



Institut des
Politiques Publiques

RAPPORT IPP N° 27 – Septembre 2020

Évaluation des effets du dispositif Cifre sur les entreprises et les doctorants participants

Arthur GUILLOUZOUIC
Clément MALGOUYRES

A large, semi-transparent version of the IPP logo is positioned in the bottom right corner of the page. It features the letters 'ipp' in a dark teal color, flanked by two lighter teal curved lines above and below, mirroring the design of the main logo.



RAPPORT IPP N° 27 – Septembre 2020

Évaluation des effets du dispositif Cifre sur les entreprises et les doctorants participants

Arthur GUILLOUZOUIC
Clément MALGOUYRES



L'Institut des politiques publiques (IPP) a été créé par l'École d'économie de Paris (PSE) et est développé dans le cadre d'un partenariat scientifique entre PSE et le Groupe des écoles nationales d'économie et statistique (GENES) avec le soutien de l'ANR au titre du programme Investissements d'avenir. L'IPP vise à promouvoir l'analyse et l'évaluation quantitatives des politiques publiques en s'appuyant sur les méthodes les plus récentes de la recherche en économie.

www.ipp.eu



LES AUTEURS DU RAPPORT

Arthur Guillouzouic est économiste à l'Institut des politiques publiques (IPP). Ancien élève de l'École Normale Supérieure de Cachan et titulaire d'un doctorat en sciences économiques de Sciences Po, ses recherches portent sur les réseaux d'innovateurs, les choix de localisation des firmes innovantes, ainsi que sur les externalités locales générées par le service public.

Page personnelle : <https://arthurguileco.wixsite.com/aguillouzouiclecorff>

Clément Malgouyres est économiste à l'Institut des politiques publiques (IPP) et chercheur associé à l'École d'économie de Paris (PSE). Il s'intéresse à des questions empiriques en économie publique, économie internationale et économie du travail. Ses travaux récents portent sur la fiscalité des entreprises et l'évaluation quantitative des politiques publiques.

Page personnelle : <https://sites.google.com/site/clementmalgouyres>

REMERCIEMENTS

Nous souhaitons remercier chaleureusement plusieurs personnes et institutions ayant permis la réalisation de ce rapport dans les meilleures conditions.

Nous remercions tout d'abord Luis Miotti d'avoir bien voulu nous faire bénéficier de sa connaissance du programme Cifre, ainsi que de son expertise sur l'enquête Génération, qui nous aurons toutes deux été précieuses.

Nous sommes ensuite très heureux d'avoir pu travailler avec la Direction Générale des Entreprises, et plus particulièrement avec Vincent Dortet-Bernadet, Kymble Christophe, Benjamin Hadjibeyli et Clémence Lenoir, que nous remercions pour leur suivi, leurs suggestions et leur relecture attentive du rapport.

Nous remercions également Pascal Giat de l'ANRT, ainsi que Sylvie Niessen et Estelle Dhont-Peltrault du Ministère de l'enseignement supérieur, de la recherche et de l'innovation, pour leur mise à disposition des données sur le programme Cifre ainsi que leurs nombreux conseils dans l'utilisation de ces données.

Nous remercions également les équipes du comité du secret statistique pour le traitement de nos demandes d'accès ainsi que les équipes du centre d'accès sécurisé aux données (CASD) pour avoir répondu à nos demandes de sorties de façon réactive.

SYNTHÈSE DES RÉSULTATS

Ce rapport vise à appréhender les caractéristiques du programme Cifre et celles de ses participants, ainsi qu'à évaluer les effets du programme sur les entreprises et les doctorants participants. Dans le cadre d'une convention Cifre, l'État apporte un soutien financier à toute structure socioéconomique établie sur le territoire français qui recrute un doctorant pour lui confier une mission de recherche qui constituera le sujet de sa thèse, pour une durée maximale de trois ans. Le doctorat est consacré entièrement au projet de recherche, avec un partage des efforts et du temps entre l'employeur et le laboratoire académique.

À partir de plusieurs appariements, nous effectuons une description du dispositif en trois parties. Nous décrivons en premier lieu l'ampleur du programme, en second lieu nous détaillons les caractéristiques des doctorants qui prennent part à ce programme et enfin, nous décrivons les entreprises qui ont recours au dispositif. Le dispositif Cifre attire un nombre croissant d'entreprises et d'étudiants. Ces doctorants sont âgés de 24 ans en moyenne, principalement français et de sexe masculin, et effectuent majoritairement de la recherche dans les sciences physiques, naturelles et de l'ingénieur.

Les employeurs des doctorants Cifre sont, quant à eux, majoritairement des entreprises, concentrées dans des secteurs industriels et scientifiques. Cette concentration est aussi géographique, dans des bassins avec une forte densité d'entre-

prises de recherche ou bien avec quelques grandes entreprises employant un nombre important de doctorants Cifre.

Nous proposons ensuite une analyse économétrique des effets de la participation au programme Cifre sur la durée des thèses et sur l'insertion sur le marché du travail pour les doctorants participants, à partir de données d'enquête. Les doctorants en Cifre ont une probabilité plus élevée de soutenir leur thèse, avec une hétérogénéité importante selon les disciplines. En nous focalisant sur les docteurs-ingénieurs, nous trouvons que le fait d'avoir réalisé une thèse Cifre génère une augmentation de la probabilité d'être en emploi stable 3 ans après la soutenance de l'ordre de 17 points de pourcentage, ainsi qu'un gain de salaire équivalent à 7 %. Nous proposons différents tests de sensibilité de ces résultats permettant d'évaluer s'ils peuvent être interprétés causalement ou s'ils reflètent au contraire la sélection des participants dans le programme. Ils s'avèrent extrêmement robustes. Nous trouvons, en revanche, que les thèses réalisées dans le cadre du programme Cifre semblent en moyenne moins fructueuses sur des critères universitaires : ces doctorants ont une probabilité plus faible de publier dans une revue à comité de lecture au cours de leur thèse, et une chance plus faible d'être qualifiés aux fonctions de maître de conférence par le Conseil national des universités (CNU) à l'issue de leur thèse.

Enfin, nous évaluons à l'aide de trois méthodes économétriques de nature similaire les effets pour les entreprises de la participation au programme Cifre, en utilisant l'entrée des entreprises dans le dispositif. La plupart des grands groupes français ayant recours à de nombreuses reprises au dispositif et ce avant le début de notre période, notre analyse ne permet pas de les inclure et nous nous concentrons donc sur des échantillons de firmes de taille plus faible. Notre analyse permet de détecter une augmentation du nombre d'ingénieurs de R&D employés par

l'entreprise autour du recours au dispositif, ce qui soutient la validité de nos approches. Outre ces effets, nous mesurons une augmentation plus générale de l'emploi R&D des entreprises traitées dans cette période, montrant que l'accueil d'un doctorant Cifre fait généralement partie d'un mouvement plus large d'expansion des équipes de R&D. Dans la deuxième approche que nous adoptons, nous détectons également une hausse de la valeur ajoutée pour les entreprises bénéficiaires sur la période suivant le recours, hausse qui semble absorbée par une augmentation de la masse salariale et n'a donc pas d'effet sur la profitabilité de l'entreprise.

SOMMAIRE

Remerciements	1
Synthèse des résultats	5
Introduction	9
1 Présentation du dispositif Cifre	13
1.1 Origine et objectifs du dispositif Cifre	13
1.2 Modalités et fonctionnement du dispositif Cifre	15
1.2.1 Conditions pour l'attribution d'une Cifre	15
1.2.2 Fonctionnement du dispositif Cifre	17
2 Description statistique du dispositif Cifre	19
2.1 Description générale du dispositif	19
2.2 Description des doctorants réalisant une thèse Cifre	23
2.3 Description des employeurs recourant au dispositif Cifre	27
2.3.1 Caractéristiques des employeurs	27
2.3.2 Évolution des caractéristiques des entreprises recourant au dispositif Cifre par rapport à des groupes comparables	33
3 Analyse au niveau des docteurs Cifre	39
3.1 Résultats sur la durée de la thèse et la probabilité de soutenance . . .	39
3.2 Insertion sur le marché du travail : résultats issus de l'enquête Géné- nération	44
3.2.1 Données	44
3.2.2 Statistiques descriptives.	46
3.2.3 Approche empirique	49
3.2.4 Effets sur l'insertion sur le marché du travail	50
3.2.5 Analyse de sensibilité	51
3.2.6 Résultats complémentaires et discussion	55

4 Analyse au niveau des entreprises Cifre	61
4.1 Analyse par étude d'événements	62
4.1.1 La méthode des études d'événements	62
4.1.2 Équation d'estimation	64
4.1.3 Résultats	65
4.2 Analyse par différence-de-différences	68
4.2.1 Sélection de l'échantillon	69
4.2.2 Équation d'estimation	71
4.2.3 Résultats	72
4.3 Analyse complémentaire : appariement et différence-de-différences .	77
4.3.1 Description de la procédure d'appariement	78
4.3.2 Statistiques descriptives	80
4.3.3 Résultats	82
Conclusion	85
Annexes	87
A.1 Discussion des hypothèses d'identification de l'étude d'événements	87
A.2 Résultats additionnels pour l'étude d'événements	90
A.3 Contrôles pour l'âge et la taille dans la différence-de-différences . . .	91
Références	95
Liste des tableaux	97
Liste des figures	99

INTRODUCTION

Contexte de l'étude

Dans le cadre de la réglementation européenne, les autorités françaises ont notifié à la Commission européenne un plan d'évaluation du régime cadre 40391 d'aides d'État à la recherche, au développement et à l'innovation (RDI), validé par la décision 2015/4445 de la Commission. Le régime inclut des aides à la R&D, des aides à l'innovation et des aides aux pôles d'innovation. L'objectif de l'évaluation est de mesurer les effets du régime tant sur les bénéficiaires que sur l'économie en général. Cette évaluation doit aboutir à un rapport intermédiaire en octobre 2019 et un rapport final en juin 2020. Le secrétariat de cette évaluation est assuré par la Direction générale des entreprises (DGE) du ministère français de l'Économie et des Finances tandis que sa direction est assurée par un comité de pilotage.

Une Convention industrielle de formation par la recherche (Cifre) est un dispositif de financement de thèse qui aide les entreprises pour le recrutement de doctorants. Au sein du régime SA 40391, le dispositif Cifre fait partie des aides à la recherche et au développement dont l'objectif global est d'augmenter l'effort de R&D des entreprises bénéficiaires. En outre, le dispositif Cifre a pour objectif de contribuer à l'emploi de docteurs par les entreprises et de favoriser les collaborations entre entreprises et laboratoires.

Ce dispositif a déjà fait l'objet d'études descriptives exploitant des résultats

d'enquêtes menées auprès des entreprises, doctorants et laboratoires bénéficiaires¹. Le dispositif Cifre n'a, en revanche, jamais fait l'objet d'une évaluation d'impact économique micro-économétrique telle qu'attendue par la Commission. Le plan d'évaluation des aides du régime SA 40391 fixe le cadre sur la méthodologie et les objectifs du plan d'évaluation. C'est dans ce cadre que s'inscrit cette évaluation économétrique d'impact causal du dispositif Cifre.

Objectifs de l'étude

L'objectif principal des évaluations menées dans le cadre du plan d'évaluation des aides d'État notifié à la Commission européenne est d'apprécier et de quantifier les effets des aides tant sur les bénéficiaires que sur l'économie en général. L'étude porte sur une période plus vaste que celle attendue par la Commission et couvre la période 2008-2018, période pour laquelle des données de bonne qualité au format numérique sont disponibles pour les conventions Cifre. Les bénéficiaires directs de la subvention accordée dans le cadre du dispositif Cifre sont les entreprises. Cette étude étudie ainsi les effets de ce programme sur deux dimensions :

- l'impact causal du dispositif Cifre sur les doctorants sous convention Cifre. Les effets étudiés portent ainsi sur la durée des thèses, l'employabilité des doctorants ayant soutenu leur thèse à la fin de leur Cifre ainsi que sur la nature et les caractéristiques de l'emploi obtenu (salaires, nature du poste). Cette analyse est réalisée sur le sous-groupe des ingénieurs-docteurs, permettant d'obtenir une réelle homogénéité des parcours entre le groupe traité et le groupe de contrôle.
- l'impact causal du dispositif Cifre sur les entreprises bénéficiaires. En particulier, il s'agit d'estimer l'impact direct du dispositif Cifre en termes de R&D dans l'entreprise : évolution du nombre d'emplois R&D, de la rémunération

1. Voir par exemple Levy (2005)

des emplois R&D ou d'autres formes de dépenses de R&D. L'étude s'attache aussi à estimer les impacts indirects sur le développement économique de l'entreprise, notamment la valeur ajoutée, le chiffre d'affaires et la profitabilité.

Un objectif complémentaire d'identification des effets sur les laboratoires participants avait été initialement formulé, mais n'a pas été poursuivi faute de données permettant d'en faire l'analyse.

Plan du rapport

Chapitre 1 : Présentation du dispositif Cifre

Chapitre 2 : Description statistique du dispositif Cifre

Chapitre 3 : Analyse au niveau des docteurs

Chapitre 4 : Analyse au niveau des entreprises

CHAPITRE 1

PRÉSENTATION DU DISPOSITIF CIFRE

Ce premier chapitre vise à donner une vue d'ensemble qualitative du dispositif Cifre. Nous présentons tout d'abord les origines et les objectifs de cette convention, dont le but est promouvoir des collaborations entre laboratoires publics et entreprises privées et stimuler la productivité de la recherche scientifique française (section 1.1). Nous précisons ensuite le fonctionnement d'une thèse Cifre, les conditions d'éligibilité des entreprises, des doctorants et des laboratoires, le déroulement annuel et les mécanismes de subvention, prenant en compte à la fois la subvention de l'ANRT et les retours du Crédit Impôt Recherche (CIR) pour le dispositif Cifre (section 1.2).

1.1 Origine et objectifs du dispositif Cifre

C'est dans le contexte d'une politique de rapprochement du système de formation, de recherche et d'innovation avec les milieux socio-économiques français que le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation (MESRI) confie à l'Association nationale recherche technologie (ANRT) la mise en oeuvre des Conventions industrielles de formation par la recherche (Cifre) en 1981.

Dans le cadre d'une convention Cifre, l'État apporte un soutien financier à toute

structure socioéconomique (hors État) établie sur le territoire français qui recrute un doctorant pour lui confier une mission de recherche qui constituera le sujet de sa thèse, pour une durée maximale de trois ans. Le doctorat est consacré entièrement au projet de recherche, avec un partage des efforts et du temps entre l'employeur et le laboratoire académique. Il y a donc trois bénéficiaires de cette convention : l'entreprise, qui reçoit un soutien de l'État, le laboratoire, qui accueille un doctorant pour mener des recherches en son sein, et le doctorant, qui reçoit une subvention pendant trois ans pour son projet et bénéficie d'une double formation académique et professionnelle. L'État subventionne donc des emplois qualifiés au sein de ces entreprises, qui en contrepartie de cette aide à la recherche s'engagent à former et rémunérer le doctorant.

Ce programme a été lancé dans les années 1980, dans une configuration de la recherche en France qui affichait une séparation franche entre le monde de la recherche universitaire publique et celui des entreprises privées. Se fondant sur l'exemple américain, le secrétaire d'État chargé de la Recherche, Pierre Aigrain, présente alors en 1980 comme condition de l'amélioration de la science et technique française « le renforcement des relations entre les universités, les établissements de recherche et les entreprises. » Le projet du dispositif Cifre voit donc le jour, dans le but de tisser et renforcer les liens entre laboratoires publics et entreprises privées par le biais des doctorants. S'il n'y a pas de limite sur le secteur d'activité de l'entreprise, et que le champ de disciplines des thèses des doctorants en Cifre s'est étendu ces dernières années, la recherche effectuée par ces doctorants est à majorité dans le domaine scientifique et technique, c'est-à-dire dans les sciences physiques, les sciences de l'ingénieur et les sciences naturelles.

Au-delà des bénéfices pour le monde de la recherche, le dispositif a des conséquences sur l'intégration professionnelle des doctorants. Les doctorants français ont en effet des débouchés limités, à la fois en terme de rémunération et de stabilité de l'emploi, comme le montrent Margolis et Miotti (2015), ce qui a entraîné

la mise en place de dispositifs incitatifs comme le dispositif "Jeunes Docteurs". La thèse Cifre représente donc un vecteur d'intégration pour ces doctorants qui profitent de trois ans d'expérience professionnelle au sein d'entreprises qui sont parfois leurs futurs employeurs. En effet, selon l'enquête de l'ANRT en 2016 sur les Cifre, 27 % des doctorants Cifre sont toujours employés au sein de l'entreprise d'accueil un an après leur thèse.

1.2 Modalités et fonctionnement du dispositif Cifre

Cette section précise les modalités des conventions Cifre. Il existe un certain nombre de conditions appliquées aux doctorants, aux entreprises et aux laboratoires qui souhaiteraient s'engager dans une Cifre. Une fois cette Cifre validée, sa continuation et le versement de la subvention, que nous détaillerons, sont aussi soumis à diverses conditions.

1.2.1 Conditions pour l'attribution d'une Cifre

Un certain nombre de conditions s'appliquent aux entreprises, aux candidats et aux laboratoires souhaitant participer à une Cifre, afin d'en limiter la portée à des projets de recherche présentant une véritable valeur ajoutée, et au sein d'entreprises capables de s'engager sur l'emploi d'un doctorant Cifre. Nous obtenons ces conditions du document sur les conditions d'octroi d'une Cifre de l'ANRT (2019).

Les doctorants postulant à une Cifre doivent être titulaires d'au moins un diplôme de niveau Master (Bac + 5). Ils s'engagent à préparer une thèse, mais ne doivent pas avoir déjà été en poste pour le même employeur depuis plus de 9 mois – il doit s'agir du début des travaux. Il n'y a pas de restriction d'âge ou de nationalité; cependant les doctorants doivent disposer d'une autorisation de travail en France. Ils ne peuvent pas avoir déjà travaillé pour l'entreprise avant la convention Cifre, sauf dans le cadre d'emplois ponctuels ou saisonniers sans rap-

port avec leur sujet de recherche. Cette restriction est présente afin de favoriser les nouveaux échanges entre doctorants et entreprises et de soutenir un projet de thèse original, et non pas subventionner un emploi qui existerait déjà.

En ce qui concerne les employeurs, ils doivent être domiciliés sur le territoire français. Le sujet de recherche du doctorant doit s'inscrire clairement dans l'objet de l'entité d'accueil (secteur d'activité, travaux en cours). Les entreprises éligibles doivent aussi témoigner d'une capacité d'accompagnement et d'encadrement suffisante (notamment selon le profil du tuteur scientifique choisi) : la décision de l'ANRT à ce sujet est prise en partie selon un rapport du délégué régional de recherche et technologie (DRRT). Ces entreprises s'engagent aussi à rémunérer les doctorants avec un salaire minimum déterminé par l'ANRT, qui était de 23 484 euros en 2018, et qui évolue chaque année, dans le cadre d'un CDD ou CDI. Les Cifre peuvent aussi être mises en place dans des organismes publics, à l'exception des services de l'État, des autorités administratives indépendantes (AAI) et des autorités publiques indépendantes (API). Sont donc éligibles les collectivités territoriales, les établissements publics industriels et commerciaux (EPIC), les établissements publics administratifs (EPA), et les groupements d'intérêts publics (GIP). Les associations sont également des employeurs éligibles.

Les laboratoires de recherche sont ceux rattachés aux établissements doctoraux dans lesquels sont inscrits les doctorants. Un laboratoire étranger peut être associé à un laboratoire français dans le cadre d'une Cifre.

Les dossiers sont présentés à l'ANRT, qui donne un avis favorable ou défavorable. Cet avis dépend de deux critères :

- Un avis sur la qualité scientifique du projet : l'ANRT statue en prenant en compte la qualité du projet du doctorant, son lien avec l'activité et la recherche de l'entreprise et l'adéquation du laboratoire de recherche choisi par rapport aux disciplines engagées. Pour chaque demande de Cifre, l'ANRT mandate une évaluation de la valeur scientifique du projet auprès d'un ex-

pert scientifique, qui rend un avis consultatif.

- La pertinence économique de la convention : la solidité financière de l'entreprise et son implication, surtout pour les entreprises n'ayant jamais bénéficié d'une Cifre, n'en ayant pas bénéficié depuis cinq ans, ou bien en situation financière précaire. C'est le DRRT qui rend un avis à l'ANRT et au comité d'évaluation et de suivi (CES), qui détient le pouvoir final de décision.

Les employeurs recrutant un nombre élevé de Cifre (plus de six par an en moyenne sur trois ans) et ayant obtenu l'accord de l'ANRT peuvent passer par une procédure simplifiée de recrutement et de convention.

1.2.2 Fonctionnement du dispositif Cifre

Lorsqu'une Cifre est validée, l'entreprise reçoit une subvention d'État d'un montant de 14 000 euros par an et par contrat Cifre (montant en 2018). Chaque convention Cifre est valable pour 36 mois dès sa date de validation, et ces subventions sont versées à échéances trimestrielles lors de ces 36 mois et ne sont pas assujetties à la TVA. Chaque année, l'entreprise doit faire parvenir un rapport annuel détaillé présentant l'avancement des travaux de thèse à l'ANRT. La subvention Cifre est conditionnelle à la réception de ce rapport et à sa cohérence par rapport au projet initial de Cifre. Ainsi, si l'ANRT juge que la recherche effectuée ne correspond pas au projet de recherche auquel elle avait donné son accord, elle peut suspendre la Cifre.

Au-delà de la subvention initiale, le dispositif Cifre représente aussi une subvention non négligeable dans le cadre du CIR. Dans le cas d'une Cifre, le CIR est égal à 30 % d'une fois et demi le salaire brut chargé (un coût d'environnement de 43 % du salaire brut chargé représentant les coûts indirects que supporte l'entreprise), dont on soustrait alors la subvention ANRT. Par exemple, un doctorant Cifre rémunéré 23 484 euros par an (le salaire plancher) rapporte 10 595 euros par

an en CIR en plus des 14 000 euros de subvention directe de l'ANRT. Ainsi, pour un salaire de 23 484 euros, le salaire brut chargé est de 32 878 euros – mais la subvention Cifre et le CIR font que le coût final du doctorant est de 8 283 euros par an. Il s'agit donc d'une subvention importante du coût du travail pour ces entreprises.

En continuation de la Cifre, le dispositif jeunes docteurs (DJD) est une incitation supplémentaire à l'emploi des doctorants dans leur entreprise d'accueil. Si le doctorant est embauché en CDI par l'entreprise, qu'il s'agit de son premier CDI et que l'effectif de l'employeur n'a pas diminué, le salaire chargé est compté double dans le calcul du CIR, et l'augmentation d'environnement passe à 100 %. Le DJD s'applique à tous les jeunes docteurs, et se présente comme la suite logique de la Cifre, où l'entreprise peut continuer à employer à moindre coût un doctorant qu'elle connaît déjà et qui est déjà intégré à son organisation.

CHAPITRE 2

DESCRIPTION STATISTIQUE DU DISPOSITIF

CIFRE

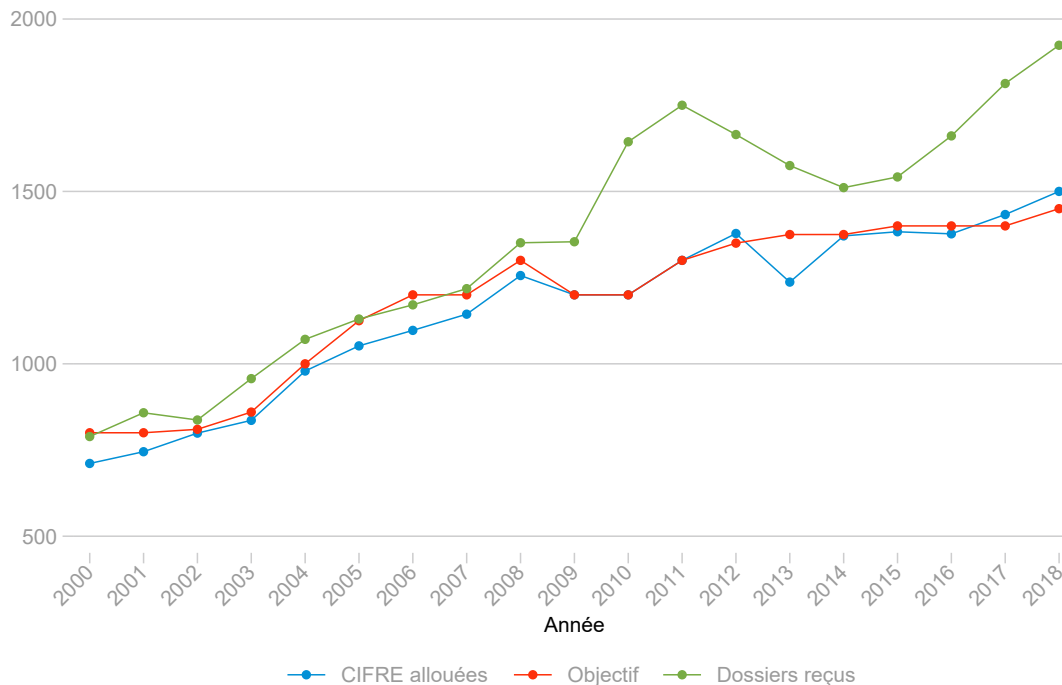
Ce deuxième chapitre est consacré à une description quantitative et statistique du dispositif Cifre, et exploite plusieurs bases de données afin de donner un portrait global du dispositif Cifre, des doctorants et des entreprises qui en font partie. Il ne s'agit pas de mettre en évidence un lien causal, mais simplement de montrer le profil des bénéficiaires du dispositif et les différences entre ceux qui en bénéficient et ceux qui n'en font pas partie.

Nous présentons donc l'évolution du programme Cifre au cours du temps, en termes de volume, de concentration et de répartition des Cifre géographiquement (section 2.1). Par ailleurs, nous décrivons les caractéristiques des doctorants participant au programme (section 2.2). De manière similaire, nous présentons le profil des entreprises qui embauchent des doctorants en Cifre (section 2.3).

2.1 Description générale du dispositif

Dans cette première section, nous nous intéressons à une description générale du dispositif Cifre et de son évolution. Depuis sa création, le nombre de conven-

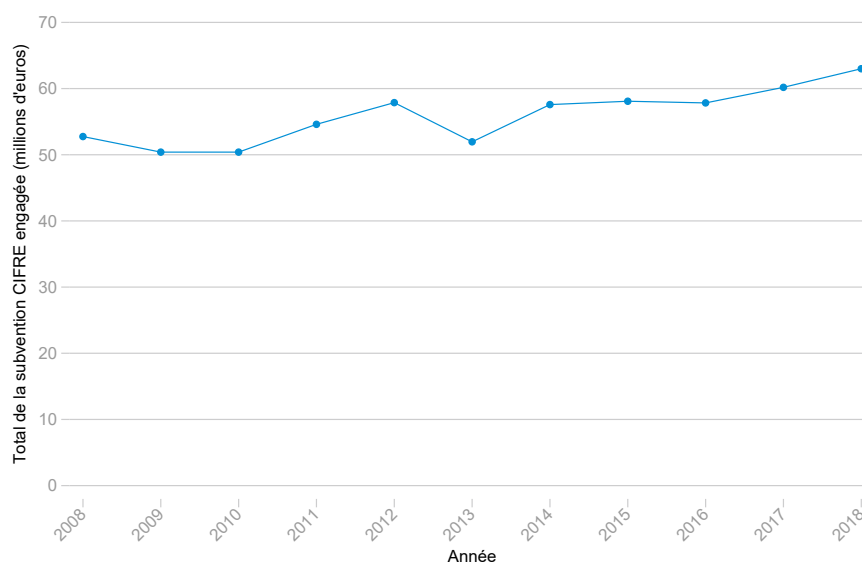
FIGURE 2.1 – Évolution du nombre de Cifre validées.



NOTES : SOURCES : Données ANRT.

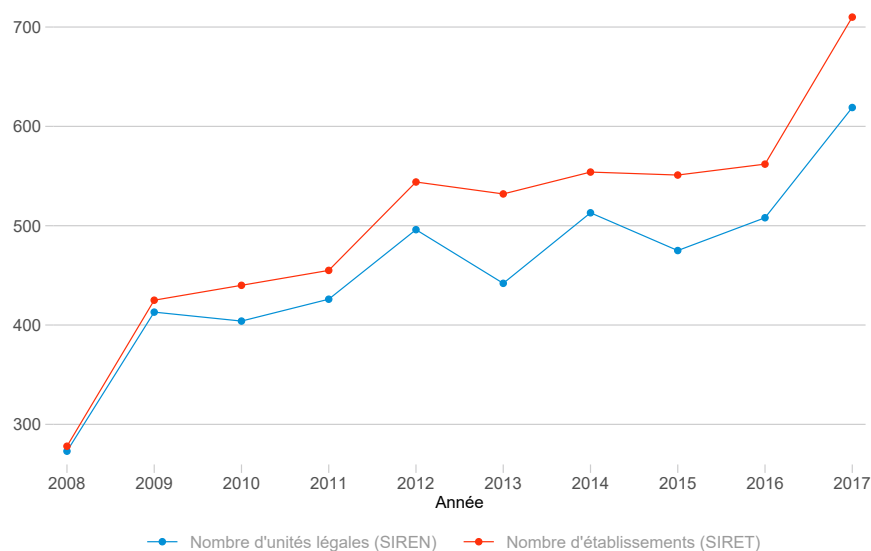
tions Cifre a connu une augmentation régulière. Si depuis ses débuts, le dispositif Cifre s'est étendu rapidement, et que depuis 2003 le nombre de dossiers envoyés a connu une forte augmentation (de 500 dossiers à près de 2000 récemment), le nombre de nouvelles conventions Cifre validées depuis 2008 connaît une croissance plus faible mais constante, comme l'indique la figure 2.1, malgré une légère baisse peu après 2008 probablement liée à la crise. De 2008 à 2018, le nombre de Cifre par an est passé d'environ 800 à près de 1 500. Le nombre de contrats actifs a donc connu une augmentation logique, et était en 2018 supérieur à 4 000 conventions. La thèse Cifre est donc un dispositif de plus en plus attractif pour étudiants, entreprises et laboratoires, et dont le coût pour l'État a augmenté. En effet, en se limitant uniquement aux subventions de l'ANRT et excluant les déductions additionnelles au titre du Crédit impôt recherche (CIR), le montant total de la subvention Cifre dépassait en 2018 les 60 millions d'euros.

FIGURE 2.2 – Évolution de la subvention ANRT.



SOURCES : Données ANRT.

FIGURE 2.3 – Évolution du nombre d'entreprises ayant recours à une convention Cifre.



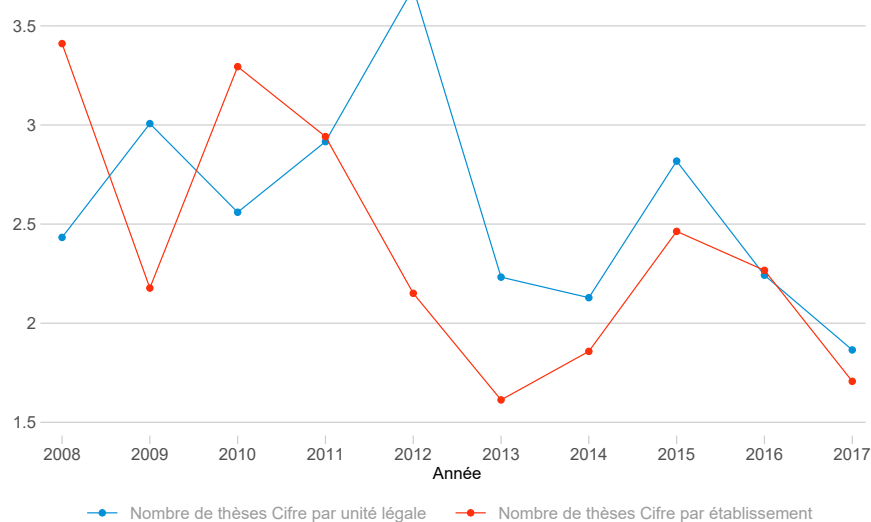
SOURCES : Base Cifre.

Deux dynamiques sous-jacentes peuvent expliquer cette croissance du côté des entreprises. Il peut s'agir d'une augmentation du nombre de conventions Cifre que

signe chaque entreprise ; dans ce cas nous devrions voir le nombre de Cifre moyen par entreprise augmenter au cours du temps. Il peut aussi être le résultat d'une augmentation du nombre d'entreprises qui accueillent un doctorant en convention Cifre.

La figure 2.3 montre l'évolution du nombre d'entreprises ayant accueilli une Cifre par année. On remarque une très nette augmentation du nombre d'entreprises signant une convention, puisque ce chiffre a plus que doublé en dix ans. La figure 2.4 montre aussi que cette augmentation n'est pas liée au nombre moyen de Cifre par entité, qui ne montre pas de croissance mais au contraire une diminution sur la dernière décennie. En 2017, plus de 700 entités (niveau SIRET, c'est-à-dire établissement) ont signé une convention Cifre, et chacune accueillait en moyenne 1,7 Cifre. La valeur médiane du nombre de Cifre par SIRET est aujourd'hui d'un seul doctorant : la majorité des établissements n'hébergent qu'une Cifre.

FIGURE 2.4 – Évolution du nombre de Cifre validées par entreprise.



SOURCES : Base Cifre.

La figure 2.4 représente le montant total de subvention engagé par l'ANRT par an sur la période 2008-2017. Pour déterminer le montant total engagé, on suppose que chaque Cifre correspond à un montant de 42 000 euros d'engagement, c'est-à-

dire 3 ans de subvention Cifre (le montant versé *in fine* peut différer, par exemple si la thèse est terminée en moins de 3 ans ou si la convention est rompue). Chaque convention est dénombrée l'année de sa date d'effet, disponible dans la base de l'ANRT. La série est relativement stable sur la période 2008-2016, le montant de subvention évoluant généralement entre 50 et 60 millions d'euros.

2.2 Description des doctorants réalisant une thèse Cifre

Cette section se concentre sur les caractéristiques des doctorants qui entreprennent une Cifre telles que définies dans la base de données de l'ANRT.

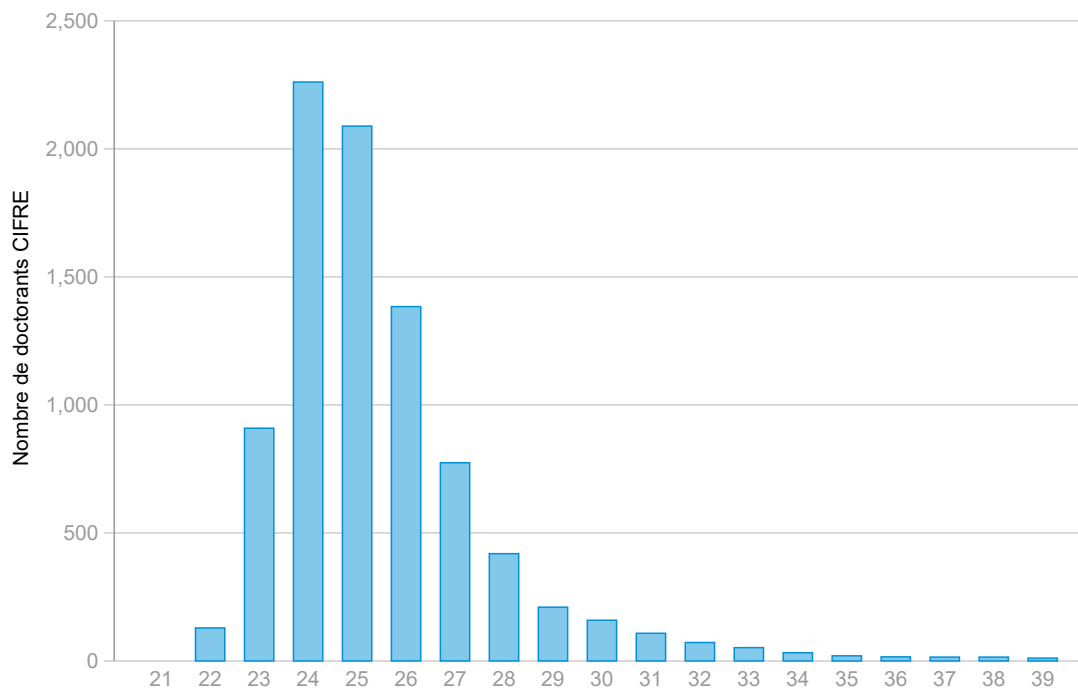
Les statistiques de la base Cifre de l'ANRT permettent en effet de dresser un portrait initial des caractéristiques des doctorants Cifre et de leur évolution. La figure (2.5) montre la distribution de l'âge chez les doctorants Cifre. Ces derniers ont principalement entre 23 et 28 ans à la date d'effet de leur Cifre, avec un pic à 24 - 25 ans. Cependant, on voit quelques étudiants plus âgés, ayant autour de trente ans, bénéficier chaque année de la Cifre.

Les doctorants Cifre sont en majorité français, comme en témoigne la figure 2.6, même si depuis 2008 la proportion d'étrangers effectuant une Cifre augmente et se trouve maintenant au-dessus de 20 %. Comme nous l'avons précisé dans la première partie de ce rapport, il est nécessaire de disposer d'une autorisation de travailler sur le sol français afin d'effectuer une thèse Cifre.

La part d'hommes dans les Cifre est nettement supérieure à celle des femmes. Cependant, la part des femmes montre une légère augmentation depuis 2008 et se situe juste en dessous de la barre des 40 % en 2018, comme le montre la figure 2.7.

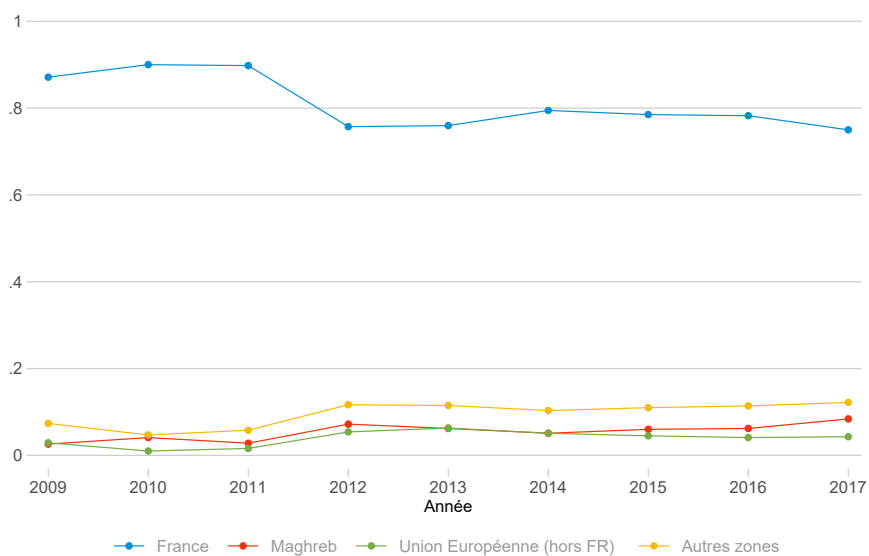
Cette base nous donne aussi des indications sur les disciplines que suivent les doctorants Cifre (figure 2.8). Cette spécification générale montre une concentra-

FIGURE 2.5 – Distribution de l'âge des doctorants (conventions 2011-2018)



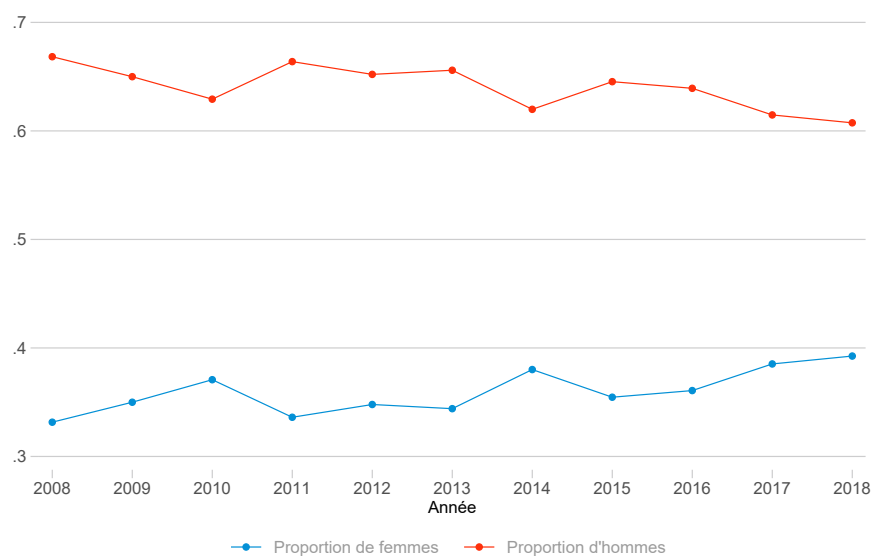
SOURCES : Base Cifre.

FIGURE 2.6 – Nationalité des doctorants Cifre



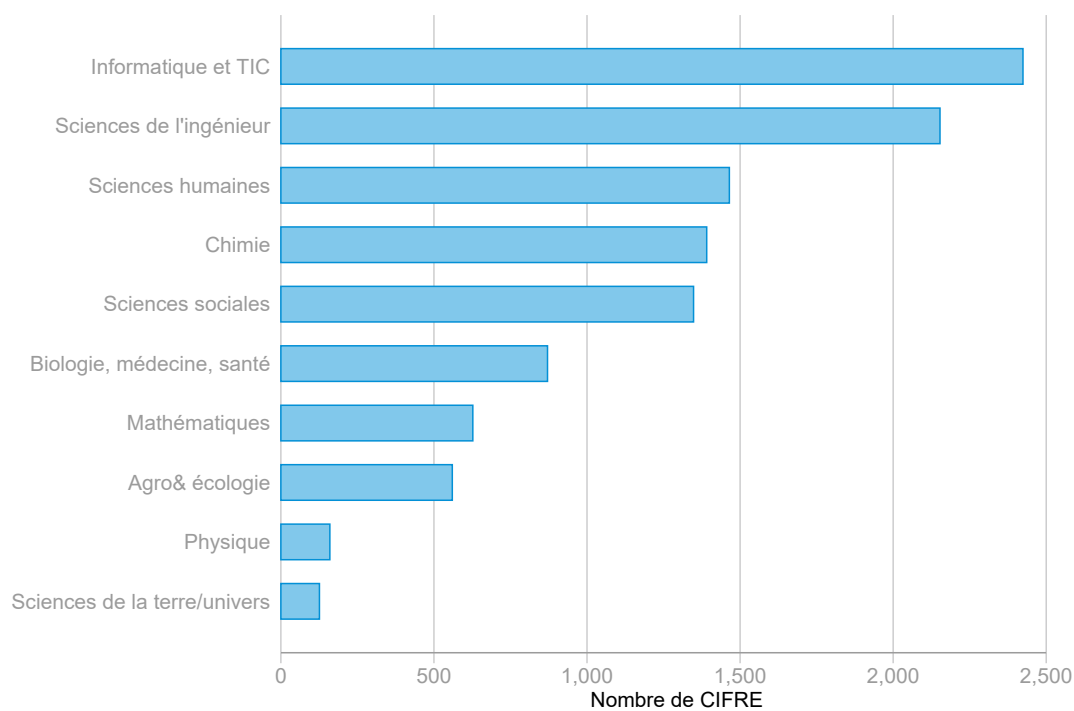
SOURCES : Base Cifre.

FIGURE 2.7 – Proportion de chaque genre parmi les doctorants



SOURCES : Base Cifre.

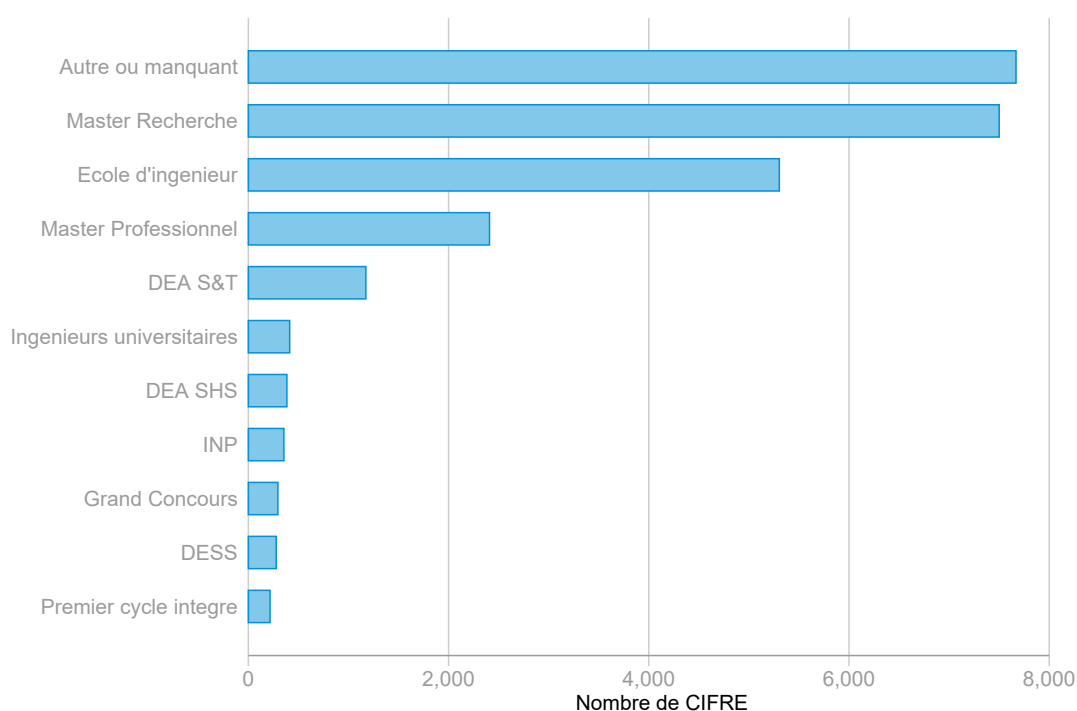
FIGURE 2.8 – Disciplines des doctorants Cifre



SOURCES : Base Cifre.

tion forte autour des sciences dites « STEM », avec les sciences de l'ingénieur et l'informatique en tête. Cependant, il semblerait que le caractère industriel de la recherche présent aux origines du dispositif ne soit plus aussi prégnant, comme en témoigne la proportion non-négligeable des sciences humaines et sociales, qui correspondent à des thèses en droit, sociologie, urbanisme, sciences économiques ou encore démographie.

FIGURE 2.9 – Diplômes des doctorants Cifre



SOURCES : Base Cifre.

Nous observons aussi les différents diplômes que détiennent les doctorants qui s'engagent dans une Cifre, dont la distribution est montrée en figure 2.9. Nous observons ainsi un grand nombre de masters recherche, suivi d'un nombre élevé de diplômes d'ingénieur. Ces diplômes sont suivis d'un plus petit nombre de masters professionnels et d'un DEA science et technologie. L'orientation vers la recherche est donc bien présente vu forte présence de masters recherche et DEA. Il faut néanmoins noter que de nombreux doctorants ne déclarent pas de second diplôme, et qu'il est probable que les détenteurs d'un master recherche soient éga-

lement diplômés, par exemple, d'écoles d'ingénieurs mais ne le mentionnent pas dans l'inscription en thèse. La forte présence des ingénieurs est en effet confirmée par l'analyse que nous menons sur l'enquête Génération, et avait déjà été identifiée dans les travaux de Margolis et Miotti (2015) sur le DJD.

2.3 Description des employeurs recourant au dispositif Cifre

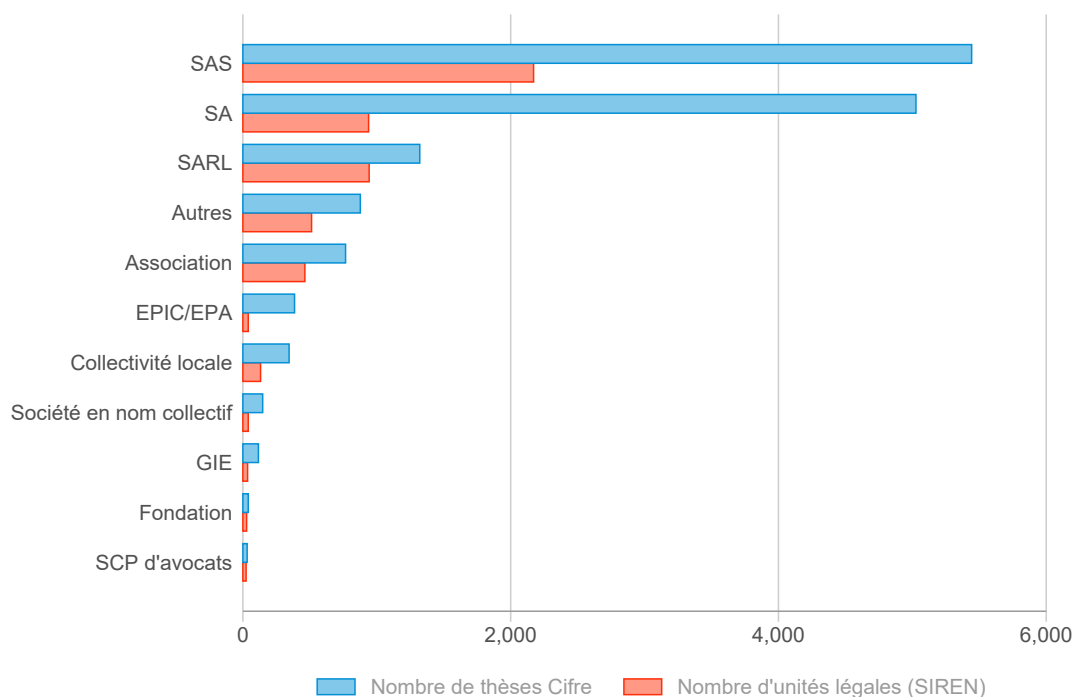
Cette section se concentre sur une description des employeurs recourant à des Cifre. Bien qu'il puisse s'agir de collectivités locales ou encore d'associations autant que d'entreprises, nous désignons pour simplifier ces employeurs comme des entreprises dans ce qui suit. Cette section présente dans un premier temps le profil de ces entreprises, à la fois juridique, d'activité mais aussi géographique. Nous décrivons aussi l'évolution de certains résultats d'intérêt, comme le chiffre d'affaires ou l'emploi, suivant le recrutement de doctorants en Cifre. Nous présentons ensuite une comparaison initiale entre les entreprises traitées et notre groupe de comparaison.

2.3.1 Caractéristiques des employeurs

Nous présentons ici nos résultats concernant les caractéristiques des entreprises accueillant des Cifre. La figure 2.10 montre la distribution des employeurs accueillant des Cifre par catégorie juridique. La vaste majorité des entités d'accueil sont des entreprises (SAS, SA, SARL) : elles représentent un total d'environ 4 000 unités légales, soit près de 80 % du total des employeurs de doctorants Cifre. Les associations ainsi que les collectivités locales (communes, départements) sont également assez nombreuses à recourir au dispositif.

Il est également intéressant de comparer ces chiffres à ceux de la distribution

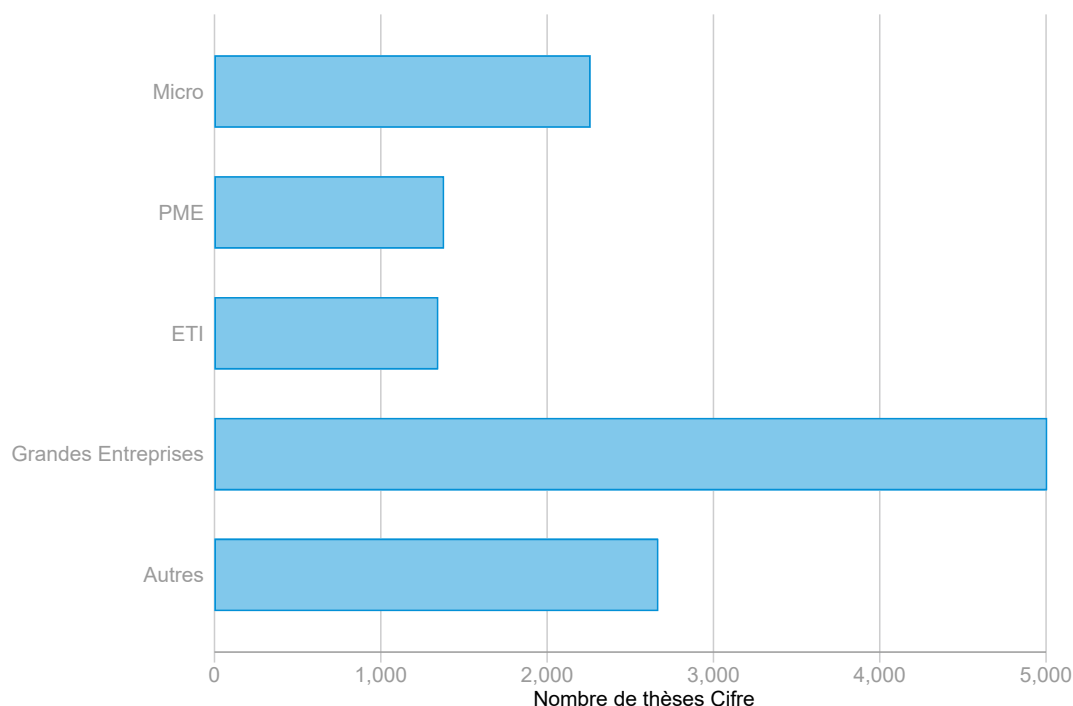
FIGURE 2.10 – Catégorie juridique des employeurs



SOURCES : Base Cifre.

des doctorants dans ces différentes structures. On note notamment que les SA, bien que moitié moins nombreuses que les SAS, prennent presque autant de doctorants Cifre, alors que le nombre de doctorants est très bas pour les SARL qui prennent moins de deux doctorants par entité. Il y a une véritable différence de structure entre ces entreprises, qui fait qu'à nombre égal elles ont des demandes de doctorants très différentes. On peut aussi noter le nombre élevé de doctorants dans les EPIC et EPA, malgré le faible nombre de ces entités.

En complément, la figure 2.11 montre la taille des entreprises accueillant des doctorants Cifre. On constate ainsi l'importance prépondérante des grandes entreprises dans le contingent total des thèses Cifre, malgré des effectifs néanmoins importants dans les petites entreprises. À noter que la catégorie « Autres » contient notamment les associations et l'ensemble des acteurs publics, ainsi que des entreprises dont la taille est inconnue.

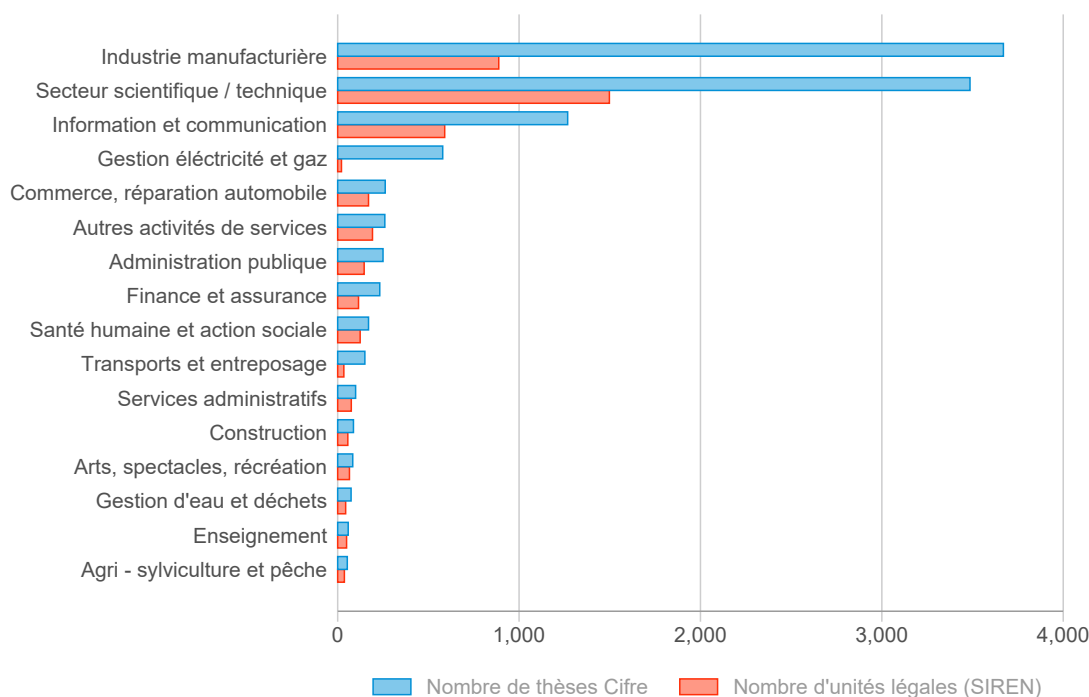
FIGURE 2.11 – Taille des entreprises accueillant des doctorants Cifre

SOURCES : Base Cifre, liasses fiscales des sociétés, Lifi.

La figure 2.12 détaille les secteurs d'activité des entreprises employant des doctorants Cifre. Nous voyons que si les SIREN sont assez bien répartis, une majorité écrasante des Cifre se trouve dans les industries manufacturières et les activités spécialisées et scientifiques. On observe aussi que la production et distribution d'électricité, secteur avec très peu d'entreprises, emploie un grand nombre de doctorants par rapport au nombre de SIREN.

Nous nous intéressons aussi à la distribution géographique des entreprises accueillant une Cifre. Tout d'abord, nous observons la distribution géographique des doctorants Cifre et des entreprises accueillant des Cifre par zone d'emploi, présentées respectivement dans les figures 2.13 et 2.14. Ces cartes montrent notamment les "pôles" de Cifre : l'Île-de-France, notamment autour de Paris, les alentours de Grenoble et de Lannion, ainsi que certaines zones d'emploi dans le Sud de la France. Ces pôles correspondent souvent à des pôles d'innovation spécialisés.

FIGURE 2.12 – Secteur d’activité des employeurs.

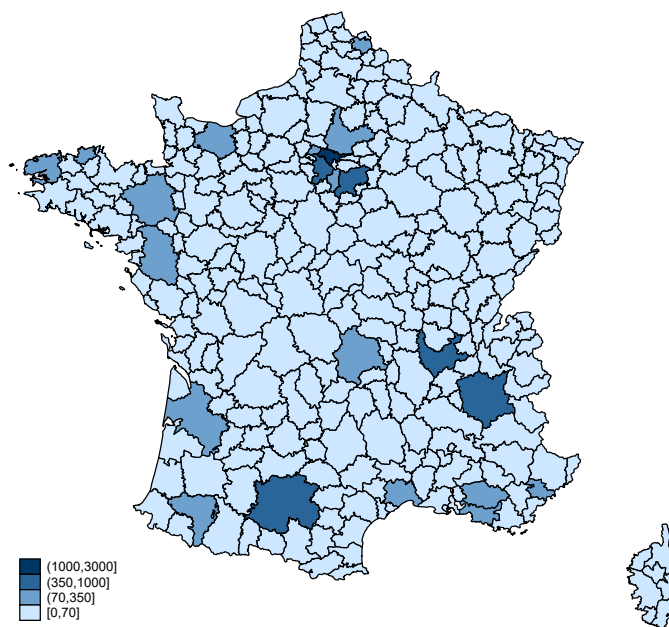


SOURCES : Base Cifre.

Afin d’approcher une hétérogénéité dans le recours au dispositif, nous représentons aussi le ratio de Cifre divisé par le nombre d’entreprises employant des ingénieurs (PCS commençant par 38) dans la figure 2.15, ou employant de la main d’œuvre hautement qualifiée (PCS commençant par 3) dans la figure 2.16. Nous définissons celles-ci comme des entreprises dont la moitié ou plus des emplois sont ingénieurs ou hautement qualifiés. Ceci permet de neutraliser en partie les effets de concentration géographique des Cifre générés mécaniquement par la plus grande concentration d’entreprises innovantes dans ces zones, et permet de mettre en valeur des zones de recours élevé au programme indépendamment de leur composition. Ceci modifie légèrement l’image donnée par les figures 2.13 et 2.14, qui faisait apparaître la plupart des zones d’emploi correspondant à des grandes villes françaises comme des zones recourant beaucoup au programme. Elle confirme en revanche la très haute concentration dans certaines zones, comme

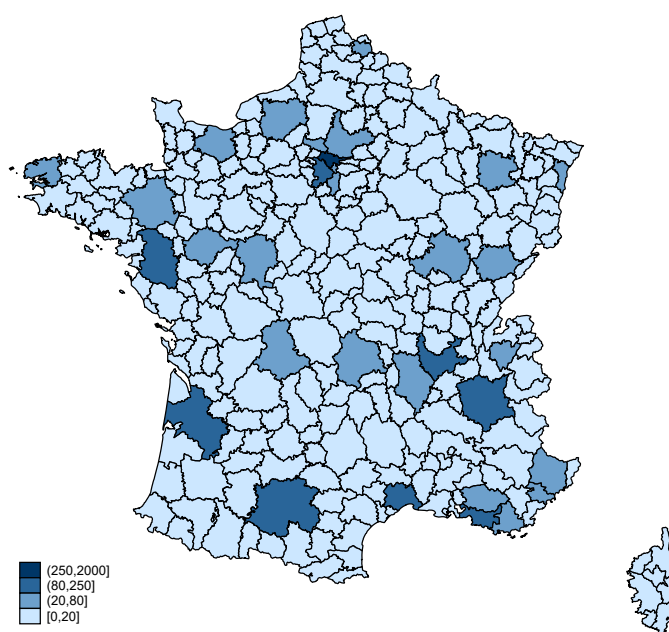
Lannion, Grenoble ou Pau, ainsi que le Sud de l'Île de France.

FIGURE 2.13 – Nombre de doctorants Cifre par ZE.



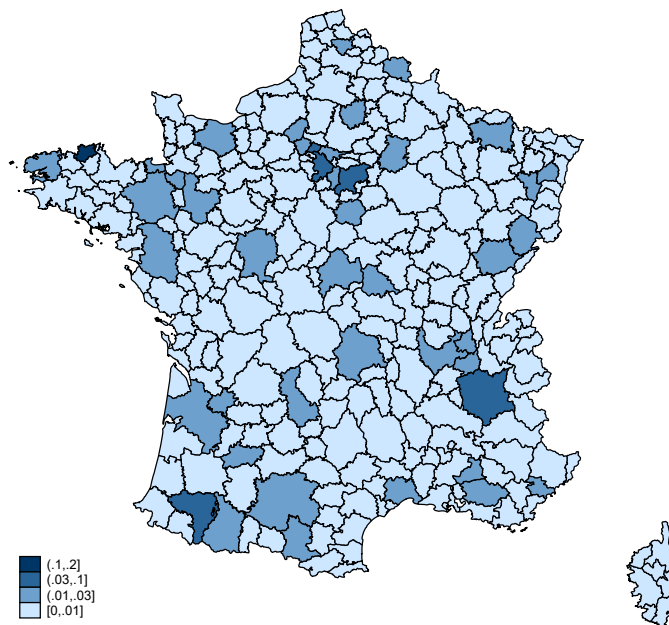
SOURCES : Base Cifre.

FIGURE 2.14 – Nombre d'établissements ayant accueilli une Cifre par ZE.



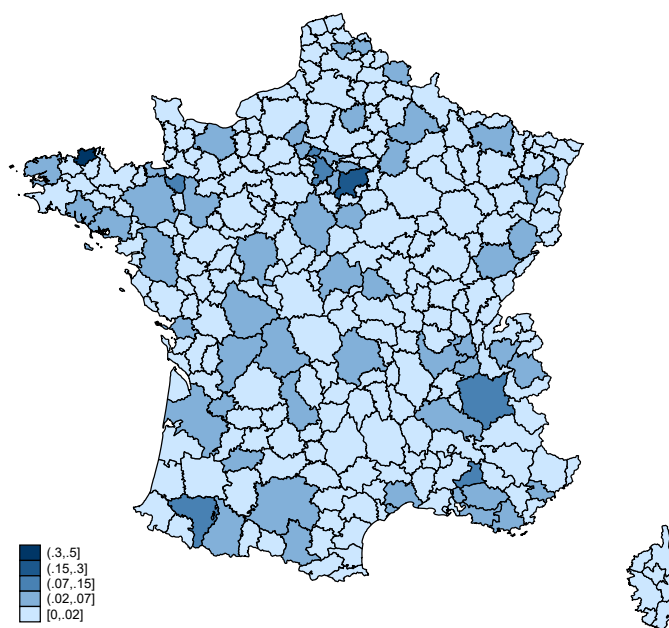
SOURCES : Base Cifre.

FIGURE 2.15 – Ratio du nombre de doctorants Cifre / entreprises avec emplois hautement qualifiés .



SOURCES : Base Cifre, DADS Postes.

FIGURE 2.16 – Ratio du nombre de doctorants Cifre / entreprises employant des ingénieurs.



SOURCES : Base Cifre, DADS Postes.

2.3.2 Évolution des caractéristiques des entreprises recourant au dispositif Cifre par rapport à des groupes comparables

Cette section propose une comparaison des évolutions sur la période étudiée (2008-2016) des caractéristiques des entreprises ayant recours au dispositif Cifre, par rapport à deux groupes d'entreprises susceptibles d'avoir des traits communs avec le groupe d'entreprises traitées. Le premier groupe de comparaison est naturellement celui des entreprises ayant recours au Crédit Impôt Recherche au cours de la période : ce crédit d'impôt porte sur l'ensemble des dépenses de recherche-développement, et est très complémentaire du dispositif Cifre puisque les salaires versés à des doctorants Cifre sont éligibles au versement du Crédit Impôt Recherche. Le second groupe d'entreprises avec lequel on peut tenter d'établir une comparaison est l'ensemble des entreprises employant des ingénieurs : il s'agit d'un groupe beaucoup plus vaste que les seules entreprises innovantes, mais il permet d'approcher le plus possible le groupe d'entreprises susceptibles d'avoir recours au dispositif Cifre.

Comme le montre le tableau 2.1, les entreprises ayant recours au dispositif Cifre sont en moyenne beaucoup plus importantes en termes de nombre d'employés que les entreprises ayant recours au CIR mais pas au Cifre : le ratio des effectifs moyens était d'environ 15 en 2008, et la différence s'est accrue au cours de la période avec un ratio supérieur à vingt en 2015 sur l'emploi salarié total entre firmes Cifre et firmes recourant au CIR seulement. La différence de taille s'explique très simplement par le fait que la grande majorité des géants technologiques français utilisent à la fois les dispositifs Cifre et CIR. Cette différence ne se traduit en revanche pas par une composition très différente des effectifs salariés, puisque les parts d'employés ingénieurs et techniciens sont très comparables entre les deux groupes. L'ensemble des firmes employant des ingénieurs et ne recourant ni au

TABLEAU 2.1 – Tableau comparant les caractéristiques des entreprises Cifre à celles ayant seulement recours au CIR, et à celles employant seulement des ingénieurs, en 2008 et 2015

En 2008	CIFRE	CIR seulement	Ingé. seulement
Emploi	3641.03	245.18	37.10
Part emploi R&D	0.10	0.06	0.01
Part emploi R&D : ingénieur	0.20	0.18	0.10
Part emploi R&D : techniciens	0.10	0.10	0.06
Observations	296	9472	130811
En 2015			
Emploi	4202.65	183.84	38.51
Part emploi R&D	0.13	0.13	0.03
Part emploi R&D : ingénieur	0.14	0.13	0.11
Part emploi R&D : techniciens	0.09	0.08	0.07
Observations	445	11905	116723

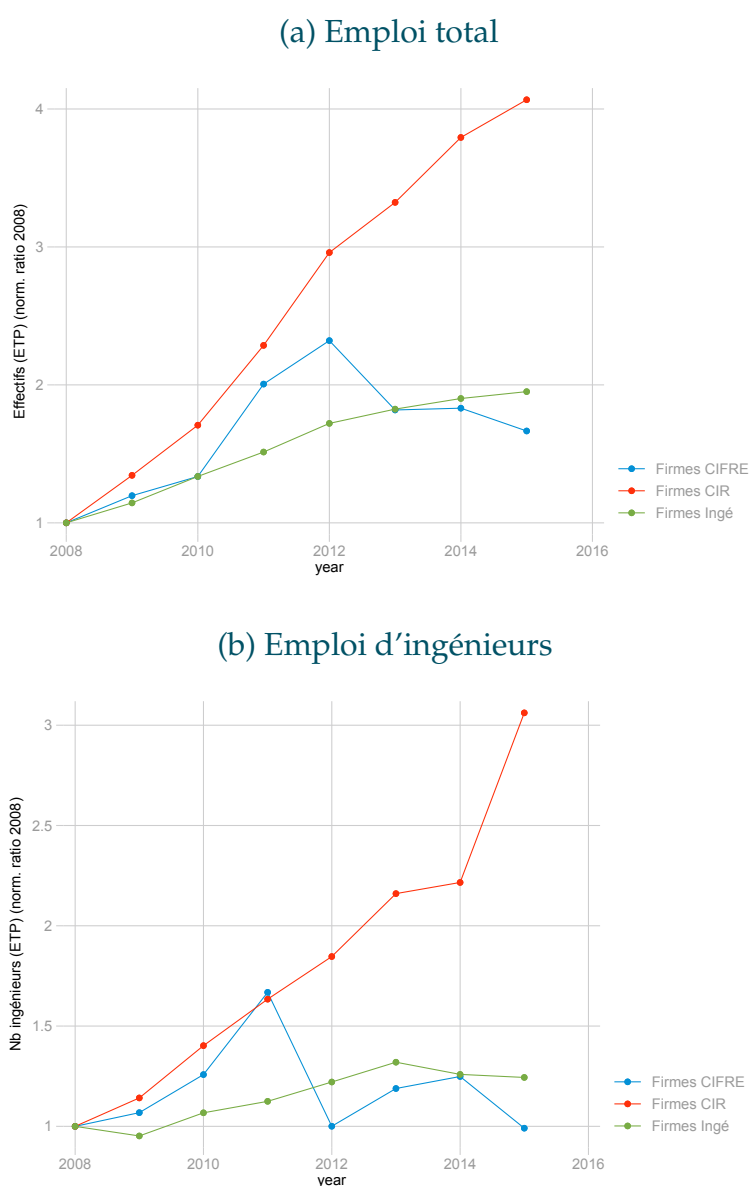
NOTES : La première colonne contient le groupe des entreprises ayant recours au dispositif Cifre pendant l'année concernée, la seconde colonne le groupe des entreprises ayant recours au CIR mais pas au dispositif Cifre pendant cette même année, et la troisième colonne le groupe des entreprises employant des ingénieurs mais ne recourant à aucun des deux dispositifs précités cette même année.

SOURCES : Bases Cifre, GECIR, DADS Postes.

CIR ni au Cifre, présenté dans la troisième colonne, est en revanche très différent des deux autres sur l'ensemble de la période : les entreprises qui le composent sont en moyenne cent fois plus petites en termes d'effectifs, et ont également des parts de l'emploi consacré à la R&D beaucoup plus faibles.

La figure 2.17 représente l'évolution sur la période 2008-2015 de l'emploi total (panneau a) et de l'emploi d'ingénieurs (panneau b), en base 100 au début de la période. Il montre que les trois groupes d'entreprises (Cifre, CIR seulement, employeurs d'ingénieurs seulement) sont sur des trajectoires ascendantes d'emploi sur la période. Ces groupes sont définis comme l'ensemble des entreprises respectant le critère les caractérisant au moins une année dans la période. La dynamique d'emploi des firmes ayant recours au dispositif Cifre est positive mais plus modérée que celle des entreprises ayant recours au CIR seulement, qui contient vraisemblablement le plus d'entreprises jeunes à très forte croissance. La figure 2.18

FIGURE 2.17 – Évolution moyenne de l’emploi total et de l’emploi ingénieur des firmes Cifre, CIR seulement et employeuses d’ingénieurs seulement sur la période 2008-2015, normalisé à 1 en 2008.

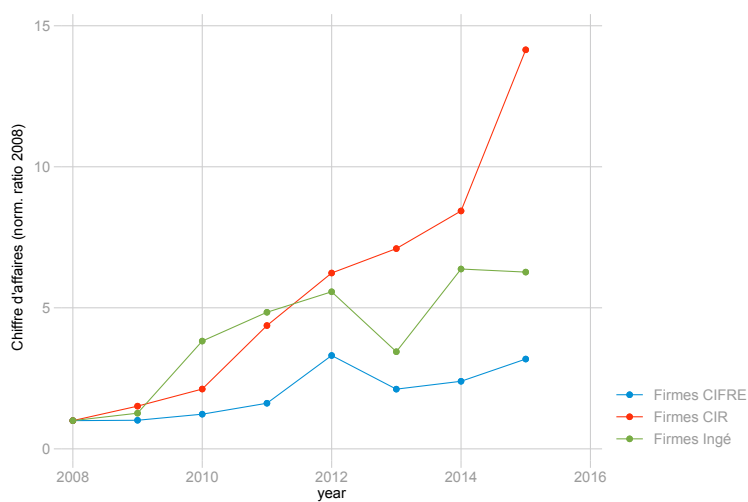


NOTE : La moyenne est non-pondérée.

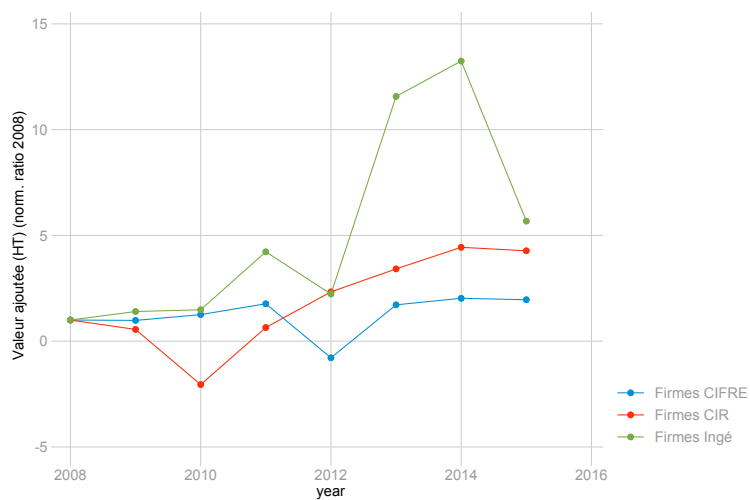
SOURCE : Bases Cifre, GECIR, DADS Postes.

FIGURE 2.18 – Évolution moyenne du chiffre d'affaires et de la valeur ajoutée (HT) des firmes Cifre, CIR seulement et employeuses d'ingénieurs seulement sur la période 2008-2015, normalisé à 1 en 2008.

(a) Chiffre d'affaires



(b) Valeur ajoutée



NOTE : La moyenne est non-pondérée.

SOURCE : Bases Cifre, GECIR, DADS Postes, FARE.

confirme cette image d'une évolution plutôt positive des firmes ayant recours au Cifre au cours de la période, mais la nuance légèrement. En effet, les firmes Cifre sont en moyenne en croissance en termes de chiffre d'affaires et de valeur ajoutée, mais ces tendances sont moins marquées que pour les firmes CIR seulement ou employeuses d'ingénieurs seulement.

CHAPITRE 3

ANALYSE AU NIVEAU DES DOCTEURS

CIFRE

Ce chapitre s'intéresse à l'effet du dispositif CIFRE sur les docteurs issus du programme. Notre analyse s'intéresse dans un premier temps à la capacité du dispositif à permettre aux doctorants à effectivement terminer leur doctorat. À cette fin, nous regardons deux indicateurs : la probabilité de soutenir en moins de 5 ans et la durée de la thèse (section 3.1). Dans un second temps, nous utilisons les données de l'enquête Génération afin d'estimer l'effet du programme sur l'insertion sur le marché du travail trois ans après la soutenance de la thèse (section 3.2). Nous considérons également des indicateurs de qualité universitaire des thèses soutenues, en l'occurrence le nombre de publications dans des revues à comité de lecture avant la soutenance et l'obtention de la qualification au CNU.

3.1 Résultats sur la durée de la thèse et la probabilité de soutenance

Afin de comparer les doctorants Cifre avec le reste des doctorants, nous effectuons un appariement avec la base de données de l'ABES, correspondant à des

informations librement disponibles sur le site internet theses.fr. Les thèses validées depuis 2008 dans la base Cifre sont appariées avec les thèses de ce site, se basant sur la correspondance entre les noms et prénoms du doctorant, des directeurs de thèse, et le titre de la thèse. Nous apparions environ 9,000 thèses sur les 12,000 de la base Cifre validées sur la période 2008 - 2018. Nous y associons les thèses non-Cifre effectuées pendant la même période dans les mêmes disciplines, composant notre sous-échantillon d'analyse.

Soutenance de thèse sous 5 ans. Dans la base de données de l'ABES, les thèses peuvent afficher l'un de deux statuts : « soutenues » ou « en cours ». Les thèses abandonnées n'ont pas de statut spécial mais demeurent simplement « en cours » pour une période indéterminée. Nous nous intéressons aux thèses débutées entre 2008 et 2013 afin de limiter le nombre de thèses marquées « en cours » qui n'auraient pas été abandonnées (on s'intéresse donc non pas à la part des thèses finissant par être soutenues, mais à la part des thèses soutenues dans les cinq années suivant l'inscription). Plus précisément, on s'intéresse à la part des thèses débutées entre 2008 et 2013 et non soutenues dans l'union de cet ensemble et de l'ensemble des thèses soutenues entre 2013 et 2018 (que l'on suppose donc implicitement avoir été débutées au plus 5 années auparavant). Sur ce sous-échantillon, nous régressons le statut de la thèse sur le fait d'avoir bénéficié d'un financement Cifre, en contrôlant pour la discipline (spécification plus large) et l'établissement de la thèse. Le tableau 3.1 montre les résultats de cette régression, et la présence d'un effet positif de la convention Cifre sur le taux de soutenance. Cet effet diminue mais persiste lorsque nous contrôlons pour l'établissement de thèse et de la discipline. Nous ne sommes cependant pas en mesure de contrôler pour l'âge, ni pour l'année de début de la thèse, ni pour le sexe, car ces informations ne sont pas fournies dans la base de données theses.fr.

La figure 3.1 montre les taux de soutenance par discipline et par statut Cifre.

TABLEAU 3.1 – Régressions linéaires du taux de soutenance sur le statut Cifre

	(1)	(2)	(3)
Convention CIFRE	0,0995*** (21,30)	0,0344*** (7,95)	0,0176*** (4,12)
Constante	0,671*** (449,04)	0,675*** (537,97)	0,677*** (549,98)
R^2	0,003	0,273	0,298
Etablissement		✓	✓
Discipline			✓
Observations	93339	93314	93314

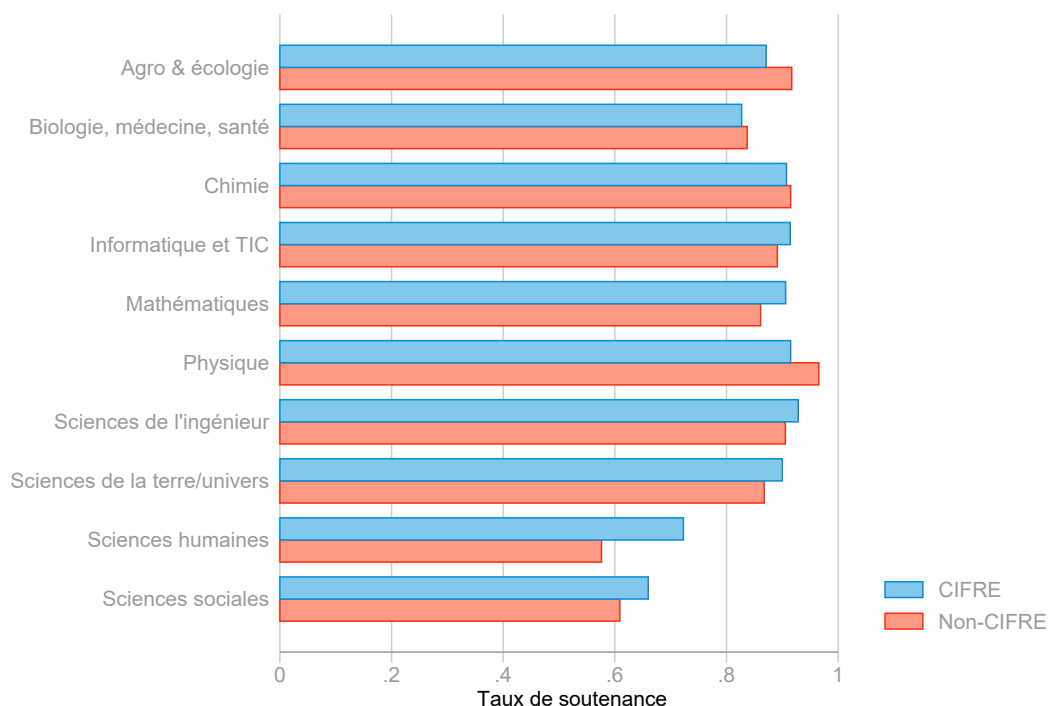
t score entre parenthèses
 * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

SOURCES : Bases Cifre, theses.fr.

Il est utile de noter que dans son enquête, l’ANRT trouvait un taux de soutenance qui avoisinait les 98 %. La différence avec notre résultat souligne sans doute une latence dans la mise à jour de la base ; cependant, nous pensons que cet effet n’est pas hétérogène entre les Cifre et non-Cifre, et que les résultats du tableau 3.1 ne sont pas biaisés par cette latence. Nous observons que même si la régression indique l’existence d’un effet positif pour les doctorants Cifre, il existe beaucoup d’hétérogénéité entre disciplines dans cet effet. Il est en effet très fort dans les sciences humaines et sociales, l’ingénierie et les TIC, mais ne l’est pas pour autant en agronomie et écologie, physique ou chimie.

Durée de thèse. Dans quelques rares cas (environ 8000), la base theses.fr contient à la fois la date de début et la date de soutenance de thèse. Sous l’hypothèse que ces cas ne sont pas sélectionnés avec un biais susceptible de fausser nos résultats, il est possible de les exploiter afin d’en extraire des informations sur la durée moyenne des thèses Cifre en comparaison aux autres. Le tableau 3.2 montre ainsi l’effet de la réalisation d’une thèse dans le cadre du programme Cifre

FIGURE 3.1 – Taux de soutenance par discipline theses.fr (2011-2018)



sur sa durée : on constate ainsi que, sans aucune variable de contrôle (colonne 1), les Cifre soutiennent leur thèse environ quatre mois plus tôt que les autres. Cette effet demeure significatif lorsque l'on introduit un ensemble d'effets fixes contrôlant pour la discipline de la thèse, l'établissement dans laquelle elle est réalisée, ainsi qu'une variable indiquant si celle-ci est réalisée en langue anglaise (qui pourrait être indicatif d'un type de thèses particulier). Avec ces variables de contrôle (colonne 3), l'effet demeure très significatif et correspond environ à une durée de thèse inférieure d'un trimestre pour les thèses Cifre relativement aux thèses classiques.

Ces résultats, qu'il s'agisse de ceux sur la probabilité de soutenir dans les 5 ans ou sur la durée de thèse, doivent néanmoins être relativisés par le fait que nous n'observons pas, parmi les thèses classiques, lesquelles correspondent à une thèse ayant reçu un financement. Ainsi, toutes les thèses Cifre sont des thèses financées,

tandis qu’au total seules 72 % des thèses sont financées¹. Par conséquent, outre les problèmes éventuels de sélection dans le programme Cifre, il n’est pas possible à partir de ces estimations d’affirmer que l’augmentation de la probabilité de soutenir ou la baisse de la durée des thèses Cifre est attribuable au programme en lui-même plutôt qu’au simple fait que ces thèses soient financées. Si la robustesse des résultats à l’introduction d’effets fixes discipline et établissement de rattachement plaide plutôt en faveur de la première hypothèse, une analyse plus approfondie de l’hétérogénéité de l’effet entre disciplines rapportée à la part des doctorants financés dans ces disciplines permettrait de soutenir cette hypothèses avec plus de confiance.

TABLEAU 3.2 – Régressions linéaires de la durée de thèse sur le statut Cifre

	(1)	(2)	(3)
Convention Cifre	-0,657*** (0,056)	-0,321*** (0,052)	-0,271*** (0,055)
Durée moyenne	4,260*** (0,022)	4,242*** (0,020)	4,247*** (0,020)
Discipline		✓	✓
Etablissement			✓
Langue anglaise			✓
<i>N</i>	7913	7913	7805
<i>R</i> ²	0,0056	0,186	0,268
<i>erreurs standards</i> entre parenthèses			
* <i>p</i> < 0,05, ** <i>p</i> < 0,01, *** <i>p</i> < 0,001			

SOURCES : Bases Cifre, theses.fr.

1. Beaufort et al. (2018)

3.2 Insertion sur le marché du travail : résultats issus de l'enquête Génération

3.2.1 Données

L'Enquête Génération à 3 ans (années 2010 et 2013). L'Enquête Génération est une base de données produite par le Centre d'études et de recherche sur la qualification (CEREQ) depuis 1997. Elle consiste en un suivi très riche des caractéristiques des jeunes sortant du système scolaire pendant une durée de 3 ans suivant leur dernier diplôme, permettant d'observer leur insertion professionnelle. Les enquêtes Génération disposent d'un module thèse qui renseigne les spécificités des doctorants durant leur thèse, ce qui permet de suivre les doctorants qui ont bénéficié d'une Cifre.

Les doctorants sont aussi sur-échantillonnés par ces enquêtes, ce qui nous permet d'exploiter un nombre significatif de Cifre dans chaque vague; nous corrigeons nos résultats avec les pondérations fournies. L'analyse de ces données s'effectue sur la base de plusieurs échantillons et variables, que nous définissons ci-dessous.

Nous utilisons les enquêtes pour les années 2010 et 2013 car elles disposent d'un échantillon conséquent sur les Cifre ainsi que sur le type de formation suivi, comme celui de l'école d'ingénieur.

Identifier les docteurs CIFRE et les groupes potentiels de comparaison.

- Étudiant en STEM : cette classification regroupe des étudiants en Sciences Techniques, Ingénierie et Mathématiques (STEM selon la dénomination anglaise). Ceci regroupe donc les étudiants dont le plus haut diplôme détenu est un doctorat (hors santé) en Sciences, Technologies et Mathématiques ; une école d'ingénieur ; une formation Bac+5 en Sciences, Technologies et Mathé-

matiques ; une formation Bac+4 en Sciences, Technologies et Mathématiques ; ou un M2 Sciences, Technologies et Mathématiques (déclinaisons de la variable PHD)

- Ingénieur : Les ingénieurs sont définis comme ceux dont le plus haut diplôme provient d'une école d'ingénieur (PHD = 16 ou PHINSEE = 170), dont la classe de sortie est une école d'ingénieur (NISOR = 16) ou dont le niveau le plus élevé avant la fin des études en 2013 était une école d'ingénieur (Q39H = 24)
- Docteur : ayant effectué comme dernier diplôme un doctorat (sortant de niveau thèse, hors santé, MODTH = 1)
- Cifre : docteurs ayant bénéficié du dispositif Cifre (TH4 =5) et ayant soutenu leur thèse.

Nombre d'observations :

- Etudiants en STEM : 6 176
- Ingénieurs (y compris docteurs) : 2 503
- Cifre (STEM) : 342
- Docteurs : 2 540
- Docteurs – ingénieurs (y compris Cifre) : 705
- Cifre – ingénieurs : 158
- Ingénieurs non docteurs : 1 798

Variables dépendantes d'intérêt. Nous nous intéressons à deux groupes de variables. En premier lieu nous cherchons à mesurer l'insertion sur le marché du travail des docteurs CIFRE. À cette fin nous nous concentrerons sur l'obtention d'emploi stable et le salaire mensuel. En second lieu, nous examinons des variables qui

rendent compte de la qualité académique de la thèse produite. La variable principale que nous étudions concerne la publication dans une revue à comité de lecture avant la soutenance de la thèse.

- CDI : variable binaire, indique si le contrat de travail est un CDI ou un contrat de fonctionnaire. Valeur de 0 pour les individus en non-emploi.
- Salaire : variable correspondant au salaire en fin de séquence d'emploi ; si la fin n'est pas disponible, le salaire de début est utilisé.
- Secteur : secteur professionnel NAF de l'entreprise, 10 possibilités².
- CSP – père : classe socioprofessionnelle du père, 9 possibilités
- PHNSF : spécialité du plus haut niveau de formation, 86 possibilités

3.2.2 Statistiques descriptives.

Nous présentons dans cette section les différences entre les docteurs Cifre et non-Cifre. L'analyse se concentre sur une comparaison entre les Cifre-ingénieurs et le reste des docteurs-ingénieurs. Cette sélection est motivée par le rôle déterminant du diplôme d'ingénieur dans l'insertion des nouveaux diplômés sur le marché du travail – tel que mis en lumière par Margolis et Miotti (2015)³. Les résultats sont présentés dans le tableau 3.3. Nous considérons plusieurs ensembles de variables que nous commentons ci-dessous.

Informations socio-économiques. Nous commençons par considérer des variables socio-économiques qui ont pour caractéristique commune d'être déterminées avant le début du doctorat. Elles ne sont donc pas susceptibles d'être causalement affectées par le dispositif Cifre lui-même. Nous ne constatons aucune différence si-

2. Agriculture, Manufacture, Construction, Commerce, Informatique et Communication, Finance, Immobilier, Métiers Spécialisés, Public et Social.

3. Les résultats obtenus sur la base de groupes de comparaisons différents sont disponibles dans le rapport intermédiaire et très proches de ceux présentés ici.

gnificative, que ce soit en termes d'âge, de réussite scolaire (mesurée par la mention au baccalauréat) ou encore relative à la profession des parents. Le signe des différences suggère néanmoins que les docteurs Cifre tendent à avoir des caractéristiques légèrement moins favorables en termes de CSP du père ou encore de réussite au baccalauréat. On remarque que la part d'hommes au sein du groupe de Cifre est plus élevée.

Situation sur le marché du travail. La différence de salaire mesurée entre les groupes est significative et favorable aux docteurs-ingénieurs Cifre. Les individus Cifre travaillent bien plus fréquemment dans le secteur privé que les non-Cifre, à hauteur de 71 % par rapport à 56 %. La probabilité d'avoir un emploi stable (CDI ou poste titulaire au sein de la fonction publique) est plus élevée pour ces individus Cifre que pour les docteurs-ingénieurs non Cifre – 55 % par rapport à 41 %. Le groupe Cifre présente également une plus forte probabilité de rapporter être satisfait de sa position professionnelle. On voit également que les Cifre-ingénieurs ont une durée d'emploi maximale plus longue en moyenne que le groupe non-Cifre. Cette différence est d'environ 3 mois, ce qui correspond à une différence d'environ 12 % de la moyenne du groupe des docteurs-ingénieurs non-Cifre.

Les différences observées concernant l'insertion sur le marché du travail pourraient refléter un biais de sélection, c'est-à-dire des différences en termes de situation potentielle qui auraient tout de même lieu en l'absence de traitement (thèse classique plutôt que Cifre). Il est peu probable que ce biais de sélection soit expliqué par les variables socio-économiques qui ne semblent pas présenter de différences majeures entre Cifre et non-Cifre.

Variables concernant les aspirations des docteurs. On voit ici que docteurs-ingénieurs Cifre et non-Cifre désirent au moment de leur soutenance occuper un poste de recherche dans les mêmes proportions. En revanche, là où les non-Cifre sont très susceptibles de désirer un poste de recherche dans le public, les Cifre

TABLEAU 3.3 – Statistiques descriptives

	CIFRE	Doc, non CIFRE	Différence
<i>Informations socio-économiques</i>	ref.		
Âge	27,64	27,73	-0,09
Genre : masculin	0,73	0,66	0,06
Père ouvrier/employé	0,25	0,26	-0,00
Père cadre	0,54	0,62	-0,08
Mention Très Bien	0,15	0,19	-0,03
<i>Situation marché du travail</i>	ref,		
Secteur Privé	0,71	0,56	0,15***
CDI	0,55	0,41	0,15**
Salaire	2647,38	2454,36	193,03**
Satisfaction	0,85	0,78	0,06
Longeur max, d'emploi	28,41	25,30	3,10***
<i>Aspirations post-soutenance</i>	ref,		
Position de recherche	0,75	0,78	-0,04
Rech, dans le sec, public	0,17	0,40	-0,23***
Rech, dans le sec, privé	0,58	0,39	0,19***
Dépôt dossier qualif,	0,15	0,37	-0,23***
<i>Qualité du doctorat</i>	ref,		
Publication académique (av, soutenance)	0,66	0,80	-0,14***
Validation qualif,	0,11	0,34	-0,23***
Validation qualif, condt, à dépôt	0,78	0,91	-0,13
Observations	157	542	

SOURCES : Enquêtes Génération 1ère vague; cohortes 2010 et 2013. L'échantillon se restreint aux individus docteurs et ingénieurs, interrogés 3 ans après l'obtention de leur diplôme.

ont une plus forte probabilité de vouloir occuper un poste de recherche dans le secteur privé. On remarque également que les docteurs Cifre déposent moins fréquemment une demande de qualification aux fonctions de maître de conférences au CNU.

Qualité universitaire du doctorat. Au-delà de la situation sur le marché du travail, il est intéressant d'étudier dans quelle mesure les Cifre produisent des thèses de bonne qualité universitaire. On remarque que les Cifre sont moins susceptibles d'être qualifiés aux fonctions de maître de conférences par le CNU. De plus, cet écart persiste lorsque l'on se restreint aux individus qui ont effectivement déposé un dossier pour qualification – même si cet écart n'est pas significatif compte

tenu du faible nombre de demandes effectuées par le groupe Cifre. La probabilité d’avoir une publication à comité de lecture pendant la durée de la thèse est également plus faible parmi les docteurs-ingénieurs Cifre.

Conclusion. Le portrait global que l’on peut dresser sur les diplômés Cifre-ingénieurs en STEM qui bénéficie de taux très élevés d’emploi stable, de séquences d’emploi plus longues, et d’une prime de salaire par rapport aux docteurs-ingénieurs non Cifre. Bien qu’on ne retrouve pas de différences dans la catégorie socio-professionnelle du père, ou la proportion de mentions « Très Bien » au Baccalauréat, on remarque des différences fortes en matière d’aspirations professionnelles (recherche dans le public par rapport au privé) et une moindre qualité universitaire du doctorat telle que mesurée par la probabilité d’obtenir une publication dans une revue à comité de lecture avant la soutenance ainsi que la probabilité d’être qualifié au CNU.

3.2.3 Approche empirique

Nous effectuons ensuite une série de régressions afin de mesurer des effets associés à la participation au programme Cifre. Nous étudions deux variables dépendantes d’intérêt : le salaire (log du salaire) et le taux d’emploi stable (variable CDI). Nous régressons ces variables sur le statut de traitement (avoir bénéficié d’un financement de thèse Cifre). Afin d’affiner nos résultats nous contrôlons progressivement pour plusieurs variables observables : le secteur de l’entreprise, la CSP du père, le genre, l’âge (fonction quadratique), la génération (2010 ou 2013) et la spécialité du plus haut diplôme (spécialité NSF).

Notre équation d’estimation de base pour l’échantillon le plus large est la suivante :

$$Y_i = \beta_1 \text{Cifre}_i + \beta_2 \text{Docteur}_i + \beta_3 \text{Ingenieur}_i + \beta_4 \text{Secteur}_i + \beta_5 \text{CSP}_i + \beta_6 \text{NSF}_i + \beta_7 \text{Genre}_i + \beta_8 \text{Age}_i + \beta_9 \text{Age}_i^2 + \beta_{10} \text{Generation}_i + \epsilon_i \quad (3.1)$$

3.2.4 Effets sur l'insertion sur le marché du travail

Nous nous intéressons au sous-échantillon le plus petit, celui des docteurs-ingénieurs, toujours dans le but de comparer les Cifre et non-Cifre. Nous régressons le statut de CDI et le salaire sur le fait d'avoir pris part à une Cifre seulement pour ceux ayant à la fois effectué des études d'ingénieur et obtenu un doctorat.

Le fait d'avoir réalisé une thèse Cifre a un effet considérable sur la probabilité d'être en CDI trois ans plus tard, estimé à 18,1 points de pourcentage dans la spécification la plus conservatrice (colonne 4 du tableau 3.4), applicable à une moyenne de 41 % pour la référence non-Cifre. Ceci correspond à une probabilité environ 44 % plus élevée pour des docteurs-ingénieurs d'être en CDI à l'issue d'une Cifre qu'à l'issue d'une thèse classique. Cet effet très important reflète probablement la conjonction de deux facteurs : les doctorants Cifre tendent à rester dans l'entreprise après la fin de leur doctorat (nous revenons sur ce point par la suite), et il est dès lors presque obligatoire légalement pour ces entreprises de leur proposer un CDI compte tenu de la durée déjà effectuée en CDD au sein de l'entreprise.

TABLEAU 3.4 – Effets du dispositif Cifre sur l'insertion professionnelle des docteurs

	CDI				Salaire			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Docteur ingénieur non-CIFRE								
CIFRE	0,225*** (0,0707)	0,223*** (0,0693)	0,218*** (0,0614)	0,181*** (0,0622)	0,104*** (0,0363)	0,0904*** (0,0342)	0,0860*** (0,0292)	0,0597** (0,0299)
R^2	0,167	0,229	0,267	0,312	0,064	0,170	0,194	0,237
Contrôles de base	✓	✓	✓	✓	✓	✓	✓	✓
PHNSF		✓	✓	✓		✓	✓	✓
CSP père			✓	✓			✓	✓
Mention Bac			✓	✓			✓	✓
Projet Pro Sout,				✓				✓
Observations	643	643	643	643	574	574	574	574

SOURCES : Enquêtes Génération 1ère vagues ; cohortes 2010 et 2013. L'échantillon se restreint aux individus docteurs et ingénieurs, interrogés 3 ans après l'obtention de leur diplôme.

Nous trouvons également un effet sur le salaire, qui est cependant plus faible dans sa magnitude que l'effet sur les CDI. L'effet est néanmoins très robuste, et représente environ un gain de 6 % de salaire dans la spécification incluant le plus de variables de contrôle (colonne 8 du tableau 3.4).

3.2.5 Analyse de sensibilité

3.2.5.1 Spécification saturée

Dans la spécification principale, nous supposons qu'il n'existe pas d'effet d'interaction entre ces variables. Par exemple, on suppose que l'effet de la mention au bac sur la situation sur le marché du travail est le même pour les hommes et les femmes.

L'ensemble des variables de contrôle utilisées dans cette analyse sont des variables discrètes. Il est donc en principe possible d'estimer une spécification saturée, c'est-à-dire une spécification qui comprend autant de paramètres qu'il existe de combinaisons uniques des variables de contrôle. Cette spécification présente l'avantage de contrôler de façon flexible pour l'ensemble des variables observables retenues dans l'analyse, en autorisant notamment des effets d'interaction arbitraires entre variables. L'estimateur présente également l'avantage d'être facile à interpréter. Il converge en probabilité vers une moyenne pondérée des effets de traitement pour chaque combinaison unique des variables de contrôles, où les poids utilisés dans la pondération reflètent la variance du traitement au sein de cette combinaison.

Une limite de cette approche est qu'elle exclut de l'estimation l'ensemble des observations associées à une valeur des variables de contrôle pour laquelle toutes les unités font partie du groupe de contrôle ou de traitement – et pour lesquelles la variance du traitement est par définition nulle.⁴

4. Formellement, on note X la variable qui prend une valeur distincte pour chacune des combinaisons uniques des variables discrètes incluses dans la spécification de base présente dans

TABLEAU 3.5 – Spécification avec variables de contrôle pleinement interagies

	CDI				Salaire			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Docteur ingénieur non-CIFRE								
CIFRE	0,239*** (0,0646)	0,259*** (0,0588)	0,197* (0,101)	0,0128 (0,168)	0,0814** (0,0320)	0,0726** (0,0339)	0,0933 (0,0569)	0,0816 (0,0632)
R^2	0,257	0,452	0,546	0,502	0,190	0,389	0,547	0,742
Contrôles de base	✓	✓	✓	✓	✓	✓	✓	✓
PHNSF		✓	✓	✓		✓	✓	✓
CSP père			✓	✓			✓	✓
Mention Bac			✓	✓			✓	✓
Projet Pro Sout,				✓				✓
Observations	694	611	210	102	615	530	173	84

SOURCES : Enquêtes Génération 1ère vague; cohortes 2010 et 2013. L'échantillon se restreint aux individus docteurs et ingénieurs, interrogés 3 ans après l'obtention de leur diplôme.

Les résultats sont présentés dans le tableau 3.5. Nous faisons le choix de laisser l'échantillon varier en fonction des spécifications. On constate donc que l'approche avec contrôles pleinement interagis limite fortement le nombre d'observations effectivement utilisées dans l'estimation qui passe de 694 avec les contrôles de base à 102 avec l'ensemble des contrôles inclus lorsque l'on s'intéresse à l'emploi stable. Cette réduction drastique de l'échantillon s'accompagne d'une perte de significativité statistique. La perte de précision est la plus marquée lorsque l'on étudie le salaire. On constate néanmoins une forte stabilité des coefficients entre les colonnes (5) et (8), en dépit d'une modification de l'échantillon d'estimation qui passe de 615 à 84. Le constat est plus nuancé concernant la probabilité d'obtenir un CDI/emploi stable. Le coefficient estimé est stable et demeure significatif dans les colonnes (1) à (3). Néanmoins l'inclusion de la variable sur le projet professionnel réduit à la fois la significativité mais également la taille du coefficient. Le changement de coeffi-

l'équation 3.1. L'espace des variables de contrôle est donc partitionné en un ensemble fini de sous-ensembles : $X = 1, 2, \dots, N_x$, où N_x dénote le nombre de combinaisons uniques que prennent l'ensemble des variables discrètes incluses dans l'analyse. La spécification estimée s'écrit simplement :

$$Y_i = \beta \times \text{Cifre}_i + \sum_{x=1}^{N_x} \mathbf{1}\{X_i = x\} \times \delta_x + \epsilon_i \quad (3.2)$$

cient complique néanmoins l'interprétation de ce résultat. Globalement, bien que cette spécification soit en principe supérieure à la spécification de base, elle est en pratique excessivement non-paramétrique compte tenu des données disponibles. En conclusion, il nous semble que cet exercice souligne une certaine robustesse de nos résultats principaux. Si l'inclusion du projet professionnel comme variable de contrôle change le résultat sur la stabilité de l'emploi, la réduction de l'échantillon d'estimation effectif qui suit l'inclusion de cette variable est tellement forte que ce résultat apparaît plutôt comme non-informatif qu'invalidant les résultats principaux présentés dans le tableau 3.4.

3.2.5.2 Prise en compte des caractéristiques inobservables.

L'analyse économétrique repose que l'hypothèse d'indépendance conditionnelle afin d'être interprétable causalement : conditionnellement aux variables de contrôle incluses, on suppose qu'appartenir au dispositif Cifre est indépendant des inobservables susceptibles de déterminer la performance sur le marché du travail. Pour reprendre les notations de l'équation (3.1) cela suppose que la variable $Cifre_i$ soit indépendante du terme d'erreur ε_i .

Ici, sur la base d'Oster (2019), nous effectuons un test de robustesse qui vise à estimer les effets programme Cifre en prenant en compte le rôle de facteurs non observables omis qui affectent à la fois l'inscription au programme et les résultats sur le marché du travail. L'hypothèse de non-sélection sur des inobservables – que l'on peut résumer à un covariance nulle entre $Cifre$ et ε : $\sigma_{Cifre,\varepsilon} = 0$ – est fondamentalement non-testable. Nous proposons un exercice de bornage qui relâche cette hypothèse et qui suppose un degré de sélection proportionnel sur la base des variables observables que nous incluons comme variables de contrôle et sur la base des inobservables. Nous fixons ce degré de proportionnalité à 1 ($\delta = 1$ pour reprendre les notations d' Oster, 2019). Cela implique que le degré de sélection sur les facteurs inobservables est supposé égal à celui que l'on observe sur la base

des facteurs observables. Nous formulons également une hypothèse sur le pouvoir explicatif d'une régression – son R^2 – dans laquelle l'ensemble des facteurs pertinents (observables et non-observables) seraient inclus. Ce R^2 , dénoté R_{max} , est possiblement inférieur à 1, en présence par exemple d'erreur de mesure. Sur la base de ces hypothèses on peut approximer le biais de notre estimation en MCO.

Les résultats du test sont présentés dans le tableau 3.6. Les colonnes (1) à (3) s'intéressent à l'accès à un emploi stable et les colonnes (4) à (6) au salaire. Nous les commentons tour à tour. La colonne (1) présente le résultat d'une régression qui n'inclut que les variables de contrôle basiques. La colonne (2) présente les résultats de la même spécification mais estimée sur l'échantillon des observations pour lesquelles l'ensemble des contrôles (et non pas seulement ceux de base) est bien renseigné. On constate un effet positif associé à la thèse Cifre de +18.5 points de pourcentage. L'introduction de l'ensemble des variables de contrôle réduit le coefficient estimé à +18.1 pp, alors même que la R^2 augmente par un facteur de 2 en passant de 0.154 à 0.33. Ces deux faits – stabilité des coefficients et hausse du pouvoir explicatif de la régression – suggèrent une robustesse de nos résultats à la sélection sur les inobservables sous l'hypothèse maintenue que cette sélection soit égale à celle observée sur les observables. Ceci est confirmé par la valeur des bornes montrées en colonne (3). Sous l'hypothèse d'un $R_{max} = 0.5$, l'effet estimé est de +17.7 pp. En autorisant R_{max} à avoir sa valeur théorique maximale (1), on trouve un effet de +15.2 pp. Les résultats sur l'accès à un emploi stable apparaissent donc robustes à un relâchement des hypothèses d'identification initiales.

On constate que cela est moins vrai pour le résultat sur les salaires, qui est divisé par 4 sous l'hypothèse de $R_{max} = 0.5$ et change de signe pour $R_{max} = 1$. On note cependant que $R_{max} = 1$ correspond à une multiplication par 4 du R^2 observé dans la colonne (6), ce qui peut dès lors être considéré comme une hypothèse particulièrement défavorable. On en conclut que l'effet sur les salaires demeure plausible mais moins robuste que celui sur l'accès à un emploi stable.

TABLEAU 3.6 – Analyse de sensibilité

	CDI			Salaire		
	(1)	(2)	(3)	(4)	(5)	(6)
CIFRE	0,207*** (0,0747)	0,185** (0,0764)	0,181*** (0,0625)	0,0954** (0,0386)	0,0886** (0,0390)	0,0597** (0,0301)
R^2	0,175	0,154	0,330	0,085	0,049	0,254
Contrôle de Base	✓	✓	✓	✓	✓	✓
PHNSF			✓			✓
CSP pere			✓			✓
Mention au baccalauréat			✓			✓
Projet Pro Sout,			✓			✓
Bornes sur $\hat{\beta}$ ($R_{max} = 0,5, \delta = 1$)			,177			,0159
Bornes sur $\hat{\beta}$ ($R_{max} = 1, \delta = 1$)			,1519			-,1191
Observations	699	650	650	620	580	580

SOURCES : Enquêtes Génération 1ère vague; cohortes 2010 et 2013. L'échantillon se restreint aux individus docteurs et ingénieurs, interrogés 3 ans après l'obtention de leur diplôme.

3.2.6 Résultats complémentaires et discussion

Le rôle des aspirations. La corrélation positive entre la participation au programme Cifre et la probabilité d'avoir un emploi stable et un salaire plus élevé apparaît robuste. On constate que cette corrélation demeure même lorsque l'on contrôle pour le projet professionnel au moment de la soutenance de la thèse. Cela suggère que les différences en termes d'insertion observable sur le marché du travail ne sont pas entièrement déterminées par des différences de préférences entre docteurs Cifre et non-Cifre. Cependant on peut questionner la pertinence de ces indicateurs pour des docteurs dont le projet professionnel est de travailler dans le secteur de la recherche publique. Ces individus sont susceptibles de choisir d'effectuer des cursus post-doctoraux de longue durée, ce qui sera reflété par un faible accès à l'emploi stable et un salaire peu élevé. Il s'agit d'un problème potentiellement important car, comme on peut le voir sur le tableau 3.3, une part non négligeable des docteurs ont pour projet de faire de la recherche dans le secteur public (université, CNRS) et celle-ci est très inégale entre groupes : elle s'élève à 40 % parmi les docteurs non-Cifre et à 17% parmi les docteurs Cifre.

Afin de clarifier la nature du contrefactuel qui permet d’interpréter causalement les corrélations (conditionnelles) estimées, nous estimons la spécification de base en nous concentrant sur les docteurs qui formulent le projet professionnel de travailler dans le secteur privé. Les résultats de cette estimation sur un sous-échantillon sont présentés dans le tableau 3.7. Les résultats sont très proches de l’estimation principale et très stables lorsque l’on inclut des contrôles observables additionnels. Ces résultats renforcent l’interprétation causale de la corrélation obtenue et montrent que le dispositif Cifre confère un avantage considérable parmi des individus aspirant à une poursuite de carrière de R&D dans le secteur privé.

Bien entendu, il s’agit là d’effets de moyen terme (3 ans). Il est possible que les docteurs non-Cifre rattrapent et dépassent la performance des docteurs Cifre dans le long-terme, si par exemple ils sont en mesure d’accéder à des emplois très rémunérateurs dans des entreprises très exigeantes d’un point de vue académique. Nous ne disposons pas des données suffisantes pour tester cette hypothèse directement en examinant les trajectoires sur le plus long terme. Néanmoins, nous nous intéressons maintenant à une comparaison de la qualité des thèses entre Cifre et non-Cifre, en fonction de leur projet professionnel – ce qui permettra d’examiner indirectement la plausibilité du scénario précédent.

TABLEAU 3.7 – Estimation parmi les docteurs aspirant à travailler dans le secteur privé au moment de la soutenance

	CDI				Salaire			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Docteur ingénieur non-CIFRE								
CIFRE	0,212** (0,0824)	0,238*** (0,0760)	0,176** (0,0699)	0,181*** (0,0696)	0,0907** (0,0366)	0,0771** (0,0391)	0,0583* (0,0303)	0,0593** (0,0301)
R^2	0,136	0,241	0,309	0,314	0,058	0,142	0,240	0,241
Contrôles de base	✓	✓	✓	✓	✓	✓	✓	✓
PHNSF		✓	✓	✓		✓	✓	✓
CSP père			✓	✓			✓	✓
Mention Bac			✓	✓			✓	✓
Projet Pro Sout,				✓				✓
Observations	366	366	366	366	329	329	329	329

SOURCES : Enquêtes Génération 1ère vague; cohortes 2010 et 2013. L’échantillon se restreint aux individus docteurs et ingénieurs, interrogés 3 ans après l’obtention de leur diplôme.

Le rôle de la séniorité (*tenure*). Dans ce paragraphe, nous analysons si le dispositif Cifre permet en moyenne aux docteurs d’atteindre une plus forte séniorité chez leur employeur au moment de l’enquête, c’est-à-dire une plus grande durée d’emploi dans la même entreprise. La colonne 1 du tableau 3.8 montre l’effet du dispositif sur la séniorité exprimée en année, parmi les docteurs ayant pour projet professionnel de travailler dans le secteur privé. On constate un effet positif et significatif d’environ 0.3 années (soit 3 mois et demi) qui est robuste à l’inclusion de l’ensemble des contrôles usuels (voir colonne 3). Ceci signifie que les docteurs Cifre travaillent en moyenne depuis 3 mois et demi de plus chez leur employeur actuel que leurs homologues non-Cifre. La colonne 2 montre l’effet sur le salaire déjà documenté dans le tableau 3.7⁵. La colonne 3 montre les coefficients de régression du salaire sur l’indicatrice Cifre tout en contrôlant pour la séniorité. On constate de façon très intuitive que la séniorité est positivement associée au salaire : une année supplémentaire avec le même employeur est associée à une hausse d’environ 5% du salaire mensuel. On note en revanche que l’inclusion de ce contrôle atténue mais ne réduit pas à zéro le coefficient associé à l’indicatrice Cifre. Cela suggère qu’une partie de l’effet du dispositif Cifre sur le salaire transite via un effet sur la séniorité sans que ce canal ne capture l’ensemble de l’effet du dispositif. L’inclusion de contrôles observables (colonnes 4 à 6) aboutit à des résultats qualitativement similaires.

5. Les coefficients ne sont pas rigoureusement identiques du fait de l’exclusion de certaines observations pour lesquelles la séniorité n’était pas définie.

TABLEAU 3.8 – Effet du dispositif sur la « séniorité »

	Pas de contrôles			Contrôles		
	(1)	(2)	(3)	(4)	(5)	(6)
CIFRE	0,282** (0,143)	0,0824** (0,0366)	0,0690* (0,0382)	0,298** (0,133)	0,0521* (0,0311)	0,0351 (0,0304)
Tenure (année)			0,0495*** (0,0181)			0,0592*** (0,0155)
R^2	0,036	0,054	0,093	0,220	0,250	0,296
Contrôles de base	✓	✓	✓	✓	✓	✓
PHNSF				✓	✓	✓
CSP père				✓	✓	✓
Mention Bac				✓	✓	✓
Projet Pro Sout,				✓	✓	✓
Observations	352	352	352	327	327	327

SOURCES : Enquêtes Génération 1ère vague ; cohortes 2010 et 2013. L'échantillon se restreint aux individus docteurs et ingénieurs, interrogés 3 ans après l'obtention de leur diplôme.

Effets sur la qualité universitaire des thèses. Nous nous intéressons dans cette partie à plusieurs mesures de la qualité universitaire des thèses de docteurs Cifre versus non-Cifre. Le tableau 3.9 montre les résultats concernant quatre indicateurs, en se concentrant sur l'échantillon de docteurs qui formulaient le projet professionnel de travailler dans le secteur privé au moment de leur soutenance. La colonne 1 indique que les docteurs Cifre présentent une probabilité de déposer un dossier de candidature pour la qualification auprès du CNU qui est de 14 points de pourcentage inférieure à celle des docteurs non-Cifre. Cet écart n'est pas sensible à l'inclusion de l'ensemble des variables de contrôle (colonne 5). On constate également que les docteurs Cifre sont moins susceptibles de publier dans une revue à comité de lecture avant leur soutenance (colonnes 2 et 6) et qu'ils sont moins susceptibles d'obtenir la qualification (colonnes 3 et 7). Si l'on considère la probabilité d'obtenir la qualification *conditionnellement* au fait de postuler on constate également une différence en défaveur des docteurs Cifre, dont la significativité et l'ampleur est cependant sensible à l'inclusion des contrôles observables usuels. La différence est non-significative en l'absence de contrôle (colonne 4) et largement

plus importante et significative une fois les contrôles observables introduits. Le tableau 3.10 montre les mêmes coefficients obtenus à partir de l'ensemble de notre échantillon de docteurs-ingénieurs, et non plus seulement souhaitant poursuivre un carrière dans le secteur privé. Les effets sur l'échantillon global sont ainsi qualitativement très similaires, à l'exception de l'obtention de la qualification CNU conditionnellement à une soumission, pour laquelle l'estimation manque de puissance statistique.

TABLEAU 3.9 – Mesures de qualité de la thèse – parmi les docteurs aspirant à travailler dans le secteur privé

	Pas de Contrôles				Contrôles			
	(1) Candit, CNU	(2) Publi	(3) Qualif	(4) Qualif condit,	(5) Candit, CNU	(6) Publi	(7) Qualif	(8) Qualif condit,
Docteur ingénieur non-CIFRE								
CIFRE	-0,139** (0,0625)	-0,160** (0,0742)	-0,116* (0,0603)	-0,128 (0,255)	-0,128** (0,0518)	-0,186*** (0,0710)	-0,105** (0,0478)	-0,455** (0,212)
R^2	0,122	0,036	0,133	0,127	0,311	0,149	0,325	0,679
Contrôles de base	✓	✓	✓	✓	✓	✓	✓	✓
PHNSF					✓	✓	✓	✓
CSP père					✓	✓	✓	✓
Mention Bac					✓	✓	✓	✓
Projet Pro Sout,					✓	✓	✓	✓
Observations	364	364	364	44	364	364	364	44

SOURCES : Enquêtes Génération 1ère vague; cohortes 2010 et 2013. L'échantillon se restreint aux individus docteurs et ingénieurs, interrogés 3 ans après l'obtention de leur diplôme. La variable « Candit. CNU » est une variable binaire égale à 1 si le docteur a postulé à la qualification au CNU. La variable « Publi » est égale à 1 si le docteur a publié dans une revue à comité de lecture avant sa soutenance de thèse. La variable « Qualif » est égale à 1 si le docteur a obtenu la qualification CNU. La colonne « Qualif condit. » est définie de façon équivalente à l'exception du fait qu'elle n'est définie que si le docteur a effectivement postulé à la qualification.

TABLEAU 3.10 – Mesures de qualité de la thèse (ensemble des docteurs de l'échantillon de base)

	Pas de Contrôles				Contrôles			
	(1) Candit, CNU	(2) Publi	(3) Qualif	(4) Qualif condit,	(5) Candit, CNU	(6) Publi	(7) Qualif	(8) Qualif condit,
Docteur ingénieur non-CIFRE								
CIFRE	-0,295*** (0,0515)	-0,145** (0,0609)	-0,281*** (0,0501)	-0,127 (0,108)	-0,187*** (0,0550)	-0,162*** (0,0575)	-0,155*** (0,0485)	-0,157 (0,114)
<i>R</i> ²	0,108	0,028	0,101	0,072	0,383	0,128	0,394	0,338
Contrôles de base	✓	✓	✓	✓	✓	✓	✓	✓
PHNSF					✓	✓	✓	✓
CSP père					✓	✓	✓	✓
Mention Bac					✓	✓	✓	✓
Projet Pro Sout,					✓	✓	✓	✓
Observations	641	641	641	205	641	641	641	205

SOURCES : Enquêtes Génération 1ère vague; cohortes 2010 et 2013. L'échantillon se restreint aux individus docteurs et ingénieurs, interrogés 3 ans après l'obtention de leur diplôme. La variable « Candit. CNU » est une variable binaire égale à 1 si le docteur a postulé à la qualification au CNU. La variable « Publi » est égale à 1 si le docteur a publié dans une revue à comité de lecture avant sa soutenance. La variable « Qualif » est égale à 1 si le docteur a obtenu la qualification. La colonne « Qualif condit. » est définie de façon équivalente à l'exception du fait qu'elle n'est définie que si le docteur a effectivement postulé à la qualification.

CHAPITRE 4

ANALYSE AU NIVEAU DES ENTREPRISES

CIFRE

Ce chapitre s'intéresse aux effets produits par le dispositif Cifre sur une deuxième unité d'analyse : les entreprises participant au programme. Il vise ainsi à caractériser l'évolution de variables susceptibles d'être modifiées par le recrutement d'un doctorant en Cifre dans les années précédant et suivant ce recrutement. Nous appliquons trois méthodes très similaires. Tout d'abord, dans la section 4.1, nous adoptons une stratégie d'étude d'événements échelonnée, c'est-à-dire que nous ne nous intéressons qu'aux entreprises traitées et comparons des périodes relatives au recours au reste des périodes dans l'échantillon. Nous complétons cette stratégie en section 4.2 en appliquant une approche par différence-de-différences. Nous ajoutons à l'analyse un groupe de contrôle (c'est-à-dire d'entreprises pouvant être considérées comme comparables mais jamais traitées), et produisons une analyse analogue à la première sur cet échantillon plus large, permettant un gain de précision. La section 4.3 présente une analyse proche en appliquant une méthode d'appariement avant de procéder à l'estimation d'une spécification de type différence-de-différences.

4.1 Analyse par étude d'événements

Cette section décrit en premier lieu la méthodologie des études d'événements (sous-section 4.1.1), puis présente l'équation d'estimation que nous utilisons (sous-section 4.1.2), et enfin présente les résultats obtenus grâce à cette méthode (sous-section 4.1.3).

4.1.1 La méthode des études d'événements

Comme suggéré dans les clauses techniques de l'appel d'offre et dans la réponse à l'appel d'offre, une première stratégie d'identification est menée, en suivant la méthodologie des études d'événements. Celle-ci restreint l'échantillon aux entreprises entrant dans le dispositif au cours de la période, et compare l'évolution de différentes variables d'intérêt de ces entreprises en fonction du moment de leur entrée dans le programme (recrutement d'un doctorant ou d'une doctorante en Cifre). Il s'agit donc d'une étude d'événements comparant uniquement les traités sur une fenêtre temporelle, en exploitant la variation entre entreprises de la date du traitement.

Par définition, les entreprises traitées – c'est-à-dire ayant recours au dispositif Cifre – dès le début de la période de l'échantillon sont exclues. Cette restriction amène ainsi à exclure l'essentiel des grands groupes français. S'ils occupent une place importante et ont a priori tous recruté en Cifre depuis la création de la mesure, le grand nombre de PME recrutant des doctorants Cifre (environ 60 % des entreprises pour 40 % des financements) permet néanmoins de construire l'étude d'événement sur un nombre suffisant d'entreprises. Dans les résultats présentés ci-après, nous procédons ainsi à des exclusions de l'échantillon selon deux critères. En premier lieu, **nous excluons l'ensemble des entreprises supposément traitées avant 2009** (le Siren des entreprises traitées est disponible avant cette date, sans toutefois que nous disposions d'informations fiables sur la réalisation de la

thèse Cifre). En outre, afin de mesurer l'effet sur des entreprises pour lesquelles le recrutement d'un premier doctorant Cifre est susceptible d'avoir un impact détectable statistiquement, **nous nous restreignons aux entreprises de moins de 100 salariés**. Nous montrons également en Annexe les résultats obtenus lorsque nous relâchons ou accentuons cette contrainte.

Une autre restriction potentielle concerne les entreprises traitées de multiples fois. Si l'on considère un échantillon allant de 2008 à 2016, la question se pose de savoir s'il est préférable de conserver ou bien d'exclure de l'échantillon les entreprises ayant recours aux Cifre plusieurs fois, par exemple en 2010 et 2012. Inclure des entreprises correspondant à ce cas de figure peut être problématique car l'intensité du traitement serait alors très hétérogène entre unités. Néanmoins les exclure reviendrait implicitement à conditionner sur la trajectoire future des entreprises présentes dans l'échantillon, créant potentiellement une sélection négative. En effet, conditionnellement au fait d'avoir recours au Cifre au moins une fois, le fait de n'y avoir recours qu'une seule fois est potentiellement associé à une croissance et une dynamique moindre de l'entreprise, voire à une expérience infructueuse la première fois. Dans les résultats de base, nous procédons à une sélection d'échantillon composée des entreprises **ayant eu recours à un maximum de 3 thèses Cifre sur la période** (le choix de ce seuil vise à éviter de sélectionner uniquement des entreprises qui n'auraient recruté qu'un seul doctorant Cifre car le résultat s'est avéré décevant). Nous montrons également nos résultats lorsque nous relâchons cette restriction.

En plus des restrictions évoquées précédemment, nous construisons un **panel d'entreprises cylindré**, c'est-à-dire que nous ne gardons dans notre échantillon d'estimation que les entreprises qui sont observées chaque année entre 2009 et 2016 dans nos données. Au total l'échantillon inclut 321 contrats Cifre distincts au cours de la période 2009-2016. Il s'agit d'une faible proportion des contrats Cifre associés à un employeur présent de façon continue dans les DADS sur la même période

(environ 7000). Comme décrit plus bas, l'estimation en différence-de-différences et appariement puis double différence sont fondées sur des échantillons plus larges.

4.1.2 Équation d'estimation

Nous estimons une spécification dynamique qui autorise l'effet du traitement à varier en fonction du temps écoulé depuis le traitement.

L'année de traitement d'une entreprise i est notée t_{i0} . On peut alors indexer le temps écoulé par rapport au traitement avec $d = t - t_{i0}$ où t est une année calendaire. L'indice d est négatif avant le traitement et positif après. Notre échantillon d'estimation couvre les années 2009 à 2016, ce qui correspond à la disponibilité jointe des données Cifre et du détail des postes occupés dans les données DADS Postes. La fenêtre d'estimation est définie par le biais d'une borne inférieure m_0 égale à -3 et supérieure m_1 égale à 3. Les observations incluses dans l'échantillon d'estimation sont donc telles que : $d \in \{-3, -2, \dots, 2, 3\}$.

L'équation d'estimation de base s'écrit alors :

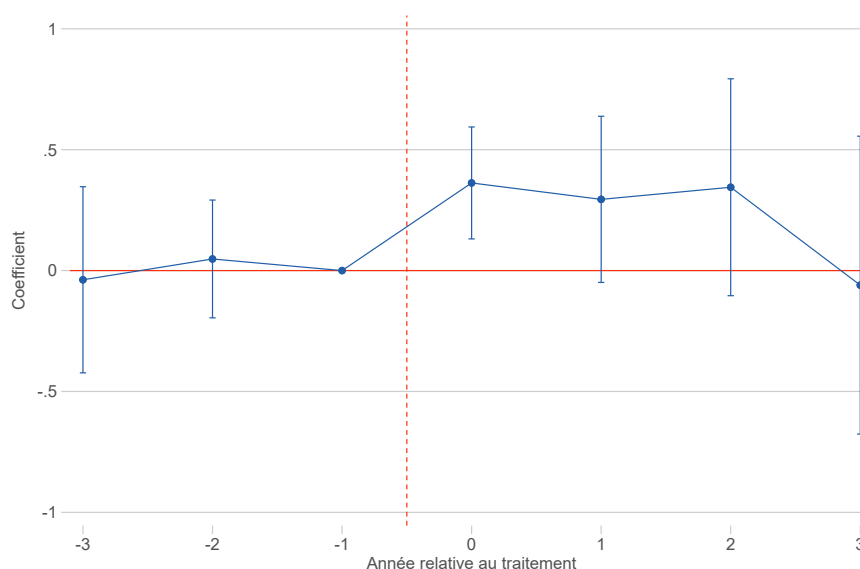
$$Y_{it} = \sum_{\substack{d=m_1 \\ d=-m_0+1 \\ d \neq -1}} \beta_d \times 1\{t = d + t_{0i}\} + \mathbf{x}'_{it} \boldsymbol{\delta} + \alpha_i + \psi_t + \varepsilon_{it}, \quad (4.1)$$

où $1\{\}$ est la fonction indicatrice ; α_i et ψ_t sont des effets fixes année et entreprise ; \mathbf{x}'_{it} est vecteur de variables de contrôle qui varient au cours du temps.

On propose ici d'omettre la variable indicatrice pour $d = -1$, et de combiner les variables indicatrices correspondant aux périodes comprises entre -5 et -3, et entre 3 et 5, c'est à dire en les contraignant à être égaux comme proposé par Schmidheiny et Siegloch (2019). Cette restriction est nécessaire pour éviter la multicollinéarité et pour identifier le processus de génération de données dynamiques sous-jacent si l'ensemble des unités incluses dans l'échantillon d'estimation est traité (Borusyak et Jaravel, 2017; Gross et al., 2019).

4.1.3 Résultats

FIGURE 4.1 – Évolution (de l'asinh) des heures travaillées par des ingénieurs de R&D autour du recours à la première thèse Cifre



NOTES : Cette figure présente les coefficients de l'estimation d'une spécification de différence-de-différences – voir équation (4.1). La sélection de l'échantillon d'estimation est détaillée dans la section 4.1.1.

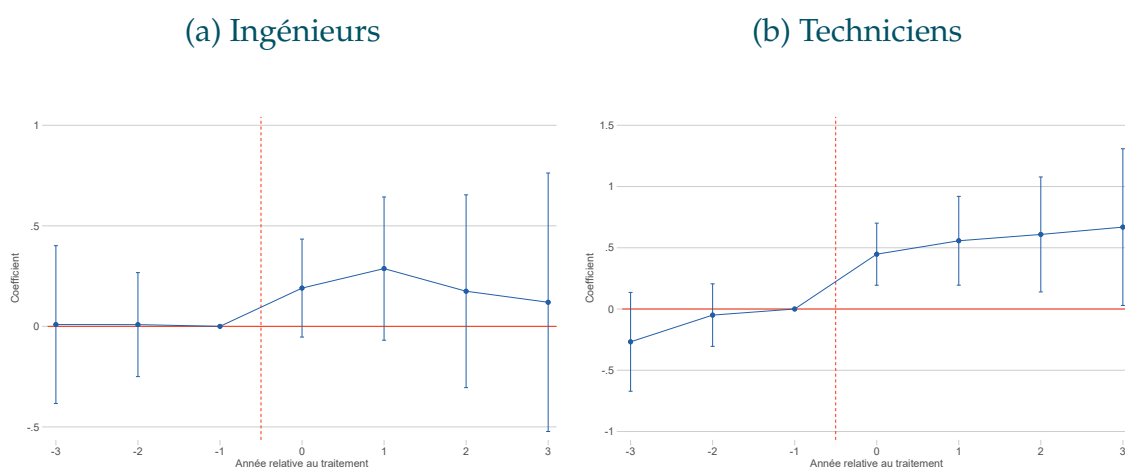
SOURCES : Bases Cifre, GECIR, FARE, DADS Postes.

La figure 4.1 montre les résultats de l'étude d'événements sur des variables d'emploi d'ingénieurs de R&D. Il représente les coefficients des variables indicatrices traitement \times année par rapport au recours, encadrés par leurs intervalles de confiance à 95 %, sur le asinh (réciproque du sinus hyperbolique, soit un équivalent du logarithme naturel pour les valeurs importantes mais défini en 0). Ces résultats constituent une régression de première étape : ils permettent de vérifier si le cadre empirique est à-même de détecter la variation générée a priori mécaniquement par le recours au programme¹, à savoir l'intégration d'un nouvel employé dont le poste devrait être qualifié comme ingénieur de R&D dans l'immense majorité des cas. La figure montre une évolution du nombre d'ingénieurs (total ou R&D) très stable avant le recours au dispositif, et une hausse importante au mo-

1. À moins d'une substitution systématique avec l'emploi d'autres ingénieurs de R&D.

ment du recours (environ 35 % de hausse du nombre d'ingénieurs de R&D). Cette hausse est significative dans les années suivant le recours, mais s'estompe néanmoins dans la troisième année après le recours à la thèse, qui correspond à l'année de fin du contrat (la première année étant l'année 0 dans la figure). Ceci suggère donc que cette hausse des effectifs de R&D est seulement temporaire, et dure le temps du contrat de travail liant le doctorant et son entreprise.

FIGURE 4.2 – Évolution (de l'asinh) des heures travaillées par des ingénieurs et des techniciens autour du recours à la première thèse Cifre



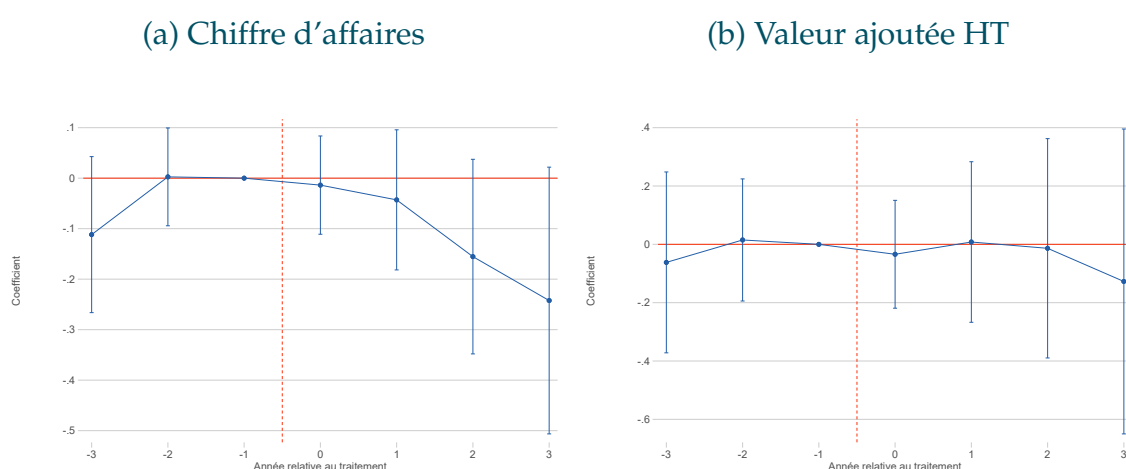
NOTES : Cette figure présente les coefficients de l'estimation d'une spécification de différence-de-différences – voir équation (4.1). La sélection de l'échantillon d'estimation est détaillée dans la section 4.1.1.

SOURCES : Bases Cifre, GECIR, FARE, DADS Postes.

La figure 4.2 montre l'évolution de l'asinh (arc sinus hyperbolique) du nombre d'heures travaillées par des ingénieurs (panneau a), et par des techniciens (panneau b). L'évolution des heures travaillées par des ingénieurs ne témoigne pas de tendance croissante pré-traitement, mais son évolution post-traitement est croissante mais moins nette que celle des ingénieurs de R&D. Ceci suggère que la catégorie des ingénieurs de R&D est une meilleure approximation du type de poste occupé par les doctorants Cifre, ce qui rend l'effet moins net. Le panneau b témoigne en revanche d'une augmentation très nette et en apparence permanente du nombre de techniciens employés par l'entreprise, avec une tendance pré-

traitement non-significative mais néanmoins légèrement croissante. Ceci suggère que le recrutement d'un doctorant en Cifre s'accompagne en réalité de mutations dans la structure de l'emploi dans l'entreprise, qui ne nous semblent pas interprétables causalement, mais nous paraissent plutôt révéler l'augmentation générale des activités de R&D à ce moment de la vie de l'entreprise.

FIGURE 4.3 – Évolution (du log) du chiffre d'affaires et de la valeur ajoutée



NOTES : Cette figure présente les coefficients de l'estimation d'une spécification de différence-de-différences – voir équation (4.1). La sélection de l'échantillon d'estimation est détaillée dans la section 4.1.1.

SOURCES : Bases Cifre, GECIR, FARE, DADS Postes.

La figure 4.3 montre l'évolution du logarithme du chiffre d'affaires (panneau a) et de la valeur ajoutée hors taxes (panneau b). Ces deux graphiques montrent très peu de différences entre les évolutions pré et post traitement des firmes traitées. L'aspect marquant de ces graphiques est la grande imprécision d'estimation qui caractérise les périodes tardives de l'estimation ($t = 2$ et $t = 3$), ce qui suggère que la stratégie d'estimation mise en place est relativement inadaptée pour détecter des effets du programme qui seraient supposés se manifester à plus long terme que les variables étudiées dans la régression de première étape (emploi d'ingénieurs de R&D). Ce qui semble être une baisse du chiffre d'affaires dans le panneau a ne nous semble ainsi pas interprétable en tant que tel, mais appelle plutôt à conduire une stratégie d'estimation alternative permettant d'ajouter de la précision aux co-

efficents que nous estimons.

4.2 Analyse par différence-de-différences

La méthode des études d'événements, dans laquelle le groupe témoin est constitué d'entreprises qui sont traitées à un moment différent, ne fonde l'estimation des effets du dispositif que sur la comparaison d'entreprises qui, à un moment ou un autre au cours de la période couverte par l'échantillon, vont effectivement être traitées. Cela peut être perçu comme un avantage si l'on suppose que le choix de rejoindre à un moment ou à un autre le dispositif Cifre est déterminé par une forme d'hétérogénéité inobservable qui est stable au cours du temps et difficilement captée par des variables observables. Néanmoins, il apparaît que cette technique souffre d'une puissance statistique limitée à cause du faible nombre d'entreprises ayant effectivement recours au programme Cifre.

Compte tenu des limites de cette stratégie d'estimation, nous mettons donc en place une stratégie d'estimation plus classique de différence-de-différences, où nous comparons des entreprises traitées autour du recours au dispositif à un autre groupe d'entreprises qui sont suffisamment comparables à notre groupe traité mais n'ont pas eu recours au dispositif dans la période d'étude. Une hypothèse sous-jacente à cette méthode d'estimation est que les entreprises appartenant au groupe de contrôle ne seront pas traitées dans une durée de moins de trois ans après la fin de notre période d'estimation, de manière à ce qu'aucune des variables d'année relative au traitement prise en compte dans notre estimation ne soit différente de zéro. Les coefficients estimés sont associés exactement aux mêmes variables que dans l'étude d'événements (variables indicatrices traitement \times année relative au traitement), mais relativement à un groupe d'entreprises plus important, et donc a priori avec plus de précision.

Cette section décrit en premier lieu la sélection d'échantillon à laquelle nous

procédons (sous-section 4.2.1), puis présente l'équation d'estimation que nous utilisons (sous-section 4.2.2), et enfin présente les résultats obtenus grâce à cette méthode (sous-section 4.2.3).

4.2.1 Sélection de l'échantillon

La sélection d'échantillon que nous réalisons est la suivante : afin de garantir l'activité de R&D des entreprises contrôles, nous nous concentrons sur les **entreprises demandant le Crédit Impôt Recherche (CIR)**. Afin d'obtenir, parmi les entreprises ayant recours au CIR, des entreprises les plus proches possibles des entreprises demandant des financements Cifre, une solution envisageable consiste à réaliser un appariement (par score de propension par exemple). Cet appariement serait réalisé sur des caractéristiques observables des entreprises ayant recours au CIR mais pas au Cifre, afin de sélectionner pour chaque entreprise traitée le ou les plus proches voisins de cette entreprise n'étant cependant pas traités. Cet appariement devrait néanmoins intégrer le fait que le traitement a lieu à des dates différentes pour chaque unité traitée, et qu'il est donc nécessaire d'apparier ces unités à une date précédant leur traitement avec d'autres unités ayant à un moment de leur existence été comparables avec cette entreprise traitée à la date précédant le traitement. En outre, si le recours au CIR nous assure de l'activité de R&D des entreprises incluses, il souffre néanmoins d'un problème majeur pour notre période d'étude, liée au fait qu'il a été réformé de manière très importante en 2008, et que les entreprises qui le demandaient déjà avant 2008 sont potentiellement de nature très différente de celles qui ont commencé à en bénéficier après la réforme. Nous réalisons cette méthode par appariement dans la section suivante (section 4.3).

Comme dans la section précédente, nous construisons un **panel d'entreprises cylindré**, c'est-à-dire que nous ne gardons dans notre échantillon d'estimation que les entreprises qui sont observées chaque année entre 2009 et 2016 dans nos don-

nées.

Ainsi, de manière à obtenir des entreprises les plus comparables possibles à nos entreprises traitées (dont le premier recours au programme Cifre ne peut avoir lieu qu'à partir de 2009) à la fois sur leurs caractéristiques observables et dans leur trajectoire, nous nous concentrons sur le **groupe des entreprises ayant demandé le Crédit Impôt Recherche pour la première fois à partir de 2009**, c'est-à-dire après sa réforme. Parmi ces entreprises demandant le CIR à une date postérieure à 2009, certaines entreprises accueillent également un doctorant en Cifre à une date postérieure à 2009, et nous estimons les coefficients d'interaction entre le fait d'avoir recours au dispositif Cifre et les années relativement au traitement.

TABLEAU 4.1 – Statistiques descriptives pour l'année 2009 de l'échantillon d'estimation par méthode de différence-de-différences

	(1) Groupe de traitement				(2) Groupe de contrôle			
	Moyenne	Médiane	1er décile	9e décile	Moyenne	Médiane	1er décile	9e décile
Première année Cifre	2012,25	2012	2009	2015	-	-	-	-
Âge	7,57	4,00	1,00	19,00	12,54	9,00	1,00	28,00
Chiffre d'affaires (M€)	2,2	0,64	0,44	0,51	2,61	0,94	0,16	5,6
Production totale (M€)	2,44	0,77	0,10	5,52	2,42	0,81	0,07	5,47
Valeur ajoutée (HT, M€)	1,20	0,44	0,01	3,03	1,17	0,49	0,08	2,59
EBE (M€)	0,07	0,03	-0,29	0,51	0,17	0,05	-0,10	0,52
Résultat net (M€)	0,07	0,04	-0,22	0,41	0,16	0,04	-0,08	0,40
Salaires et traitements (M€)	0,80	0,36	0,79	1,86	0,68	0,31	0,07	1,45
Nombre d'employés	17,74	8,00	2,00	42,50	16,16	7,00	1,50	35,75
Nombre d'ingénieurs	5,08	1,11	0,00	12,67	3,34	0,25	0,00	6,63
Nombre d'ingé, R&D	2,49	0,00	0,00	4,90	1,47	0,00	0,00	2,13
Capitaux propres (k€)	1,03	0,32	0,04	2,28	1,54	0,28	0,02	2,18
Impôt s/ bénéfices (k€)	-0,05	-0,02	-0,23	0,05	0,01	0,00	-0,07	0,09
Investissement (k€)	0,21	0,04	0,00	0,48	0,16	0,02	0,00	0,25
Nombre d'entreprises	402				8399			
Observations	3216				67192			

SOURCES : Bases Cifre, FARE, DADS Postes, CIR MVC.

Le tableau 4.1 permet de comparer les caractéristiques observables en 2009 des groupes de traitement et de contrôle pour notre analyse. Un premier élément qu'il faut noter est que le groupe de traitement auquel nous nous intéressons est constitué d'un total de 400 entreprises, soit environ 50 % de l'effectif présent dans l'étude d'événements précédente (739 entreprises). Nous comparons ces entreprises traitées à un groupe de contrôle d'environ 8000 entreprises. On constate que, malgré la restriction que nous imposons sur l'année de recours au CIR censée sélectionner

des tendances comparables, les entreprises ayant recours au programme Cifre sont légèrement plus jeunes que les autres. Elles ont en revanche des niveaux de chiffre d'affaires, de production, de valeur ajoutée, de main d'œuvre et de profitabilité très comparables. Au sein de leur main d'œuvre, les entreprises traitées accueillent un nombre d'ingénieurs quelque peu supérieur à leurs équivalents dans le groupe de contrôle. Au final, notre échantillon contient 642 contrats Cifre distincts sur un total d'environ 7,000 au cours de la période 2009-2016, soit une couverture d'environ 10 %.

4.2.2 Équation d'estimation

Nous notons T l'ensemble des entreprises i telles qu'elles font partie du groupe de traitement sur la période considérée. Afin de corriger des différences de tendances liées à la date de recours au CIR ou à l'appartenance sectorielle, notre spécification de base introduit des effets fixes année \times année de premier recours au CIR \times secteur (2 premiers chiffres de la NAF rev. 2). L'équation d'estimation est très proche de (4.1), elle s'écrit :

$$Y_{it} = \sum_{\substack{d=-m_0 \\ d \neq -1}}^{d=m_1} \beta_d \times \mathbf{1}\{t = d + t_{i0}\} \times \mathbf{1}\{i \in T\} + \mathbf{x}'_{it} \boldsymbol{\delta} + \alpha_i + \psi_{t, \tau(i), \iota(i)} + \varepsilon_{it}, \quad (4.2)$$

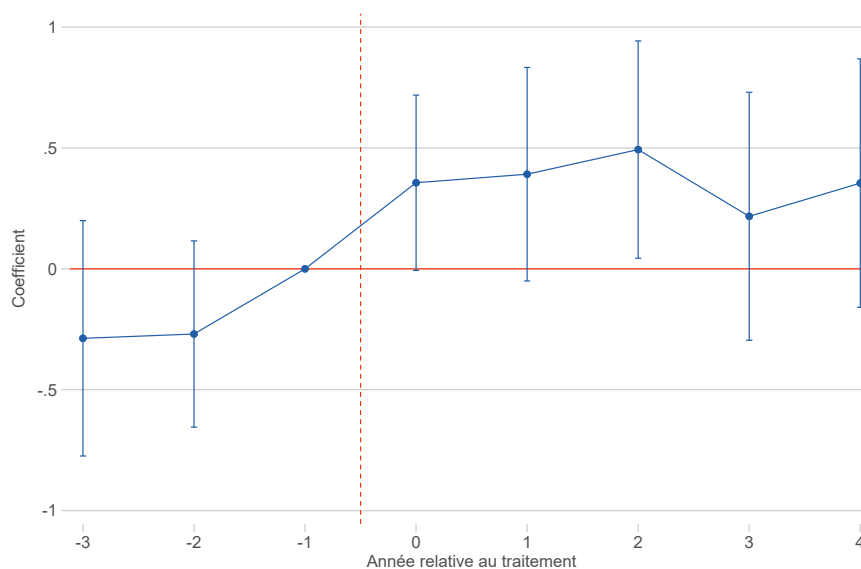
où $\tau(i)$ est l'année de premier recours au CIR et $\iota(i)$ le secteur de l'entreprise i . Nous regroupons les erreurs-types au niveau de la firme afin de répondre aux problèmes d'auto-corrélation des erreurs entre périodes (Bertrand et al., 2004).

Cette configuration nous offrant a priori plus de puissance statistique, nous retenons cette fois la fenêtre de $(-3; 4)$ années autour du traitement. Comme dans la stratégie précédente, nous excluons la variable indicatrice associée au traitement pour l'année -1.

4.2.3 Résultats

La figure 4.4 montre l'évolution autour du recours à un doctorant en thèse Cifre de l'asinh des heures travaillées par des ingénieurs de R&D, de manière exactement similaire à la figure 4.1 pour l'étude d'événements sur les firmes traitées uniquement. Elle confirme largement l'image donnée par cette dernière : l'emploi des ingénieurs de R&D est stable dans les années précédant le recours à une thèse Cifre (même s'il semble y avoir une augmentation dans l'année précédant le recours), et une augmentation marquée et significative a lieu au moment du recours et persiste sur toute la durée de la thèse, pour baisser légèrement ensuite et devenir non-significative dans l'année 3 suivant le recours, soit à la fin du contrat de thèse.

FIGURE 4.4 – Évolution (de l'asinh) des heures travaillées par des ingénieurs de R&D autour du recours à la première thèse Cifre

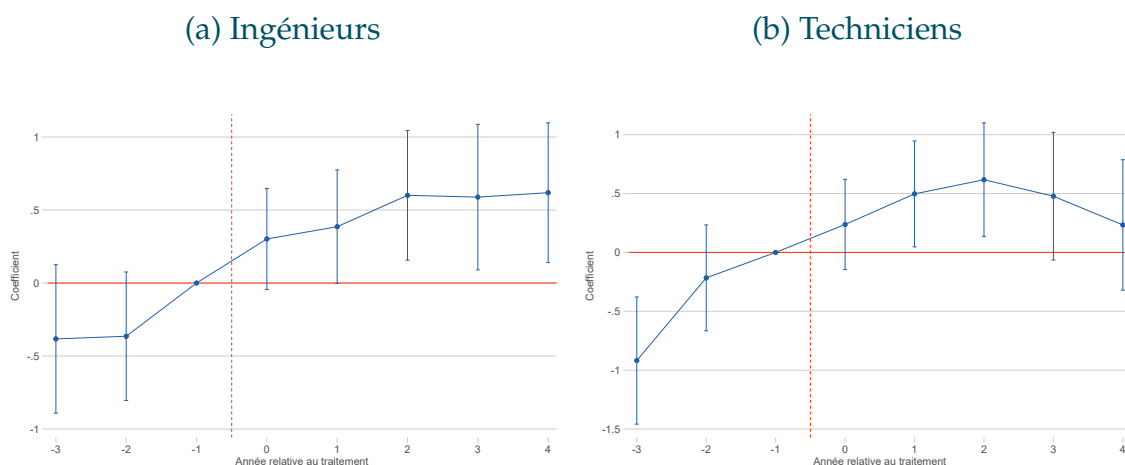


NOTES : Cette figure présente les coefficients de l'estimation d'une spécification d'étude d'évènements – voir équation (4.2). La sélection de l'échantillon d'estimation est détaillée dans la section 4.2.1.

SOURCES : Bases Cifre, GECIR, FARE, DADS Postes.

La figure 4.5 montre quant à elle l'évolution des coefficients traitement \times année relative au traitement pour les variables d'emploi d'ingénieur et de techni-

FIGURE 4.5 – Évolution (de l'asinh) des heures travaillées par des ingénieurs et des techniciens autour du recours à la première thèse Cifre



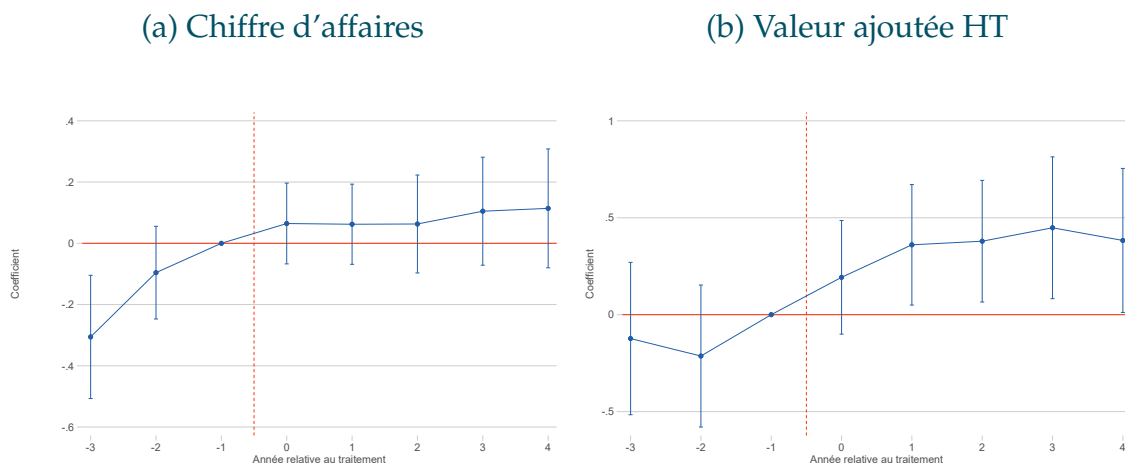
NOTES : Cette figure présente les coefficients de l'estimation d'une spécification d'étude d'évènements – voir équation (4.2). La sélection de l'échantillon d'estimation est détaillée dans la section 4.2.1.

SOURCES : Bases Cifre, GECIR, FARE, DADS Postes.

ciens, de manière similaire à la figure 4.2 pour l'étude d'événements. Elle confirme également l'image donnée par cette dernière. Dans le panneau a, on observe une augmentation du nombre d'ingénieurs employés par les firmes traitées autour du recours, qui semble cette fois permanente. Dans le panneau b, on observe une augmentation du nombre de techniciens, qui est néanmoins marquée par une tendance pré-traitement significative, et semble s'interrompre à la fin de la fenêtre temporelle étudiée.

La figure 4.6 reproduit la figure 4.3 réalisée pour l'étude d'événements, en exposant l'évolution du logarithme du chiffre d'affaires (panneau a) et de la valeur ajoutée hors taxes (panneau b). Le panneau a montre, comme dans le cadre de l'étude d'événements, une absence de variation du chiffre d'affaires sur la période, malgré l'existence dans ce cadre d'une tendance précédant le recours. Dans le panneau b, la valeur ajoutée semble en revanche connaître une croissance très marquée après le recours et qui paraît persistante. Afin de développer une meilleure

FIGURE 4.6 – Évolution (du log) du chiffre d'affaires et de la valeur ajoutée hors taxe



NOTES : Cette figure présente les coefficients de l'estimation d'une spécification d'étude d'événements – voir équation (4.2). La sélection de l'échantillon d'estimation est détaillée dans la section 4.2.1.

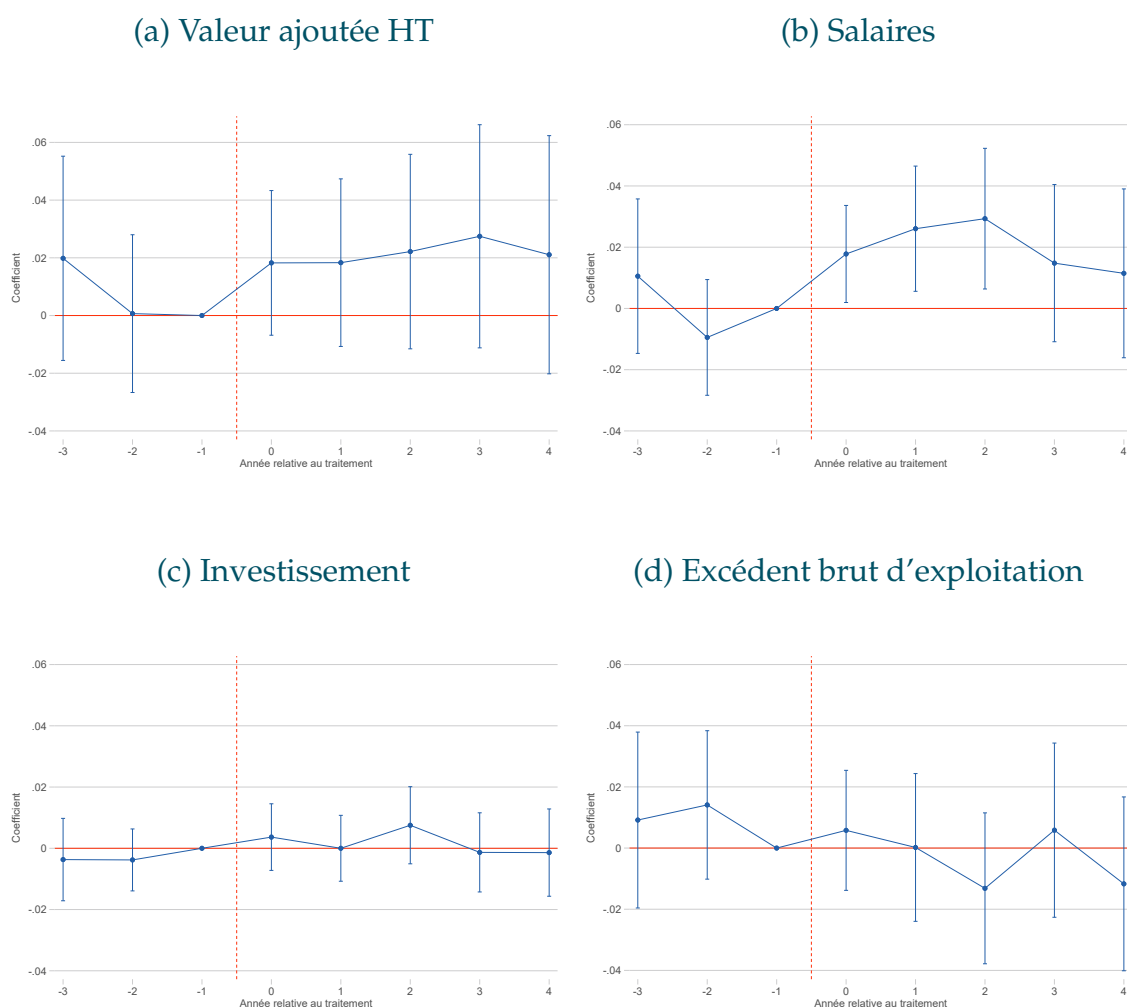
SOURCES : Bases Cifre, GECIR, FARE, DADS Postes.

compréhension de l'évolution des variables réelles de l'entreprise, nous optons ci-après pour une approche alternative. Au lieu d'étudier le logarithme ou l'asinh des variables d'intérêt, nous rapportons toutes ces variables au total du bilan de l'entreprise, afin de donner une échelle commune en niveau aux variations que nous considérons².

Les quatre panneaux de la figure 4.7 représentent respectivement la valeur ajoutée hors taxes (panneau a), les salaires (panneau b), l'investissement dans des immobilisations non financières (panneau c) et l'excédent brut d'exploitation (panneau d), tous rapportés au total du bilan de l'entreprise, et représentés sur une échelle commune allant de - 4 centimes à + 6 centimes par euro de bilan. Le panneau a confirme l'augmentation de la valeur ajoutée observée dans le panneau b de la figure 4.6 après le recours au dispositif Cifre, malgré une significativité du

2. Les deux dénominateurs les plus communs dans ce type d'approches sont les capitaux propres et le bilan, qui diffèrent donc dans l'inclusion ou non du montant de dette contractée par l'entreprise. Si la normalisation par les capitaux propres est relativement plus classique, la variable de bilan est mieux renseignée dans les données FARE, ce qui nous amène à préférer cette dernière.

FIGURE 4.7 – Évolution de variables comptables rapportées au total du bilan



NOTES : Cette figure présente les coefficients de l'estimation d'une spécification d'étude d'événements – voir équation (4.2). La sélection de l'échantillon d'estimation est détaillée dans la section 4.2.1.

SOURCES : Bases Cifre, GECIR, FARE, DADS Postes.

résultat moindre. Une telle augmentation, associée à l'augmentation tendancielle et lente du chiffre d'affaires chez les traités observée en figure 4.6, implique une baisse des consommations intermédiaires de l'entreprise particulièrement forte autour du recours à une thèse Cifre. Les panneaux b et c de la figure 4.7 sont informatifs sur l'usage fait par l'entreprise de cette augmentation de valeur ajoutée : on observe une hausse significative et de magnitude similaire à la hausse de valeur ajoutée des salaires versés aux employés, tandis que l'investissement ne connaît

aucune évolution différentielle entre nos groupes autour du recours au dispositif. En conséquence de cette augmentation des salaires versés, la profitabilité de l'entreprise, représentée ici par l'excédent brut d'exploitation (EBE) rapporté au bilan, ne connaît pas de variation différentielle autour du recours au programme Cifre. Deux précisions nous semblent nécessaires à l'interprétation de ces résultats. D'une part, si cet effet sur la valeur ajoutée peut sembler important (puisqu'il correspond à une hausse de près de 50 %), il concerne des entreprises dont la valeur ajoutée moyenne en 2009 était de 1.2 millions d'euros, c'est-à-dire un montant assez faible. D'autre part, plus qu'un effet causal de l'accueil d'un doctorant Cifre, ces résultats nous paraissent révéler des différences fondamentales de développement des entreprises recourant au programme relativement aux entreprises ne recourant qu'au CIR. Celles-ci n'en sont pas moins informatives : pour ce groupe d'environ 400 entreprises, l'accueil d'un doctorant Cifre accompagne un développement réel des fonctions de recherche-développement, et ne semble pas souffrir d'un problème de capture qui pourrait survenir si l'on voyait à l'inverse la profitabilité de l'entreprise augmenter suite au recours.

Il est au premier abord peu aisé d'interpréter un effet positif sur la valeur ajoutée sans hausse du chiffre d'affaires. En outre, développer l'activité des entreprises est certainement plus proche de l'esprit des objectifs du dispositif Cifre que ne l'est la restructuration de leurs charges. Une possible interprétation serait que les entreprises ayant recours au dispositif Cifre tendent à vouloir internaliser un ensemble d'activités auparavant sous-traitées. Les entreprises sont en moyenne plus jeunes, plus intensives en R&D et plus dynamiques (si l'on en juge les prétendances du chiffre d'affaire). Il est alors plausible que le recours à un contrat Cifre soit un premier pas vers la création d'un service R&D. Cette internalisation s'accompagne de l'embauche/rétention de personnel complémentaire (ingénieurs et techniciens) même au-delà de la durée du contrat Cifre, ce qui explique également une hausse de la masse salariale de l'entreprise.

En annexe, nous présentons les résultats correspondant à des spécifications qui incluent un ensemble plus riches d'effets fixes temporels. Les spécifications présentent des effets fixes année-secteur-taille-âge où la taille et l'âge sont mesurés en quintiles en 2009. Cette spécification permet de prendre en compte les différences d'âge notamment qui sont visibles dans le tableau 4.1. Les résultats sont présentés dans l'annexe A.3 – voir Figures A.4, A.5 et A.6.

4.3 Analyse complémentaire : appariement et différence-de-différences

Les résultats présentés dans la section précédente reposent sur une approche en différence-de-différences. Nous appliquons ici une méthode proche. Nous commençons par associer plusieurs entreprises contrôle à chaque entreprise traitée sur la base de variables observables via un appariement exact sur des caractéristiques discrètes – par exemple, le secteur d'activité – puis un appariement sur la base de la distance Mahalanobis en utilisant cette fois-ci des variables continues – par exemple, la part de la R&D dans les heures travaillées.

Une fois cette appariement effectué, nous procédons à l'estimation d'une équation différence-de-différences standard en prenant soin de comparer chaque entreprise traitée, avant et après la mise en place de la politique, au groupe d'entreprises contrôles auxquelles elle a été associée par la procédure d'appariement.

Cette approche étant proche de celle de la section 4.2, il est utile de préciser ici les points principaux qui les différencient. L'appariement associe un groupe de contrôle explicite pour chaque entreprise traitée – sous réserve de support commun. Un avantage qui découle de la constitution de ces groupes de contrôle dans le cas des traitements *staggered* est de pouvoir explicitement comparer les dynamiques des entreprises traitées et contrôles autour de la date du début du trai-

tement des entreprises traitées. En l'absence d'appariement, il est en effet peu évident d'attribuer à chaque entreprise contrôle une période par rapport au traitement, dans la mesure où ces entreprises contrôles ne sont pas explicitement associées à une entreprise traitée donnée³. Un autre avantage potentiel est de pouvoir contrôler de façon non-paramétrique pour les variables incluses dans le processus d'appariement exact⁴.

Nous décrivons en détail la procédure d'appariement dans la section 4.3.1, nous présentons des statistiques descriptives ainsi que de l'équation d'estimation dans la section 4.3.2. Enfin, la section 4.3.3 contient les résultats.

4.3.1 Description de la procédure d'appariement

La procédure d'appariement est réalisée sur la base de l'échantillon d'estimation de l'analyse en différence-de-différences – voir Section 4.2.1 pour une explication des règles déterminant sa constitution. L'entrée dans le dispositif Cifre, le traitement, a lieu à des moments différents pour chaque entreprise. Il est donc approprié de mettre en place une procédure d'appariement séparément pour chaque événement. La date pertinente est donc l'année relative à l'adhésion de l'entreprise au programme Cifre. Nous indiquons t_{i0} la date à laquelle l'entreprise i est traitée et d le nombre de périodes relatives à cet événement lors de l'année calendaire t (on a donc $d \equiv t - t_{i0}$).

Nous procédons à un appariement sur la base d'observations, une entreprise observée une année donnée, associée à $d = -1$. Nous excluons du groupe témoin potentiel toute entreprise traitée lors de la période d'estimation (2009-2016). Dès

3. Par exemple, Kleven et al. (2019) alloue une date de traitement de façon aléatoire aux unités de leur groupe de contrôle. Dans la section 4.2, nous avons plutôt fait l'hypothèse qu'aucune entreprise du groupe de contrôle ne serait traitée pour une durée d'au moins 3 ans après la fin de notre période d'estimation, de telle sorte que les variables indicatrices associées au période relativement au traitement soient toutes égales à zéro

4. Néanmoins il n'est pas totalement clair dans quelle mesure cette approche est supérieure à la simple inclusion d'effets fixes associées aux variables catégorielles utilisées dans l'appariement exact.

lors, le groupe témoin pour l'entreprise i inclut potentiellement l'ensemble des entreprises jamais traitées (par CIFRE), ayant demandé le Crédit Impôt Recherche pour la première fois à partir de 2009 et présente dans la base de donnée issus des appariements des bases FARE et DADS lors de l'année calendaire $t_{i0} - 1$.

En ce qui concerne les détails de l'appariement, nous combinons deux approches : (i) un appariement exact sur des variables catégorielles ou bien discrétisées – c'est-à-dire initialement continues mais rendues catégorielles via discrétisation (*coarsening*)–, (ii) un appariement sur la base de la distance de Mahalanobis entre variables continues.

L'appariement exact sur variables catégorielles s'appuie sur les caractéristiques suivantes :

1. Participation au programme JEI et CIR
2. Taille de l'entreprise (5 quantiles)
3. Âge de l'entreprise (5 quantiles)
4. Secteur

Suite à cette procédure, certaines entreprises traitées sont associées à un grand nombre d'entreprises de contrôle. Au contraire certaines entreprises traitées présentent une combinaison de variables discrètes qui ne correspond à aucune entreprise contrôle. Dans ce cas d'absence de « support commun », les entreprises ne sont pas incluses dans l'échantillon d'estimation.

Une fois, l'appariement exact réalisé nous limitons le nombre d'entreprise contrôles à 5 en utilisant les plus proches voisins sur la base du score de Mahalanobis. Les variables utilisées pour évaluer la distance sont :

1. La part des ingénieurs R&D dans les heures travaillées à $t_{i0} - 1$ et $t_{i0} - 2$.
2. La part de l'ensemble des ingénieurs dans les heures travaillées à $t_{i0} - 1$.

L'appariement sur la part des ingénieurs R&D aux dates $t_{i0} - 1$ et $t_{i0} - 2$ permet de

prendre compte des différences de niveau mais également d'évolution de l'emploi R&D pré-Cifre.

4.3.2 Statistiques descriptives

Le tableau 4.2 présente des statistiques descriptives concernant l'échantillon d'estimation en distinguant groupe de traitement (colonnes de gauche) et groupe de contrôle (colonnes de droite).

On constate que le nombre d'entreprises incluses dans le groupe de traitement est plus faible que dans la stratégie précédente. Cela est lié au fait que certaines entreprises traitées présentes une combinaison des valeurs des variables catégorielles utilisées dans l'appariement exact qui ne correspondent à aucune entreprise contrôle potentielle. Comme mentionné ci-dessus, en l'absence de support commun ces entreprises sont exclues de l'analyse⁵.

On constate qu'il y a moins d'entreprises que dans l'échantillon de différence-de-différences – ce qui est lié à l'exclusion d'observations suite à l'appariement exact. Au final, notre échantillon contient 395 contrats Cifre distincts sur un total d'environ 7,000 au cours de la période 2009-2016, soit une couverture d'environ 6 %. Le nombre d'entreprises distinctes dans le groupe de contrôles est proche de 5 fois celui des entreprises traitées, ce qui reflète vraisemblablement le fait que peu d'entreprises servent de contrôles plusieurs reprises. Enfin on voit, logiquement, que les différences en termes notamment d'âge et de part des ingénieurs, sont atténuées en comparaison avec l'échantillon précédent (voir 4.1).

5. On note néanmoins que dans l'approche de différence-de-différences avec contrôles (voir résultats en annexe de ce chapitre), l'inclusion d'effets fixes année interagies avec un ensemble de variables catégorielles va également aboutir à l'exclusion des entreprises traitées qui présentent une combinaison de valeurs de ces variables qui ne correspond à aucune entreprise contrôle.

TABLEAU 4.2 – Statistiques descriptives pour l’année 2009 de l’échantillon d’estimation par appariement et méthode de différence-de-différences

	(1) Entreprises traitées				(2) Entreprises contrôles			
	Moyenne	Médiane	1er décile	9ème décile	Moyenne	Médiane	1er décile	9ème décile
Année de traitement	2012,85	2013,00	2010,00	2016,00	-	-	-	-
Âge	13,37	9,00	3,00	28,00	14,44	11,00	4,00	27,00
% ingénieurs R&D (heures)	0,12	0,00	0,00	0,53	0,11	0,00	0,00	0,49
% ingénieurs (heures)	0,23	0,09	0,00	0,75	0,21	0,08	0,00	0,71
Chiffre d’affaires (k€)	3407,29	965,31	132,54	7615,30	2605,15	1080,70	196,22	6765,92
Valeur ajoutée (HT, k€)	1949,65	689,19	122,73	4640,99	1386,08	676,60	127,68	3446,53
Excédent brut d’exploitation (k€)	519,91	76,53	-91,06	770,83	244,58	59,45	-64,52	670,21
Résultat net (k€)	422,06	64,30	-37,93	644,47	208,05	56,42	-52,40	531,17
Salaires et traitements	992,12	437,42	91,01	2780,25	780,65	409,36	87,83	1791,99
Nombre d’employés	22,42	9,38	2,00	52,50	17,88	9,00	2,00	43,00
Nombre d’ingénieurs	6,27	1,11	0,00	14,76	4,23	1,00	0,00	10,03
Nombre d’ingénieurs de R&D	3,67	0,00	0,00	5,87	2,28	0,00	0,00	4,55
Capitaux propres (k€)	810,62	302,13	50,93	2316,89	1391,50	293,60	25,15	2208,16
Impôt sur les bénéfices (k€)	-28,36	-1,07	-173,96	83,74	26,48	0,00	-81,47	127,67
Investissement (non financier, k€)	321,20	44,83	1,02	510,26	127,65	21,66	0,00	289,51
Nombre d’entreprises	295				1471			

SOURCES : Bases Cifre, FARE, DADS Postes, CIR MVC.

Equation d’estimation. L’équation d’estimation est proche de (4.2), elle s’écrit :

$$Y_{it} = \sum_{\substack{d=-m_0 \\ d \neq -1}}^{d=m_1} \beta_d \times \mathbf{1}\{t = d + t_{i0}\} \times \mathbf{1}\{i \in T\} + \mathbf{x}'_{it} \boldsymbol{\delta} + \alpha_i + \psi_{t,g(i)} + \varepsilon_{it}, \quad (4.3)$$

où $g(i)$ est le groupe d’entreprises appariées auquel l’entreprise i est associée. Nous voulons utiliser la variation dans le traitement au sein des groupes d’entreprises appariées g mais pas entre groupes. À cette fin, nous incluons un effet fixe groupe-année dénoté $\psi_{t,g(i)}$ ⁶. Nous regroupons les erreurs-types au niveau de la firme afin de répondre aux problèmes d’auto-corrélation des erreurs entre périodes (Bertrand et al., 2004) mais aussi de prendre en compte le fait que certaines entreprises témoins peuvent servir de contrôle plusieurs fois – même si comme évoqué ci-dessus ce cas de figure semble rare.

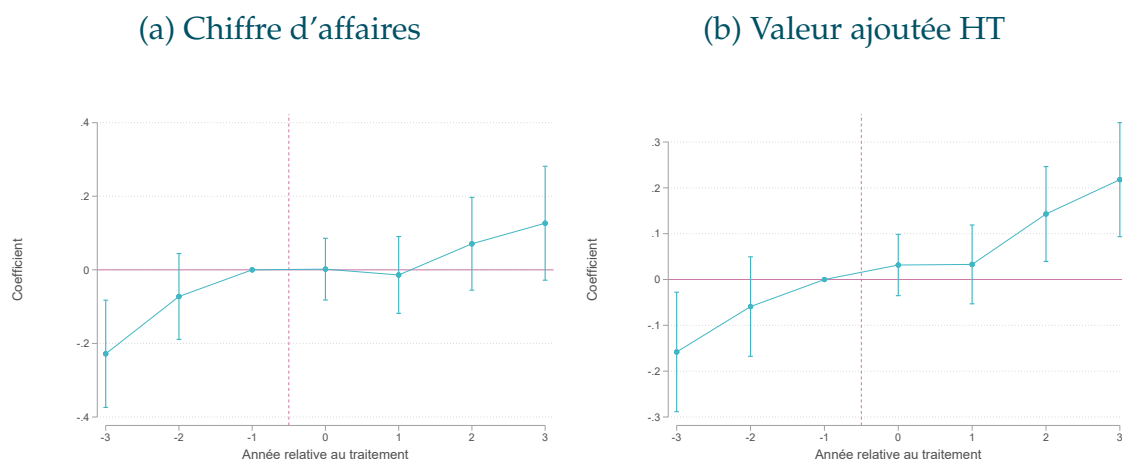
6. Voir par exemple Dube et al. (2010), équation (6), pour une spécification proche.

4.3.3 Résultats

Il est instructif de comparer la Figure 4.6 à la figure 4.8 qui présentent les résultats de l'estimation sur le chiffre d'affaire et la valeur ajoutée pour la méthode de différence-de-différences et l'appariement plus différence-de-différences respectivement.

Concernant le chiffre d'affaire les résultats sont très proches : on constate une violation des pré-tendances et une absence d'effet significatif à 95%. Dans le cas de la valeur ajoutée, on obtient dans les deux cas un effet positif avec néanmoins une violation des pré-tendances dans le cas de l'appariement. La violation provient néanmoins en grande partie d'une plus grande précision de l'estimation avec appariement plutôt que d'une ampleur très différente entre méthodes. Le coefficient positif à $d = +4$ dans le cadre de l'estimation avec appariement est plus faible, environ +20 % contre +40% dans le cadre de la simple différence-de-différences.

FIGURE 4.8 – Appariement et DD : chiffre d'affaires et de la valeur ajoutée



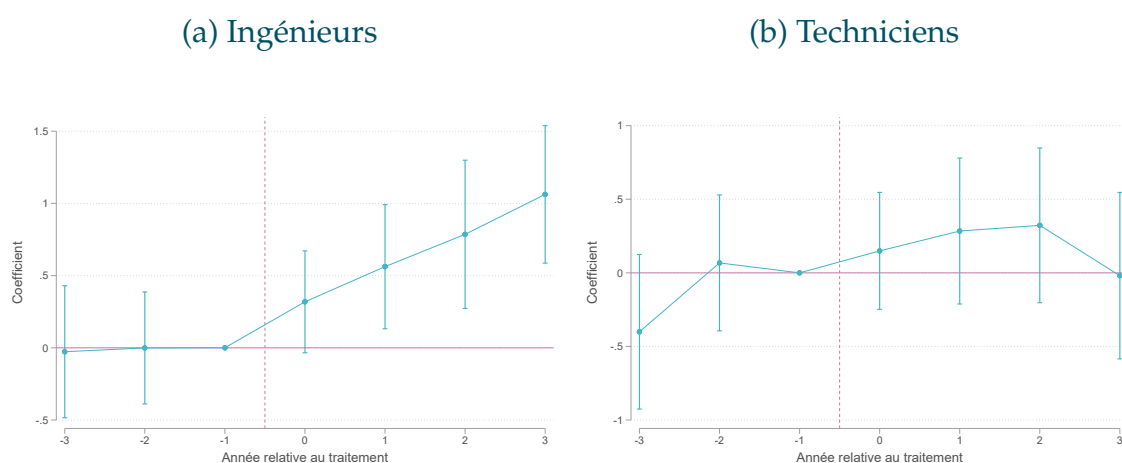
NOTES : Cette figure présente les coefficients de l'estimation d'une spécification de différence-de-différences – voir équation (4.3) – sur un échantillon apparié. La procédure d'appariement est détaillée dans la section 4.3.1. La sélection de l'échantillon d'estimation est similaire à celle détaillée dans la section 4.2.1. L'emploi est mesuré par l'asinh des heures travaillées par des ingénieurs et des techniciens. DD = différence-de-différences.

SOURCES : Bases Cifre, GECIR, FARE, DADS Postes.

En ce qui concerne l'emploi d'ingénieurs et de techniciens (Figure 4.9), on constate de meilleures pré-tendances que dans la simple différence-de-différences (Figure

4.5) et un résultat plus fort sur l'emploi d'ingénieurs, ainsi qu'une absence d'effets sur l'emploi de techniciens. Le résultat sur les techniciens en différence-de-différences simple souffre d'une violation des pré-tendances et est donc difficile à interpréter. Les résultats sur les variables rapportées au total du bilan (Figure 4.10) sont très proches de la simple différence-de-différences (Figure 4.7).

FIGURE 4.9 – Appariement et DD : Emploi par occupation

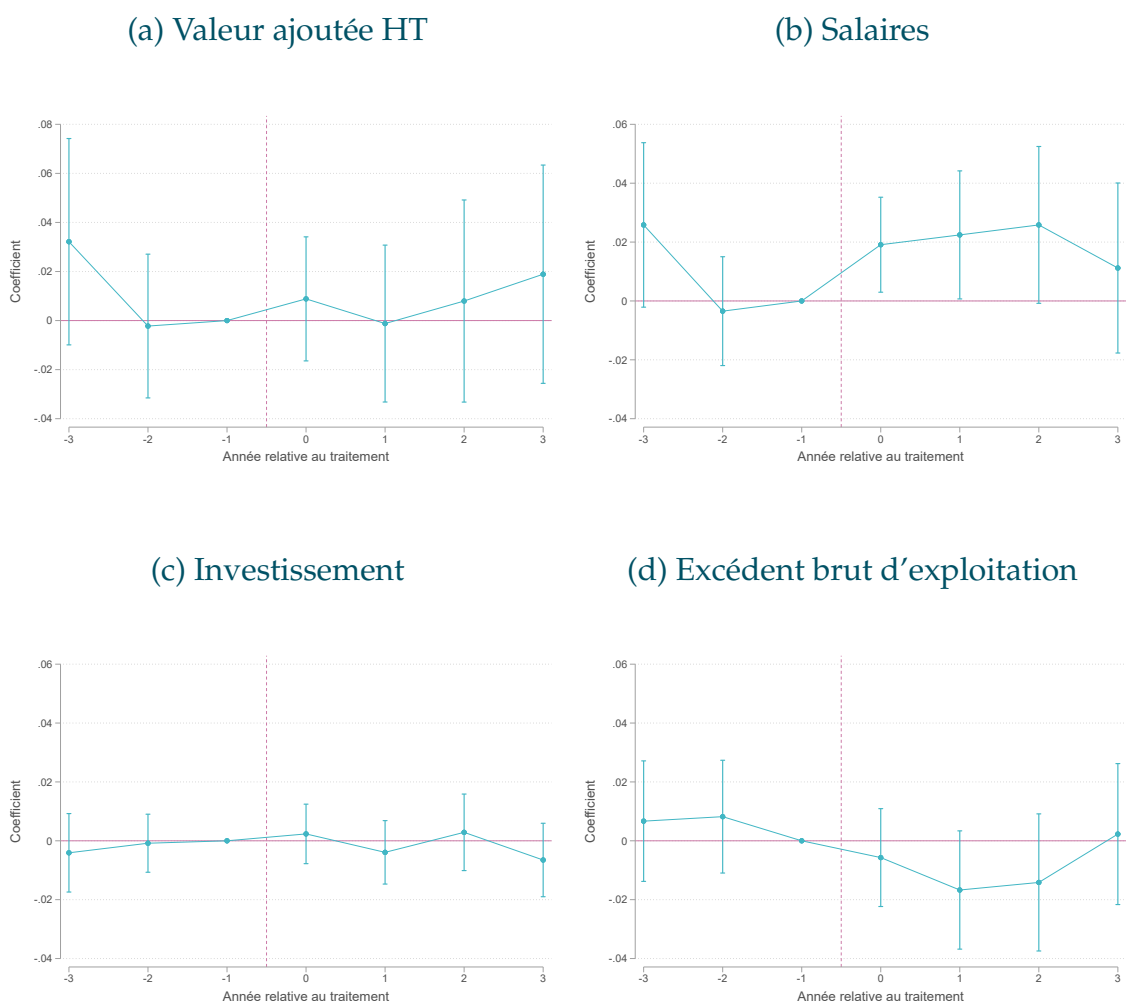


NOTES : Cette figure présente les coefficients de l'estimation d'une spécification de différence-de-différences – voir équation (4.3) – sur un échantillon apparié. La procédure d'appariement est détaillée dans la section 4.3.1. La sélection de l'échantillon d'estimation est similaire à celle détaillée dans la section 4.2.1. L'emploi est mesurée par l'asinh des heures travaillées par des ingénieurs et des techniciens. DD = différence-de-différences.

SOURCES : Bases Cifre, GECIR, FARE, DADS Postes.

Globalement, les résultats provenant de la combinaison de l'appariement et différence-de-différences présentés dans cette section sont très proches de ceux obtenus via l'approche par différence-de-différences.

FIGURE 4.10 – Appariement et DD : variables comptables rapportées au total du bilan



NOTES : Cette figure présente les coefficients de l'estimation d'une spécification de différence-de-différences – voir équation (4.3) – sur un échantillon apparié. La procédure d'appariement est détaillée dans la section 4.3.1. La sélection de l'échantillon d'estimation est similaire à celle détaillée dans la section 4.2.1. L'emploi est mesuré par l'asinh des heures travaillées par des ingénieurs et des techniciens. DD = différence-de-différences.

SOURCES : Bases Cifre, GECIR, FARE, DADS Postes.

CONCLUSION

Ce travail constitue une tentative de présentation et d'évaluation d'impact ex-post du dispositif Cifre. Il a exploité conjointement les données de l'ANRT sur le suivi du dispositif Cifre et un ensemble de bases de données administratives et issues d'enquêtes. Plusieurs enseignements se dégagent des résultats obtenus.

Un premier enseignement est que le dispositif Cifre semble avoir un effet positif sur l'intégration sur le marché du travail des docteurs, avec une plus forte probabilité d'occuper un emploi stable, et un salaire plus élevé trois ans après la soutenance. On note par ailleurs que les docteurs Cifre tendent également à soutenir leur thèse plus rapidement. Le dispositif apparaît donc présenter un certain nombre d'avantages pour les docteurs bénéficiant du programme. Notre travail souligne néanmoins, par contraste avec ces avantages, une moindre qualité des thèses sur des critères universitaires. Celle-ci semble en effet inférieure à celle des thèses non-Cifre sur la base de plusieurs indicateurs, certes imparfaits mais néanmoins informatifs (publication avant soutenance, obtention de la qualification au CNU). Ceci confirme un ancrage du dispositif dans la recherche appliquée, et donc un produit de la thèse d'une nature différente de celui des thèses classiques. Si les docteurs, qui sont partie prenante du dispositif, apparaissent effectivement en bénéficiaire via leur position sur le marché du travail, il est possible que les thèses Cifre soient à l'origine de moindres externalités positives que les thèses conventionnelles, si les résultats de ces thèses ont moins de chances d'être diffusés via des publications dans des revues scientifiques. Il s'agit là d'une considération po-

tentiellement pertinente dans l'analyse coût-bénéfice du dispositif et appelle possiblement à des études complémentaires.

Un second enseignement concerne les entreprises. Si les méthodes employées ne garantissent l'identification d'un effet causal que sous des hypothèses assez strictes, elles sont a minima informatives sur la dynamique des entreprises qui ont recours au dispositif. On constate que l'accueil d'un doctorant Cifre s'opère dans le cadre d'une augmentation de l'emploi ingénieur dans l'entreprise. Dans deux des trois cadres empiriques exploités, on constate une hausse de la valeur ajoutée générée par l'entreprise autour du recours au dispositif, qui semble intégralement redirigée vers les salaires versés et accompagne donc l'augmentation d'emploi de main d'œuvre de R&D, et n'implique aucune variation sensible d'investissement ou de profitabilité.

En conclusion, on constate que le dispositif Cifre est un dispositif de taille modeste mais dynamique au cours de la période. Il est accompagné d'effets positifs sur la vie professionnelle des docteurs, et est utilisé par des entreprises qui sont sur une tendance d'activité croissante. Au-delà des effets sur les parties prenantes privées au dispositif, il semble opportun à l'avenir d'étudier en détail les externalités associées à cette politique ainsi que les potentiels effets en matière de réseaux de collaboration entre entreprises et laboratoires publics.

ANNEXES

A.1 Discussion des hypothèses d'identification de l'étude d'événements

Bien que la méthode des études d'évènements soit très fréquemment employée, les conditions sous lesquelles elle permet d'aboutir à une estimation non-biaisée des effets qu'elle vise à mesurer ne sont pas toujours explicitées. Cette sous-section a pour objet de clarifier les hypothèses d'identification sous-tendant les études d'évènement. Pour étudier sous quelles hypothèses l'estimation de l'équation (4.1) estime l'effet de traitement moyen sans biais, il est utile de suivre Abraham et Sun (2019) et d'introduire une notation en résultat potentiel (*potential outcome*).

Dans les études d'événements, le vecteur des états de traitement d'une unité peut être caractérisé de façon unique par la période du traitement initial, la date de l'évènement. Nous désignons cette variable aléatoire discrète par $t_{i0} = \min\{t : D_{i,t} = 1\}$. En effet, le traitement est un état absorbant (une fois traitée, une unité est considérée comme toujours traitée quelque soit la valeur de d) et le vecteur des valeur de la variable de traitement $D_{i,t}$ est dès lors une séquence non-décroissante (en t , les années calendaires, mais aussi en d le temps relatif au traitement) de zéros puis de uns, c'est-à-dire $D_{i,t} < D_{i,s}$ pour $s < t$.

Nous notons $Y_{i,t}^c$ pour indiquer le résultat potentiel dans la période t sous une trajectoire de traitement hypothétique (date de l'évènement) c . Soit $Y_{i,t}^B$ pour indi-

quer le résultat si l'unité i ne reçoit aucun traitement. Il s'agit de la trajectoire de base, c'est-à-dire celle qu'aurait suivi l'unité i en l'absence de traitement.

Nous définissons l'effet du traitement pour un temps depuis la date de l'événement c comme la différence entre la trajectoire de base $Y_{i,t}^B$ et la trajectoire lorsque le traitement a eu lieu en c :

$$\alpha_{i,t}^c \equiv Y_{i,t}^c - Y_{i,t}^B. \quad (\text{A.4})$$

Bien entendu, pour chaque unité, nous observons chaque unité sous un seul chemin de traitement t_{i0} . Dès lors, le résultat observé pour l'unité i est simplement $Y_{i,t} = Y_{i,t}^{t_{i0}}$.

On peut exprimer la variable d'intérêt observée comme la somme de la trajectoire de base et des effets spécifiques à chaque date de traitement :

$$Y_{i,t} = Y_{i,t}^B + \sum_c (Y_{i,t}^c - Y_{i,t}^B) \times \mathbf{1}\{c = t_{i0}\} \quad (\text{A.5})$$

Abraham et Sun (2019) montrent que sous trois ensemble d'hypothèses, l'estimation de l'équation (4.1) par moindres carrés ordinaires aboutit à une estimation causalement interprétable de l'effet moyen.

1. La première hypothèse⁷ concerne l'homogénéité des effets moyens entre "cohortes" où une cohorte fait référence à l'ensemble des unités qui ont reçu le traitement à une date donnée (par exemple la cohorte c est définie comme : $\{i \in N : t_{i0} = c\}$). Le traitement moyen pour la cohorte c pour une distance au traitement donnée l est défini comme $CATT_l^c = E(\alpha_{i,c+l}^c | t_{i0} = c)$. L'hypothèse s'écrit donc :

$$CATT_l^c = CATT_l, \forall c$$

2. La deuxième hypothèse⁸ est une généralisation de l'hypothèse de tendance

7. Il s'agit de *Assumption 3* dans Abraham et Sun (2019).

8. Il s'agit de *Assumption 1* dans Abraham et Sun (2019).

commune entre cohortes. Formellement :

$$E(Y_{i,t}^B - Y_{i,s}^B | t_{i0} = c) = E(Y_{i,t}^B - Y_{i,s}^B), \forall c \in \text{supp}(t_{i0}), \forall t, s$$

où supp est le support de la variable aléatoire.

3. La troisième hypothèse⁹ concerne l'absence d'anticipation. Elle revient à supposer que le traitement futur n'a pas d'effet dans le passé. Formellement :

$$Y_{i,c+d}^e = Y_{i,c+d}^B, \forall d < 0$$

Elle implique qu'avant la date de traitement ($d < 0$), la variable observée correspond à la variable en l'absence de traitement : $Y_{i,c+d}^B$. Dès lors, sous cette hypothèse, le comportement pour $d < 0$ de la variable dépendante observée $Y_{i,t}$ avec $t < t_{i0}$ est informatif quant à la validité de l'hypothèse de tendance commune énoncée ci-dessus.

Dans le cadre de notre application, l'hypothèse d'homogénéité de l'effet du traitement entre cohortes ne semble pas particulièrement problématique.

Les entreprises ayant recours au dispositif Cifre pour la première fois au cours de la période 2008-2016 tendent à être des PME et auront certainement des caractéristiques assez proches, au moins entre cohortes. De plus, l'hétérogénéité liée à des évolutions structurelles de l'économie au cours de la période est vraisemblablement limitée du fait de la courte durée de la période couverte par l'échantillon.

L'hypothèse vraisemblablement la plus problématique est l'absence d'effet d'anticipation. En effet, avoir recours au dispositif Cifre est le fruit d'un choix délibéré de l'entreprise et nécessite vraisemblablement un long travail de préparation (coordination avec un laboratoire, etc.). Dès lors il n'est pas clair que l'absence ou bien la présence de tendance pré-traitement (*pre-trends*) reflète l'anticipation de la

9. Il s'agit de *Assumption 2* dans Abraham et Sun (2019).

politique ou bien une violation de l'hypothèse 2 de tendance commune.

A.2 Résultats additionnels pour l'étude d'événements

FIGURE A.1 – Évolution de l'asinh du nombre d'ingénieurs de R&D, variations autour de la limite d'effectif des entreprises

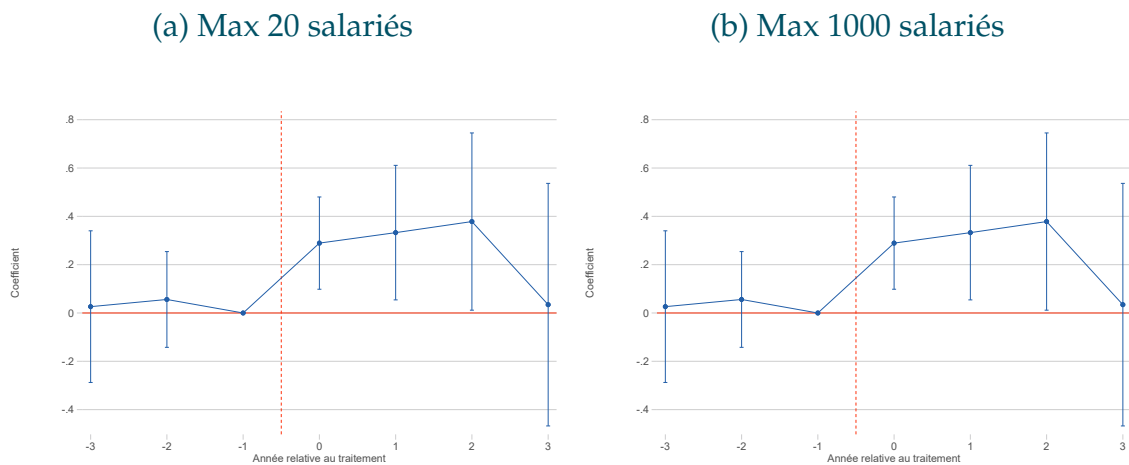
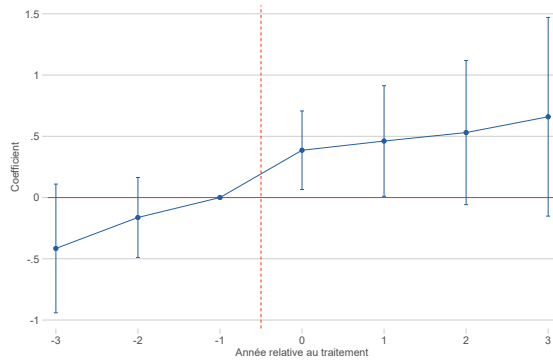


FIGURE A.2 – Évolution de l'asinh du nombre d'ingénieurs, variations autour de la limite d'effectif des entreprises

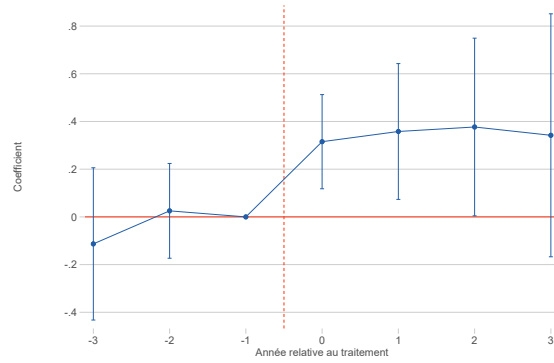


FIGURE A.3 – Évolution de l'asinh du nombre de techniciens, variations autour de la limite d'effectif des entreprises

(a) Max 20 salariés

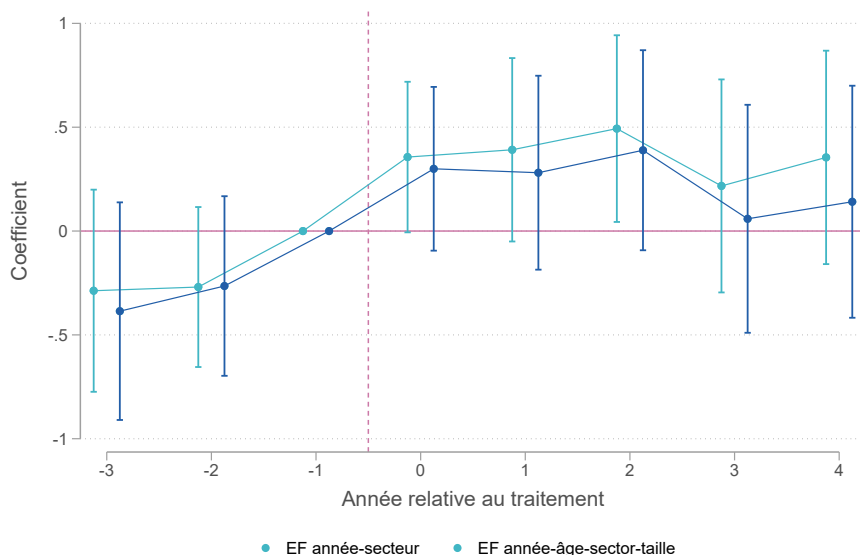


(b) Max 1000 salariés



A.3 Contrôles pour l'âge et la taille dans la différence-de-différences

FIGURE A.4 – Tests de robustesse : heures travaillées par des ingénieurs de R&D (asinh)

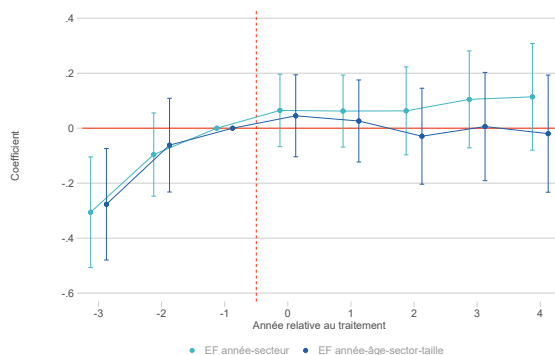


NOTES : Cette figure présente les coefficients de l'estimation d'une spécification d'étude d'évènements – voir équation (4.2). La sélection de l'échantillon d'estimation est détaillée dans la section 4.2.1.

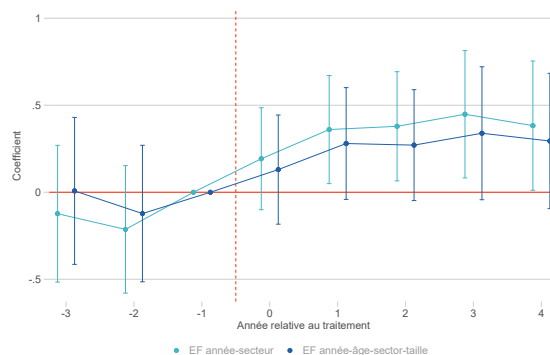
SOURCES : Bases Cifre, GECIR, FARE, DADS Postes.

FIGURE A.5 – Tests de robustesse : chiffre d'affaires et de la valeur ajoutée

(a) Chiffre d'affaires



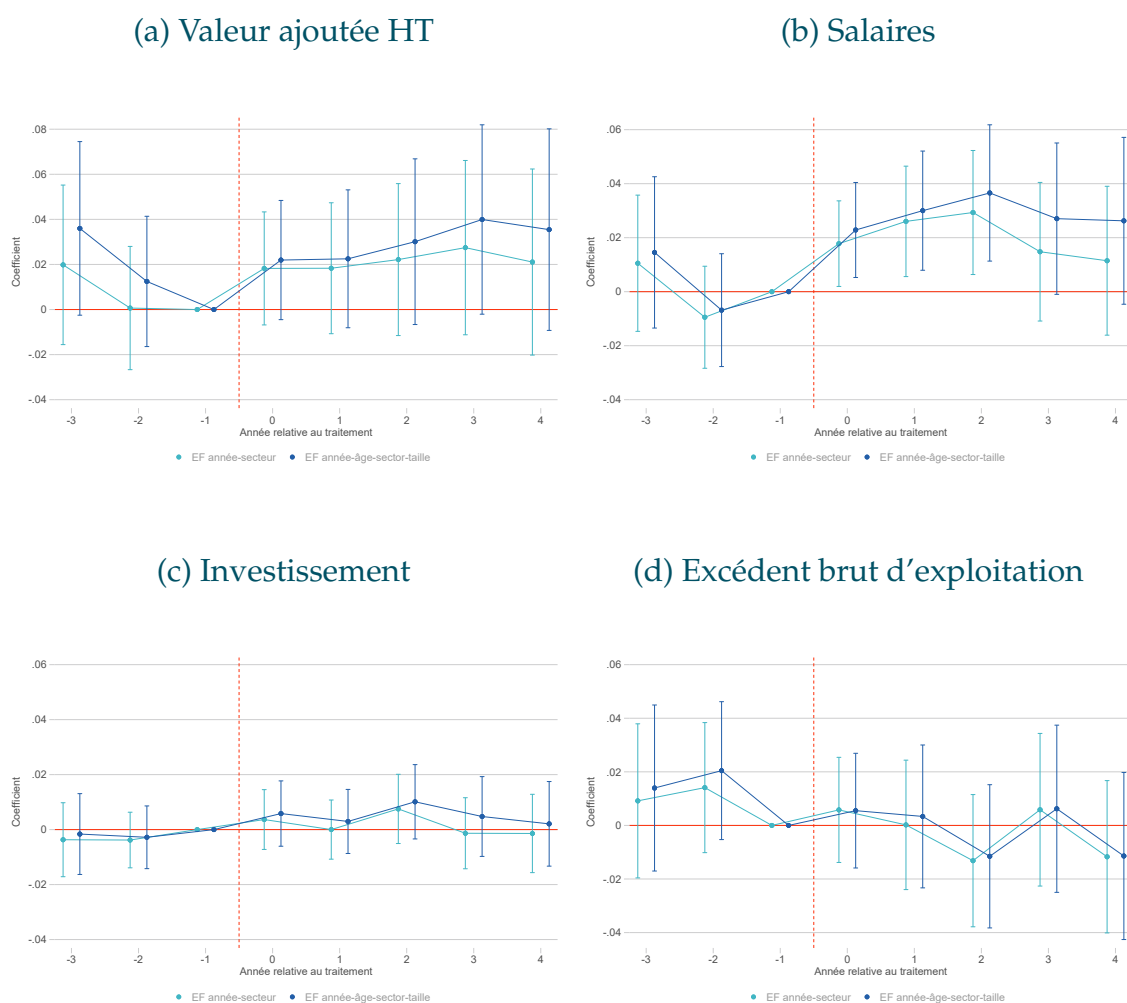
(b) Valeur ajoutée HT



NOTES : Cette figure présente les coefficients de l'estimation d'une spécification d'étude d'évènements – voir équation (4.2). La sélection de l'échantillon d'estimation est détaillée dans la section 4.2.1.

SOURCES : Bases Cifre, GECIR, FARE, DADS Postes.

FIGURE A.6 – Test de robustesse : variables comptables rapportées au total du bilan



NOTES :

Cette figure présente les coefficients de l'estimation d'une spécification d'étude d'évènements – voir équation (4.2). La sélection de l'échantillon d'estimation est détaillée dans la section 4.2.1.

SOURCES : Bases Cifre, GECIR, FARE, DADS Postes.

RÉFÉRENCES

- Abraham, S. et Sun, L. (2019). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Available at SSRN 3158747*.
- ANRT (2019). *Conditions générales d'octroi et d'éligibilité des conventions industrielles de formation par la recherche*.
- Beaufort, R., Beurenaut, A.-S., Calmand, J., Darid, H., Hammi, H., Langevin, E., Laussucq, G., Liogier, V., Marlat, D., Meuric, L., et al. (2018). L'État de l'emploi scientifique en France. Rapport 2018.
- Bertrand, M., Duflo, E., et Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, 119(1), p. 249–275.
- Borusyak, K. et Jaravel, X. (2017). Revisiting event study designs with an application to the estimation of the marginal propensity to consume. *mimeo*, p. 33.
- Dube, A., Lester, T. W., et Reich, M. (2010). Minimum wage effects across state borders : Estimates using contiguous counties. *The review of economics and statistics*, 92(4), p. 945–964.
- Gross, T., Notowidigdo, M. J., et Wang, J. (2019). The marginal propensity to consume over the business cycle. *American Economic Journal : Macroeconomics*.
- Kleven, H., Landais, C., et Søgaard, J. E. (2019). Children and gender inequality :

Evidence from denmark. *American Economic Journal : Applied Economics*, 11(4), p. 181–209.

Levy, R. (2005). Les doctorants cifre : médiateurs entre laboratoires de recherche universitaires et entreprises. *Revue d'économie industrielle*, 111(1), p. 79–96.

Margolis, D. et Miotti, L. (2015). Évaluation de l'impact du dispositif jeunes docteurs du crédit d'impôt recherche. *Rapport, Ministère de l'Éducation nationale, de l'enseignement supérieur et de la recherche*.

Oster, E. (2019). Unobservable selection and coefficient stability : Theory and evidence. *Journal of Business & Economic Statistics*, 37(2), p. 187–204.

Schmidheiny, K. et Siegloch, S. (2019). On event study designs and distributed-lag models : Equivalence, generalization and practical implications. *CEPR Discussion Paper No. DP13477*.

LISTE DES TABLEAUX

2.1	Tableau comparant les caractéristiques des entreprises Cifre à celles ayant seulement recours au CIR, et à celles employant seulement des ingénieurs, en 2008 et 2015	34
3.1	Régressions linéaires du taux de soutenance sur le statut Cifre	41
3.2	Régressions linéaires de la durée de thèse sur le statut Cifre	43
3.3	Statistiques descriptives	48
3.4	Effets du dispositif Cifre sur l’insertion professionnelle des docteurs .	50
3.5	Spécification avec variables de contrôle pleinement interagies	52
3.6	Analyse de sensibilité	55
3.7	Estimation parmi les docteurs aspirant à travailler dans le secteur privé au moment de la soutenance	56
3.8	Effet du dispositif sur la « séniorité »	58
3.9	Mesures de qualité de la thèse – parmi les docteurs aspirant à travailler dans le secteur privé	59
3.10	Mesures de qualité de la thèse (ensemble des docteurs de l’échantillon de base)	60
4.1	Statistiques descriptives pour l’année 2009 de l’échantillon d’estimation par méthode de différence-de-différences	70
4.2	Statistiques descriptives pour l’année 2009 de l’échantillon d’estimation par appariement et méthode de différence-de-différences . . .	81

LISTE DES FIGURES

2.1	Évolution du nombre de Cifre validées.	20
2.2	Évolution de la subvention ANRT.	21
2.3	Évolution du nombre d'entreprises ayant recours à une convention Cifre.	21
2.4	Évolution du nombre de Cifre validées par entreprise.	22
2.5	Distribution de l'âge des doctorants (conventions 2011-2018)	24
2.6	Nationalité des doctorants Cifre	24
2.7	Proportion de chaque genre parmi les doctorants	25
2.8	Disciplines des doctorants Cifre	25
2.9	Diplômes des doctorants Cifre	26
2.10	Catégorie juridique des employeurs	28
2.11	Taille des entreprises accueillant des doctorants Cifre	29
2.12	Secteur d'activité des employeurs.	30
2.13	Nombre de doctorants Cifre par ZE.	31
2.14	Nombre d'établissements ayant accueilli une Cifre par ZE.	31
2.15	Ratio du nombre de doctorants Cifre / entreprises avec emplois hautement qualifiées	32
2.16	Ratio du nombre de doctorants Cifre / entreprises employant des ingénieurs.	32

2.17	Évolution moyenne de l'emploi total et de l'emploi ingénieur des firmes Cifre, CIR seulement et employeuses d'ingénieurs seulement sur la période 2008-2015, normalisé à 1 en 2008.	35
2.18	Évolution moyenne du chiffre d'affaires et de la valeur ajoutée (HT) des firmes Cifre, CIR seulement et employeuses d'ingénieurs seulement sur la période 2008-2015, normalisé à 1 en 2008.	36
3.1	Taux de soutenance par discipline theses.fr (2011-2018)	42
4.1	Évolution (de l'asinh) des heures travaillées par des ingénieurs de R&D autour du recours à la première thèse Cifre	65
4.2	Évolution (de l'asinh) des heures travaillées par des ingénieurs et des techniciens autour du recours à la première thèse Cifre	66
4.3	Évolution (du log) du chiffre d'affaires et de la valeur ajoutée	67
4.4	Évolution (de l'asinh) des heures travaillées par des ingénieurs de R&D autour du recours à la première thèse Cifre	72
4.5	Évolution (de l'asinh) des heures travaillées par des ingénieurs et des techniciens autour du recours à la première thèse Cifre	73
4.6	Évolution (du log) du chiffre d'affaires et de la valeur ajoutée hors taxe	74
4.7	Évolution de variables comptables rapportées au total du bilan	75
4.8	Appariement et DD : chiffre d'affaires et de la valeur ajoutée	82
4.9	Appariement et DD : Emploi par occupation	83
4.10	Appariement et DD : variables comptables rapportées au total du bilan	84
A.1	Évolution de l'asinh du nombre d'ingénieurs de R&D, variations autour de la limite d'effectif des entreprises	90
A.2	Évolution de l'asinh du nombre d'ingénieurs, variations autour de la limite d'effectif des entreprises	90

A.3	Évolution de l'asinh du nombre de techniciens, variations autour de la limite d'effectif des entreprises	91
A.4	Tests de robustesse : heures travaillées par des ingénieurs de R&D (asinh)	92
A.5	Tests de robustesse : chiffre d'affaires et de la valeur ajoutée	92
A.6	Test de robustesse : variables comptables rapportées au total du bilan	93



L'Institut des politiques publiques (IPP) a été créé par l'École d'économie de Paris (PSE) et est développé dans le cadre d'un partenariat scientifique entre PSE et le Groupe des écoles nationales d'économie et statistique (GENES). L'IPP vise à promouvoir l'analyse et l'évaluation quantitatives des politiques publiques en s'appuyant sur les méthodes les plus récentes de la recherche en économie.

PSE a pour ambition de développer, au plus haut niveau international, la recherche en économie et la diffusion de ses résultats. Elle rassemble une communauté de près de 140 chercheurs et 200 doctorants, et offre des enseignements en Master, École d'été et Executive education à la pointe de la discipline économique. Fondée par le CNRS, l'EHESS, l'ENS, l'École des Ponts-ParisTech, l'INRA, et l'Université Paris 1 Panthéon Sorbonne, PSE associe à son projet des partenaires privés et institutionnels. Désormais solidement installée dans le paysage universitaire mondial, la fondation décloisonne ce qui doit l'être pour accomplir son ambition d'excellence : elle associe l'université et les grandes écoles, nourrit les échanges entre l'analyse économique et les autres sciences sociales, inscrit la recherche académique dans la société, et appuie les travaux de ses équipes sur de multiples partenariats. www.parisschoolofeconomics.eu



Le GENES est un établissement public d'enseignement supérieur et de recherche. Au sein du GENES, le CREST est un centre de recherche interdisciplinaire spécialisé en méthodes quantitatives appliquées aux sciences sociales regroupant des chercheurs de l'ENSAE Paris, de l'ENSAI, du département d'économie de l'École polytechnique et du CNRS. Centre interdisciplinaire spécialisé en méthodes quantitatives appliquées aux sciences sociales, le CREST est organisé en quatre thématiques : économie, statistiques, finance-assurance et sociologie. La culture commune des équipes est celle d'un attachement fort aux méthodes quantitatives, aux données, à la modélisation mathématiques, et d'allers-retours continus entre les modèles théoriques et les preuves empiriques permettant d'analyser des problématiques sociétales et économiques concrètes. <http://crest.science>

