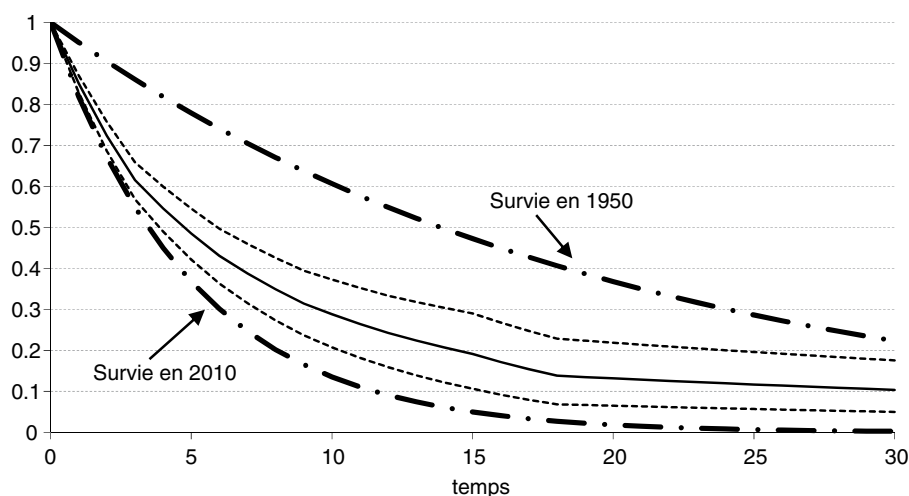
Estimation du risque instantané en présence d'une corrélation entre la variable D et la variable T



Note : on a simulé pour 100 000 personnes des durées dont la loi dépend de la date D d'entrée en famille monoparentale selon une loi de Weibull(1, $a-b.D$) où a et b sont des coefficients valant 507.5 et 0.25. Les traits en pointillés représentent l'intervalle de confiance à 95 % obtenu par la méthode du maximum de vraisemblance.

Figure A3-II

Estimation de la survie en présence d'une corrélation entre la variable D et la variable T



Note : on a simulé pour 100 000 personnes des durées dont la loi dépend de la date D d'entrée en famille monoparentale selon une loi de Weibull(1, $a-b.D$) où a et b sont des coefficients valant 507.5 et 0.25. Les traits en petits pointillés représentent l'intervalle de confiance à 95 % obtenu par la méthode du maximum de vraisemblance.

BIAIS DE L'ESTIMATEUR DE LA MOYENNE

Supposons qu'on dispose d'une variable de durée continue notée T_c et d'une autre variable de durée discrète, ne prenant que des valeurs entières, notée T_d . Supposons de plus que la fonction de survie S_d de la variable

discrète est égale, pour les durées à valeurs entières, à la fonction de survie S_c de la variable continue. Autrement dit, $P(T_d \geq t) = P(T_c \geq t) \forall t \in \mathbb{N}$. Alors on montre que $E[T_d] \approx E[T_c] - 0,5$.

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

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

Si on approxime $S_c(x)$ par une fonction linéaire entre $u-1$ et u , alors on a $S_c(x) - S_c(u) \approx (S_c(u-1) - S_c(u))(u-x)$ et donc $R(u) \approx 0,5(S_c(u) - S_c(u-1))$. D'où $R = \sum_{u \geq 1} R(u) \approx 0,5$.

On en conclut que $E[T_d] \approx E[T_c] - 0,5$.

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 1979-2001*
- 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.

Economie Statistique **ET**

Economics **AND** Statistics

Au sommaire
du prochain numéro :
La crise 10 ans après

Forthcoming:
The crisis 10 years after

