



HAL
open science

Évaluation intermédiaire des aides "Programmes d'investissements d'avenir" de l'ADEME

Sophie Cottet, Fanny Henriet, Katrin Millock, Marion Monnet, Lucile Romanello

► To cite this version:

Sophie Cottet, Fanny Henriet, Katrin Millock, Marion Monnet, Lucile Romanello. Évaluation intermédiaire des aides "Programmes d'investissements d'avenir" de l'ADEME. [Rapport de recherche] Rapport IPP n°17, Institut des politiques publiques (IPP). 2017. halshs-02514705

HAL Id: halshs-02514705

<https://shs.hal.science/halshs-02514705v1>

Submitted on 22 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Institut des
Politiques Publiques

RAPPORT IPP n°17 – Septembre 2017

Evaluation intermédiaire des aides "Programmes d'investissements d'avenir" de l'ADEME

Sophie COTTET
Fanny HENRIET
Katrín MILLOCK
Marion MONNET
Lucile ROMANELLO





L'Institut des politiques publiques (IPP) est développé dans le cadre d'un partenariat scientifique entre PSE-Ecole d'économie de Paris (PSE) et le Centre de Recherche en Economie et Statistique (CREST). 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



RAPPORT IPP n°17 – Septembre 2017

Evaluation intermédiaire des aides "Programmes d'investissements d'avenir" de l'ADEME

Sophie COTTET
Fanny HENRIET
Katrin MILLOCK
Marion MONNET
Lucile ROMANELLO

LES AUTEURS DU RAPPORT

Sophie COTTET est économiste à l'Institut des politiques publiques. Diplômée de l'École nationale des statistiques et de l'administration économique (ENSAE Paris-Tech) où elle s'est spécialisée en méthodes statistiques pour l'économie, ses intérêts portent sur la fiscalité et sur les inégalités, et plus particulièrement sur l'évaluation des effets de politiques publiques.

Page personnelle : <http://www.ipp.eu/fr/annuaire/sophie-cottet/>

Fanny HENRIET est chargée de recherche au CNRS et chercheur associé à l'École d'économie de Paris. Diplômée de l'École Polytechnique (X04), elle a soutenu sa thèse sur l'économie du changement climatique, réalisée sous la direction de Roger Guesnerie, en 2012. Elle est spécialiste en économie de l'environnement et de l'énergie et en économie publique, et ses travaux de recherche actuels sont consacrés à la fiscalité environnementale et aux outils de la transition énergétique.

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

Katrin MILLOCK PhD, est chargée de recherche CNRS à l'École d'Économie de Paris et au Centre d'Économie de la Sorbonne (Université Paris 1 Panthéon-Sorbonne). Spécialiste en économie de l'environnement et des ressources naturelles, elle travaille sur les politiques de l'environnement et a effectué des expertises pour l'OCDE et le Ministère français de l'Environnement. Plus récemment, elle a piloté un projet financé par l'ANR sur la migration et l'environnement.

Page personnelle : <http://http://www.parisschoolofeconomics.eu/fr/millock-katrin/>

Marion MONNET est économiste à l'IPP depuis septembre 2014. Diplômée de

l'École d'économie de Paris, elle a travaillé sur les politiques de l'environnement au sein du Ministère de l'écologie et du développement durable. Spécialisée dans l'évaluation quantitative des politiques publiques, ses recherches portent à présent sur les politiques de l'éducation ainsi que sur les politiques d'aides aux entreprises.

Page personnelle : <http://www.ipp.eu/fr/annuaire/marion-monnet/>

Lucile ROMANELLO est économiste à l'IPP. Après avoir fait sa thèse à Sciences-Po en économie des politiques publiques, elle a intégré la Direction générale du Trésor où ses recherches ont porté sur le système socio-fiscal français et les politiques de redistribution. Elle continue de travailler sur ces problématiques au sein de l'IPP.

Page personnelle : <http://www.ipp.eu/fr/annuaire/lucile-romanello/>



REMERCIEMENTS

Cette étude s'inscrit dans le cadre d'un marché public lancé par l'Agence de l'environnement et de la maîtrise de l'énergie (ADEME) pour la réalisation d'une étude économétrique intermédiaire du régime d'aides « Investissements d'avenir » de l'ADEME (SA.40266).

Nous souhaitons remercier tout particulièrement les équipes de l'ADEME et du Commissariat général à l'investissement pour les interactions fructueuses que nous avons pu avoir tout au long de ce projet, en particulier Mmes Isabelle Vincent, Noura Rezgui, Marie-Laure Nauleau, Laurence Ould-Ferhat, Véronique Tetu, Sylviane Gastaldo, Caroline Acosta et MM. Ivan Faucheux, Benjamin Stremsdoerfer et Sébastien Delarue.

Nous tenons ensuite à remercier l'ensemble des membres du comité de pilotage de l'évaluation, qui ont apporté de nombreux éclairages à cette étude : tout particulièrement MM. Christophe Bellégo (Insee) et François Magnien (Direction générale des entreprises).

Enfin, nous remercions chaleureusement toute l'équipe du Centre d'accès sécurisé à distance (CASD) pour la mise à disposition des données ainsi que pour leur réactivité et leur efficacité, aussi bien pour le traitement des exports CASD que pour la résolution de problèmes techniques.

SYNTHÈSE DES RÉSULTATS

- L'intervention publique en faveur du soutien aux activités de R&D est justifiée par la nature particulière de la R&D. Génératrice d'externalités positives, elle peut désinciter les acteurs économiques à entreprendre des activités de R&D puisqu'ils ne peuvent s'appropriier l'intégralité des bénéfices attendus. Les asymétries d'information entre l'entreprise et le financeur contribuent en outre à augmenter le coût d'un financement extérieur, au risque de freiner des projets de R&D. Les financeurs, moins bien informés sur la qualité du projet et sur sa potentielle réussite, prêtent des fonds à des taux plus élevés. Enfin, le manque de coordination entre les différents acteurs peut également conduire à des situations de sur- ou de sous-investissement.
- Lancé en 2010, le Programme d'investissements d'avenir (PIA) a précisément pour ambition de pallier ces défaillances de marché, en soutenant les investissements des entreprises et en stimulant leur capacité d'innovation dans des domaines porteurs de croissance, tels que la santé, l'économie numérique, le développement durable ou encore la transition énergétique. Piloté par le Commissariat général à l'investissement (CGI), sa gestion a été déléguée à différents opérateurs d'État.
- L'Agence de l'environnement et de la maîtrise de l'énergie (ADEME) a été choisie comme opérateur des PIA pour son expertise sur les thématiques environnementales. Depuis 2010, elle a apporté un soutien à près de 270 projets sous forme de subventions ou d'avances remboursables. Ce sont 930 entre-

prises, associations, collectivités territoriales et laboratoires de recherche qui ont pu bénéficier de ce soutien.

- En tant que régime de grande ampleur, le PIA est soumis à une obligation d'évaluation par la Commission européenne. L'ADEME a soumis, en août 2015, un plan d'évaluation proposant notamment une méthode en vue d'une évaluation économétrique de son régime d'aide. Suite à ce plan et au regard des exigences communautaires, l'ADEME et le CGI ont pris la décision de conduire une étude intermédiaire, menée par des évaluateurs indépendants, visant à tester la faisabilité du scénario identifié dans ce plan.
- Les évaluations économétriques de dispositifs de soutien à la R&D sont nombreuses et ont principalement cherché à répondre à deux questions : celle de l'effet d'additionnalité de ces aides par rapport aux dépenses privées de R&D, et celle de leur effet sur l'innovation et sur les performances des entreprises. Si aucun consensus n'émerge quant à l'effet d'additionnalité des aides distribuées sous forme de subventions, les études s'accordent plutôt à dire que les aides prenant la forme d'un crédit d'impôt ne viendraient pas se substituer aux dépenses privées de R&D. L'impact des subventions sur les performances économiques des entreprises varie quant à lui largement en fonction du contexte et du pays étudié. En revanche, l'effet positif sur le nombre de brevets déposés semble avéré pour les entreprises participant à un consortium de recherche bénéficiant de subventions.
- Plusieurs sources de données ont été mobilisées pour la réalisation de cette évaluation intermédiaire. Les données de gestion du PIA-ADEME, transmises par l'ADEME, ont permis de constituer une base au niveau de l'entreprise, identifiée par son numéro Siret, regroupant à la fois les partenaires de projets financés et des partenaires de projets ayant candidaté mais n'ayant pas été retenus pour bénéficier d'aides. Cette base de données a été complétée par

- des informations comptables et financières à partir des liasses fiscales FICUS-FARE et des déclarations annuelles de données sociales (DADS), ainsi que par des informations sur les activités de R&D des entreprises avec l'Enquête R&D.
- Si l'Enquête R&D est la source d'information la plus riche en ce qui concerne les activités de R&D des entreprises, son caractère non exhaustif sur les petites entreprises conduit à perdre près de deux tiers des observations de la base ADEME lors de l'appariement, et introduit ainsi un biais de sélection dans l'échantillon. Les bases administratives FICUS-FARE et DADS sont ainsi privilégiées pour l'analyse, au prix cependant d'une perte de précision concernant les variables de R&D. En appariant ces données sur les années 2007, 2008, 2009 et 2013, nous obtenons une base contenant 706 observations présentes les quatre années.
 - Au vu des données à disposition et en s'inspirant des méthodes économétriques utilisées dans la littérature, nous évaluons l'effet du PIA-ADEME par la méthode des doubles différences. Nous comparons les trajectoires des entreprises ayant bénéficié d'une aide (groupe traité) à celles des entreprises ayant postulé mais n'ayant pas été retenues (groupe témoin), et ce avant et après la réception de l'aide. Sous l'hypothèse que l'évolution dans les deux groupes est comparable du point de vue des caractéristiques observables avant l'obtention de l'aide (hypothèse des tendances parallèles), les différences observées peuvent être directement interprétées comme l'effet causal du PIA-ADEME.
 - La période de pré-traitement, sur laquelle l'hypothèse des tendances parallèles est vérifiée, est la période 2007-2009. L'effet du traitement est quant à lui calculé comme la différence entre les années 2009, qui est la première année pour le dépôt de dossiers du PIA-ADEME, et 2013, qui est la dernière année pour laquelle nous disposons des données.
 - L'hypothèse des tendances parallèles n'étant pas validée sur l'échantillon consti-

tué par l'appariement des données ADEME, FARE et DADS, nous mettons en œuvre une procédure d'appariement afin d'améliorer la comparabilité entre les entreprises traitées et témoins. Une fois calculée la probabilité de bénéficier d'une aide pour chaque partenaire à l'aide de l'estimation d'un modèle *probit*, nous testons deux procédures d'appariement : l'appariement *one to one*, qui consiste à former une paire avec les partenaires traités et témoins ayant la probabilité la plus proche ; et l'appariement *kernel* qui attribue un poids moyen à chaque unité témoin en fonction de sa distance à chaque traité.

- Outre le fait que le pouvoir prédictif de notre estimation *probit* est trop faible pour affirmer que, conditionnellement aux observables, l'attribution d'une aide peut être considérée comme aléatoire, notre stratégie d'identification est d'autant plus affaiblie que l'hypothèse des tendances parallèles n'est pas validée, même à l'issue de la procédure d'appariement. Les évolutions des entreprises traitées et témoins, au cours de la période précédant la mise en place du traitement, demeurent sensiblement différentes. L'estimation d'un effet causal du PIA-ADEME ne peut être réalisée sans la validation de ces hypothèses identifiantes.
- Bien que l'évaluation causale du PIA-ADEME n'ait pu être menée dans le cadre de cette évaluation intermédiaire, ce travail exploratoire a permis d'identifier plusieurs écueils qui pourront être évités en 2020 et de formuler des recommandations en vue de l'évaluation finale. Un premier ensemble de recommandations propose des pistes pour remédier aux deux problèmes majeurs rencontrés au cours de ce travail : la faible taille de l'échantillon d'une part, et le manque de comparabilité du groupe témoin et du groupe contrôle, d'autre part. L'utilisation de la base de gestion du crédit d'impôt recherche (GECIR), non disponible pour l'évaluation intermédiaire, semble notamment constituer une piste prometteuse. Elle est en effet susceptible d'améliorer la compara-

bilité des entreprises traitées et témoins grâce aux variables de R&D qu'elle contient. Une fois ces deux problèmes traités, des méthodologies alternatives, telles qu'un changement de la période de traitement, pourraient être testées.

SOMMAIRE

Remerciements	1
Synthèse des résultats	7
Introduction	11
1 Éléments de cadrage sur le Programme d’investissements d’avenir	17
1.1 Le Programme d’investissements d’avenir (PIA)	18
1.2 Le PIA géré par l’ADEME	24
1.3 Projets sélectionnés et profil des bénéficiaires du PIA-ADEME	27
1.4 L’évaluation des effets du PIA-ADEME	30
2 Justifications économiques et évaluations empiriques des aides à la R&D	33
2.1 Les justifications économiques d’un soutien public à la R&D et à l’innovation	33
2.2 Quels sont les effets des aides publiques à la R&D?	43
3 Présentation des données et constitution des bases d’analyses	55
3.1 Description des données sur le PIA-ADEME	55
3.2 Description des bases de données externes	63
3.3 Appariement de la base ADEME et des données externes	68
4 Revue des méthodologies utilisées dans la littérature	75
4.1 Le cadre d’une évaluation idéale	75
4.2 Revue des méthodes des évaluations réalisées	81
5 Choix méthodologiques et résultats de l’étude	95
5.1 Choix méthodologiques	95
5.2 Résultats de l’analyse	103

6 Bilan et recommandations	123
6.1 Les données transmises par l'ADEME et le CGI	123
6.2 Les données externes	130
6.3 Méthodes économétriques	136
6.4 Synthèse des recommandations	140
Annexe A. Les données du Commissariat général à l'investissement	145
Annexe B. Les emplois R&D dans la nomenclature d'activités française	147
Annexe C. Caractéristiques des entreprises traitées et témoins de la base ADEME-FARE-DADS en 2009	150
Annexe D. Vérification statistique des tendances parallèles sur la base ADEME- RD-FARE	151
Annexe E. Vérification graphique des tendances parallèles sur la base ADEME- FARE-DADS	153
Annexe F. Résultats de l'estimation des effets des aides du PIA-ADEME	162
Références	169
Liste des tableaux	179
Liste des figures	181

INTRODUCTION

Lancé en 2010, le Programme d'investissements d'avenir (PIA) a pour ambition de stimuler les activités de recherche, de développement et d'innovation dans des secteurs ciblés et porteurs de croissance, aussi divers et variés que l'enseignement supérieur et la recherche, l'économie numérique, la santé ou encore le développement durable. Près de 50 milliards d'euros de crédits ont été engagés pour les deux premiers volets des PIA qui ont été reconduits pour la troisième fois en juin 2016 et dotés de 10 milliards d'euros supplémentaires. Le PIA est ainsi devenu un instrument majeur de la politique de soutien à l'innovation en France, aux côtés du crédit d'impôt recherche (CIR).

Comme pour toute politique d'aide publique en faveur des entreprises, se pose la question de l'incidence de ces aides sur les comportements des entreprises ainsi que sur leurs performances. En effet, afin de juger de la pertinence et de l'efficacité d'une telle politique de soutien aux entreprises, il est crucial de déterminer i) si ces aides viennent s'ajouter aux dépenses en R&D des entreprises tout en les incitant à investir davantage ou si au contraire elles viennent s'y substituer ; et ii) si ces aides permettent d'améliorer les performances des entreprises bénéficiaires et dans quelle mesure elles affectent les performances des entreprises non bénéficiaires.

En se concentrant sur le PIA géré par l'Agence de l'environnement et de la maîtrise de l'énergie (ADEME), spécialisée dans les thématiques environnementales comme celles de la transition énergétique ou de la croissance « verte », l'objectif de la présente étude est d'évaluer *ex post* les effets des aides octroyées par l'ADEME sur les entreprises bénéficiaires.

Contexte de l'étude

En tant que régime de grande ampleur de par les montants consacrés à ces aides et ses potentiels effets sur la concurrence, la Commission Européenne exige que soit produite une évaluation des effets des aides accordées dans le cadre du PIA. Chaque opérateur du PIA, à l'instar de l'ADEME, est ainsi tenu de mettre en œuvre une évaluation. Cela suppose qu'un plan d'évaluation, conforme aux méthodologies préconisées par la Commission européenne, soit soumis à la validation de la Commission. C'est ainsi qu'à la suite de la notification (en juillet 2015) et de la validation (le 4 août 2015) d'un plan d'évaluation du régime d'aides IA de l'ADEME (n°40266) soumis auprès de la Commission européenne par les autorités françaises et l'ADEME, ces dernières ont décidé le lancement d'une étude intermédiaire des effets du PIA-ADEME qui a pour vocation de préparer l'évaluation finale de 2020. Ce plan stipule également que l'évaluation doit être menée par un organisme indépendant. L'Institut des politiques publiques a ainsi été sélectionné pour mener cette étude à la suite d'une procédure d'appel d'offres lancée par l'ADEME.

Objectifs de l'étude

Ce projet a pour but de réaliser une première évaluation économétrique du régime d'aides de l'ADEME. Plus précisément, nous tentons de mettre en évidence un lien causal entre les subventions et les avances remboursables reçues par les entreprises bénéficiaires du PIA-ADEME et leurs performances à travers divers indicateurs : les dépenses en R&D, les dépôts de brevets, l'emploi, les performances économiques, etc. La méthodologie d'évaluation repose sur la méthode des doubles différences, en comparant les performances des entreprises bénéficiaires à celles des entreprises non bénéficiaires avant et après l'octroi de l'aide.

Pour réaliser cette étude, nous utilisons les bases de gestion de l'ADEME, afin d'identifier les entreprises bénéficiaires ou non d'une aide, que nous complétons par des bases de données administratives (DADS, liasses fiscales) et par des données d'enquête (Enquête R&D) qui contiennent des informations détaillées sur les

entreprises et leurs activités de R&D.

À l'issue de cette première évaluation du PIA-ADEME, l'objectif est de formuler tout un ensemble de recommandations qui permettront à l'ADEME de préparer au mieux l'évaluation finale de son régime d'aide prévue pour 2020. Ces recommandations portent aussi bien sur les bases de données à mobiliser, sur les indicateurs pertinents à retenir, que sur la méthodologie à mettre en œuvre pour établir un lien causal.

Plan du rapport

Chapitre 1 : Éléments de cadrage sur le Programme d'investissements d'avenir

Nous revenons dans ce chapitre sur la genèse du Programme d'investissements d'avenir en général. Nous présentons ensuite sa gouvernance, les différents acteurs du PIA, puis précisons le processus de sélection des bénéficiaires. Les spécificités du PIA-ADEME sont exposées ainsi que quelques statistiques descriptives sur le nombre de bénéficiaires et sur les financements des projets retenus.

Chapitre 2 : Justifications économiques et évaluations empiriques des aides à la R&D

Le second chapitre est consacré à une revue de la littérature sur les dispositifs de soutien à l'activité de R&D des entreprises et cherche à comprendre quelles sont les justifications économiques de ce choix de politique publique. Nous présentons la nature particulière des activités de R&D, qui génèrent des biens dits « publics », ainsi que les défaillances de marché qui y sont associées : asymétries d'information, barrières non financières, etc. Dans un second temps, sont présentés les résultats des études ayant cherché à évaluer l'efficacité de ces dispositifs, et notamment leur capacité à générer des dépenses de R&D supplémentaires et à améliorer les performances des entreprises bénéficiaires.

Chapitre 3 : Présentation des données et constitution des bases d'analyses

Plusieurs sources de données sont mobilisées pour la réalisation de cette étude : les données de gestion de l'ADEME ainsi que plusieurs sources de données externes telles que les déclarations annuelles de données sociales (DADS), l'Enquête R&D et les liasses fiscales (FICUS-FARE). Nous les présentons dans ce troisième chapitre puis détaillons le traitement qui en a été fait pour constituer nos deux bases d'analyse.

Chapitre 4 : Revue des méthodologies utilisées dans la littérature

Ce chapitre vise dans un premier temps à décrire les conditions idéales dans lesquelles une évaluation causale des aides de l'ADEME devrait être menée, puis énonce les différentes contraintes auxquelles nous faisons face pour cette évaluation, certaines propres au contexte de notre étude, d'autres communes aux évaluations des aides à la R&D. Nous passons ensuite en revue les différentes méthodes économétriques mises en œuvre dans la littérature et insistons sur les avantages et les limites de chacune de ces méthodes. Nous jugeons ensuite de leur applicabilité à l'évaluation du PIA-ADEME.

Chapitre 5 : Choix méthodologiques et résultats de l'étude

En s'appuyant sur les enseignements tirés de la revue des méthodologies et étant données les contraintes spécifiques à notre étude, nous exposons dans ce chapitre les choix méthodologiques qui ont été faits : la temporalité de l'étude retenue, la définition des groupes traité et témoin et la méthode économétrique. La méthodologie est ensuite appliquée aux deux bases d'analyse constituées et les résultats sont reportés, avec toutes les précautions d'interprétation des estimations nécessaires.

Chapitre 6 : Bilan et recommandations

Ce chapitre vient clore le rapport en proposant un bilan du travail réalisé et des problèmes rencontrés avec les données de gestion de l'ADEME, avec les données externes et avec la mise en œuvre de la méthodologie. Ce constat nous sert ensuite

de base pour formuler des recommandations pour que certains écueils puissent être évités lors de l'évaluation finale.

CHAPITRE 1

ÉLÉMENTS DE CADRAGE SUR LE PROGRAMME D'INVESTISSEMENTS D'AVENIR

La France occupe une position particulière vis-à-vis de ses homologues européens : si elle est l'un des pays soutenant le plus la R&D, notamment en raison des montants importants accordés au dispositif de crédit d'impôt recherche (CIR), elle reste cependant dans le groupe des « pays suiveurs » en matière d'innovation, et non parmi les pays *leaders* que sont l'Allemagne et les pays d'Europe du Nord au sein desquels le niveau de dépenses privées en R&D, qui oscille entre 2,4 % et 2,9 % du PIB en 2015, est plus élevé qu'en France (2,2 % en 2015¹). Ainsi, les performances françaises ne reflètent pas cet écart de soutien public à la recherche (OCDE, 2014). Le Programme d'investissements d'avenir s'inscrit à la fois dans cette spécificité de soutien public à la R&D et dans la volonté d'améliorer l'efficacité des activités de recherche et d'innovation en France. L'objectif de ce chapitre est d'apporter quelques éléments de compréhension relatifs au PIA (partie 1.1), et plus spécifiquement au PIA-ADEME, objet de cette étude (partie 1.2). Enfin, ce chapitre présente, de façon concise, quelques premiers éléments de bilan et d'évaluation de Programme d'investissements d'avenir.

1. OCDE (2017), Dépenses intérieures brutes de R-D (indicateur).

1.1 Le Programme d'investissements d'avenir (PIA)

1.1.1 Objectifs et mise en place du PIA

Aux lendemains de la crise économique, une commission bipartisane présidée par MM. Alain Juppé et Michel Rocard a été mise en place afin de formuler des préconisations et de définir des priorités stratégiques. Ces priorités sont recensées dans le rapport « Investir pour l'avenir » rendu par la commission (Juppé, 2010). Du fait de la pression qui s'exerçait alors sur les finances publiques, la tentation était forte de limiter l'investissement public et de diminuer les budgets consacrés à ce dernier. C'est pourquoi les auteurs du rapport proposent de mettre en œuvre un plan d'investissements étatique dont l'objectif est de soutenir la R&D afin de favoriser l'innovation, d'améliorer la compétitivité des entreprises françaises et de préserver la croissance potentielle. Selon la commission, il s'agit du rôle et de la responsabilité de l'État de faire des choix d'investissement qui permettent d'assurer la croissance de long terme. Ces investissements doivent prendre en compte l'ensemble du processus d'innovation, de la recherche à la mise sur le marché de produits innovants. Ils doivent également contribuer à développer et à structurer les relations entre les différents acteurs du monde de l'industrie et celui de la recherche publique, et ont pour objectif d'engendrer un effet de levier en stimulant les fonds privés consacrés à la R&D (Levet et Mathieu, 2013).

La commission a identifié six axes stratégiques prioritaires : l'enseignement supérieur, la recherche et la formation ; la valorisation de la recherche et le transfert au monde économique ; les filières industrielles (développement des petites et moyennes entreprises et des établissements de taille intermédiaire) ; le développement durable ; l'économie numérique et enfin, la santé et les biotechnologies (Commissariat général à l'investissement, 2015). Deux priorités transversales sont également retenues : l'économie de la connaissance ainsi que l'économie verte (Levet et Mathieu, 2013).

Le Programme d'investissements d'avenir (PIA), dont la création est officialisée

par la loi de finances rectificative du 9 mars 2010, se fondent sur ces priorités, déclinées en 35 actions, et couvrant les différents champs de la recherche et de l'innovation. Ce programme, mis en place sur la période 2010-2020, ont vocation à financer des projets portés aussi bien par des universités ou des laboratoires de recherche que par des entreprises ou encore des associations. Cette initiative s'inscrit directement dans le cadre de la stratégie « Europe 2020 », stratégie de croissance de l'Union européenne sur dix ans (OCDE, 2014).

Initialement, la loi de finances rectificative du 9 mars 2010 prévoit l'engagement de 35 milliards d'euros de crédits supplémentaires afin de financer le PIA. La loi de finances initiale pour 2014 du 29 décembre 2013 prévoit l'ouverture de 12 milliards d'euros supplémentaires afin de poursuivre l'effort d'investissement avec la mise en place d'un second volet du PIA (Commissariat général à l'investissement, 2015). Il est prévu que 45 % de ces crédits soient consacrés à l'enseignement supérieur et la recherche, considérés comme une priorité absolue (Cour des comptes, 2015).

1.1.2 Gouvernance, acteurs et fonctionnement du PIA

Organisation du PIA et répartition des crédits

Le Commissariat général à l'investissement (CGI) a été créé afin de piloter le PIA. Suivant une logique interministérielle, le CGI a été placé sous l'autorité du Premier ministre. La mise en œuvre opérationnelle a, quant à elle, été déléguée à des opérateurs de l'État ainsi qu'à des ministères. Le recours à des opérateurs déjà existants a été préféré afin de limiter les coûts de fonctionnement mais aussi d'inscrire le PIA dans la durée (France Stratégie, 2016). En effet, le PIA a été pensé de façon à pouvoir se soustraire au principe de l'annualité budgétaire puisque ce dernier entraine directement en contradiction avec la mise en place d'un processus de sélection exigeant. Les Investissements d'avenir reposaient sur une réflexion de long terme et les opérateurs, à qui les 47 milliards de crédits à investir ont été transférés, devaient pouvoir gérer les montants attribués sans contrainte annuelle de dépense.

Les crédits accordés peuvent l'être sous forme de dotations consommables ou

non consommables. Les premières sont versées aux opérateurs qui les placent sur un compte au Trésor. Ils peuvent ensuite en disposer dans leur intégralité pour financer les projets retenus à l'issue d'une phase de sélection. En revanche, pour les dotations non consommables, seuls les intérêts générés par le placement de ces crédits peuvent être utilisés afin de financer des projets. Ce mode de financement a été initialement inspiré des universités américaines, chargées de gérer le capital initial attribué, et concerne, dans le cadre du PIA, principalement les domaines de la recherche et de l'enseignement supérieur. Initialement, les opérateurs, et plus spécifiquement l'Agence nationale de la recherche (ANR), devaient pouvoir disposer des intérêts du capital sur une durée illimitée. Dans un second temps, cette durée a été cependant réduite à 10 ans. Ainsi, les crédits réellement disponibles pour le financement des projets sont la somme des dotations consommables et des intérêts issus des dotations non consommables, soit 28,7 milliards d'euros au total² (Cour des comptes, 2015).

Le tableau 1.1 présente les différents opérateurs et les crédits qui leur ont été accordés pour les PIA 1 et 2. Les opérateurs recevant les plus gros budgets sont la Caisse des dépôts et des consignations (CDC) (7,4 milliards d'euros), suivie de l'ADEME (3,8 milliards d'euros). L'ANR est un cas à part, puisque sur les 22,7 milliards d'euros de crédits alloués, 18 milliards sont des dotations non consommables.

Sélection des bénéficiaires et types de soutiens accordés

Les différents opérateurs du PIA sont chargés par le CGI de la sélection des projets candidats, de la contractualisation avec les partenaires ainsi que du versement des crédits consacrés au financement des projets. Les étapes sont nombreuses avant qu'un projet puisse se concrétiser. Dans la majorité des cas, les potentiels bénéficiaires déposent un dossier de candidature en réponse aux appels à projets lancés par les opérateurs³.

Commence ensuite le processus d'examen et de sélection des candidats. Des ju-

2. Les dotations non consommables représentent 15 milliards d'euros pour le premier volet du PIA et 3,3 milliards d'euros pour le PIA 2.

3. Certaines actions ayant un objet bien précis ne relèvent toutefois pas d'une procédure d'appel à projets.

TABLEAU 1.1 – Opérateurs du PIA et crédits prévus initialement (en millions d'euros)

Opérateurs	Acronymes	Crédits attribués PIA 1 (2010-2014)	Crédits attribués PIA 2 (2014-2016)	Total
Agence de l'environnement et de la maîtrise de l'énergie	ADEME	2 850	1 000	3 850
Agence nationale de l'habitat	Anah	500	-	500
Agence nationale pour la gestion des déchets radioactifs	Andra	100	-	100
Agence nationale de la recherche	ANR	18 849	3 875	22 724
Agence nationale pour la rénovation urbaine	ANRU	500	553	1 053
OSEO	-	2 440	-	2 440
Caisse des dépôts et consignations	CDC	6 500	936	7 436
Commissariat à l'énergie atomique et aux énergies alternatives	CEA	900	1 746	2 646
Centre national d'études spatiales	CNES	500	354	854
Office national d'études et de recherches aérospatiales	ONERA	1 500	945	2 445
Banque publique d'investissement	Bpifrance	-	2 671	2 671
France-Agrimer	-	-	120	120
Total		34 639	12 500	47 139

LECTURE : Dans le cadre du PIA 1, les crédits initialement prévus pour l'ADEME s'élèvent à 2 850 millions d'euros.

NOTE : Les montants reportés sont les montants présents dans les conventions initiales signées entre l'État et les opérateurs. Les crédits réellement versés sont légèrement différents, à la suite de certaines réallocations entre actions. « - » est reporté lorsqu'aucun montant n'a été attribué, lorsqu'il n'y a pas d'acronyme, ou lorsque l'entité n'existe plus (fusion de OSEO avec BPI France en 2013). Les crédits alloués à l'ANR sont des dotations non-consommables.

SOURCE : Cour des comptes (2015)

rys ou des comités d'experts indépendants analysent les dossiers puis pré-sélectionnent un ensemble de projets qu'ils soumettent ensuite à un comité de pilotage, composé également de représentants des ministères impliqués. Ce comité de pilotage peut proposer une sélection différente de celle du jury ou du comité indépendant. La proposition finale de sélection des projets revient toutefois au CGI qui la soumet ensuite au Premier ministre, chargé de la décision définitive.

Les critères de sélection sont spécifiques à chaque appel à projets. L'accent est toutefois largement mis sur l'excellence scientifique, le caractère innovant et sur la capacité du projet à générer un surplus de dépenses privées en R&D. Par ailleurs, dans le cadre du PIA 2, les financements des projets sont nécessairement éco-conditionnés, ce qui signifie que les impacts environnementaux des projets sont systématiquement analysés lors de la phase de sélection des projets. Enfin, les cofinancements privés sont quasi systématiquement une condition pour l'octroi d'une aide, l'idée étant de limiter la substitution de fonds publics à des fonds privés et de limiter un potentiel effet d'éviction. Par ailleurs, les aides sont plafonnées par les

textes communautaires et des financements extérieurs sont ainsi nécessaires pour que le projet puisse voir le jour. La part de cofinancements requis varie cependant d'un opérateur à l'autre.

Les projets sélectionnés font ensuite l'objet d'une contractualisation entre l'opérateur et les différents partenaires du projet. La contractualisation n'a lieu qu'une fois les cofinancements réunis et les modalités de remboursement définies. Les projets pour lesquels l'aide demandée dépasse un certain seuil, fixé par les textes communautaires, doivent être notifiés à la Commission européenne⁴. La Commission européenne s'assure alors que les aides envisagées ne créent pas de distorsion de concurrence sur les marchés. Cette notification peut retarder de plusieurs mois la mise en œuvre du projet.

Enfin, lorsqu'un projet est contractualisé, les partenaires du projet reçoivent les crédits soit immédiatement après la contractualisation, soit au fur et à mesure de la réalisation des étapes du projet : c'est la phase de décaissement. Le décaissement des montants engagés tout au long de la réalisation du projet permet de garder un certain contrôle sur la qualité des projets financés. Ce décaissement progressif offre la possibilité d'arrêter les financements si les partenaires du projet ne tiennent pas leurs engagements.

Le soutien de l'État peut prendre différentes formes : subventions, avances remboursables, prêts ou prises de participation. Cette diversité des modalités de financement permet de s'adapter aux différents projets ainsi qu'aux caractéristiques des bénéficiaires. Les subventions sont des aides monétaires directement accordées à un projet sans aucune contrepartie. Les avances remboursables consistent également en une aide financière qui doit cependant être remboursée en cas de succès technique ou commercial du projet, généralement évalué à partir du chiffre d'affaires de l'entreprise bénéficiaire de l'aide mais aussi sur la base d'indicateurs définis par l'opérateur. Le montant des remboursements est défini en fonction du projet et peut

4. Les seuils dépendent du domaine dans lequel l'aide est attribuée (R&D&I ou protection de l'environnement) et de l'activité de recherche (recherche fondamentale, recherche industrielle ou développement expérimental). Dans l'encadrement des aides d'État à la R&D&I (texte communautaire 2014/C 198/01) ce seuil est fixé à 40 millions d'euros pour les activités de recherche fondamentale, à 20 millions d'euros pour les activités de recherche industrielle et à 15 millions d'euros pour le développement expérimental. Ce seuil est également fixé à 15 millions d'euros pour les aides à la protection de l'environnement (texte communautaire 2014/C 200/01).

être supérieur aux montants de l'avance remboursable dans le cas d'une réussite allant au-delà du succès défini *ex ante*. Les prêts sont une remise de fonds accordée à une entreprise contre le paiement d'intérêts et avec l'obligation de rembourser la somme prêtée. Enfin, la prise de participation consiste, pour l'opérateur, à entrer dans le capital de l'entreprise par un apport de fonds. Dans ce cas, l'État se comporte comme un investisseur avisé dont l'objectif est alors d'obtenir un retour sur investissement.

1.1.3 Les limites du PIA

Bien que le PIA constitue une démarche et un effort exceptionnels par l'ampleur des montants investis et par l'originalité de ses modalités de fonctionnement, il convient cependant d'en souligner les limites, formulées notamment par la Cour des comptes dans son rapport d'évaluation du PIA (Cour des comptes, 2015) ou encore par France Stratégie (2016).

Tout d'abord, comme nous l'avons précisé plus haut, les crédits accordés dans le cadre du PIA le sont soit en dotations consommables, soit en dotations non consommables pour lesquelles seuls les intérêts générés peuvent être investis. Cette distinction est importante puisque par ce mécanisme des dotations non consommables et surtout par la limitation à une durée de 10 ans seulement du bénéfice des intérêts du capital initialement attribué, les montants qui ont réellement été consacrés au financement de projets de recherche sont inférieurs à ceux initialement prévus. Ces dotations non consommables représentent 44 % des montants du PIA 1 et 27 % des montants du PIA 2 (Cour des comptes, 2015) et concernent principalement l'Agence nationale de la recherche.

Dans un second temps, le principe d'additionnalité du PIA n'a pas toujours été respecté, notamment du fait d'un contexte budgétaire très largement contraint. Selon la logique initiale du PIA, les montants investis devaient venir s'ajouter aux dotations déjà existantes et non se substituer au budget de l'État. Or, les crédits du PIA ont pu servir à financer des projets d'investissement décidés avant même le lancement du PIA ou à recapitaliser certains opérateurs. Ainsi, dans le cas de l'ADEME, la Cour des comptes note que l'action « Démonstrateurs des énergies renouvelables

et décarbonnées » s'est substituée à des aides pré-existantes accordées par l'ADEME, dans le cadre de programmes mis en place juste avant le lancement du PIA. L'effet du PIA a dès lors davantage été de nature qualitative que quantitative : comme le souligne la Cour des comptes, ils ont avant tout conduit à une meilleure gestion des montants investis ainsi qu'à une rationalisation de la logique d'investissement de l'État. Le rapport de France Stratégie (2016) souligne que les crédits, notamment en ce qui concerne le PIA 2, ont parfois été utilisés pour financer des projets dont la finalité et les objectifs étaient assez éloignés de ceux initialement énoncés.

Enfin, la mise en oeuvre du PIA a été très progressive et les procédures administratives souvent lourdes. Ainsi, la seule phase d'instruction prenait en moyenne neuf mois au premier trimestre 2013 pour les projets déposés auprès de l'ADEME (Cour des comptes, 2015). Bien que les délais nécessaires à la phase d'instruction aient été divisés par trois depuis, des délais supplémentaires viennent s'ajouter pendant les phases de contractualisation avec les entreprises partenaires des projets, retardant par conséquent le lancement de la recherche et de l'innovation au sein des entreprises bénéficiaires. Les retards dans les phases de contractualisation et de décaissement peuvent également s'expliquer par la faible qualité des projets déposés. En effet, certains programmes ne sont pas parvenus à faire émerger suffisamment de projets respectant le critère d'excellence scientifique caractérisant le PIA.

1.2 Le PIA géré par l'ADEME

Nous présentons dans cette partie les thématiques couvertes par le PIA géré par l'ADEME, ainsi que ses modalités d'attribution et de remboursements.

1.2.1 Les thématiques couvertes par le PIA-ADEME

L'ADEME a été choisie, afin d'être l'un des opérateurs du PIA, pour son expertise concernant les problématiques environnementales, et plus particulièrement le domaine de l'énergie et des déchets. Les aides d'État octroyées dans le cadre du PIA-ADEME relèvent du régime d'aides de l'ADEME exempté de notification n°SA.40266, relatif aux aides à la R&D&I et à la protection de l'environnement

prises en œuvre dans le cadre des « Investissements d'Avenir ». Ce régime d'aides est entré en vigueur le 1^{er} janvier 2015 et s'applique jusqu'au 31 décembre 2020.

Lors du lancement du PIA 1, des projets de R&D appartenant à quatre thématiques liées à l'environnement peuvent bénéficier d'aides accordées par l'ADEME au travers de différentes actions : « Démonstrateurs et plateformes technologiques en énergies renouvelables et décarbonées et chimie verte » (1,125 millions d'euros) ; « Réseaux électriques intelligents » (165 millions d'euros) ; « Économie circulaire » (210 millions d'euros) et « Véhicule du futur » (950 millions d'euros). En 2014, lors du lancement du PIA 2, les différentes thématiques sont regroupées en deux programmes. Les trois premières sont regroupées sous la thématique « Démonstrateur de la transition écologique et énergétique », dotée de plus de 2 milliards d'euros. Le second programme « Véhicule du futur » est doté, quant à lui, d'un milliard d'euros et reprend les thématiques du précédent programme.

Depuis 2015, l'ADEME a par ailleurs lancé des appels à projets réservés aux petites et moyennes entreprises (PME) pour un montant de 150 millions d'euros, appelés « Initiative PME » (IPME), afin de mieux soutenir les PME qui font face à des contraintes de financement plus fortes que les grandes entreprises. L'idée derrière ces appels à projets spécifiques est également de permettre aux PME de développer leur propre projet et de leur éviter d'être « noyées » au sein des consortiums de recherche plus gros qui impliquent de grandes entreprises. Les thématiques restent les mêmes que celles mentionnées ci-dessus.

1.2.2 Sélection des projets et modalités d'octroi de l'aide

Le processus de sélection pour les projets déposés auprès de l'ADEME est similaire à celui décrit pour l'ensemble des PIA (partie 1.1.2). L'ADEME étant experte sur les thématiques environnementales, l'étude des projets déposés s'appuie principalement sur une expertise interne au sein de l'ADEME, et non sur des jurys indépendants, bien que l'ADEME puisse également parfois avoir recours à une expertise externe.

Les modalités d'attribution et de l'éventuel remboursement des aides d'État varient selon s'il s'agit du programme « Démonstrateurs de la transition écologique et

énergétique » ou du programme « Véhicule du futur » (tableaux 1.2 et 1.3). Elles diffèrent également selon que le projet retenu relève de la recherche industrielle (RI), définie comme l'acquisition de nouvelles connaissances ayant un but ou un objectif pratique déterminé, ou du développement expérimental (DE), qui consiste en des travaux fondés sur des connaissances obtenues par la recherche ou l'expérience, et qui visent à concevoir de nouveaux matériaux, produits ou procédés (OCDE, 2002).

TABLEAU 1.2 – Modalités d'attribution et de remboursement - Programme « Véhicule du futur »

	Subventions	Avances remboursables
Modalité d'attribution	Applicable pour les dépenses de recherche industrielle (RI)	Applicable pour les dépenses de développement expérimental (DE)
Taux d'aide	Calculé sur la base des taux maximum en RI de la CE	Calculé sur la base des taux maximum en DE de la CE
Remboursement	Pas de remboursement	100 % de l'aide remboursée au premier euro de chiffre d'affaires 30 % supplémentaires en cas de succès commercial

LECTURE : Dans le cas de l'octroi d'une subvention, il n'y a pas de remboursement.

NOTE : CE signifie ici Commission européenne. Les taux maximum servant au calcul du taux d'aide se trouvent en page 29 du texte de la Commission européenne (2014b). Ces règles s'appliquent depuis mi-2014.

SOURCE : ADEME - présentation des IA, mars 2016

TABLEAU 1.3 – Modalités d'attribution et de remboursement - Programme « Démonstrateurs de la transition écologique et énergétique »

	Subventions	Avances remboursables
Modalité d'attribution	Le taux est fixé en amont par appel à projets en fonction de la typologie du partenaire et du montant de l'aide	
Taux d'aide	Calculé sur la base des taux maximum pour les dépenses de R&D	
Remboursement	Pas de remboursement	50 % de l'aide remboursée à la fin du projet 50 % de l'aide remboursée en cas de succès commercial

LECTURE : Dans le cas de l'octroi d'une subvention, il n'y a pas de remboursement.

NOTE : Les modalités d'attribution sont les mêmes pour les subventions et les avances remboursables. Les taux maximum servant au calcul du taux d'aide se trouvent en page 29 du texte de la Commission européenne (2014b). Ces règles s'appliquent depuis mi-2014.

SOURCE : ADEME - présentation des IA, mars 2016

Par ailleurs, en ce qui concerne les appels à projets IPME, l'aide est accordée uniquement sous la forme d'une subvention dont le montant maximum est fixé à 200 000 euros et dont 70 % peuvent être versés directement à la notifica-

tion de l'aide. Les procédures ont été adaptées afin de tenir compte au mieux des contraintes spécifiques des PME. Les délais de la phase d'instruction des dossiers et de la décision de financement ont notamment été réduits à environ six semaines afin d'informer le porteur de projet le plus rapidement possible.

1.3 Projets sélectionnés et profil des bénéficiaires du PIA-ADEME

Entre son lancement en 2010 et le premier semestre de l'année 2016, ce sont 266 projets qui ont été aidés par le PIA-ADEME représentant 929 partenaires^{5 6}. En regardant la répartition du nombre de partenaires par type d'action et par type d'entreprises (tableau 1.4), on constate que ce sont les petites et moyennes entreprises (PME) et les très petites entreprises (TPE-TPI) qui concentrent le plus grand nombre de partenaires avec 368 partenaires. Elles représentent près de 40 % des partenaires des volets « Démonstrateur de la transition écologique » et « Véhicule du futur ». Parmi ces 368 partenaires, 52 ont reçu une aide dans le cadre de l'« Initiative PME », réservée uniquement aux PME. Les grandes entreprises sont également bien représentées avec 305 partenaires, dont plus de la moitié ont un projet « Véhicule du futur ». Enfin, les établissements publics d'État sont également d'importants bénéficiaires des aides de l'ADEME avec 214 projets, soit un cinquième des projets financés.

En ce qui concerne le financement des projets sélectionnés, la forme d'aide la plus courante est le financement mixte – qui combine une part de subvention et une part d'avances remboursables – qui concerne 178 projets (67 % des projets), suivi par le financement par subventions uniquement, qui concerne 85 projets. Les projets aidés uniquement par avances remboursables font plutôt figure d'exception par leur faible nombre et par le montant moyen accordé (14 millions d'euros) qui se

5. Ces statistiques sont calculées sur les projets qui entrent dans le champ de notre étude. Ne sont pas présents les projets financés autrement que par des subventions ou des avances remboursables ainsi que les projets de l'AMI « Déploiement des infrastructures territoriales ». Les projets de l'Initiative PME sont en revanche comptés dans ces chiffres.

6. Ce chiffre ne prend pas en compte le fait qu'un partenaire peut participer à plusieurs projets. Il s'agit donc en réalité de projet*partenaires.

TABLEAU 1.4 – Profil des partenaires des projets bénéficiaires

Volet d'action	Grandes Entreprises et ETI	PME et TPE-TPI	Étab. publics	Collectivités territoriales	Total
Démonstrateur de la transition écologique	165	227	140	5	561
Véhicule du futur	140	141	74	6	368
Total	305	368	214	11	929

LECTURE : Sur les 929 partenaires de projets financés par le PIA-ADEME, 165 sont des grandes entreprises partenaires de projets du volet « Démonstrateur de la transition écologique ».

NOTE : ETI : Entreprises de taille intermédiaire ; PME : Petites et moyennes entreprises ; TPE-TPI : Très petites entreprises. Le nombre de partenaires des colonnes ne somme pas à 929 en raison des 31 partenaires classifiés dans la catégorie « Autres » (partenaires étrangers, associations, partenaires fictifs).

SOURCE : Données SISE au premier trimestre 2016.

trouve largement au-dessus de la moyenne sur l'ensemble des projets. De manière générale, les montants accordés au titre de subventions sont inférieurs aux montants autorisés pour les avances remboursables : les projets financés uniquement par subvention reçoivent en moyenne près d'un millions d'euros et ceux bénéficiant d'un financement mixte ont une subvention moyenne de 2,1 millions d'euros, contre 3,5 millions d'euros pour la partie financée en avances remboursables. Le volet d'action concentrant les montants moyens d'aide les plus élevés est le volet « Démonstrateurs de la transition » avec 4,7 millions d'euros contre 3,3 millions d'euros pour les projets du volet « Véhicule du futur ».

TABLEAU 1.5 – Répartition des types de financement des projets PIA-ADEME par programme (en millions d'euros)

Action	Subvention		AR		Subvention et AR			Total	
	Nb	Moy.	Nb	Moy.	Nb	Moy. subv	Moy. AR	Nb	Moy.
Démonstrateur de la transition écologique	17	1,8	3	14,1	122	1,9	3,2	149	4,7
Véhicule du futur	68	0,7	0	0	54	2,6	4,2	124	3,3
Total	85	0,9	3	14,1	178	2,1	3,5	266	4,1

LECTURE : 85 projets reçoivent des aides du PIA-ADEME sous forme de subventions uniquement dont le montant moyen est de 0,9 millions d'euros.

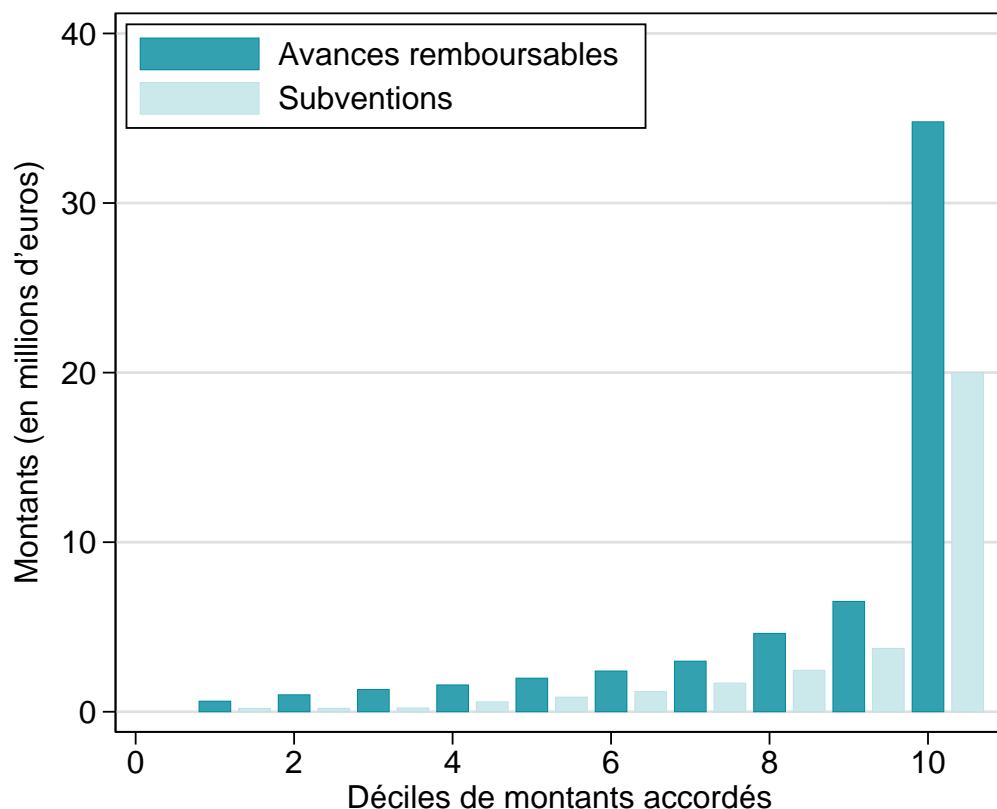
NOTE : Nb : « Nombre de projets » ; Moy. : « Moyenne du montant accordé ».

SOURCE : Données SISE au premier trimestre 2016.

Si l'on regarde la distribution des montants autorisés entre subventions et avances

remboursables en fonction du décile de montant autorisé total dans lequel se situent les projets, on constate une distribution très hétérogène des subventions et des avances remboursables octroyées (graphique 1.1). Pour les aides en subventions comme pour les aides en avances remboursables, les montants autorisés augmentent assez peu et de façon relativement linéaire dans les sept premiers déciles, puis augmentent un peu plus à partir du huitième décile avant de se démarquer nettement dans le dixième décile, dans lequel les montants moyens sont de 20 millions pour les subventions et de 35 millions pour les avances remboursables. Ce constat suggère que les montants accordés totaux sont concentrés en réalité sur une minorité de projets.

FIGURE 1.1 – Distributions des subventions et des avances remboursables par déciles de montants autorisés



LECTURE : Les 10 % des projets les mieux financés reçoivent en moyenne 20 millions d'euros de subventions et 35 millions d'euros d'avances remboursables.

SOURCE : Données SISE au premier trimestre 2016.

Enfin, une dernière caractéristique du financement des projets du PIA-ADEME à analyser est celle du cofinancement. En effet, un des objectifs du PIA est de constituer un effet de levier en incitant d'autres acteurs à participer au financement des projets. Parmi les 266 projets financés, 246 bénéficient d'un cofinancement, aussi bien par des acteurs publics que privés. Les montants de cofinancements moyens vont de 10,6 millions d'euros pour le volet « Démonstrateurs de la transition écologique », à 15,5 millions d'euros pour les projets de l'action « Véhicule du futur ». Dans tous les cas, la part des cofinancements dans le total des financements extérieurs reçus par le projet est de 60 % en moyenne, et ce quel que soit le volet considéré.

1.4 L'évaluation des effets du PIA-ADEME

Dès la création du PIA, un budget spécifique a été prévu afin d'évaluer les effets des aides et l'évaluation d'impact indépendante a été définie comme une composante même du PIA (Cour des comptes, 2015; OCDE, 2014). Le budget consacré à l'évaluation est de 31 millions d'euros pour le PIA 1 et de 16 millions d'euros pour le PIA 2. Il s'agit d'une exigence de la Commission européenne, puisque en tant que régime d'aides de grande ampleur⁷, les aides « Investissements d'avenir » sont soumises à l'obligation d'évaluation⁸. Les plans d'évaluations proposés par les opérateurs et l'État à la Commission européenne doivent s'inspirer de la méthodologie commune développée par la Commission européenne (2014a), qui promeut l'utilisation des méthodes quasi-expérimentales.

En termes d'évaluation, il convient de distinguer les activités de suivi, le reporting et les audits, d'une part, des évaluations qui cherchent à établir un lien de cause à effet entre les aides et les performances des entreprises (évaluations dites « causales »), menées par des experts indépendants, d'autre part. Les bilans annuels et le reporting, qui sont réalisés en interne par les opérateurs du PIA eux-mêmes,

7. Un régime d'aide est qualifié de « régime de grande ampleur » au sens de l'article 1^{er}, paragraphe 2a. du RGEC lorsque son budget annuel moyen consacré aux aides d'État excède 150 millions d'euros.

8. Les applications et conditions d'obligation d'évaluation pour les régimes d'aides de grande ampleur sont prévues dans le Règlement Général d'Exemption par Catégorie (RGEC 2014)

sont nécessaires au suivi et à la gestion du PIA au quotidien. Les audits, comme ceux réalisés par la Cour des Comptes, servent quant à eux à contrôler la qualité de la gestion des sommes investies dans le cadre du PIA. Les différents rapports publiés sur le PIA, que ce soit par le CGI, les ministères concernés ou des organismes extérieurs, permettent de dresser un état des lieux du PIA (fonctionnement du PIA, types et montants des projets financés, thématiques abordées, *etc.*) et d'en avoir une vision globale. Il s'agit d'une description purement qualitative qui ne nous permet cependant pas d'appréhender de manière causale l'efficacité du PIA. L'évaluation d'impact doit, quant à elle, être effectuée par des experts indépendants, qui auront la liberté totale de publier les résultats de leurs travaux. Les opérateurs sont chargés de la rédaction des cahiers des charges pour l'appel d'offre ainsi que de la sélection des experts chargés de réaliser l'évaluation.

Concernant le PIA-ADEME, les évaluations conduites jusqu'à présent restent majoritairement de nature qualitative ou descriptive. Un premier bilan, réalisé par l'ADEME à la fin de l'année 2013, conclut à la capacité du PIA à transformer les filières et identifie plus précisément quatre types d'impact. Tout d'abord, selon ce bilan, le PIA joue un rôle déterminant sur les acteurs, notamment dans les filières émergentes comme le solaire photovoltaïque par exemple. En l'absence d'aides, compte tenu de risques très élevés, certains projets n'auraient pu être réalisés. Pour les filières en construction, les aides reçues dans le cadre du PIA ont permis d'accélérer la structuration de ces filières. En ce qui concerne des secteurs en mutation rapide, le PIA a permis d'accompagner le changement et d'aider les entreprises à y faire face. Ce fut par exemple le cas pour l'apparition de l'hybridation des véhicules de transport. Enfin, pour ce qui est des filières déjà structurées et au sein desquelles les mutations sont plus lentes, le PIA a davantage un impact de long terme et a permis la diffusion des innovations. L'ADEME souligne que le dispositif est parvenu à mobiliser les entreprises sur les différentes thématiques dont elle a la responsabilité et à stimuler la capacité à innover de ces dernières (ADEME, 2013). D'après les entreprises bénéficiaires interrogées pour dresser ce bilan, les aides accordées auraient engendré un chiffre d'affaires de plus de 10 milliards d'euros et entraîneraient la création de plus de 10 000 emplois d'ici à 2020.

Par ailleurs, l'ADEME et le CGI ont également commandé, auprès d'un *consortium* associant des cabinets de conseil spécialisés en évaluation des politiques publiques (Quadrant Conseil et Euréval), des experts des problématiques environnementales (I-care environment) ainsi que des chercheurs spécialisés en innovation (CGS, Mines ParisTech) une étude visant à développer une méthodologie pour évaluer les effets du PIA opéré par l'ADEME (ADEME, 2015). Les objectifs de cette étude étaient de parvenir à estimer les changements attribuables au bénéfice d'une aide ainsi qu'à développer une méthode qui pourrait être réutilisée lors des évaluations futures du programme. Ce travail vise principalement à clarifier les indicateurs à privilégier afin d'évaluer l'atteinte des objectifs des aides octroyées dans le cadre du PIA-ADEME. Trois effets doivent être étudiés en priorité : l'activité générée grâce à l'octroi d'une aide, les effets en matière d'emploi ainsi que les impacts environnementaux. Si les méthodes d'évaluation quantitative sont en théorie adaptées à une évaluation du PIA-ADEME, l'étude conclut en revanche que ces méthodes ne peuvent être mises en place en pratique car les bénéficiaires des aides ne sont pas suffisamment nombreux et les données mobilisables trop peu fiables. En effet, l'étude retient le *consortium* comme niveau d'observation pertinent, afin notamment de regarder l'effet sur la coopération, qui sont trop peu nombreux pour pouvoir observer des effets. En ce qui concerne l'analyse quantitative menée dans cette étude, seule une comparaison de la situation avant et après l'obtention d'une aide dans le cadre du PIA-ADEME a été réalisée. Ces résultats ne peuvent cependant en aucun cas être interprétés de manière causale.

À ce jour, aucune évaluation causale du PIA n'a encore été menée, notamment parce que le recul temporel nécessaire à la détection d'un effet est relativement faible. L'évaluation intermédiaire du régime d'aide de l'ADEME est ainsi la première évaluation conduite. Cette évaluation économétrique intermédiaire est prévue dans le plan d'évaluation soumis par l'ADEME et validé par la Commission européenne en août 2015⁹. Elle vise à préparer l'évaluation finale, dont les conclusions doivent être remises en juin 2020.

9. Décision C(2015) 5394 du 4 août 2015.

CHAPITRE 2

JUSTIFICATIONS ÉCONOMIQUES ET ÉVALUATIONS EMPIRIQUES DES AIDES À LA R&D

Dans ce chapitre, nous revenons sur les justifications économiques d'un soutien public à la R&D (partie 2.1), puis résumons ensuite les résultats des différents travaux d'évaluation menés pour juger de l'efficacité de l'aide publique, qu'elle prenne la forme d'une subvention, d'une avance remboursable ou d'un crédit d'impôt (partie 2.2).

2.1 Les justifications économiques d'un soutien public à la R&D et à l'innovation

Dans cette partie, nous revenons sur les justifications économiques d'un soutien des pouvoirs publics à la R&D et à l'innovation. Nous présentons en quoi la R&D s'apparente à un bien public, et nous décrivons les défaillances de marché résultant de la nature même de la R&D, telles que l'asymétrie d'information, l'imperfection des marchés de capitaux ou encore les barrières non-financières pouvant décourager la mise en œuvre de certains projets. Nous abordons aussi le coût social de la non-coopération dans les activités de R&D. Enfin, nous terminons cette partie en

évoquant les relations qu'entretiennent la R&D, la compétitivité des entreprises et la croissance, et comment celles-ci viennent justifier une intervention de l'État dans l'économie.

2.1.1 R&D, bien public et externalités positives

L'activité de recherche et développement (R&D) crée un type de bien particulier qui peut être considéré comme un bien public. Les connaissances produites sont d'une part non-rivales : leur utilisation par une entreprise ne diminue pas la quantité disponible de ce savoir pour les autres entreprises. Elles sont, d'autre part, non-exclusives dans le sens où toutes les entreprises peuvent, en théorie, en bénéficier gratuitement. La nature même de l'activité de R&D fait qu'elle est susceptible de générer des externalités positives, c'est-à-dire de bénéficier aussi bien à celui qui l'a initiée qu'à d'autres acteurs. Ceci implique que le rendement pour la société dépasse le rendement privé de l'entreprise investissant en R&D, qui par conséquent ne s'approprie qu'une partie des bénéfices générés par sa propre recherche. Or, l'entreprise ne s'engagera dans un projet de R&D que si son rendement privé est au moins égal au montant investi, même si les bénéfices pour la société peuvent être bien plus conséquents. Le niveau d'investissement privé est susceptible d'être inférieur au niveau qui serait socialement optimal, venant ainsi justifier une intervention des pouvoirs publics.

Ce risque de sous-investissement serait d'autant plus fort pour la R&D réalisée dans le domaine environnemental, qui est caractérisée par un problème de double externalité (Rennings, 2000). En plus de créer des connaissances supplémentaires dont peuvent bénéficier d'autres acteurs, la R&D et les innovations vertes qui en résultent produisent également des externalités environnementales positives. Ainsi, les bénéfices liés à une réduction de la pollution ou le développement de technologies moins consommatrices en énergies profitent également aux autres acteurs sous forme d'une meilleure qualité de l'environnement.

Si sa nature de bien public fait de la R&D une ressource accessible à tous, s'approprier et être capable de mobiliser le savoir produit par d'autres entreprises a un coût et nécessite que les entreprises aient déjà un certain niveau de dépenses

en R&D (Cohen et Levinthal, 1989; Zahra et George, 2002). Ainsi, une entreprise aura intérêt à investir en R&D pour augmenter sa capacité d'absorption des connaissances produites par d'autres entreprises. Dans les secteurs d'activité où les externalités positives¹ sont conséquentes, on peut atteindre une situation paradoxale dans laquelle les entreprises sur-investissent pour augmenter leur capacité d'absorption des connaissances produites par leurs concurrents. Dans ce cas-là, le niveau d'investissement privé excéderait le niveau socialement optimal. Certaines situations de sur-investissement peuvent également entraîner une perte sèche pour la société. Dasgupta et Stiglitz (1989) montrent que lorsque plusieurs entreprises d'un même secteur entrent en compétition pour l'obtention d'un brevet, et donc d'un pouvoir de monopole, les coûts importants investis dans cette course au brevet ne peuvent pas être récupérés par les entreprises perdantes. Ces montants constituent une perte sèche et une perte de bien-être pour la société, et auraient pu servir à financer la R&D d'une manière plus coopérative par l'intermédiaire de *consortium* de recherche ou d'autres stratégies visant à favoriser la coopération.

Par ailleurs, d'autres travaux de recherche suggèrent que la R&D ne doit pas être vue comme un bien public pur puisqu'il existe plusieurs moyens pour une entreprise de limiter l'accès au produit de sa recherche, par des dépôts de brevets par exemple. Ainsi, les bénéfices pour la collectivité peuvent être très limités, et le savoir produit très spécifique à un pan de l'industrie en particulier. Dans un tel contexte, subventionner la R&D peut constituer une dépense injustifiée (Cerulli, 2010).

Si, du point de vue de la théorie, il semble difficile de conclure à un sur-investissement ou à un sous-investissement en R&D par rapport à l'optimum social, un consensus semble en revanche émerger dans les travaux empiriques, qui tendent plutôt à supporter l'hypothèse d'un investissement sous-optimal en R&D. Ces travaux, en estimant les taux de rendement privé et social des investissements en R&D, trouvent, dans leur très grande majorité, que les retombées pour la société vont bien au-delà des simples retombées pour l'entreprise. Ainsi, dans sa revue de

1. On parle d'externalité lorsque l'activité d'un agent économique bénéficie, ou au contraire nuit, à un autre agent sans qu'il y ait de compensation monétaire. La R&D est source d'externalités positives puisque la production de connaissances d'une entreprise peut être utilisée librement par d'autres sans qu'il y ait de compensation monétaire (sauf dans le cas où il y a un dépôt de brevet).

littérature, Nadiri (1993) trouve des taux de rendement privé situés entre 20 % et 30 % pour l'entreprise, contre 20 % et 100 % pour la société, avec une moyenne de 50 %. Dans une revue de littérature plus récente, Sveikauskas (2007) montre que sur 57 innovations considérées, le taux privé de rendement médian est de 28 % contre 71 % pour les rendements sociaux (voir également Hall (1996), Hall et Mohnen (2010) et Griffith (2000)). En partant du constat que les entreprises fondent leur décision d'investir en R&D sur les rendements privés escomptés, ces résultats suggèrent un sous-investissement en R&D puisque les rendements privés sont inférieurs aux rendements sociaux.

Les taux de rendements de la R&D dépendent toutefois de la source de financement. Sveikauskas (2007) montre ainsi que, d'après les études considérées dans sa revue de littérature, les taux de rendement calculés ne sont valables que pour la R&D financée par des fonds privés, et que les rendements privés et sociaux de la R&D financée par les pouvoirs publics sont proches de zéro. Hall et Mohnen (2010) concluent également que les études qu'ils incluent dans leur revue de littérature trouvent des taux de rendements plus faibles pour la R&D financée par des fonds publics. Les auteurs avancent plusieurs explications. La première explication est la difficulté à mesurer les externalités générées par la R&D financée par des fonds publics, qui concerne des secteurs bien précis². Deuxièmement, les aides publiques à la R&D ciblent des secteurs souvent plus risqués, où le rendement est potentiellement moindre. Enfin, la R&D publique peut stimuler la R&D privée et n'aurait donc qu'un rendement indirect, difficile à appréhender. Les auteurs soulignent toutefois que la R&D publique peut avoir de forts rendements dans des conditions très spécifiques, notamment en recherche fondamentale ou lorsque les entreprises ont des budgets importants en R&D et interviennent sur des marchés où l'État est un des principaux acheteurs.

2. En effet, dans les études considérées, la R&D publique concerne surtout le secteur des services dans l'industrie manufacturière, secteur dans lequel les externalités sont difficilement mesurables.

2.1.2 Asymétrie d'information et marché des capitaux imparfaits

Une autre raison souvent invoquée pour expliquer le niveau sous-optimal d'investissement en R&D est le problème d'asymétrie d'information, notamment entre l'entreprise qui souhaite investir en R&D et le financeur, lorsque les deux sont distincts (Hall, 2002). En effet, les entreprises ont une meilleure information que le financeur sur la potentielle réussite de leurs projets de recherche ainsi que sur les bénéfices escomptés. N'étant pas capables de distinguer les bons des mauvais projets, les financeurs prêtent à un taux plus élevé que pour un investissement classique en capital, incluant une prime de risque liée à la réussite incertaine du projet. L'entreprise pourrait cependant réduire cette asymétrie d'information en divulguant son projet de R&D. Or, ses idées risqueraient fortement d'être reprises par des concurrents, ce qui viendrait affaiblir une potentielle position dominante dans le cas où les recherches s'avéreraient fructueuses. La conséquence directe est que les financements extérieurs seront toujours plus coûteux qu'un financement par fonds propres.

Plus l'écart entre le coût d'un financement extérieur et un financement interne est élevé, moins les entreprises fortement dépendantes de financements externes se lanceront dans des projets potentiellement porteurs d'innovations et de croissance. Ceci est d'autant plus vrai que l'entreprise est petite et ne dispose pas de fonds propres, à l'instar des petites et moyennes entreprises (PME) et des start-ups, comme l'ont montré de nombreuses études empiriques. Mohnen et Rosa (2000) analysent des données d'enquêtes menées auprès d'entreprises canadiennes sur les perceptions des entraves à l'innovation, et montrent que les difficultés de financement sont perçues comme un obstacle majeur à la conduite de R&D pour les petites entreprises, alors que cela ne fait pas partie des préoccupations principales des grandes entreprises. En étudiant l'effet d'un financement public sur l'investissement en R&D et sur les perspectives de croissance de PME finlandaises, Hyytinen et Toivanen (2005) montrent que les entreprises les plus dépendantes de financements extérieurs dépensent significativement plus en R&D quand des fonds publics sont mis à leur disposition. Ces résultats viennent justifier, une fois de plus, une intervention des pouvoirs publics dans le domaine de la R&D.

2.1.3 Autres obstacles à l'innovation et à la R&D

Au-delà de la barrière que constitue l'accès à des financements extérieurs, plusieurs autres barrières non financières peuvent soit dissuader d'investir en R&D, soit rendre plus difficile l'atteinte des objectifs une fois le projet lancé. Dans leurs revues de littérature sur les freins à l'innovation et à la R&D, Tiwari et Buse (2007) et Galia et al. (2012) recensent notamment le manque d'informations relatives aux nouvelles technologies et aux marchés sur lesquels des activités innovantes peuvent être menées, des rigidités organisationnelles au sein de l'entreprise et des rigidités institutionnelles (normes, législation, *etc.*) ainsi que le manque de main d'œuvre qualifiée pour mener des travaux de R&D.

Les résultats de Mohnen et Rosa (2000) montrent que l'importance accordée à certaines barrières varie en fonction du secteur industriel, de la taille de l'entreprise, de la perception de l'environnement concurrentiel et du niveau d'investissement en R&D. Leurs analyses suggèrent que les rigidités organisationnelles sont d'autant plus fortes que l'entreprise est grande. Les petites entreprises ont plutôt tendance à souligner que, au-delà de la contrainte de financement, le manque d'équipements spécialisés constitue l'une des barrières les plus importantes. En revanche, le manque de personnel qualifié est perçu comme une entrave à la R&D aussi bien pour les petites que pour les grandes entreprises. Le manque de personnel qualifié comme barrière à l'innovation est invoqué dans de nombreuses travaux empiriques. Ainsi, Segarra et al. (2008) montrent que les coûts d'entrée et la difficulté à trouver des employés et des partenaires qualifiés sont les facteurs les plus importants affectant l'innovation pour les entreprises catalanes. Galia et Legros (2004) parviennent à des résultats similaires sur la base d'une enquête menée auprès des PME françaises. Par ailleurs, les dirigeants des PME engagées dans l'innovation « verte » perçoivent des barrières plus nombreuses et d'une intensité plus importante (Pinget et al., 2015). En effet, la nature plus complexe de la R&D dans le domaine environnemental nécessite de repenser l'activité de l'entreprise en profondeur, multipliant ainsi les barrières à l'innovation. Cela va du changement des matières premières et des composants utilisés à la reconception des produits, et exige des étapes de contrôles plus fréquentes et plus contraignantes que pour une activité classique

(De Marchi, 2012). La R&D « verte » nécessite par conséquent plus de coordination entre acteurs, plus de connaissances et est plus risquée, ce qui appelle à une attention toute particulière de la part des pouvoirs publics concernant ce domaine de recherche.

Plutôt que d'étudier l'effet de ces obstacles de manière séparée sur l'activité de R&D, de nombreux travaux cherchent à montrer que c'est la complémentarité entre ces différentes barrières qu'il est important d'appréhender. Comprendre dans quelles mesures ces barrières se renforcent mutuellement est essentiel pour assurer l'efficacité des politiques publiques relatives à l'innovation. En effet, à quoi serviraient les crédits d'impôt en faveur de l'innovation si le pré-requis de la main d'œuvre qualifiée n'était pas rempli ? Ainsi, Mohnen et Röller (2000) utilisent les données de l'enquête européenne *Community Innovation Survey* et trouvent que le manque de capital humain vient renforcer les autres obstacles à l'innovation. Ce résultat est également présent dans les travaux de Galia et Legros (2004). Mohnen et Rosa (2000) identifient, quant à eux, des paires d'obstacles : les forts coûts à l'entrée et les risques liés à l'activité de R&D d'une part, et les problèmes de gouvernance interne et externe d'autre part, semblent se renforcer mutuellement. Ces complémentarités révèlent un problème systémique (Chaminade et Edquist, 2010) et appellent à considérer l'environnement dans lequel prend place l'innovation, afin de concevoir des politiques publiques efficaces.

2.1.4 Favoriser la coordination entre les entreprises

Une autre défaillance de marché qui peut justifier l'intervention publique est celle du manque de coordination entre les différents acteurs de la R&D et de l'innovation, qui peut une fois de plus conduire à un niveau d'investissement sous-optimal de R&D. Comme nous l'avons déjà évoqué plus haut (voir 2.1.1), les externalités positives générées par la R&D peuvent entraîner des comportements de « passager clandestin » si le coût de la R&D pour une entreprise excède les bénéfices attendus, et conduire à une situation de sous-investissement. À l'inverse, les entreprises peuvent surinvestir en R&D si les perspectives de profits sont élevées, avec le risque de générer des doublons de recherche qui constituent une perte sèche pour la so-

ciété (voir par exemple Dasgupta et Stiglitz (1989) ou Baron et al. (2011)).

Ces situations de surinvestissement ou de sous-investissement reflètent un manque de coordination entre les différents acteurs. La coopération, notamment entre les universités ou laboratoires de recherche et les entreprises, est un facteur important qui permet de maximiser le rendement des investissements réalisés en R&D, qu'ils soient publics ou privés. La coopération permet en outre de mutualiser les risques liés à l'incertitude de la réussite de la R&D et de lever les contraintes financières (Abramovsky et al., 2005). Enfin, cela permet de réduire les dépenses en R&D de chacun des acteurs qui coopèrent en évitant les investissements excessifs en R&D qui peuvent conduire à des doublons de recherche (Irwin et Klenow, 1996).

Par ailleurs, Klette et Møen (1999) considèrent qu'une partie des activités de R&D, appelée « technologies d'application générale » (*general purpose technologies*) dans la littérature, doit faire l'objet d'une attention particulière, puisque sa nature spécifique est susceptible d'engendrer des problèmes de coordination plus marqués. Ces technologies sont caractérisées par leur potentiel en termes d'améliorations techniques et par leur complémentarité forte avec d'autres innovations. Les technologies de l'information en sont une bonne illustration³. La coordination entre ceux qui développent la technologie de base et ceux qui développent ses applications concrètes accélère l'échange d'informations et, *in fine*, l'éventuelle commercialisation de ces applications.

Si les différents acteurs de la R&D coopèrent parfois d'eux-mêmes en formant des *consortiums* de recherche (D'Aspremont et Jacquemin, 1988), l'action publique est néanmoins justifiée. Abramovsky et al. (2005) montrent que les incitations publiques à la coopération augmentent la probabilité des entreprises à s'engager dans des stratégies de coopération, et ce d'autant plus lorsqu'il s'agit de coopération avec la recherche fondamentale.

3. Par exemple, le développement des micro-ordinateurs – eux-mêmes développés grâce à l'invention des semi-conducteurs et des circuits intégrés – a servi de base à la conception de nombreux logiciels et applications qui ont complètement modifié l'organisation de plusieurs secteurs d'activité.

2.1.5 Stimuler la compétitivité des entreprises et la croissance économique

Dans les récents modèles théoriques de croissance, l'accumulation de R&D joue un rôle fondamental puisque, par le biais de l'innovation, elle contribue à augmenter la productivité des entreprises et, *in fine*, le stock de richesses produites.

Alors que dans les modèles de croissance classiques, le progrès technologique est perçu comme un élément complètement exogène et résiduel, les modèles de croissance endogène, formalisés au début des années 1990 (voir par exemple Romer (1990) et Sala-i-Martin (1990)), intègrent le progrès technologique explicitement comme un déterminant de la croissance⁴. Griliches (1992) souligne deux points centraux des modèles de croissance endogène. D'une part, le changement technologique induit par la R&D est le résultat d'un choix conscient et d'une décision explicite de la part des agents économiques, et notamment des entreprises. D'autre part, les connaissances produites par la R&D ne sont pas sujettes à des rendements décroissants, puisqu'elles génèrent des externalités positives et bénéficient ainsi à d'autres acteurs. Ceci ouvre des perspectives de croissance quasi illimitées venant ainsi justifier, en théorie, l'intervention des pouvoirs publics.

Fagerberg (1994) résume les implications de ces modèles pour la croissance comme suit : le taux de croissance dépend i) du montant investit en R&D, ii) du degré avec lequel une nouvelle technologie peut être appropriée (perspectives de monopole) et iii) de la patience des investisseurs.

En plus d'être un déterminant direct de la croissance économique, Baumol (2002) voit dans l'investissement en R&D la clé de la compétitivité des entreprises. D'après lui, la survie des entreprises dans les économies capitalistes actuelles repose davantage sur leur capacité à innover que sur leur compétitivité-prix. Ainsi, une entreprise prospère peut ne pas être efficiente dans un équilibre de court terme, mais devenir rentable si elle poursuit son investissement en R&D ; et inversement, une entreprise rentable aujourd'hui peut être amenée à quitter le marché si l'effort de R&D n'est pas poursuivi.

4. Voir Fagerberg (1994) pour un très bon résumé des différents modèles de croissance endogène avec une perspective historique.

Enfin, au-delà même de l'effet positif de la recherche sur la croissance, les pouvoirs publics peuvent avoir un intérêt stratégique à intervenir afin de favoriser l'innovation et la croissance dans certains domaines particuliers, comme le domaine environnemental par exemple. Si, pendant longtemps, l'intervention publique visait simplement à augmenter le niveau de recherche dans l'ensemble de l'économie, la R&D et l'innovation ont ensuite davantage été perçues comme un moyen d'orienter l'économie vers des secteurs pouvant être considérés comme stratégiques (Institut des Hautes Etudes pour la Science et la Technologie, 2013). Une littérature abondante s'est penchée sur la façon dont l'investissement en R&D « propre » (ou verte) plutôt que « sale » peut aboutir à une croissance verte. Dans un modèle théorique, Acemoglu et al. (2012) montrent que, pour éviter un désastre environnemental, il faut, en plus d'une taxe carbone, mettre en place des subventions aux innovations vertes, et ce le plus tôt possible. Ce résultat provient de l'hypothèse sous-jacente selon laquelle les innovations se font plus facilement dans les secteurs les plus avancés d'un point de vue technologique. En d'autres termes, les externalités positives de la R&D sont de nature intra-sectorielle. Plus on investit dans l'innovation verte, plus les innovations se feront spontanément dans ce secteur et plus la croissance sera « propre ».

Cette hypothèse théorique de dépendance au sentier historique des innovations a été testée empiriquement pour certains types de technologie. Aghion et al. (2016) construisent de nouvelles données de panel de brevets au niveau de l'entreprise dans l'industrie automobile en distinguant les innovations « sales » (par exemple un moteur à combustion interne) et « propres » (moteurs électriques et hybrides par exemple) pour 80 pays sur plusieurs décennies. Ils montrent une dépendance au type d'innovations passées (propre et sale), à la fois au niveau global et à l'échelle de l'entreprise. On peut retrouver des résultats similaires dans les travaux de Popp (2002) et Dechezleprêtre et al. (2014), ainsi que dans la revue de littérature réalisée par Popp et al. (2010).

2.2 Quels sont les effets des aides publiques à la R&D ?

Après avoir souligné les justifications économiques d'un soutien public à la R&D, il s'agit de s'intéresser aux effets de ce soutien et à la capacité de cette dépense publique à atteindre les objectifs qui lui sont assignés. En effet, cette dépense vise à amener les efforts de R&D à un niveau socialement optimal et à stimuler la recherche privée. Elle doit aussi permettre une amélioration des performances des entreprises, de leur compétitivité et de leur capacité à innover. Nous nous concentrons d'abord et principalement sur l'efficacité des subventions, qui sont l'objet de cette étude⁵, puis comparons celle-ci avec l'efficacité des incitations fiscales à la R&D, une aide indirecte qui constitue un autre instrument, aujourd'hui majeur, de la politique de soutien à la R&D.

2.2.1 Additionnalité ou substitution entre subventions et dépenses privées de R&D ?

Un des objectifs des aides publiques est de parvenir à accroître le niveau global de R&D, afin de le rapprocher de ce que serait le niveau optimal socialement. De nombreuses études ont ainsi cherché à déterminer si le soutien public, sous la forme de contrats publics ou de subventions, venait se substituer à la dépense en recherche privée (effet de *crowding-out* de la dépense privée par la dépense publique), au contraire, stimuler les efforts privés de R&D (effet d'additionnalité, ou effet de levier), ou encore simplement s'ajouter à cette dernière. En théorie, chacun de ces effets peut prévaloir. En effet, on pourrait s'attendre à ce que les aides publiques à la R&D viennent simplement s'ajouter aux dépenses privées, et ainsi accroître le volume total de dépenses sans affecter les dépenses privées de R&D. Au-delà d'un simple effet de complémentarité, un effet d'additionnalité pourrait apparaître si l'obtention d'une aide publique agit comme un signal de rendements d'innovations élevés et vient baisser le coût d'un financement externe. Dans le cas de subventions

5. A notre connaissance, il n'existe pas d'évaluation d'aides à la R&D prenant la forme d'avances remboursables.

accordées à condition que des entreprises soient réunies dans un *consortium* de recherche, l'aide financière peut stimuler l'activité de R&D de l'entreprise à plus long terme en impulsant une coopération dans les activités de recherche. Inversement, les aides financières pourraient tout aussi bien venir se substituer aux dépenses privées que les entreprises aidées auraient de toutes façons effectuées (*crowding-out effect*), par un pur effet d'aubaine.

Bien que nombreuses, les études relatives aux effets des subventions sur les dépenses privées de R&D ne parviennent pas à un consensus (David et al., 2000). Ainsi, quand Leyden et Link (1991) trouvent un effet positif des dépenses publiques de R&D sur le budget total de R&D des entreprises, tout en contrôlant de la simultanéité de l'accord de subventions et des décisions de dépenses privées, Lichtenberg (1984) trouve, en revanche, que si à première vue les contrats publics semblent induire une hausse des dépenses de R&D, cet effet disparaît dès lors que l'on introduit des effets fixes au niveau des entreprises. En étudiant les aides financières à la R&D dans le secteur de l'électronique en Norvège, Klette et Møen (2011) montrent que l'absence d'effet de *crowding-out* s'explique par la part d'autofinancement requise pour bénéficier de ces aides ; pour autant, ces aides n'ont pas eu l'effet escompté de stimuler les efforts privés de R&D. Ce résultat (ni additionnalité, ni substitution) semble robuste : il ne varie pas avec l'échantillon et les périodes considérés. A long terme, en revanche, avoir bénéficié de ce soutien public pourrait se traduire par une augmentation des dépenses de R&D, par un effet dit d'apprentissage. Au Japon, appartenir à l'un des *consortiums* de recherche, fortement subventionnés, dans les hautes technologies est associé à des dépenses privées de R&D plus élevées, après avoir contrôlé de certaines caractéristiques des entreprises (Branstetter et Sakakibara, 1998).

Si les résultats semblent légèrement pencher en faveur d'une complémentarité des dépenses publiques et privées de R&D, David et al. (2000), dans leur revue de la littérature sur les travaux empiriques publiés depuis 1965, montrent que la plupart des études concluant à un effet de complémentarité, voire d'additionnalité, souffrent d'importantes limites méthodologiques et leurs résultats sont souvent biaisés. Les entreprises bénéficiant d'aides publiques en général, et de subventions

en particulier, ne sont en effet pas comparables aux entreprises ne bénéficiant pas de ces aides, ce qui rend difficile l'identification d'un effet causal des subventions publiques sur le niveau de recherche privée des entreprises. Lorsque l'on corrige, même partiellement, d'un effet de sélection⁶, que ce soit par l'utilisation de variables instrumentales ou par l'estimation d'un modèle structurel, l'effet d'additionnalité des dépenses publiques et privées de R&D disparaît le plus souvent. En effet, l'hypothèse selon laquelle l'attribution de subventions serait exogène peut mener à une mauvaise identification de l'effet des aides : il est probable que les opportunités technologiques, tout comme d'autres caractéristiques des entreprises bénéficiant de subventions pour la R&D, expliquent le fait de bénéficier de celles-ci, tout en étant liées au niveau de dépenses privées en R&D des entreprises.

Face au nombre important d'articles et à la forte variabilité de leurs résultats, García-Quevedo (2004) a réalisé une méta-analyse des différents travaux réalisés sur cette question. Prenant en compte 39 articles traitant de la relation entre subventions et dépenses privées de R&D, tant au niveau des entreprises, des secteurs qu'au niveau national, García-Quevedo (2004) trouve qu'aucune des caractéristiques de l'analyse (niveau d'analyse, modélisation économétrique, variables de contrôle, *etc.*) n'est associée à un résultat particulier, à l'exception de l'utilisation de données au niveau des entreprises : les études utilisant ce type de données ont une probabilité légèrement plus forte de conclure en faveur d'un effet de substitution.

2.2.2 Effets des subventions sur l'innovation et les performances des entreprises

Outre l'accroissement des dépenses de R&D au sein de l'économie, un objectif visé par les dispositifs de soutien public à la R&D est l'amélioration des performances des entreprises aidées. En effet, ces dispositifs d'aide doivent permettre d'accroître la capacité à innover des entreprises, sur laquelle repose aujourd'hui largement leur compétitivité (Baumol, 2002). Dans quelle mesure le soutien public

6. Les différences de caractéristiques entre entreprises bénéficiant d'un dispositif de soutien public à la R&D, et celles n'en bénéficiant pas, peut biaiser les résultats : c'est ce qu'on appelle un biais de sélection (voir chapitre 4).

à la R&D sous la forme de subventions atteint-il cet objectif?

Effet sur l'innovation

Des travaux plus récents s'intéressent à l'effet sur l'innovation des subventions à la R&D en utilisant les brevets déposés comme mesure de l'innovation. L'étude de la littérature permet de conclure que les résultats sont relativement hétérogènes en fonction du pays étudié.

Ainsi, Jaffe et Le (2015) montrent que les entreprises néo-zélandaises bénéficiant de subventions à la R&D sont davantage dépositaires de brevets que les entreprises ne bénéficiant pas de ce type d'aides. De même, en évaluant un programme d'aides dans une région du nord de l'Italie, Bronzini et Piselli (2016) montrent que le fait de recevoir une subvention pour un projet de R&D accroît le nombre de brevets déposés par entreprise. Kilponen et Santavirta (2007) parviennent à des résultats similaires en Finlande : les subventions accordées dans le cadre d'un programme en faveur de la R&D auraient un effet positif sur le nombre de brevets déposés. En revanche, s'intéressant au dispositif SkatteFUNN de soutien financier à la R&D en Norvège⁷, Cappelen et al. (2012) ne trouvent pas d'effet des subventions sur les brevets déposés. Czarnitzki et al. (2007) montrent que les entreprises finlandaises soutenues financièrement dans leurs activités de R&D connaîtraient une probabilité moins grande de déposer des brevets si elles ne bénéficiaient pas de cette aide financière, mais trouvent que cette relation n'est en revanche pas vérifiée en Allemagne.

Outre le fait de s'intéresser à l'impact du soutien à la R&D sur l'innovation au niveau de l'entreprise, une littérature plus abondante a étudié l'effet des subventions accordées à des *consortiums* de recherche sur le nombre de brevets déposés. Ainsi, Czarnitzki et al. (2007) montrent que les entreprises finlandaises et allemandes impliquées dans des *consortiums* de recherche en général, et celles appartenant à des *consortiums* de recherche subventionnés en particulier, connaîtraient des probabi-

7. Ce dispositif prend la forme d'un crédit d'impôt. Toutefois, celui-ci n'est accordé que pour des projets sélectionnés par l'agence nationale de la recherche, et tout ou partie du crédit d'impôt est reversé comme une subvention si l'entreprise paie un impôt d'un montant inférieur au crédit accordé ou aucun impôt. Pour ces deux raisons, ce dispositif s'apparente aux subventions pour projets sélectionnés (*grants*) décrits par la littérature que nous considérons ici.

lités plus faibles de déposer des brevets si elles n'appartenaient pas à un *consortium* de recherche et ne bénéficiaient pas non plus de financements publics. Dans un précédent article, Czarnitzki et Fier (2003) montraient, en se concentrant sur les données allemandes, que participer à une collaboration de recherche en général ou prendre part à une collaboration subventionnée augmente les probabilités de déposer un brevet dans un délai court ou à plus long terme. Les résultats indiquent d'autre part que l'effet sur le dépôt de brevets est d'autant plus fort que le *consortium* de recherche auquel participe l'entreprise est en partie financé par des fonds publics. Les deux auteurs trouvent, par ailleurs, un effet également positif des deux types (subventionnée ou non-subventionnée) de collaboration de R&D sur le nombre de brevets *accordés*. De la même façon, l'appartenance à l'un des *consortiums* de recherche existant dans le secteur des semi-conducteurs au Japon, largement subventionnés, est associée à une augmentation de 5 % du nombre de brevets accordés (Branstetter et Sakakibara, 1998). En revanche, plus récemment, les subventions accordées à des *consortiums* dans le cadre du dispositif européen « Eurostars » n'ont pas stimulé la production de connaissances, appréhendée ici par les dépôts de brevets. Toutefois, cela est probablement dû au fait que le dispositif concerne des projets de R&D aux applications très « proches du marché », et s'adresse aux petites et moyennes entreprises, qui sont moins dépositaires de brevets dans l'absolu (Hünermund et Czarnitzki, 2015).

Effets sur les performances des entreprises

Tout comme pour l'innovation, les effets des subventions à la R&D sur les indicateurs de performance des entreprises, parmi lesquels la productivité ou l'évolution des ventes et de l'emploi, ne font pas aujourd'hui l'objet d'un consensus dans la littérature économique.

Ainsi, à partir de données israéliennes, Griliches et Regev (1998) estiment une fonction de production avec, comme déterminant de la productivité du travail, la part de R&D accumulée grâce à des financements publics, incluant à la fois les subventions pour des projets de R&D et des contrats publics de R&D. Les auteurs trouvent que ces financements ont un effet positif et d'une ampleur importante sur

la productivité du travail. Cet effet demeure lorsque l'on contrôle par les autres caractéristiques des entreprises afin de limiter les possibles biais de sélection.

À l'inverse, Klette et Møen (1999) montrent que les entreprises ayant bénéficié d'importants dispositifs d'aides à la R&D pour le secteur des technologies de la communication en Norvège ont de moins bonnes performances en termes de productivité du travail et de productivité globale des facteurs de production que les entreprises n'ayant pas bénéficié de ces aides. Toutefois, les auteurs expliquent ces résultats par la sélection des entreprises les moins performantes dans ce secteur. En effet, ces dispositifs avaient avant tout pour objectif de soutenir les entreprises d'un secteur faisant alors face à de grandes difficultés.

Dans leur évaluation des effets du programme américain SEMATECH qui s'adressait à l'industrie manufacturière des semi-conducteurs, Irwin et Klenow (1996) mettent en évidence un effet neutre des aides publiques sur la productivité. Toutefois, leur analyse montre que les entreprises ayant bénéficié de ce programme ont connu une amélioration de leurs performances économiques et notamment une croissance de leurs ventes. Ces effets positifs peuvent s'expliquer par les modalités d'attribution des subventions qui avaient tendance à favoriser les entreprises leaders dans leur domaine. De la même façon, les subventions accordées dans le cadre du programme américain SBIR destiné aux PME ou du programme FEDER en Finlande, moins largement axés sur les hautes technologies, ont conduit à une amélioration des performances des entreprises bénéficiaires par une augmentation de leurs ventes ainsi que de l'emploi (Lerner, 1999; Einiö, 2014). Dans le premier cas, Lerner (1999) explique ces résultats par un effet signal sur les marchés des capitaux de l'attribution d'une subvention.

Ces différents travaux soulignent les possibles biais de sélection dans l'attribution des aides qui peuvent conduire à une surestimation ou une sous-estimation des effets des aides sur les différents indicateurs de performance des entreprises, rappelant ainsi la précaution avec laquelle ces résultats doivent être interprétés.

2.2.3 Efficacité des incitations fiscales à la R&D

Enfin, après s'être concentrés sur l'efficacité des dépenses publiques sous forme de subventions, nous résumons brièvement les résultats des études évaluant un autre instrument désormais majeur de la politique de soutien à la R&D : les crédits d'impôt, qui viennent réduire l'impôt des entreprises ayant des activités de R&D⁸. Ceci permet de comparer l'efficacité de ces deux modes de soutien public.

Effets d'additionnalité ou de substitution des crédits d'impôt pour la R&D

La littérature relative à l'impact des crédits d'impôt sur les dépenses de R&D est relativement abondante, et contrairement aux études sur ce thème pour les subventions à la R&D, le constat est plutôt unanime : si un effet d'additionnalité n'est pas toujours mis en évidence, l'ensemble des travaux semble rejeter l'idée que la dépense fiscale de R&D se substituerait aux dépenses privées de R&D.

Une partie des travaux conclut largement en faveur d'un effet d'additionnalité. Ainsi, dans leur revue de la littérature, Hall et Van Reenen (2000) mettent en évidence, que ce soit à partir de données macro (au niveau d'un pays) ou micro (au niveau des entreprises), sur des périodes différentes ou en utilisant différentes méthodologies, que les travaux réalisés convergent vers l'idée que les entreprises qui bénéficient de crédits d'impôt engagent significativement plus de ressources en R&D. En mobilisant les données de neuf pays de l'OCDE sur une période de 19 ans allant de 1979 à 1997, Bloom et al. (2002) utilisent un modèle d'économétrie structurelle⁹ dans lequel le niveau d'investissement en R&D d'une entreprise est déterminé par le coût du capital en R&D, pour estimer l'effet des incitations fiscales sur l'intensité en R&D. Ils estiment qu'une baisse de 10 % du coût de la R&D entraîne une augmentation de 1 % du niveau de R&D sur le court terme et de près de 10 % sur le long terme. Au niveau de l'entreprise, les élasticités de long

8. Selon les dispositifs, le montant du crédit d'impôt est calculé de manière incrémentale (un taux s'applique à la variation des dépenses en R&D entre l'année $n - 1$ et l'année n), et/ou en volume (un taux s'applique au volume total des dépenses en R&D).

9. A la différence de méthodes quasi-expérimentales (voir chapitre 4), l'économétrie structurelle suppose une modélisation explicite du phénomène d'intérêt, c'est-à-dire une formalisation de la relation entre une variable et, par exemple, un comportement. Les méthodes quasi-expérimentales sont, elles, « agnostiques » concernant la façon dont se produit l'effet qui sera éventuellement observé.

terme sont nettement plus faibles et estimées à 0,45 pour les Pays-Bas sur la période 1996-2004 (Lokshin et Mohnen, 2013b), et à 0,4 en France sur la période 1981-2007 (Mulkay et Mairesse, 2013). Toujours pour le cas de la France, mais en utilisant des méthodes quasi-expérimentales, Duguet (2010) évalue l'effet du crédit d'impôt recherche (CIR) sur la période 1993-2003 et montre un effet de un pour un : un euro de crédit d'impôt entraîne une hausse de un euro des dépenses privées en R&D, résultats également confirmés par l'étude de Lhuillery et al. (2013), qui étend l'analyse à la période 1993-2009. Enfin, en comparant les entreprises françaises ayant bénéficié d'un crédit d'impôt après la réforme de 2008 à celles n'y ayant pas eu recours, Bozio et al. (2016) trouvent que les bénéficiaires ont significativement augmenté leurs dépenses en R&D : pour les entreprises qui demandaient déjà le CIR avant 2008, les auteurs trouvent qu'un euro supplémentaire de crédit d'impôt est associé à une augmentation de 2,30 euros de dépenses privées de R&D ; lorsqu'on prend aussi en compte les entreprises bénéficiant du CIR seulement depuis la réforme étudiée, l'effet multiplicateur est moindre mais toujours positif, égal à 1,8.

Ces résultats semblent plaider en faveur d'une additionnalité de la dépense publique lorsque celle-ci prend la forme d'un crédit d'impôt. Dans leur revue de littérature, Martínez-Ros et Corchuelo (2009) montrent qu'une écrasante majorité des travaux concluent en ce sens. Toutefois, d'autres travaux rejettent l'hypothèse d'additionnalité, sans pour autant remettre en question l'absence d'effet de *crowding-out*. C'est le cas de l'étude de Martínez-Ros et Corchuelo (2009) qui montre que s'il semble y avoir un effet d'additionnalité pour les grandes entreprises ou les entreprises intervenant dans le secteur des hautes technologies, cet effet n'apparaît pas comme significatif pour les petites et moyennes entreprises. Une explication avancée par les auteurs est que les petites et moyennes entreprises font face à d'autres obstacles à la R&D, rendant plus aléatoire l'effet des dépenses publiques sur les efforts privés en R&D.

Si l'effet d'additionnalité sur les efforts privés de R&D de dispositifs fiscaux visant à réduire leurs coûts n'est pas toujours avéré, la littérature converge toutefois vers l'idée que les incitations fiscales à la R&D n'induisent pas d'effet de *crowding-*

out des dépenses privées.

Pour les mêmes raisons qui font que les travaux sur les effets des subventions sur les dépenses privées de R&D souffrent de limites méthodologiques, ceux portant sur l'effet des crédits d'impôt peuvent être critiqués pour leur stratégie d'identification. Toutefois, ces travaux étant dans l'ensemble plus récents, ils s'appuient davantage sur des méthodes économétriques permettant de mieux contrôler du biais de sélection, comme les techniques de *matching* permettant de créer un contrefactuel « crédible ». De plus, ces travaux n'étaient pas l'hypothèse d'un effet de *crowding-out* et semblent au contraire davantage converger vers l'idée d'une additionnalité des dépenses publiques de R&D sous la forme de crédits d'impôt, et des dépenses privées de R&D. On note une bien moindre variabilité des résultats avec les méthodes et les dispositifs étudiés que dans le cas des subventions à la R&D.

Effets des incitations fiscales à la R&D sur l'innovation et les performances

Si les évaluations des crédits d'impôt portant sur l'effet d'additionnalité des incitations fiscales sont très nombreuses, les travaux sur leur impact en termes d'innovation et de performances des entreprises sont plus rares et beaucoup plus récents. En utilisant des données canadiennes pour l'année 1999, Czarnitzki et al. (2011) trouvent un effet positif des crédits d'impôt sur l'innovation des entreprises bénéficiaires. En créant un groupe de contrôle approprié à l'aide d'un *matching*, ils montrent que, relativement à des entreprises similaires mais ne demandant pas à bénéficier de ce dispositif fiscal, les entreprises bénéficiant du crédit d'impôt mettent sur le marché plus de nouveaux produits, et augmentent leurs ventes grâce à ces derniers. Elles connaissent aussi une probabilité plus grande d'apporter une innovation inédite, au niveau national (canadien) ou mondial. En revanche, l'étude précédemment citée de Bozio et al. (2016) ne conclut pas à un effet significatif sur l'innovation, mesurée ici par le nombre de brevets déposés deux ans après l'entrée en vigueur de la réforme. Les auteurs soulignent toutefois que ce délai est certainement trop court pour observer un effet, et ne permet pas de conclure de façon robuste à une inefficacité du crédit d'impôt sur cet indicateur.

Enfin, Lokshin et Mohnen (2013a) s'intéressent à l'impact sur le salaire des em-

ployés en R&D aux Pays-Bas sur la période 1997-2004 d'un crédit d'impôt sur les dépenses en R&D. Les auteurs trouvent qu'une augmentation de 1 % de la part du salaire financée par le dispositif d'incitation fiscale se traduit par une hausse de 0,2 % du salaire des travailleurs dans le court terme, et 0,24 % à plus long terme.

L'étude de la littérature évaluant les effets des subventions à la R&D sur les efforts privés en R&D et leurs résultats, en termes de capacité à innover et de performances des entreprises, ne laisse pas présager, pour notre étude, de l'effet des aides accordées dans le cadre du PIA-ADEME sur ces indicateurs. En effet, les résultats semblent fortement liés au contexte, au dispositif étudié ainsi qu'à la méthode utilisée.

D'autre part, si les résultats et l'interprétation des études sur les subventions à la R&D portent autant à discussion, tandis que les travaux sur les incitations fiscales à la R&D font davantage l'objet d'un consensus, ceci s'explique probablement par la nature même des subventions, qui impliquent nécessairement une sélection des bénéficiaires. En effet, candidater à une subvention pour un projet et être sélectionné sont autant d'étapes qui invitent à interroger les résultats des études, en raison d'un possible biais d'endogénéité.

CHAPITRE 3

PRÉSENTATION DES DONNÉES ET CONSTITUTION DES BASES D'ANALYSES

La réalisation de cette étude repose sur l'exploitation de plusieurs sources de données, ici décrites. Des données de gestion ont été transmises par le CGI ainsi que par l'ADEME concernant les PIA. Les données du CGI n'ayant servi qu'à des fins de statistiques descriptives, elles ne sont pas présentées dans ce chapitre. Nous les présentons succinctement dans l'Annexe A. Les données brutes de l'ADEME sont présentées ainsi que le traitement qui en a été fait afin de constituer des bases de données exploitables pour l'analyse. Ces données ont également été complétées par des sources de données externes, afin de mieux caractériser notre échantillon tant du point de vue de leur activité de R&D que du point de vue de leurs caractéristiques comptables et financières.

3.1 Description des données sur le PIA-ADEME

L'ADEME nous a transmis l'ensemble de ses fichiers de gestion relatifs au PIA sous format Excel. Ces différents fichiers ont été retraités afin de pouvoir être appariés entre eux pour obtenir une base dont le niveau d'observation est celui du partenaire.

3.1.1 Les différents fichiers transmis par l'ADEME

L'ADEME recense tous les projets pour lesquels un dossier de candidature a été déposé, qu'il ait ou non été sélectionné, dans le fichier « DIA Projets ». Les fichiers « LOCO » ne concernent quant à eux que les projets contractualisés et contiennent au niveau du projet*partenaires des informations très précises sur les différents coûts, les versements de l'aide octroyée et sur les modalités de remboursement. Les différents fichiers couvrent la période 2009 jusqu'au premier semestre 2016.

Les données « DIA Projets »

Le fichier « DIA Projets », dont l'unité d'observation est le projet, contient des informations sur l'ensemble des projets pour lesquels a été déposé un dossier de demande d'aide auprès de l'ADEME. Cette base recense donc notamment des informations sur les projets n'ayant pas été retenus et n'ayant par conséquent pas bénéficié d'aides de l'ADEME. Il est possible d'identifier les projets sélectionnés grâce au numéro d'accord-cadre, un identifiant unique renseigné lorsque le projet est passé en Copil de validation.

Cette base nous renseigne sur les caractéristiques générales du projet déposé : l'action dans laquelle s'inscrit le projet (« Véhicule du futur » ou « Démonstrateurs de la transition écologique »), la durée du projet, sa localisation géographique, le nombre de partenaires impliqués, ainsi que son coût initial et les montants d'aide demandés. Plusieurs variables nous informent également sur le processus de sélection (date de dépôt du dossier, date du Copil de validation, décision du Premier ministre, date de contractualisation), ainsi que sur les montants engagés et la nature de l'aide apportée (subvention et/ou avance remboursable) lorsque le projet est instruit à la suite d'un passage en Copil de sélection. Enfin, les dates et les motifs de suspension sont reportés le cas échéant.

Les données « LOCO »

Les données « LOCO » ne concernent que les projets sélectionnés et contractualisés, et sont directement extraites de l'outil de gestion des PIA de l'ADEME. On y

retrouve tous les projets contractualisés, ainsi que les partenaires de chaque projet. Chaque partenaire au sein d'un projet est identifiable grâce à un code contrat unique. Un même partenaire peut en revanche avoir plusieurs code contrat s'il participe à différents projets financés.

Les données LOCO se présentent sous la forme de dix fichiers excels séparés, chacun apportant des informations sur les différentes phases d'avancement du projet. L'information contenue dans chacun de ces fichiers est résumée dans le tableau 3.1. Les six premiers fichiers portent sur l'ensemble des projets tandis que les quatre derniers ne concernent que les projets ayant bénéficié d'avances remboursables puisqu'ils détaillent les conditions de remboursement.

TABLEAU 3.1 – Description des fichiers LOCO

Nom du fichier	Variables du fichier	Nombre d'obs. (projets*partenaires)
<i>Pour tous les projets</i>		
LOCO 1 Opérations	Accord-cadre ; Programme ; Siret ; date de notification ; durée du projet ; statut du contrat ; coût total du projet pour le partenaire ; dépenses éligibles	1 253
LOCO 2 Mt contract	Type d'aide ; montants engagés par l'ADEME ; dépenses éligibles ; coûts admissibles	1 154
LOCO 3 Financement	Types de cofinancement ; montant des cofinancements ; autofinancement ; montant des opérations	149
LOCO 4 Ech. contract	Dates d'échéances et montants à mandater (contractuels)	1 051
LOCO 5 Ech. paiement	Dates d'échéances et montants à mandater (actualisés)	1 063
LOCO 6 Acteurs	Siret ; secteur NAF ; Ville du siège social	1 243
<i>Uniquement pour les avances remboursables</i>		
LOCO retour 1 Entête	Accord-cadre ; Montant de l'avance et date de versement du solde	358
LOCO retour 2 CG et LOCO retour 3 CD	Dates de début et de fin de remboursement ; montants à rembourser	334
LOCO retour 4 Echéanciers	Dates et montants de remboursements ; montants remboursés	331
LOCO appariée	Accord cadre ; code contrat ; nom du projet ; Siret du partenaire ; nom de l'AAP ; date de notification ; coût du projet ; statut du contrat ; type d'aide reçue ; date d'échéance ; montant des cofinancements ; montant à mandater	1 092

LECTURE : Le fichier LOCO 1 Opérations donne, entre autres, des informations sur les caractéristiques du projets au moment de son dépôt (coût total du projet, dépenses éligibles , etc.).

NOTE : Tous ces fichiers ont également en commun la variable « code contrat », qui est un identifiant unique renseigné pour chaque partenaire et qui permet les appariements entre les différents fichiers.

SOURCE : Fichiers issus de l'outil de gestion « LOCO » de l'ADEME.

Ces bases ont parfois des structures différentes et ont dû être retraitées de manière à ce qu'une ligne corresponde à un code contrat, soit une ligne par couple projet*partenaire. En effet, les codes contrats sont dupliqués dans la plupart des fichiers, en particulier pour les montants contractuels et les échéanciers de versement, distinguant notamment selon le type de montant (subvention ou avance

remboursable). Une fois cette opération réalisée, neuf fichiers ont été appariés sur le code contrat. La base LOCO 6 a été mise de côté puisque plusieurs partenaires partageaient le même code contrat sans que nous puissions identifier la raison de cette anomalie. L'information contenue dans cette base n'est en outre pas cruciale pour l'analyse.

À l'issue de ces différents appariements, nous obtenons une base avec 1 092 observations au niveau projet*partenaire, identifiables par le numéro d'accord cadre et par le code contrat. Une grande majorité des variables contenues dans les fichiers LOCO sont utiles à des fins de gestion comptable des projets mais ne sont pas nécessairement pertinentes du point de vue de l'analyse économique. Nous faisons ainsi le choix de ne garder que les variables essentielles à l'analyse, détaillées dans la dernière ligne du tableau 3.1.

Les données des annexes 4 et 5

La dernière source de données transmise sont des fichiers au format Excel ou Pdf correspondant aux annexes 4 et 5 des dossiers de candidature remplis par les partenaires des projets déposés en réponse à un appel à manifestation d'intérêt ou à un appel à projets. Les entreprises candidates aux aides accordées dans le cadre du PIA-ADEME doivent, entre autres et dans la majorité des cas, compléter deux documents : l'annexe 4 et l'annexe 5 du dossier de candidature¹. Dans la première, les coûts du projet soumis sont détaillés. La seconde se présente le plus souvent sous la forme d'un document Excel à compléter, dans lequel les partenaires renseignent un certain nombre d'informations : nom du partenaire, adresse, numéro de Siret, contact, informations bancaires, coût du projet pour le partenaire et montant d'aide demandé par le partenaire.

L'information de ces fichiers, notamment pour les candidats rejetés, n'est pas présente dans les données de gestion décrites ci-dessus (la base « DIA Projets » ne donnant des informations qu'au niveau du projet) et est nécessaire pour la consti-

1. Il convient de noter que cette exigence ne s'applique pas à l'ensemble des dossiers de candidature. Ainsi, l'ADEME a informé l'équipe de l'IPP que l'annexe 5 ne fait plus partie des dossiers de candidature pour les appels à projets du programme « Véhicule du futur » depuis deux ans. L'ADEME s'est toutefois dotée en 2016 d'un nouvel outil de suivi des projets qui intègre désormais directement les informations des projets déposés au niveau du partenaire.

tution d'un groupe de partenaires témoins à partir des partenaires non contractualisés². Le numéro Siret des partenaires des projets constitue une information indispensable puisqu'il permet l'appariement des données de l'ADEME avec les données externes. En récupérant également dans ces annexes tout un ensemble de caractéristiques sur les partenaires des projets déposés, tels que les coûts et les montants d'aide demandés, cette extraction doit également nous permettre d'améliorer, en théorie, la qualité de l'appariement entre les partenaires de projets ayant été retenus et les partenaires de projets non contractualisés qui leur sont similaires.

Le nombre de fichiers étant relativement conséquent (546 fichiers pour les seules annexes 5), les données ont été extraites de manière automatique en créant un programme d'extraction dit « parseur », qui faisait l'objet de la Tranche conditionnelle 1 du marché. Il est apparu assez rapidement que l'extraction d'informations à partir des annexes 4 serait un processus très coûteux : la structure de ces annexes présente une hétérogénéité très importante, rendant impossible l'automatisation de l'extraction. En revanche, l'exploitation des annexes 5 est faisable puisque les différents formats des fichiers Excel ont pu être regroupés en trois grandes catégories. Les différentes étapes du processus d'extraction sont détaillées dans le *Rapport final de la tranche conditionnelle 1*.

La base construite à partir de l'extraction d'informations des annexes 5 comporte 2 896 observations au niveau partenaire*projet, concernant à la fois des projets acceptés et des projets refusés, les projets décrits dans la base pouvant ainsi être financés comme non-financés. La base constituée contient six variables : nom du partenaire, numéro Siret du partenaire, montants du coût du projet et de l'aide demandée par partenaire, nom du projet et nom du fichier à partir duquel l'information a été extraite.

Si ces informations récupérées à la suite de l'extraction nous servent pour identifier les partenaires de projets déposés mais non retenus afin de constituer un groupe témoin, la qualité des variables extraites, notamment sur le coût du projet et sur le montant de l'aide demandée, est médiocre. D'une part, pour les observations des projets non décrits dans l'outil de gestion LOCO, ces variables sont manquantes

2. Ceci n'est toutefois plus le cas puisque depuis 2016, l'ADEME transmet au CGI les informations sur tous les déposants pour le suivi du programme.

pour une part non négligeable de l'échantillon (tableau 3.2). Ainsi pour 11,7 % des observations, le coût du projet par partenaire n'est pas renseigné. Cette proportion va jusqu'à 16 % pour l'aide demandée par partenaire.

TABLEAU 3.2 – Nombre de valeurs manquantes pour les partenaires rejetés

Variable	Nombre de valeurs manquantes	En % des observations
Numéro Siret du partenaire	249	13,5 %
Coût au niveau du partenaire	216	11,7 %
Aide demandée au niveau du partenaire	297	16,2 %
Coût et aide demandée	208	11,3 %

LECTURE : Pour les partenaires de projets déposés mais non retenus, et donc non présents dans les fichiers de gestion LOCO, l'aide demandée au niveau du partenaire est manquante dans 16,2 % des cas.

SOURCE : Données PIA-ADEME, Annexes 5 des projets déposés.

D'autre part, il s'est avéré que la fiabilité de ces données était relativement faible. En effet, la comparaison du nombre de partenaires par projet, du coût initial et de l'aide demandée par partenaire entre les modalités renseignées dans les annexes et celles indiquées dans DIA Projets révèle d'importantes différences (tableau 3.3). Par exemple, pour 149 projets, le coût du projet renseigné dans DIA Projets est supérieur de 7,9 millions d'euros en moyenne à celui renseigné dans les annexes.

TABLEAU 3.3 – Comparaison de valeurs entre les Annexes et DIA Projets pour les partenaires rejetés

Cas	Nombre de projets concernés	Écart moyen
Nombre de partenaires Annexes > DIA	34	2,3 partenaires
Nombre de partenaires DIA > Annexes	38	3,0 partenaires
Coût du projet Annexes > DIA	137	1,6 M €
Coût du projet DIA > Annexes	149	7,9 M €
Aide demandée Annexes > DIA	140	1,1 M €
Aide demandée DIA > Annexes	123	2,7 M €

LECTURE : L'aide demandée renseignée dans la base DIA Projets est, pour 123 projets, supérieure de 2,7 millions d'euros en moyenne par rapport à ce qui est renseigné dans les annexes 5.

NOTE : Afin de pouvoir être comparés aux montants de DIA Projets, renseignés au niveau du projet, les montants des partenaires d'un même projet dans les Annexes 5 ont été agrégés.

SOURCE : Données PIA-ADEME, Annexes 5 des projets déposés.

Au vu des valeurs manquantes d'une part, et des différences importantes entre

les montants indiqués dans DIA Projets et ceux agrégés au niveau du projet à partir des Annexes 5 d'autre part, l'utilisation de ces variables pour la procédure d'appariement est compromise.

3.1.2 La constitution d'une base ADEME unique

Afin de pouvoir évaluer l'effet des aides de l'ADEME sur les performances des entreprises, il est nécessaire de constituer une base dont l'unité d'observation est le partenaire, ce qui implique de passer de plusieurs bases au niveau projet*partenaire à une base unique au niveau du partenaire. Pour cela, le cas des entreprises ou organismes participant à plusieurs projets doit être pris en compte pour qu'ils n'apparaissent qu'une seule fois dans la base.

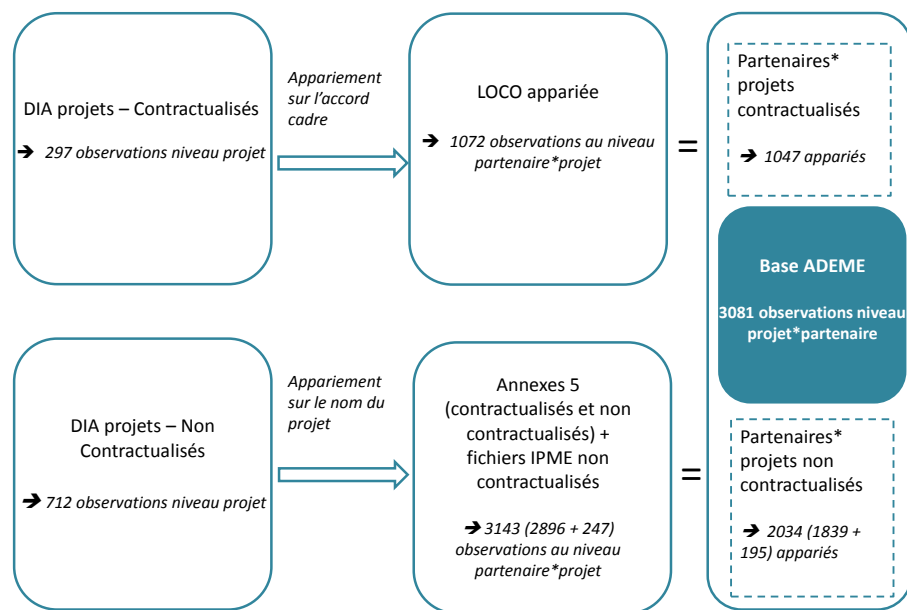
Nous procédons à la constitution de cette base unique en plusieurs étapes. Pour les partenaires contractualisés, nous les identifions et récupérons certaines de leurs caractéristiques en appariant la base « DIA Projets » avec la base LOCO sur le numéro d'accord-cadre³. Pour les partenaires non contractualisés, notre point de départ est également la base « DIA Projets » à laquelle nous ajoutons les informations de la base des annexes 5 en nous servant du nom du projet comme clé d'appariement. Ces appariements entre les différentes bases ont nécessité un travail important sur les variables « accord-cadre » et « nom du projet ». L'information a souvent dû être recherchée par l'ADEME pour les accord-cadres à cause des nombreuses valeurs manquantes, et les noms de projets ont dû être harmonisés entre les différents fichiers.

Les bases intermédiaires obtenues pour les partenaires contractualisés et non contractualisés sont ensuite assemblées pour constituer une seule base, toujours au niveau du projet*partenaire à ce stade. Au total, la base ADEME contient 3 081 projets*partenaires. La Figure 3.1 résume les différentes étapes de la constitution de la base de données ADEME.

La dernière étape consiste à passer d'une base au niveau projet*partenaire à

3. Les annexes 5 qui sont disponibles pour les partenaires contractualisés ne sont pas utilisées d'une part car l'information est redondante avec l'information contenue dans LOCO, et d'autre part car l'information des annexes est moins fiable car elle a été renseignée par les partenaires eux-mêmes.

FIGURE 3.1 – Constitution de la base ADEME



LECTURE : Cette figure illustre la constitution de la base ADEME au niveau projet*partenaire.

une base au niveau du Siret ou du Siren, ce qui implique d'identifier les partenaires participants à plusieurs projets. Nous disposons du numéro de Siret (ou a minima du numéro de Siren) pour 97 % des couples projet*partenaire. Nous supprimons les 3 % pour lesquels le Siret n'est pas renseigné puisqu'aucun appariement avec les bases externes n'est possible dans ces cas. 77,5 % des numéros de Siret de notre base apparaissent une fois uniquement. Pour les 22,5 % restants, nous faisons la moyenne des différentes variables numériques⁴, hormis les montants d'aide perçus qui sont, eux, sommés, au niveau du Siret et gardons la première occurrence des variables non numériques⁵. Après cette opération, notre échantillon est ramené à 1 838 observations si l'on se place au niveau du Siret et à 1 633 si le niveau retenu est celui du Siren.

Ces nombreux fichiers de gestion ont été assemblés en une base unique et à des fins d'exploitation statistique pour la première fois dans le cadre de ce projet.

4. Nombre de partenaires par projet, durée du projet, coût initial du projet, coût retenu au Copil, coût total au niveau du partenaire et montant d'aide demandée.

5. Nom du partenaire, action, type d'aide reçue, année de dépôt du dossier, la note de présélection et le numéro de Siret.

Il est important de souligner qu'un lourd travail de près de six mois a été nécessaire afin de mettre en cohérence l'ensemble des fichiers, chaque manipulation de base ou appariement entre bases ayant soulevé un certain nombre de problèmes. En collaboration avec l'ADEME, nous avons entre autres dû améliorer la qualité des variables servant d'identifiants lors des différents appariements, à savoir les numéros d'accord cadre, les codes contrat, les noms des partenaires ainsi que les numéros de Siret. Ces variables présentaient parfois des problèmes de format qui différaient d'une base à l'autre (fréquent pour le nom du partenaire), pouvaient ne pas avoir le nombre de caractère requis (le numéro de Siret ne comportait pas toujours 14 chiffres), étaient manquantes, ou encore pouvaient éventuellement être mises à jour dans un fichier mais pas dans l'autre. La question des appels à projet se situant dans le champ de l'étude a aussi donné lieu à de nombreuses réactualisations des fichiers sources. Régler ces problèmes a permis de maximiser le nombre d'observations se trouvant dans l'échantillon final⁶.

3.2 Description des bases de données externes

Afin de pouvoir construire nos différentes variables d'intérêt (tableau 3.6) et récupérer des caractéristiques sur notre population d'entreprises, nous avons mobilisé plusieurs bases de données externes que nous décrivons ici : l'Enquête R&D, les déclarations annuelles de données sociales (DADS) et les liasses fiscales FICUS-FARE.

3.2.1 L'Enquête R&D

Description de l'Enquête R&D

L'Enquête annuelle R&D est la principale base de données externes nécessaire à notre étude. Elle contient en effet toutes les variables relatives aux activités de R&D des entreprises qui constituent nos principales variables d'intérêt. Nous disposons de ces données d'enquête sur la période 1992-2013.

6. Ces différents problèmes soulevés feront l'objet de recommandations, afin qu'ils ne se posent plus lors de l'évaluation finale (voir chapitre 6).

L'inconvénient majeur de cette base de données est qu'elle n'est, par définition, pas exhaustive puisqu'il s'agit d'une enquête. En effet, en 2013, seules les entreprises ayant des dépenses de R&D supérieures à 400 000 € ont été interrogées exhaustivement⁷. Les autres entreprises ne sont présentes dans le panel que deux années de suite maximum, le total constituant un échantillon de près de 10 000 entreprises, se répartissant à part égale entre les entreprises sondées chaque année et celles présentes que deux années dans le panel⁸. Nous verrons dans quelle mesure ce point pose problème au moment de l'appariement avec la base ADEME, notamment du point de vue de la représentativité de l'échantillon (voir partie 3.3).

Un autre inconvénient de cette enquête est le décalage qu'il y a entre le moment où une entreprise exerçant des activités de R&D est créée, celui où elle entre dans la base de sondage de l'Enquête R&D et le moment où elle est enquêtée. En effet chaque année, la base de sondage de l'Enquête R&D est complétée par les entreprises repérées comme exerçant des activités de R&D à partir de fichiers administratifs. Sont incluses les entreprises bénéficiaires du crédit d'impôt recherche, les entreprises bénéficiaires du statut de jeune entreprise innovante et celles bénéficiaires de l'aide d'incubateur d'entreprises. Une fois incluses dans la base, seules un peu plus de 10 % d'entre elles seront interrogées l'année suivante. Ainsi, deux ans minimum peuvent s'écouler entre la création d'une entreprise et le moment où elle est enquêtée. Ceci est particulièrement problématique dans notre contexte, puisqu'une bonne partie des entreprises aidées dans le cadre du PIA-ADEME sont des *start-up* qui risquent de ne pas être présentes dans l'Enquête R&D.

L'Enquête R&D collecte des informations très riches sur l'activité de R&D des entreprises. Un premier groupe de variables permet de caractériser la manière dont se partagent les dépenses intérieures de R&D (DIRD) entre les dépenses courantes, qui contiennent principalement les rémunérations des salariés, et les dépenses de capital (terrains, matériel, achat de logiciel, etc.). Il est également possible de savoir comment se répartit la dépense intérieure entre les dépenses de recherche fondamentale, de recherche appliquée et de développement expérimental. Les entreprises

7. Ce seuil vaut pour l'année 2013 et est susceptible de varier d'une enquête à l'autre. Il était par exemple de 350 000 € en 2009.

8. Par exemple en 2013, 6105 entreprises font partie des entreprises sondées exhaustivement, et 5099 ne sont interrogées que deux années de suite.

sont ensuite interrogées sur leurs dépenses extérieures de R&D (DERD), qui correspondent aux travaux de R&D financés par l'entreprise interrogée mais exécutés par un tiers. Il peut s'agir d'un organisme public, d'une entreprise domiciliée en France ou encore d'une entreprise étrangère. Viennent ensuite les questions relatives aux sources de financement de la R&D. Les entreprises reportent les financements publics qu'elles reçoivent (y compris les financements PIA), les financements reçus d'autres entreprises ainsi que le montant d'autofinancement de l'entreprise. Enfin, l'enquête comporte des questions relatives aux effectifs consacrés à la R&D : nombre de chercheurs, de techniciens, d'ouvriers et d'administratifs. Le nombre de brevets déposés est également disponible.

Constitution d'un panel Enquête R&D 2009 et 2013

Afin de pouvoir mettre en œuvre la méthodologie des doubles différences détaillée dans la partie 4, nous constituons un panel cylindré⁹ avec les Enquêtes R&D de 2009 et celle de 2013, qui constituent respectivement nos points de départ et d'arrivée de l'analyse (voir la partie 4 pour plus de précisions sur la méthodologie de l'étude).

En appariant l'Enquête R&D de 2009 (7 966 observations) avec celle de 2013 (9 512 observations) sur le numéro de Siren, nous retrouvons au total 3 498 entreprises présentes les deux années. Pour les raisons évoquées plus haut, un biais était prévisible en faveur des grandes entreprises, puisque seules celles ayant un certain montant de DIRD sont enquêtées exhaustivement. Le tableau 3.4 compare, sur certaines variables de l'Enquête R&D, les moyennes des entreprises présentes une des deux années à celles présentes en 2009 et en 2013. Il apparaît très clairement que les entreprises constituant notre panel cylindré sont des grandes entreprises : leur budget total moyen de R&D est de 8,6 millions d'euros supérieur à celui des autres entreprises, les effectifs consacrés à la R&D sont nettement supérieurs avec près de 50 salariés équivalent temps plein (ETP) de plus. Les écarts sont également considérables pour ce qui est du nombre de brevets déposés ou encore du chiffre

9. Un panel cylindré est un panel qui contient à chaque moment d'observation toutes les entreprises présentes initialement dans la base.

d'affaires. Ceci introduira un biais de sélection au moment de l'appariement avec la base ADEME et posera des questions de représentativité de l'échantillon pour l'analyse finale.

TABLEAU 3.4 – Comparaison des entreprises présentes en 2009 et 2013 dans l'Enquête R&D à celles présentes seulement une année

Variables	Entreprise enquêtée en 2009 OU 2013	Entreprise enquêtée 2009 ET 2013	Différence
Budget total de R&D (milliers d'euros)	772,3	9 405,3	8 633***
Dépenses extérieures de R&D (milliers d'euros)	129,2	2 311,2	2 182***
Dépenses intérieures de R&D (milliers d'euros)	643,1	7 094,1	6 450,9***
Effectifs ETP de R&D	6,7	54,5	47,8***
Nombre de brevets déposés	0,7	7,4	6,6***
Chiffre d'affaires HT (milliers d'euros)	169 300,6	272 702,2	103 401,6*
Chiffre d'affaires à l'export (milliers d'euros)	80 980,1	230 915,1	149 934*
Nombre d'observations	10 482	3 498	

LECTURE : Les entreprises enquêtées uniquement en 2009 ou en 2013 ont un budget total de R&D inférieur de 8,6 millions d'euros à celui des entreprises présentes dans l'échantillon les deux années. Cette différence est significative à 1 %.

NOTE : Les étoiles indiquent la significativité de la différence. *** : significatif à 1 % ; ** : significatif à 5 % ; * significatif à 10 %.

SOURCE : Enquête R&D 2009 et 2013.

3.2.2 Les liasses fiscales des entreprises (fichiers FICUS-FARE)

Les fichiers FICUS-FARE regroupent l'ensemble des déclarations comptables et des informations fiscales des entreprises soumises à l'impôt sur les bénéfices industriels commerciaux et non commerciaux. Ces données administratives sont ainsi exhaustives pour ces entreprises et permettent de contrôler de nombreuses caractéristiques et de construire des indicateurs de performances économiques. Nous disposons des fichiers FICUS pour les années 1996 à 2007 et des fichiers FARE pour les années 2008 à 2013.

Les variables peuvent se regrouper principalement en deux catégories. D'une part, les variables du compte de résultat – qui reflètent l'activité de l'entreprise et permettent de savoir si, à l'issue de l'exercice comptable, l'entreprise s'est enrichie ou appauvrie – contiennent entre autres le total des ventes, le chiffre d'affaires, les subventions d'exploitation, les traitements et salaires, les cotisations sociales et les impôts sur les bénéfices. D'autre part, les variables du bilan comptable permettent

d'appréhender à un instant donné le patrimoine d'une entreprise et la manière dont il est financé. Il comprend des variables telles que les immobilisations corporelles, incorporelles et financières ainsi que le total des actifs et du passif. Plusieurs soldes utiles à l'analyse des performances économiques peuvent être construits à partir de ces variables : l'excédent brut d'exploitation, les investissements ou encore la valeur ajoutée. Enfin, la date de création de l'entreprise est également présente dans cette base.

De la même manière que pour l'Enquête R&D, nous constituons un panel cylindré avec les fichiers FARE. Nous apparions sur le numéro de Siren les données de FARE 2009 (2,9 millions d'observations) avec celles de 2013 (4,1 millions d'observations) et retrouvons 2,1 millions d'entreprises présentes les deux années. Les entreprises présentes uniquement une année sont soit des entreprises créées après 2009, soit des entreprises ayant cessé leur activité avant 2013. Il peut également s'agir d'entreprises ayant changé de numéro de Siren ¹⁰.

3.2.3 Les déclarations annuelles de données sociales (DADS)

La déclaration annuelle de données sociales (DADS) est une formalité administrative qui doit être réalisée chaque année par toute entreprise employant des salariés. Les employeurs sont tenus de renseigner les effectifs de chacun de leurs établissements ainsi que la masse des salaires versés. Ils doivent également indiquer pour chaque salarié tout un ensemble de caractéristiques sur le poste occupé : le montant des rémunérations et le nombre d'heures rémunérées, le type de contrat et le type de poste. Les DADS constituent ainsi une source d'information riche et très complète sur les salaires et l'emploi des entreprises, et permettent d'enrichir les informations déjà contenues dans les autres bases de données.

Ces fichiers sont diffusés sous plusieurs formats : au niveau de l'entreprise (du groupe), de l'établissement, ou au niveau du poste. Pour cette étude, nous utilisons les DADS au niveau du poste, puisque seuls ces fichiers donnent une information

10. Dans le chapitre 6, nous proposons quelques pistes pour mieux documenter les raisons de cette perte d'observations lors des différents appariements.

précise sur les types de postes occupés, grâce à la nomenclature des professions et des catégories socio-professionnelles (PSC-ESE). Les fichiers DADS « poste » sont diffusés sous forme de fichiers régionaux pour lesquels l'information est exhaustive, accompagnés d'un fichier national représentatif qui ne contient qu'un 12^{ème} des observations. Ayant besoin de l'ensemble des entreprises, nous reconstituons une base unique contenant l'ensemble des fichiers régionaux. Nous réalisons cette étape en ne conservant que les variables nécessaires à l'analyse, c'est-à-dire les numéros Siret et Siren, les salaires bruts, l'effectif total et le type de poste, qui utilise la nomenclature PCS-ESE. À partir de cette dernière variable, nous identifions également les emplois R&D regroupés sous les catégories d'ingénieurs, de cadres d'étude et de techniciens en recherche et développement¹¹. Cette variable, bien que beaucoup moins précise que les différentes variables relatives à l'emploi R&D de l'Enquête R&D a l'avantage d'être disponible pour toutes les entreprises et pas uniquement pour les entreprises enquêtées dans le cadre de l'Enquête R&D. Toutes ces variables DADS sont ensuite agrégées au niveau du Siren. Enfin, nous apparions les DADS de 2009 à celles de 2013 sur le numéro de Siren et retrouvons près de 2,3 millions d'entreprises présentes les deux années dans les fichiers.

3.3 Appariement de la base ADEME et des données externes

Nous présentons les deux bases d'analyse constituées dans le cadre de cette étude. Nous présentons dans un premier temps l'appariement entre la base ADEME et les panels FARE et Enquête R&D 2009 et 2013, puis dans un second temps l'appariement entre la base ADEME et les panels FARE et DADS 2009 et 2013.

3.3.1 Appariement ADEME avec les panels FARE et Enquête R&D

11. Plus précisément, nous considérons les catégories 383a, 384a, 385a, 386b, 386c, 388a, 473b, 474b, 475a et 479a comme étant des postes en R&D.

Appariement ADEME et FARE

Nous faisons le choix de commencer par l'appariement entre la base ADEME et le panel FARE 2009 et 2013 afin de récupérer les caractéristiques comptables et fiscales des partenaires qui nous permettront de mieux caractériser les partenaires de notre échantillon que l'on ne retrouve pas au moment de l'appariement avec l'Enquête R&D.

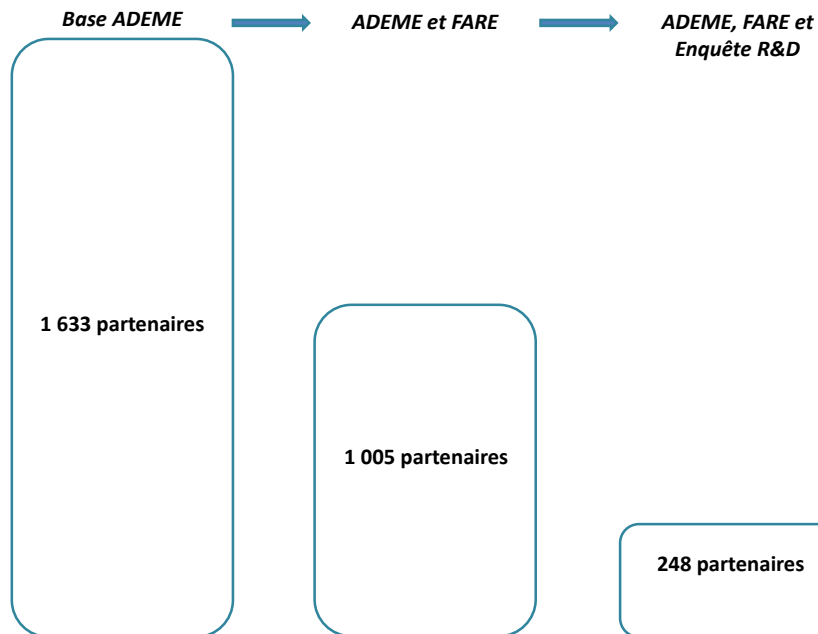
En appariant la base ADEME et le panel FARE sur le numéro de Siren, nous retrouvons 1 005 partenaires sur les 1 633 initiaux. Déterminer les raisons pour lesquelles 628 partenaires de la base ADEME ne se retrouvent pas dans la base FARE demanderait un travail d'investigation supplémentaire qui n'est pas l'objet de cette étude. En récupérant les numéros Siren de ces partenaires, il serait possible de les retrouver dans le répertoire SIRENE des entreprises ou dans les bases FARE ultérieures, et de savoir si ces entreprises ne se retrouvent pas dans notre panel FARE car elles ont été créées après 2009 ou parce qu'elles ont cessé leur activité avant 2013. Une telle analyse pourrait être intéressante notamment pour voir si les entreprises financées par le PIA-ADEME présentent un taux de cessation d'activité anormalement plus élevé que les autres entreprises. Toutefois, la raison principale à cette perte d'observations semble être le fait que, d'après l'ADEME, près de 80 % des numéros de Siren renseignés initialement dans les bases ADEME sont susceptibles d'avoir changé et de ne pas avoir été mis à jour. Ce problème fait l'objet d'une recommandation dans le chapitre 6.

3.3.1.1 Appariement ADEME-FARE avec le panel Enquête R&D

Nous apparions ensuite ces 1 005 partenaires aux observations du panel 2009 et 2013 de l'Enquête R&D sur le numéro de Siren : seuls 248 partenaires se retrouvent dans notre base à l'issue de ce triple appariement. Ces 248 partenaires sont très susceptibles de présenter des caractéristiques spécifiques puisque comme nous l'avons montré plus haut, les partenaires présents dans l'Enquête R&D en 2009 et 2013 sont principalement des grandes entreprises.

Nous répétons ici l'exercice de comparaison de moyennes des partenaires se retrouvant dans les trois bases et dans les bases ADEME et FARE uniquement (ta-

FIGURE 3.2 – Appariement des bases ADEME, FARE et R&D 2009 et 2013



LECTURE : 1 005 entreprises sont présentes à la fois dans la base ADEME et dans les données FARE 2009 et 2013.

bleau 3.5). Le constat est similaire : les partenaires de l'échantillon d'analyse sont des établissements de taille significativement plus importante. Les projets financés de ces partenaires sont des projets plus coûteux (+ 17 millions d'euros) et les montants d'aide attribués par l'ADEME sont plus conséquents (+ 3,6 millions d'euros). Du point de vue des caractéristiques des partenaires, ils mobilisent en moyenne plus de main d'œuvre pour leurs activités et produisent beaucoup plus de valeur. Ceci pose un problème de représentativité de l'échantillon d'analyse : les partenaires de la base résultant de l'appariement ADEME-FARE-Enquête R&D ne sont pas représentatifs de l'ensemble des partenaires déposant un dossier de candidature pour le PIA-ADEME.

Afin d'améliorer l'appariement entre la base ADEME-FARE et l'Enquête R&D et pallier ce problème de représentativité de l'échantillon d'analyse, nous avons envisagé de nous inspirer du travail de Béllego et Dortet-Bernadet (2014) qui, pour construire leur échantillon, utilisent le fait que les petites entreprises sont inter-

TABLEAU 3.5 – Comparaison des partenaires ADEME-FARE présents dans l'Enquête R&D en 2009 et 2013 aux autres

Variables	Entreprises ADEME-FARE non appariées (avec l'Enquête R&D)	Entreprises ADEME-FARE appariées (avec l'Enquête R&D)	Différence
Coût initial du projet (M€)	27	28,7	1,7
Coût validé en Copil(M€)	15,2	33	17,8***
Aide demandée (M€)	1,4	2,3	0,7*
Montant attribué (M€)	1,8	5,4	3,6***
Nombre de salariés	200,7	2 261,5	2 060,8 ***
Valeur ajoutée HT (m€)	19 863,7	251 020,2	231 156,5***
Chiffre d'affaires à l'export (m€)	23 234,9	235 549,5	212 314,5***
Nombre d'observations	757	248	

LECTURE : Les partenaires qui sont présents dans la base ADEME-FARE et que l'on retrouve dans l'Enquête R&D en 2009 et en 2013 déposent des projets dont le coût initial est de 1,7 millions d'euros supérieurs à ceux déposés par les partenaires uniquement présents dans ADEME-FARE. Cette différence n'est toutefois pas significative.

NOTE : Ce tableau compare les moyennes des partenaires ADEME-FARE présents dans l'Enquête R&D à celles des partenaires ADEME-FARE uniquement. Les étoiles indiquent la significativité de la différence. *** : significatif à 1 % ; ** : significatif à 5 % ; * significatif à 10 %.

SOURCE : Base ADEME ; données FARE 2009 et 2013 ; Enquête R&D 2009 et 2013.

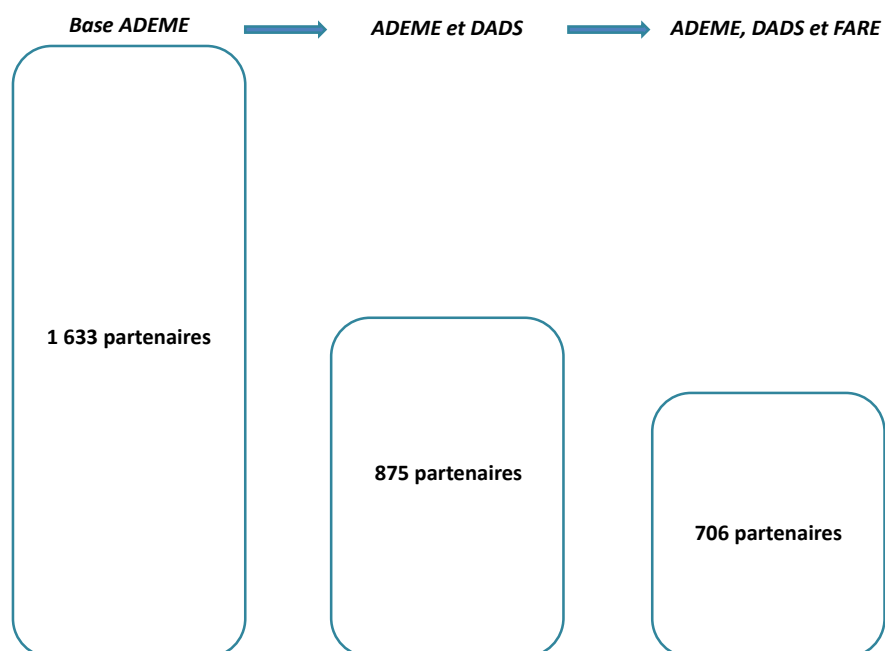
rogées deux années consécutives dans l'Enquête R&D. Plutôt que de contraindre les petites entreprises à être présentes en 2009 et en 2013 dans l'Enquête R&D, l'idée était d'apparier les enquêtes d'une année sur l'autre, et d'identifier les effets des aides du PIA-ADEME sur la période 2009-2013 comme la somme des effets sur deux années consécutives (calcul par chaînage des évolutions). Cette proposition a néanmoins été rapidement écartée puisque l'essentiel de la perte d'observations se fait lors du premier appariement, c'est-à-dire entre la base ADEME et l'Enquête R&D 2009. L'ajout d'une année supplémentaire de l'Enquête R&D ne conduit en réalité pas à perdre beaucoup d'observations. Ce mauvais appariement initial est principalement dû au décalage entre la création d'une entreprise et le moment où celle-ci apparaît dans la base de sondage de l'Enquête R&D.

3.3.2 Appariement ADEME avec les panels FARE et DADS

Afin de remédier au problème de représentativité de l'échantillon évoqué précédemment, nous proposons également de mener nos analyses sur l'échantillon ADEME-FARE-DADS. Les fichiers FARE et DADS étant des données exhaustives, le nombre d'observations présentes dans les trois bases est nécessairement plus élevé

et l'échantillon moins sujet aux problèmes de sélection. Ainsi, à l'issue de l'appariement entre la base ADEME et le panel DADS, nous retrouvons 1 750 observations, soit 875 partenaires par année (figure 3.3). En appariant avec les données FARE, notre échantillon d'analyse est constitué de 706 partenaires, soit près de trois fois plus qu'avec l'échantillon précédent. Les raisons de la perte d'observations résultant des différents appariements sont les mêmes que celles évoquées à la partie 3.3.1.

FIGURE 3.3 – Appariement des bases ADEME, FARE et DADS 2009 et 2013



LECTURE : Lors de l'appariement entre les bases ADEME et DADS 875 observations sont présentes dans les deux bases.

Les données sont en revanche moins précises pour nos variables d'intérêt. Il existe ainsi un arbitrage entre la qualité de l'échantillon et la précision des informations disponibles. En effet les dépenses de R&D sont approchées par les investissements incorporels disponibles dans les données FARE, et les emplois en R&D le sont grâce aux types de postes ayant été identifiés comme des emplois R&D dans les DADS (partie 3.2.3). Tout un ensemble d'indicateurs, tels que le nombre de brevets déposés ou encore le détails des dépenses de R&D, ne sont pas disponibles. Le

tableau 3.6 reporte les différents indicateurs disponibles dans l'une et l'autre des bases.

TABLEAU 3.6 – Variables d'intérêt disponibles dans les bases constituées

Type d'indicateur	Base ADEME-FARE-RD	Base ADEME-FARE-DADS
Dépenses de R&D	Investissement en immobilisation incorporelles (FARE) ; dépenses intérieures de R&D (salaires, matériel, etc.) (Enquête R&D) ; dépenses extérieures de R&D (Enquête R&D)	Investissement en immobilisation incorporelles (FARE)
Emploi en R&D	Effectif total (FARE) ; effectif R&D (Enquête R&D) ; Nombre de chercheurs, de techniciens, d'ouvriers et d'administratifs en R&D(Enquête R&D)	Effectif total, effectif d'emploi R&D selon la nomenclature PCS-ESE (DADS)
Performance économique	Valeur ajoutée ; marge commerciale ; excédent brut d'exploitation ; dividendes versés ; chiffre d'affaires ; chiffres d'affaires à l'export (variables FARE)	Valeur ajoutée ; marge commerciale ; excédent brut d'exploitation ; dividendes versés ; chiffre d'affaires ; chiffres d'affaires à l'export (variables FARE)
Produit de la R&D	Production totale (FARE) ; nombre de brevets (Enquête R&D) ; innovation de produit et innovation de procédé (Enquête R&D)	Production totale

LECTURE : Le seul indicateur disponible pour quantifier le produit de la R&D dans la base ADEME-FARE-DADS est la production totale.

NOTES : La base dont est issu l'indicateur est précisé entre parenthèses.

CHAPITRE 4

REVUE DES MÉTHODOLOGIES UTILISÉES DANS LA LITTÉRATURE

Dans un premier temps, ce chapitre vise à présenter les conditions idéales dans lesquelles une évaluation causale des effets des aides accordées par l'ADEME devrait être menée (partie 4.1). Au vu des contraintes d'une telle analyse, nous verrons, dans un second temps, comment les autres études sont parvenues à surmonter les difficultés méthodologiques inhérentes à l'identification d'un effet causal des aides à l'innovation (partie 4.2).

4.1 Le cadre d'une évaluation idéale

L'évaluation rigoureuse d'une politique publique requiert principalement deux éléments : trouver une bonne manière d'identifier un effet causal et disposer d'informations suffisantes pour pouvoir en mesurer l'effet sur les variables d'intérêt. Nous abordons dans cette partie les conditions nécessaires à l'évaluation et rappelons les contraintes de la présente évaluation.

4.1.1 Question de recherche et identification causale

La question de recherche à laquelle ce travail doit parvenir à répondre est celle des effets des aides, accordées par l'ADEME dans le cadre du PIA, sur différents

indicateurs. Les variables d'intérêt considérées ici sont principalement les dépenses d'investissement en R&D, les effectifs affectés à la R&D, les dépôts de brevets ou encore les performances économiques des entreprises. Une des questions de second ordre à laquelle il faudra également répondre est celle de la substitution entre les dépenses publiques et privées de R&D : en effet, celle-ci conditionne l'effet des aides financières à la R&D sur les dépenses privées.

Une fois la question à laquelle l'évaluation doit répondre définie, se pose alors celle de l'identification causale de l'effet des aides accordées à des projets de R&D sur ces indicateurs. Idéalement, il faudrait pouvoir comparer la situation actuelle des entreprises ayant bénéficié des aides accordées dans le cadre du PIA à ce qui se serait passé en l'absence du dispositif. Cependant cette situation, qualifiée de « contrefactuelle », est par définition inobservable. Tout l'enjeu de l'évaluation consiste à reconstruire un contrefactuel crédible (groupe de contrôle), en identifiant un groupe d'entreprises n'ayant pas bénéficié d'aides mais ayant des caractéristiques similaires à celles en ayant bénéficié (groupe traité).

Un protocole d'évaluation idéal consisterait à tirer au sort, parmi un large échantillon de projets déposés, la moitié des projets qui recevrait alors une aide de l'ADEME. La loi des grands nombres et le tirage aléatoire garantissent que les deux groupes aient des caractéristiques comparables en moyenne *avant* l'attribution de l'aide. Afin de mesurer l'effet des aides, il suffirait alors de comparer les variables d'intérêt entre les deux groupes *après* l'octroi d'une subvention aux entreprises du groupe traité. Les différences constatées entre les deux groupes après l'attribution de l'aide peuvent alors être directement interprétées comme un effet causal des aides accordées.

Si un tel cadre expérimental présente l'avantage de réduire tout risque de biais et permettrait d'évaluer de manière rigoureuse l'effet d'aides publiques sur les variables d'intérêt, il est bien souvent difficile de le mettre en place. L'utilisation d'expériences dites « naturelles » permet toutefois de se rapprocher de cet idéal et, sous certaines hypothèses, d'identifier un effet causal. Par expérience naturelle, nous entendons tout changement ou variation des dispositifs d'aide à la R&D qui soit exogène et qui ait affecté une partie seulement des entreprises, tel un changement

législatif ou l'existence d'une discontinuité dans les critères d'attribution de l'aide. Nous verrons dans la revue de littérature sur les méthodes d'évaluation utilisées (partie 4.2) que certaines études y ont recours.

4.1.2 Les contraintes de l'évaluation

Les contraintes auxquelles nous faisons face pour cette évaluation sont de deux types. D'une part, le processus de sélection des projets bénéficiant de l'aide de l'ADEME, la simultanéité des dépenses publiques et privées de R&D ainsi qu'un potentiel effet d'anticipation rendent difficile l'*identification* d'un effet causal, ce qui nécessite de s'éloigner du cadre idéal d'évaluation précédemment décrit. D'autre part, les informations disponibles ainsi que les externalités liées à la R&D compliquent la *mesure* de l'effet causal. Enfin, de manière générale, choisir un bon indicateur qui soit représentatif des efforts de R&D est relativement complexe et fait débat.

4.1.2.1 Des difficultés d'identification

Comme nous venons de le préciser, une évaluation idéale nécessiterait que les entreprises bénéficiaires d'aides le soient de manière totalement aléatoire afin de pouvoir utiliser les non bénéficiaires comme groupe de contrôle valide. Or, le mode de sélection des aides accordées dans le cadre du PIA-ADEME n'est pas aléatoire, mais dépend au contraire de la qualité des projets candidats et des caractéristiques intrinsèques des entreprises. En effet, les projets considérés comme les plus prometteurs sont sélectionnés afin de bénéficier d'une aide publique. En réalité, il y a même une double sélection : une auto-sélection des entreprises qui vont choisir de candidater ou non, et une sélection des projets par les autorités. Dans de telles conditions, utiliser les entreprises non bénéficiaires comme groupe de contrôle sans recourir à des méthodes statistiques adaptées ne peut constituer une approximation satisfaisante de la situation contrefactuelle à cause d'un biais de sélection : les différences dans les variables d'intérêt entre bénéficiaires et non-bénéficiaires pourront en effet être attribuables à des différences intrinsèques entre les entreprises.

Au-delà du fait que les entreprises bénéficiant des dispositifs d'aide financière à la R&D diffèrent de manière structurelle des entreprises n'en bénéficiant pas, au moins deux autres phénomènes – communs à toutes les études visant à évaluer des aides publiques à la R&D – peuvent compliquer l'identification d'un effet causal. Comme le notent David et al. (2000), les décisions de dépenses de R&D à la fois publiques et privées peuvent être prises de manière simultanée en répondant à une même variable inobservable, telle que les opportunités technologiques. Ainsi, si les pouvoirs publics et les entreprises privées décident en même temps d'investir en R&D car ils perçoivent un fort potentiel sur un pan particulier de l'industrie, l'effet spécifique de l'aide publique sera beaucoup plus difficile à identifier, à cause de ce biais de simultanéité. À l'heure de la transition écologique, les opportunités technologiques sont nombreuses dans le domaine de la R&D environnementale et il est très probable que ce biais de simultanéité soit présent dans le cadre de notre étude, et doive par conséquent être traité avec une attention particulière.

Le second phénomène venant compliquer l'identification de l'effet causal des aides est un effet d'anticipation de l'aide publique. En effet, les dépenses privées de R&D peuvent être déterminées en amont par les aides publiques dont bénéficie l'entreprise, et ce non pas seulement par stratégie (effet d'aubaine), par un effet de signal auprès des financeurs ou par une augmentation des rendements de la R&D, mais aussi par anticipation de la subvention : une entreprise peut investir en R&D pour mieux pouvoir prétendre à une subvention ou à la participation à un *consortium* subventionné (Lichtenberg, 1984; David et al., 2000). L'effet anticipé généré par l'aide publique, qui rend également compte de l'effet causal de l'aide, est relativement difficile à identifier. Ce biais d'anticipation semble également susceptible d'exister dans le contexte de notre étude. En effet, depuis 2013, les entreprises candidates aux aides de l'ADEME peuvent inclure des dépenses déjà engagées en R&D dans les coûts estimés de leurs projets¹. Cet aspect rétroactif de l'aide peut ainsi inciter les entreprises à augmenter leurs dépenses de R&D.

1. Ces dépenses déjà engagées sont déclarées comme « dépenses éligibles ».

4.1.2.2 Des difficultés de mesure de l'effet causal

Au-delà de cette difficulté fondamentale qu'est l'identification de l'effet des aides, la mesure même de cet effet apparaît également complexe. Premièrement, la taille de l'échantillon final constitué dont nous disposons pour mener l'analyse est relativement faible. En effet, le nombre de partenaires de notre échantillon varie entre 248 et 706, ce qui implique que la puissance statistique, c'est-à-dire notre capacité à détecter un effet du dispositif, sera nécessairement faible car elle dépend directement du nombre d'observations : plus le nombre d'observations est élevé, plus l'évaluateur est en mesure de détecter un effet, même de faible ampleur. Dans le cas présent, les effets devront être suffisamment importants pour pouvoir être détectés de façon significative.

Enfin, le fait que les aides à la R&D soient susceptibles de générer des externalités positives, en bénéficiant également aux autres entreprises du secteur, rend compliquée l'interprétation de l'effet causal. Ces externalités sont en effet difficilement quantifiables, et peuvent induire en erreur lors de l'interprétation des résultats, comme le notent Klette et al. (2000). En effet, si les externalités positives sont un objectif poursuivi par les dispositifs d'aide à la R&D, l'absence de différences significatives de performance entre entreprises bénéficiaires et non-bénéficiaires peut être interprétée à la fois comme une réussite ou un échec du programme d'aide. Elle pourra en effet être le résultat d'externalités positives importantes – la performance de l'ensemble des entreprises ayant été améliorée –, ou bien être au contraire la conséquence d'un échec à stimuler la production de connaissances ou la performance des entreprises aidées, dont les résultats demeurent alors similaires à ceux des entreprises non-subventionnées.

4.1.2.3 Les difficultés à choisir un bon indicateur de R&D

Enfin, une dernière difficulté est celle de la mesure de la performance des entreprises en termes d'innovation. On distingue plusieurs indicateurs utilisés dans la littérature que l'on peut regrouper en deux catégories : les intrants de la R&D (*input*), tels que les dépenses privées en R&D ou le nombre d'employés en R&D ; et le produit de la R&D (*output*), comme le nombre de brevets déposés ou accordés, le

nombre de citations de brevets, le nombre de nouveaux produits mis sur le marché ou encore le chiffre d'affaires en produits innovants². Alors que les études utilisent généralement un unique indicateur, certains travaux proposent de les regrouper au sein d'un indicateur composite, plus complet, qui englobe les différentes étapes du processus de R&D (Mohnen et Dagenais, 2002; Hagedoorn et Cloudt, 2003; Carayannis et Provan, 2008).

L'utilisation d'un indicateur unique est souvent critiquée puisqu'elle ne permet de rendre compte que d'une partie spécifique de l'activité d'innovation des entreprises, qui doit plutôt être pensée comme un système global allant des dépenses en R&D à la commercialisation d'un produit fini. Par ailleurs, certains indicateurs sont plus pertinents pour les grandes entreprises que pour les petites et moyennes entreprises. Ainsi, pour Carayannis et Provan (2008), l'utilisation du nombre de brevets déposés comme mesure de l'innovation est *a priori* moins pertinente pour les petites et moyennes entreprises que pour les entreprises de plus grosse taille. Bien que Griliches (1998) considère également que les brevets constituent un indicateur imparfait de la production de connaissances d'une entreprise, notamment parce que les brevets n'ont pas tous la même valeur et que toutes les innovations ne sont pas brevetables, il conclut néanmoins que les brevets donnent une certaine mesure de la production de connaissances, aussi bien pour les petites que pour les grandes entreprises. En effet, sa revue de littérature montre l'existence d'une relation très forte entre les dépenses de R&D et l'obtention d'un brevet, suggérant que les brevets sont un bon indicateur à la fois des *inputs* et des *outputs* en R&D.

Enfin, en comparant tout un ensemble d'indicateurs entre eux³, Hagedoorn et Cloudt (2003) montrent également que, dans le secteur des hautes technologies, ces indicateurs se recoupent largement et capturent ainsi la même variable latente qui est la capacité à innover. Il n'y a donc *a priori* pas de préférence à avoir dans le choix des indicateurs.

2. Ce dernier indicateur est relativement peu utilisé dans la littérature mais a l'originalité d'être à l'intersection des indicateurs d'*output* de la R&D et des indicateurs de performances économiques des entreprises. Il est notamment proposé par Negassi et Hellier (2011).

3. Les différents indicateurs utilisés sont : les intrants en R&D, le nombre de brevets, les citations de brevets, l'annonce de nouveaux produits et un indicateur synthétisant ces quatre mesures.

4.2 Revue des méthodes des évaluations réalisées

Il s'agit, dans cette partie, de passer en revue les précédentes évaluations causales des aides à la R&D. Dans la littérature économique, les évaluations des aides divergent principalement selon trois dimensions : le niveau d'analyse, les données utilisées ainsi que les méthodes mises en œuvre afin d'identifier un effet causal.

4.2.1 Les niveaux d'analyse et les types de données

Différents niveaux d'analyse

L'étude de la littérature quantitative fait émerger trois niveaux d'analyse différents : l'entreprise (ou parfois la chaîne de production), le secteur d'activité et enfin, le pays⁴. Selon le niveau d'analyse retenu, les problèmes de biais de sélection ou de simultanéité sont plus ou moins présents.

Le niveau d'analyse qui semble être le plus pertinent pour l'étude des effets des aides à la R&D sur les différentes mesures de l'innovation est celui de l'entreprise, puisqu'il s'agit de l'entité d'intérêt dont on cherche à analyser les comportements en matière de R&D (García-Quevedo, 2004). Cependant, l'attribution de l'aide est rarement exogène puisque les entreprises sont sélectionnées en fonction de leurs caractéristiques. Selon les dispositifs étudiés, ce sont les projets les plus prometteurs en termes d'innovation qui peuvent bénéficier d'un soutien public ou, au contraire, les entreprises les plus en difficulté. Une analyse menée à ce niveau devra prendre soin de traiter explicitement les possibles biais de sélection et de simultanéité qui peuvent biaiser les estimations de l'effet des aides à la hausse ou à la baisse (Lichtenberg, 1984; David et al., 2000).

Au niveau du secteur d'activité, l'hypothèse d'exogénéité dans l'attribution de l'aide semble davantage crédible, bien que le soutien public à la R&D puisse également cibler des industries particulièrement intensives en R&D, ou au contraire

4. À notre connaissance, aucune étude quantitative n'utilise le *consortium* comme niveau d'analyse. Les études qui cherchent à analyser l'effet d'appartenir à un *consortium* de recherche ou non sont menées au niveau de l'entreprise. Une variable indicatrice est ajoutée dans les régressions pour prendre en compte le fait d'appartenir ou non à un *consortium*. Voir par exemple Irwin et Klenow (1996); Czarnitzki et Fier (2003); Czarnitzki et al. (2007).

des secteurs industriels rencontrant des difficultés (Pottelsberghe et Capron, 1997). En revanche, une des principales limites d'une analyse au niveau du secteur est que l'estimation d'un effet causal des aides peut être biaisée par les différentes opportunités technologiques entre secteurs, les rendant ainsi difficilement comparables entre eux (García-Quevedo, 2004). Par exemple, en ce qui concerne la question de la substitution entre la recherche publique et la recherche privée, la réponse semble différer selon le secteur considéré. Ainsi, Guellec et van Pottelsberghe (1997) montrent l'existence d'un effet de substitution dans le domaine de la R&D militaire, pouvant s'expliquer par la rareté des spécialistes et leurs salaires particulièrement élevés. En revanche, dans d'autres secteurs, de tels effets ne semblent pas être à l'œuvre.

Enfin, au niveau du pays, l'hypothèse d'exogénéité est plus acceptable, puisque la mise en place d'aides à la R&D relève directement d'une décision politique. Certaines analyses ont alors consisté à comparer les indicateurs de performances en R&D⁵ entre pays et à attribuer les différences constatées aux politiques nationales mises en place. Ces travaux macro-économiques présentent cependant de nombreuses limites dont la première est la faible comparabilité des pays entre eux. Tout comme entre secteurs, il existe des différences technologiques importantes entre les pays pouvant biaiser les estimations.

Comme le souligne García-Quevedo (2004), en s'intéressant à la question d'un effet de substitution entre la R&D publique et privée, le niveau d'analyse choisi peut conduire à des conclusions et à des recommandations de politiques publiques différentes. Les analyses menées au niveau de l'entreprise semblent davantage conclure à un effet de substitution quand celles conduites à un niveau plus agrégé tendent plutôt à démontrer une complémentarité entre la R&D publique et privée.

Différents types de données mobilisées

On peut principalement distinguer deux types de données utilisées dans les évaluations des aides publiques à la R&D : des données en coupe (*cross-sections*), qui

5. Il s'agit principalement des dépenses privées de R&D, voir les études listées dans le tableau 4 de David et al. (2000).

consistent en un échantillon observé à un instant donné, et des données de panel, formées par un même échantillon observé à plusieurs périodes.

L'utilisation de données de panel présente de nombreux avantages par rapport aux simples données en coupe, notamment parce que les techniques économétriques sur données de panel permettent de mieux gérer les différents problèmes de biais en intégrant des « effets fixes » au niveau des unités d'observation. Ces données offrent ainsi la possibilité de réduire le biais de sélection lié aux caractéristiques inobservables des entreprises, des secteurs ou encore des pays lorsque ces caractéristiques sont fixes au cours du temps. Elles permettent également d'introduire une dimension dynamique au modèle économétrique, en ajoutant par exemple les dépenses privées de R&D des années antérieures comme variable explicative des dépenses actuelles de dépenses privées de R&D. L'ampleur et la significativité des résultats semblent d'ailleurs changer lorsque l'on introduit un caractère dynamique au modèle (Pottelsberghe et Capron, 1997). Enfin, les données de panel permettent de détecter des effets des aides à la R&D qui ont lieu à plus long terme, tel que le dépôt de brevet ou la commercialisation de nouveaux produits, et d'appréhender la persistance des effets dans le temps.

4.2.2 Les différentes méthodes utilisées dans la littérature

Nous présentons dans cette sous-partie les méthodes utilisées dans les études menées au niveau de l'entreprise, qui est le niveau d'analyse retenu pour la présente étude⁶.

Les régressions linéaires classiques

Les premiers travaux ayant essayé de quantifier la relation entre les dépenses publiques et privées de R&D ont eu recours à la méthode des moindres carrés ordinaires, en régressant un indicateur des dépenses privées de R&D sur une mesure de la dépense publique. Plus précisément, ces travaux estiment une équation de la

6. Pour des revues très complètes des méthodologies utilisées aussi bien au niveau de l'entreprise qu'aux autres niveaux d'analyse mentionnés plus haut, voir David et al. (2000), Klette et al. (2000) ou Cerulli (2010). Pour aller plus loin sur les aspects techniques des méthodologies présentées, se référer à Givord (2010), Cameron et Trivedi (2005) et Angrist et Pischke (2008)

forme suivante :

$$Y_i = \alpha + \beta_1 D_i + \beta_2 X_i + \epsilon_i \quad (4.1)$$

avec :

- Y_i la variable que l'on cherche à expliquer, pour l'entreprise i , par exemple les dépenses privées de R&D ;
- D_i la variable explicative principale représentant les aides publiques de R&D reçues par l'entreprise i ;
- X_i un vecteur de variables explicatives, capturant principalement les caractéristiques spécifiques de l'entreprise ;
- β_j les coefficients associés à ces variables, que l'estimation vient déterminer ;
- ϵ_i le résidu, c'est-à-dire ce qui, de Y_i , n'est pas expliqué par les variables incluses dans la régression. Autrement dit, ce terme regroupe les variables « inobservables », qui ne figurent pas parmi les variables explicatives.

Hamberg (1968) est le premier à avoir utilisé une régression linéaire pour répondre à cette question de recherche. En utilisant des données en coupe d'entreprises américaines, il définit huit groupes d'entreprises appartenant à des industries similaires et conduit des régressions séparément pour chacun des groupes. L'avantage de procéder ainsi est de limiter le biais de variables omises lié aux différences d'opportunités technologiques entre industries, puisque l'on peut plus raisonnablement faire l'hypothèse qu'au sein des huit groupes définis, ces opportunités sont les mêmes. Dans la même lignée, Howe et McFetridge (1976) étudient l'impact des subventions à la R&D au Canada sur les dépenses privées de R&D (voir également les études de Higgins et Link (1981), Link (1982) et de Scott (1984)).

Une des hypothèses à vérifier pour que les estimations d'un effet causal par des régressions linéaires soient valides est celle de l'exogénéité des variables explicatives. En d'autres termes, les variables explicatives sont exogènes si elles ne sont pas corrélées aux variables inobservables représentées par ϵ_i dans l'équation 4.1. Or, il est très peu probable que cette hypothèse soit validée dans notre cas, puisque

cette spécification ne permet pas de contrôler pour les caractéristiques inobservables des entreprises, telles que le potentiel d'innovation par exemple. Ce dernier est à la fois corrélé au fait d'être sélectionné pour recevoir des aides ou au montant des aides reçu, ainsi qu'à la variable que l'on cherche à expliquer Y_i . Dans ce cas, on attribuera au fait de bénéficier d'aides un effet en réalité dû à la performance intrinsèque de l'entreprise en matière d'innovation. Ainsi, l'effet de bénéficier d'aides sur la dépense de R&D des entreprises est dans ce cas mal identifié à cause d'un biais de sélection et d'un biais de simultanéité remettant directement en cause l'hypothèse d'exogénéité des variables explicatives.

Les régressions avec variables instrumentales

Afin de corriger ce biais de sélection, plusieurs travaux ont par la suite eu recours à l'utilisation de variables instrumentales. Les aides publiques D_i reçues par l'entreprise sont alors explicitement considérées comme endogènes. L'utilisation de la méthode des variables instrumentales implique de trouver une variable appelée instrument (Z_i) qui soit corrélée aux aides publiques D_i mais qui ne détermine pas directement la variables expliquée Y_i . Autrement dit, l'instrument Z_i influence la variable dépendante Y_i uniquement par la variable représentant les aides publiques. L'estimation se fait également avec la méthode des moindres carrés mais en deux étapes. Premièrement, on estime l'équation qui détermine les aides publiques à la R&D (équation 4.2), puis, dans un second temps, on estime une équation similaire à 4.1 à un détail près : D_i est remplacé par sa valeur prédite calculée à l'étape précédente, \hat{D}_i (équation 4.3).

$$D_i = \delta_0 + \delta_1 Z_i + \mu \quad (4.2)$$

$$Y_i = \alpha + \beta_1 \hat{D}_i + \beta_2 X_i + \epsilon_i \quad (4.3)$$

Lichtenberg (1988) et Wallsten (2000), dans un contexte relativement similaire au nôtre puisqu'ils disposent d'un échantillon composé des entreprises retenues et rejetées pour bénéficier d'un programme de subventions publiques, ainsi que les

entreprises éligibles n'ayant pas postulé, instrumentent par exemple le montant des aides dont bénéficie une entreprise par le montant des contrats publics ou des subventions auxquels l'entreprise peut potentiellement prétendre. Bien que ces travaux représentent une avancée en traitant explicitement une partie du biais de sélection, l'endogénéité liée au fait que les aides sont susceptibles d'être accordées en fonction d'opportunités technologiques qui diffèrent d'une industrie à l'autre n'est, quant à elle, pas prise en compte. Conscient de cette limite, Einiö (2014) utilise la variation géographique de l'aide publique à la R&D potentiellement disponible comme instrument. En effet, les régions en Finlande octroient des aides à la R&D en fonction des fonds qu'elles reçoivent du Fonds européen de développement régional (FEDER), les régions les moins densément peuplées recevant le plus de soutien du FEDER. L'avantage de cette approche est que le montant maximum auquel une entreprise peut prétendre dépend d'un critère de population et non d'un critère lié directement à des considérations économiques, comme des niveaux faibles de R&D ou un manque de dynamisme économique, ce qui réduit le risque de biais de sélection.

De manière générale, les limites de l'évaluation par la méthode des variables instrumentales résident dans le fait que lorsqu'un unique instrument est utilisé, l'hypothèse d'exogénéité de ce dernier ne peut être vérifiée⁷. La crédibilité de l'évaluation dépend par conséquent très fortement de la qualité de l'instrument utilisé.

L'évaluation par contrefactuel

Une autre approche de l'évaluation d'une politique publique est celle de la comparaison entre la situation actuelle et celle qui aurait prévalu en l'absence de la politique, que nous avons déjà évoquée dans la partie 4.1. Il existe plusieurs méthodes pour reconstruire cette situation contrefactuelle. Nous les présentons ici.

Les méthodes d'appariement (*matching methods*) : de nombreux travaux ont utilisé les techniques d'appariement pour évaluer l'effet des aides à la R&D sur différents indicateurs (voir notamment González et Pazó (2008) ; Duguet (2010) ;

7. Lorsque plusieurs instruments sont disponibles, l'exogénéité combinée des instruments peut être testée.

Czarnitzki et al. (2011) ; Jaffe et Le (2015)). Cette technique consiste à retrouver, pour chaque entreprise bénéficiant d'une aide (entreprise traitée), une entreprise non bénéficiaire ayant les caractéristiques observables les plus proches (entreprise témoin). Bien souvent, les caractéristiques des entreprises sont résumées en un seul indice : on calcule la probabilité pour chaque entreprise de bénéficier d'une aide, puis les entreprises traitées et témoins ayant la probabilité la plus proche forment alors une paire (*propensity score matching*). Former ces paires de traités et de témoins suppose évidemment que pour chaque unité traitée, un témoin existe (hypothèse du support commun). Par exemple, si toutes les grandes entreprises bénéficiaient automatiquement du PIA-ADEME, nous ne pourrions pas appliquer cette méthode sur ce sous-groupe car nous n'aurions aucune entreprise non bénéficiaire comparable.

Au-delà de l'appariement « un pour un » décrit ci-dessus, plusieurs méthodes d'appariement existent⁸. On peut choisir d'apparier une entreprise traitée avec seulement une entreprise témoin comme nous venons de le décrire. On peut également choisir d'utiliser toutes les observations situées à une distance prédéfinie de l'entreprise traitée, c'est-à-dire toutes les observations ayant un score de propension situé à plus ou moins ϵ de celui de l'unité traitée (*caliper* ou *radius matching*). On peut également construire un contrefactuel synthétique en utilisant toutes les observations du groupe témoin, qui sont pondérées en fonction de la distance à laquelle elles se situent de l'unité traitée. Plus l'entreprise témoin est proche de l'entreprise traitée, plus elle aura un poids important (*kernel matching*).

Lorsque cet appariement est réalisé, il suffit de comparer les variables d'intérêt du groupe traité à celles du groupe témoin. Sous l'hypothèse que les deux groupes sont effectivement comparables du point de vue de leurs caractéristiques, comparer la moyenne de la variable d'intérêt du groupe traité à celle du groupe témoin permet d'estimer l'effet causal du traitement.

Une fois les groupes traité et témoin constitués, cette méthode a l'avantage de n'imposer aucune hypothèse sur la structure de la relation entre l'aide à la R&D et la

8. Voir Cameron et Trivedi (2005) section 25.4 pour une description complète des différentes méthodes d'appariement, ou Caliendo et Kopeinig (2005) pour les détails pratiques de la mise en œuvre des méthodes d'appariement.

variable d'intérêt puisqu'il s'agit d'une simple comparaison de moyenne entre deux groupes. Cependant, l'appariement entre les unités traitées et témoins, qui permet la constitution du groupe témoin, ne se fait que sur les caractéristiques observables, c'est-à-dire celles dont disposent l'évaluateur pour appréhender l'effet causal. Or, la qualité de l'évaluation peut être mise à mal si des caractéristiques inobservables jouent un rôle prépondérant. L'hypothèse sous-jacente très forte permettant d'identifier un effet causal est que les différences observables entre les bénéficiaires et les non-bénéficiaires captent l'ensemble des déterminants de la sélection des bénéficiaires. Autrement dit, une fois que l'on contrôle pour les variables observables l'attribution de l'aide peut-être considérée comme aléatoire (hypothèse d'indépendance conditionnelle).

Jaffe et Le (2015), qui utilisent un *propensity score matching* sur données néo-zélandaises, précisent dans quel cas cette méthode peut être fiable. Si le pouvoir prédictif des variables utilisées pour calculer le score de propension est élevé, alors le risque de sélection causé par les inobservables sera minime. Dans ce cas, l'effet estimé pourra être interprété comme un effet causal. González et Pazó (2008) recommandent, quant à eux, d'inclure comme variables dans le calcul du score de propension les valeurs prises par la variable dépendante Y_i au cours des périodes précédentes (*lagged variable*). En effet, il est très probable que les dépenses privées de R&D en $t-1$ déterminent celles de la période t . Cette inclusion permet d'augmenter le pouvoir prédictif du score de propension⁹.

Le recours aux expériences naturelles : lorsque des discontinuités claires existent dans le processus d'attribution des aides ou que d'importants changements législatifs le permettent, des méthodes d'évaluation reposant sur l'exploitation de ces expériences dites « naturelles », comme les doubles différences ou la régression sur discontinuités (RDD), sont utilisées.

Si des données de panel sont disponibles, la méthode des doubles différences est souvent utilisée puisqu'elle permet de mieux contrôler pour le biais de sélection lié aux caractéristiques inobservables des entreprises. En effet, puisque l'on observe les

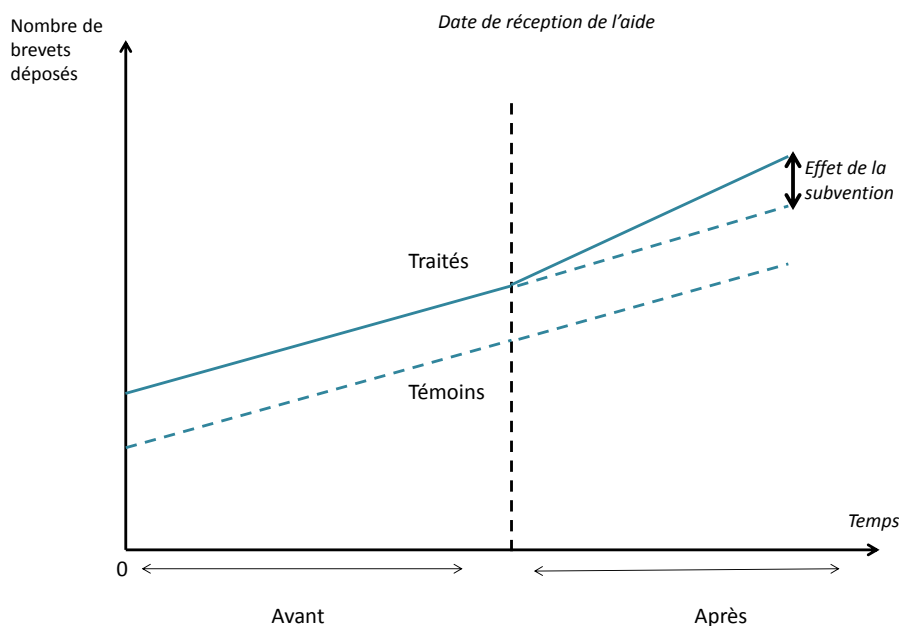
9. L'inclusion de ces variables n'est possible que dans le cas où l'on dispose de données de panel pour l'analyse.

entreprises traitées et témoins avant et après l’octroi des aides à la R&D, les caractéristiques inobservables des entreprises fixes au cours du temps « disparaissent » lorsque l’on calcule la différence entre les deux périodes. Il s’agit de la première différence, qui consiste à comparer la variable d’intérêt Y_i à la période t à celle de la période $t - 1$, séparément pour les deux groupes. On obtient ainsi l’évolution de la variable au cours du temps pour les deux groupes. La soustraction de ces deux évolutions nous donne l’effet du traitement qui, sous l’hypothèse que les deux groupes auraient évolué de la même manière en l’absence du traitement, peut être interprété comme un effet causal. L’hypothèse identifiante selon laquelle les deux groupes évolueraient de la même manière si l’intervention n’avait pas eu lieu est bien évidemment impossible à vérifier. Le test des tendances parallèles, qui vérifie si les évolutions des deux groupes sont similaires sur la période pré-traitement, permet ainsi de tester une hypothèse nécessaire mais pas suffisante de l’identification de l’effet causal du traitement par l’estimateur en doubles différences (Givord, 2010).

Klette et Møen (1999), en estimant des régressions OLS avec des effets fixes au niveau de l’entreprise, appliquent la méthode des doubles différences pour évaluer un programme gouvernemental norvégien soutenant la R&D réalisée par les entreprises dans le domaine des technologies de la communication. En comparant ainsi les entreprises bénéficiaires aux entreprises non bénéficiaires avant et après la mise en place du programme, ils trouvent que les entreprises traitées n’ont pas de meilleures performances que les entreprises témoins. Or, l’absence de différence entre les deux groupes ne doit pas être interprétée comme un échec du programme mais plutôt comme le résultat d’une sélection « négative » des entreprises bénéficiaires : les entreprises en difficulté sont celles qui ont eu la plus forte probabilité de bénéficier du programme.

L’étude de Klette et Møen (1999) souligne les limites de la méthode des doubles différences. Si cette méthode permet de réduire les biais de sélection liés aux caractéristiques inobservables des entreprises, elle ne parvient pas à prendre en compte les chocs temporaires en période $t - 1$ qui influencent la probabilité de bénéficier d’une aide. Klette et al. (2000) préconisent d’enrichir la méthode des doubles diffé-

FIGURE 4.1 – Illustration de la méthode des doubles différences



LECTURE : Ce graphique illustre la méthode des doubles différences. Avant l'obtention d'une aide, les entreprises traitées et non traitées suivent une même trajectoire. Après l'obtention de l'aide, les entreprises traitées déposent plus de brevets par rapport à l'évolution qu'elles auraient connu en l'absence d'aide.

rences par l'utilisation des méthodes d'appariement décrites ci-dessus, afin d'améliorer la comparabilité initiale des deux groupes et de mieux corriger ce biais. Estimer l'effet causal du traitement en combinant les méthodes d'appariement à celle des doubles différences permet en outre de relâcher certaines hypothèses identifiantes. Ces estimateurs ont été introduits par Heckman et al. (1998) et font l'objet d'une présentation détaillée dans Blundell et Costa Dias (2002).

Pour citer un autre exemple, Lach (2000) utilise également la méthode des doubles différences pour évaluer l'impact d'un programme de subvention à la R&D en Israël et trouve un effet positif mais non significatif du programme sur les dépenses privées en R&D. Il explique ces résultats en soulignant une autre limite de la méthode des doubles différences : cette méthode est peu appropriée pour capturer les effets dynamiques des aides.

Enfin, bien qu'elle soit très attractive car très proche d'une expérience aléatoire, la méthode de la régression sur discontinuité (RDD) est encore relativement peu

utilisée. À notre connaissance, seuls Bronzini et Iachini (2014) et Bronzini et Piselli (2016) y ont recours pour évaluer l'effet d'un programme de soutien à l'innovation en place dans le nord de l'Italie sur les dépenses en R&D et sur le dépôt de brevets. Le processus de sélection des entreprises bénéficiaires se prête parfaitement à l'utilisation de la RDD : les projets déposés reçoivent une note et seuls les projets recevant une note supérieure à un certain seuil bénéficient d'une aide. Sous l'hypothèse que les entreprises candidates ne peuvent pas influencer complètement leur note, hypothèse vérifiable et plausible dans le contexte de l'étude, l'attribution des subventions autour du seuil est considérée comme aléatoire. La méthode de la RDD s'apparente alors à une expérience aléatoire. L'hypothèse centrale de l'identification est que les entreprises se situant de part et d'autre de la discontinuité ont des caractéristiques similaires et la comparaison des performances des bénéficiaires à celles des non bénéficiaires autour de la discontinuité donne l'effet causal du programme de subvention. Les auteurs soulignent en outre que les entreprises sont d'autant plus comparables qu'elles sont situées et investissent dans la même région, ce qui réduit le risque de biais lié à l'hétérogénéité inter-entreprises. Ils montrent un effet positif du programme sur le nombre de brevets déposés (0,87 brevets déposés) et un effet sur les dépenses d'investissement en R&D uniquement pour les petites entreprises.

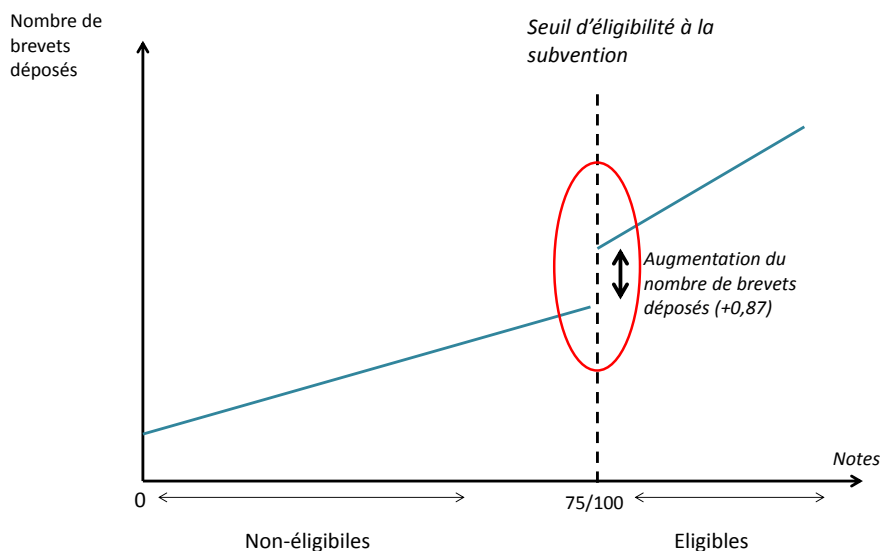
La limite majeure de la méthode des régressions sur discontinuité est que l'estimation des effets est très localisée, puisque l'on s'intéresse uniquement à ce qui se passe autour de la discontinuité. Les résultats sont par conséquent difficilement généralisables.

4.2.3 Applicabilité des différentes méthodes à l'évaluation du PIA-ADEME

À l'issue de cet état des lieux sur les différentes méthodes économétriques utilisées dans les évaluations relatives aux aides publiques à la R&D et des contextes dans lesquels elles sont mises en œuvre, nous jugeons de leur applicabilité à l'évaluation du PIA-ADEME (tableau 4.1).

L'estimation de régressions linéaires, si elle a pour avantage d'être très facile

FIGURE 4.2 – Illustration de la méthode de la RDD - L'étude de Bronzini et Piselli (2016)



LECTURE : Ce graphique illustre la méthode de la régression sur discontinuité (RDD) en prenant l'exemple de l'étude de Bronzini et Piselli (2016). Les entreprises ayant une note légèrement au-dessus de 75/100 sont considérées comme traitées, et celles légèrement en-dessous comme contrôles.

à mettre en œuvre, ne permet pas de contrôler pour les différents types de biais spécifiques à notre cas, à savoir les biais de sélection et de simultanéité. Elles peuvent toutefois servir d'estimation de référence, permettant ainsi d'appréhender dans quelle mesure les méthodes cherchant à réduire ces différents biais améliorent la qualité des estimations.

La méthode des variables instrumentales est une technique attrayante pour corriger du biais de sélection, mais elle suppose qu'il existe un (ou des) instrument(s) valide(s), ce qui n'est pas le cas dans le contexte du PIA-ADEME. En effet les études y ayant eu recours utilisent comme instrument le montant maximum d'aide auquel peut prétendre l'entreprise, or nous ne disposons pas d'une telle variable pour la présente évaluation. Même si cette variable existait, il serait difficile de valider l'hypothèse de l'exogénéité de l'instrument puisque cela nécessite de disposer de plusieurs variables instrumentales. Par ailleurs, l'utilisation de variables instrumen-

tales ne permettrait pas de traiter du biais de simultanéité.

Les méthodes qui semblent le mieux se prêter à l'évaluation des aides du PIA-ADEME sont ainsi les méthodes d'appariements et des doubles différences, même si elles ne nous permettent pas de prendre en compte correctement l'effet des chocs temporaires qui influencent la probabilité de recevoir une aide. En suivant les recommandations de Klette et al. (2000) nous proposons de combiner ces deux méthodes, afin d'améliorer la comparabilité des groupes traité et témoin, hypothèse cruciale pour valider l'utilisation des doubles différences. Cette méthode nous permettra de corriger le biais de sélection dû aux observables et aux inobservables fixes dans le temps.

Enfin, l'évaluation par régressions sur discontinuité aurait été une méthode intéressante à mettre en œuvre puisque c'est la méthode qui s'approche le plus d'une évaluation par assignation aléatoire et qui, par conséquent, aurait pu permettre de traiter les différents types de biais. Or, ceci nécessiterait de disposer d'une note attribuée en fonction de la qualité du projet ainsi que d'un seuil au-dessus duquel le projet serait financé. L'absence d'un tel système de notation rend impossible l'utilisation de cette méthode.

TABLEAU 4.1 – Récapitulatif des méthodes d'évaluations disponibles et de leur applicabilité

Méthodes	Hypothèses identifiantes	Avantages	Limites	Applicabilité au PIA-ADEME	Etudes de références
Les régressions linéaires	Exogénéité des variables explicatives : variables inobservables ϵ_i non corrélées aux variables explicatives X_i et D_i	Facile à mettre en œuvre	Problèmes d'endogénéité liés au biais de sélection et de simultanéité	Oui	Hamberg (1968) Howe et McFetridge (1976) Higgins et Link (1981)
Variables instrumentales	Exogénéité de l'instrument Z_i : Z_i très fortement corrélé au fait de recevoir une aide D_i mais non corrélé directement à Y_i	Correction du biais de sélection	Difficile de trouver un instrument valide ; Exogénéité de Z_i testable uniquement si plusieurs instruments sont disponibles ; Ne traite pas le biais de simultanéité	Non	Lichtenberg (1988) Wallsten (2000) Einiö (2014)
Evaluation contrefactuelle : méthodes d'appariement	Indépendance conditionnelle : conditionnellement aux variables observables, l'attribution de l'aide peut être considérée comme aléatoire (i.e. le problème de sélection est entièrement expliqué par les observables) Support commun : pour chaque unité traitée, il existe une unité non traitée	Corrige la partie du biais de sélection due aux observables Pas d'hypothèse sur la structure de la relation entre D_i et Y_i	La qualité de l'appariement et la validité de l'hypothèse d'indépendance conditionnelle dépendent très fortement de la richesse des données disponibles	Oui	González et Pazó (2008) Jaffe et Le (2015) Duguet (2010) Czarnitzki et al. (2011)
Evaluation contrefactuelle : doubles différences (DiD)	Hypothèse des tendances parallèles : les entreprises traitées et témoins auraient évolué de la même manière en l'absence de l'aide	Corrige la partie du biais de sélection due aux observables et aux inobservables fixes au cours du temps	Ne corrige pas de la sélection due aux inobservables variables au cours du temps ; Peu appropriée pour capturer des effets dynamiques ; Capture mal l'effet des chocs temporaires influençant la probabilité d'être traité	Oui	Klette et Møen (1999) Lach (2000)
Evaluation contrefactuelle : régressions sur discontinuités (RDD)	Existence d'une variable de sélection (une note) pour laquelle la probabilité de recevoir une aide est discontinue (existence d'un seuil) Les autres déterminants de l'obtention d'une aide sont continus autour de ce seuil	S'apparente à une évaluation randomisée : annule tout type de biais	Effet estimé très local donc difficilement généralisable	Non	Bronzini et Iachini (2014) Bronzini et Piselli (2016)

NOTE : Ce tableau présente les différentes méthodologies utilisées dans la littérature. Il résume les hypothèses identifiantes, les avantages, les limites ainsi que l'applicabilité de ces méthodes à l'évaluation du PIA-ADEME.

CHAPITRE 5

CHOIX MÉTHODOLOGIQUES ET RÉSULTATS DE L'ÉTUDE

Dans ce chapitre, nous détaillons les choix méthodologiques qui ont été faits pour cette étude, guidés par la revue des méthodologies présentée au chapitre précédent ainsi que par les contraintes imposées par les données. Nous présentons ensuite l'analyse menée et les résultats de cette étude.

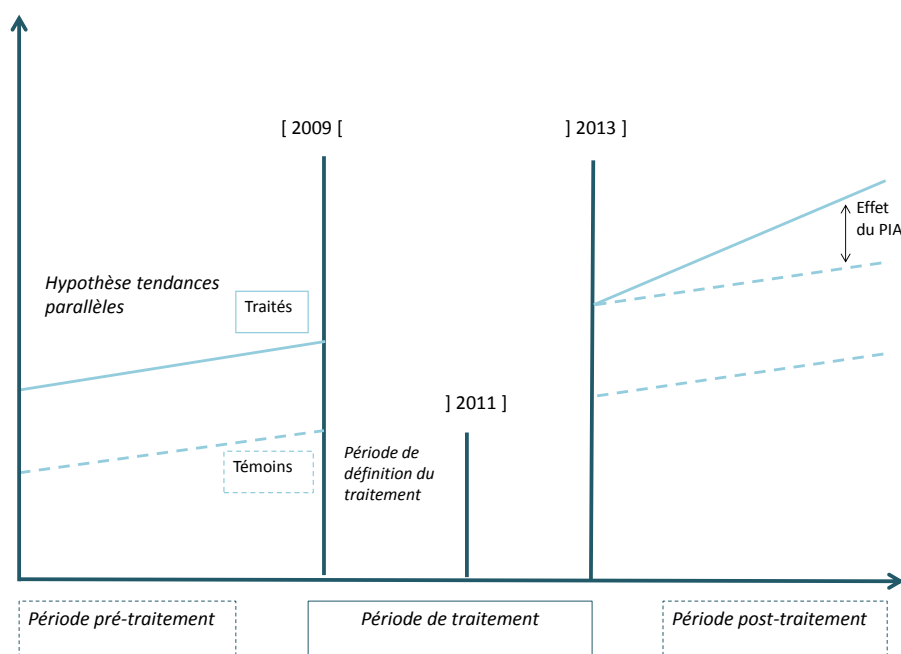
5.1 Choix méthodologiques

Pour évaluer l'effet des aides de l'ADEME sur diverses variables d'intérêt, la méthodologie envisagée est celle des doubles différences, qui consiste à comparer les bénéficiaires et les non-bénéficiaires, avant et après l'obtention de l'aide. Alors que de manière générale l'utilisation des doubles différences s'applique lorsque le traitement a lieu à un instant t pour tous les traités, notre étude a la spécificité d'avoir un traitement qui ne prend pas effet à la même date pour tous les partenaires. Cette partie présente la manière dont nous traitons cette spécificité en précisant la temporalité de l'étude, la définition d'un partenaire traité, la constitution des groupes traité et témoin ainsi que la méthode économétrique retenue.

5.1.1 Définition de la temporalité de l'étude

Alors que les données de l'ADEME sont disponibles jusqu'au premier trimestre 2016, les données externes ne le sont que jusqu'en 2013 inclus. Cela pose la question de la fenêtre temporelle à retenir pour l'étude. Plus précisément, cela implique de déterminer en quelle année nous arrêtons de considérer comme « traité » un partenaire contractualisé, pour avoir ensuite assez de recul pour pouvoir détecter un effet.

FIGURE 5.1 – Temporalité de l'étude



LECTURE : La période de traitement s'étend de 2009 à 2013, mais la période retenue pour la définition du traitement est restreinte à la période 2009-2011 inclus.

Nous faisons le choix de retenir une période de recul de deux ans minimum entre l'entrée en traitement et l'année d'observation des variables d'intérêt. Comme le résume la figure 5.1, ce choix implique que la période de traitement commence en 2009 (première année pour laquelle des dossiers ont été déposés pour le PIA-ADEME) pour aller jusqu'en 2011 inclus, de sorte qu'on bénéficie en 2013 d'un recul de deux années pour les partenaires entrant en traitement en 2011. La période de pré-traitement, sur laquelle l'hypothèse de tendances parallèles sera vérifiée,

couvre la période antérieure à 2009 et la période post-traitement est la période postérieure à 2013. La durée de traitement est ainsi de deux ans minimum, mais peut aller jusqu'à cinq ans pour les partenaires traités dès 2009. L'effet causal est mesuré en faisant la différence entre les valeurs observées en 2013 et 2009 pour les variables d'intérêt, pour l'ensemble des traités et des témoins. Si d'autres options sont envisageables (voir 5.1.4), nous optons pour une période d'observation fixe pour toutes les observations, afin d'essayer de mettre en évidence des effets de moyen terme, horizon qui paraît raisonnable pour les activités de R&D.

5.1.2 Définition du traitement et constitution des groupes

Une fois la période de traitement fixée, reste à définir ce que nous considérons être un partenaire traité ou témoin. Un partenaire est considéré comme traité s'il fait partie d'un projet contractualisé au sens de l'ADEME, c'est-à-dire s'il dispose d'une date de notification¹, puisqu'il est en effet très rare que le partenaire ne reçoive pas de financement une fois notifié. L'année retenue pour définir le statut du partenaire est la date de dépôt de dossier. L'année du Copil de sélection et l'année de la date de notification avaient également été proposées mais l'une comme l'autre restreignaient la taille de l'échantillon d'analyse et ne permettaient pas de limiter le biais d'anticipation pour les partenaires sélectionnés. En effet, entre la date de dépôt et la date de notification, certaines entreprises peuvent avoir engagé des dépenses : le financement PIA-ADEME, s'il est accordé, porte sur les dépenses dites « éligibles », parmi lesquelles entrent les dépenses réalisées avant la contractualisation. Retenir la date de dépôt permet de réduire ce biais.

En prenant en compte la fenêtre temporelle choisie et la définition du traitement retenue, le statut d'un partenaire peut être résumé comme suit :

- Un partenaire est considéré comme traité s'il fait partie d'un projet ayant une date de notification et ayant été déposé entre 2009 et 2011 inclus (395 cas) ;
- Il sera considéré comme témoin s'il ne fait pas partie d'un projet sélectionné sur la période 2009-2013. Il peut, pour autant, faire partie d'un projet sélec-

1. La date de notification est reportée lorsque la convention d'aide signée par le partenaire a été reçue par l'ADEME.

tionné après 2013 (1 443 cas) ;

- Un partenaire est exclu du champ de l'étude s'il fait uniquement partie d'un projet sélectionné entre 2011 et 2013 (94 cas).

En appliquant cette définition à la base ADEME, nous obtenons ainsi un échantillon composé de 395 traités et 1 443 témoins si l'on se met au niveau du Siret du partenaire, et respectivement 355 et 1 278 si l'on considère le Siren, c'est-à-dire si on agrège les partenaires constituant différents établissements d'une même entreprise (tableau 5.1). Une fois que l'on a apparié la base ADEME avec les données de l'Enquête R&D et les données FARE, le nombre de traités n'est plus que de 90, contre 173 dans la base appariée ADEME-FARE-DADS². Si la base ADEME-FARE-RD est la principale base d'intérêt pour l'analyse en raison des informations apportées par l'Enquête R&D, le trop faible nombre d'observations nous limite fortement dans notre capacité à mener une évaluation rigoureuse des aides de l'ADEME et a des conséquences pour la méthodologie privilégiée.

TABLEAU 5.1 – Effectifs de partenaires traités et témoins

Base de données	Niveau d'observation	Traités	Témoins	Total
ADEME	Siret	395	1 443	1 838
	Siren	355	1 278	1 633
ADEME-FARE-RD	Siren	90	158	248
ADEME-FARE-DADS	Siren	173	533	706

LECTURE : Dans la base de données appariée ADEME-FARE-RD, il y a 90 traités et 158 témoins.

NOTE : Le niveau d'agrégation le plus élevé pour les bases ADEME-FARE-RD et ADEME-FARE-DADS est le Siren, c'est pourquoi nous n'avons pas le nombre d'observations pour le Siret.

SOURCE : Données ADEME ; FARE 2009 et 2013 ; DADS 2009 et 2013 ; R&D 2009 et 2013.

2. Nous reportons en annexe un tableau de comparaison de moyennes entre les entreprises traitées et témoins afin de caractériser un peu mieux les entreprises de l'échantillon ADEME-FARE-DADS (Annexe C)

5.1.3 Choix de la méthode économétrique

L'évaluation des aides du PIA-ADEME se prête, en théorie, plutôt bien à l'utilisation de la méthode des doubles différences présentée dans la revue des méthodologies (voir chapitre 4). En effet, nous disposons d'un panel d'entreprises observables sur plusieurs années ainsi que d'une identification claire des entreprises traitées et témoins. Les données sur les projets non-retenus et leurs partenaires permettent de constituer un groupe témoin *a priori* plus pertinent que si les entreprises témoins étaient choisies parmi l'ensemble des entreprises ayant des activités de R&D.

En revanche, la qualité des données nous permettant de caractériser les activités de R&D des partenaires d'une part, et les projets déposés par les partenaires, comme le montant de l'aide demandée et le coût du projet, d'autre part, est limitée. Or ces variables sont cruciales pour tester les hypothèses justifiant l'utilisation des doubles différences. En effet, comme nous l'avons décrit au chapitre 3, l'Enquête R&D, qui est la source de données la plus détaillée sur les activités de R&D des entreprises, ne couvre qu'une petite partie de notre échantillon, nous contraignant ainsi à utiliser des variables *proxy* issues d'autres bases comme les DADS. Les informations extraites à partir des Annexes 5 sur les montants d'aide demandés et sur les coûts du projet par partenaire sont quant à elles assez peu fiables et manquantes pour une part non négligeable des observations (voir partie 3.1.1). Nous disposons néanmoins de caractéristiques financières et comptables assez précises grâce aux liasses fiscales FICUS et FARE.

Nous nous appuyons ainsi principalement sur les variables issues des DADS et des liasses fiscales afin de tester si l'hypothèse des tendances parallèles est validée sur la période pré-traitement. Il s'agit de l'hypothèse que les partenaires traités et témoins suivent des trajectoires parallèles sur les variables d'intérêt *avant* l'obtention de l'aide. Afin d'améliorer la comparabilité des groupes, nous utilisons une procédure d'appariement (*matching*) des partenaires traités avec les partenaires témoins les plus proches du point de vue des caractéristiques observables. Nous calculons ainsi la probabilité de bénéficier d'une aide (*propensity score*) pour chaque partenaire, puis formons une paire avec les deux partenaires traité et témoin ayant la probabilité la plus proche (*one-to-one matching*). Alternativement, nous utilisons

l'ensemble des partenaires témoins en attribuant à chacun un poids, calculé en fonction de la distance entre les scores de propension de chaque témoin à chaque traité (*kernel matching*) : plus les témoins ont un score de propension proche (resp. éloigné) de celui des traités, plus ils auront un poids important (resp. faible)³. Les partenaires aux caractéristiques trop spécifiques, et donc pour lesquels aucun « jumeau » ne peut être trouvé car leur score est trop éloigné des autres, sont alors exclus de l'échantillon (hypothèse dite du « support commun »). Cela revient à exclure les partenaires traités qui, au vu de leurs caractéristiques observables, ont une probabilité trop élevée de bénéficier de l'aide par rapport au traité. Cette procédure d'appariement n'est réalisée que pour l'analyse sur la base ADEME-FARE-DADS, le nombre d'observations étant trop faible dans la base ADEME-FARE-RD.

Si la comparabilité des partenaires traités et témoins avant la réception d'une aide est avérée et qu'il semble raisonnable d'affirmer que, conditionnellement aux observables, l'obtention d'une aide est aléatoire (hypothèse d'indépendance conditionnelle)⁴, nous proposons d'estimer l'effet des aides du PIA-ADEME en comparant la situation des partenaires bénéficiaires et non-bénéficiaires d'aides avant et après réception de l'aide. Plus précisément, nous estimons l'équation de doubles différences suivante :

$$Y_i = \alpha + \beta_1 Post_i + \beta_2 Trait_i + \beta_3 Post_i * Trait_i + \beta_4 X_i + \epsilon_i \quad (5.1)$$

où

- Y_i est la variable d'intérêt (voir tableau 3.6 pour les différentes variables d'intérêt utilisées) ;
- $Post_i$ est une indicatrice pour la période post-traitement : elle vaut 1 pour l'année 2013 ;
- $Trait_i$ est une indicatrice de traitement : elle vaut 1 pour les partenaires bénéficiaires ;

3. Dans les régressions, les observations seront ainsi pondérées par leur poids issu de la procédure *kernel*. L'effet du traitement peut ainsi se concevoir comme une moyenne pondérée de l'effet causal, en donnant plus d'importance aux témoins les plus proches.

4. L'ensemble des hypothèses identifiantes (tendances parallèles, indépendance conditionnelle et support commun) sont détaillées dans la partie 4.2.2 du chapitre 4.)

- $Post_i * Trait_i$ est l'interaction des deux. Le coefficient associé à cette indicatrice, β_3 , mesure ainsi l'effet des aides accordées ;
- X_i est un ensemble de variables de contrôle représentant les caractéristiques des entreprises (voir tableau 5.2 pour le détail des contrôles utilisés). Ces contrôles sont des variables susceptibles d'être corrélées avec la variable d'intérêt, Y_i ;
- ϵ_i est le terme d'erreur qui regroupe les variables inobservables.

Les variables de contrôle utilisées varient selon que la base d'analyse utilisée est ADEME-FARE-RD ou ADEME-FARE-DADS (cf. tableau 5.2). De la même manière que pour les variables d'intérêt, les variables de contrôle sont beaucoup moins nombreuses et moins précises lorsque l'on utilise la base de données ADEME-FARE-DADS. Par exemple, nous ne sommes pas en mesure de contrôler pour le fait de bénéficier de financements accordés par d'autres opérateurs du PIA et nous utilisons le fait de recevoir des subventions d'investissement comme *proxy* pour cette variable. Les dépenses de R&D sont approximées par les investissements en immobilisations incorporelles, qui incluent entre autres les frais de R&D mais comprennent aussi les dépenses liées à l'obtention d'un brevet, d'une marque ou d'une licence.

Enfin, lors de l'estimation, les erreurs-types (*standard errors*) sont calculées en prenant en compte le fait que les observations appartenant au même secteur industriel ne sont pas indépendantes mais sont au contraire corrélées. En effet, les partenaires d'un même secteur sont soumis à des chocs conjoncturels et à d'autres caractéristiques inobservables similaires : cela a pour conséquence que le calcul des erreurs-types doit être corrigé du fait que chaque observation ne nous apporte pas une information unique. Cette correction, qui consiste à « clusteriser » les erreurs-types, conduit à augmenter les erreurs-types et donc *in fine* à diminuer la significativité des coefficients, mais elle est nécessaire afin d'éviter toute erreur d'inférence⁵.

5. « Clusteriser » les erreurs-types permet notamment d'éviter de faire des erreurs d'inférence de type 1, c'est-à-dire de considérer qu'un coefficient est significatif alors qu'en réalité il ne l'est pas.

TABLEAU 5.2 – Les variables de contrôle utilisées

Justification de la variable	Base ADEME-FARE-RD	Base ADEME-FARE-DADS
Effet de recevoir d'autres aides	Bénéficie d'un autre PIA (Enquête R&D)	Montant des subventions d'investissement (DADS)
Expérience de l'entreprise	Age de l'entreprise (FARE)	Age de l'entreprise (FARE)
Effet du type d'activité de recherche	Activités de recherche fondamentale / de développement expérimental (Enquête R&D)	
Effet du niveau de l'activité de R&D	Ratio des dépenses de R&D sur les ventes totales (Enquête R&D)	Investissements en immobilisations incorporelles (FARE)
Taille de l'entreprise	Nombre de salariés (FARE)	Nombre de salariés (FARE)
Ouverture commerciale	Montant des exports (FARE ou Enquête R&D)	Montant des exports (FARE)

LECTURE : L'analyse menée sur la base ADEME-FARE-RD utilise le fait de bénéficier d'autres PIA comme contrôle dans les régressions. Dans le cas de l'utilisation de la base ADEME-FARE-DADS, cette variable est approximée par le montant des subventions d'investissement reçues.

NOTES : Les bases dont sont issues les variables sont précisées entre parenthèses.

5.1.4 Discussion méthodologique

Au cours des différents Comités de pilotage, la question de la période de traitement à retenir a été largement abordée. Si dans la présente étude nous utilisons les années 2009 et 2013 comme points de départ et de fin pour toutes les observations, avec des durées de traitement allant de deux à cinq années, d'autres options seraient envisageables pour l'évaluation finale.

Une première option proposée serait de fixer la période de traitement à deux ans pour toutes les observations, avec une date de pré-traitement qui n'aurait pas été 2009 pour tout l'échantillon mais qui aurait varié en fonction de l'année où le partenaire reçoit effectivement l'aide. Par exemple pour un partenaire bénéficiaire d'une aide en 2010, l'année de pré-traitement retenue aurait été 2010 et l'année de post-traitement 2012. Les tendances parallèles seraient vérifiées sur les années antérieures à la réception de l'aide, ce qui implique nécessairement que l'entreprise doit exister avant de bénéficier du PIA-ADEME. Cette méthode aurait permis d'avoir une période d'observation similaire pour tous les partenaires et de mieux capter les effets de court terme. Cette option a toutefois été écartée puisqu'il est également possible d'observer des effets de court terme avec la méthode retenue : en fonction de la taille de l'échantillon, cette méthode permet de conduire une analyse causale sur des sous-groupes, qui peuvent être caractérisés par exemple par la durée de traitement.

Une autre option, inspirée de Mauroux (2012), qui s'intéresse à l'effet du changement du taux du crédit d'impôt pour le développement durable sur le recours à ce dispositif, aurait été de faire varier la date d'entrée en traitement en fonction de l'année de réception de l'aide, sans toutefois imposer de contrainte sur la date de fin de traitement. Si nous avons retenu cette définition plus flexible du traitement, une estimation plus raffinée de l'équation 5.1 aurait pu être proposée. L'équation 5.1 pourrait ainsi être enrichie des termes d'interaction entre les années et l'indicatrice de traitement, comme suit :

$$Y_i = \alpha + \sum_{y=2009}^{2013} \beta_y \text{Année}_y + \beta_2 \text{Trait}_i + \sum_{z=2009}^{2013} \beta_z \text{Année}_z * \text{Trait}_i + \beta_4 X_i + \epsilon_i \quad (5.2)$$

Ceci permettrait d'une part de mieux prendre en compte le fait que les partenaires entrent en traitement à des dates différentes, et d'autre part d'utiliser l'information des années intermédiaires au lieu de n'utiliser que l'information des années 2009 et 2013. Cela permettrait en outre de mieux décomposer l'effet du traitement année par année, donné par l'estimation des β_z , plutôt que de considérer l'effet moyen sur toute la période, comme nous proposons de le faire pour cette étude.

Si cette option est intéressante, elle requière néanmoins un travail sur les données bien plus important puisqu'elle multiplie le nombre d'appariements entre les différentes bases. En effet, on ne considère plus uniquement les années 2009 et 2013, mais également les années intermédiaires. Faute de temps, cette piste n'a pu être explorée mais fera l'objet d'une recommandation pour l'évaluation dans le chapitre 6.

5.2 Résultats de l'analyse

Nous présentons dans cette partie les résultats des estimations économétriques sur les deux échantillons d'analyse constitués : les bases ADEME-FARE-RD et ADEME-FARE-DADS. Nous étudions dans un premier temps si l'hypothèse des tendances parallèles est vérifiée, graphiquement et statistiquement, puis commentons les ré-

sultats des régressions.

5.2.1 L'analyse sur la base ADEME-FARE-RD

La base de données contenant les informations les plus détaillées sur les activités de R&D étant la base ADEME-FARE-RD, nous commençons par vérifier l'hypothèse des tendances parallèles sur celle-ci, et lui ajoutons deux années supplémentaires de données⁶. Si idéalement nous souhaiterions vérifier que les partenaires des groupes traité et témoin évoluent de la même manière sur les indicateurs de R&D avant réception de l'aide, nous ne sommes pas en mesure de le faire en pratique, étant donné que les appariements avec les Enquêtes R&D nous font perdre un nombre conséquent d'observations pour les raisons évoquées précédemment (voir chapitre 3). Ainsi l'appariement de notre base ADEME-FARE-RD avec l'Enquête R&D 2007 nous fait perdre à lui seul 55 observations sur les 248 présentes dans la base.

Nous testons donc cette hypothèse uniquement sur des indicateurs de performances économiques issus de la base FARE, et apparions la base ADEME-FARE-RD avec les années 2007 et 2008 des données FARE. Cette étape ne nous fait perdre que 13 observations supplémentaires, ce qui permet de tester l'hypothèse sur un échantillon de 86 traités et 149 témoins. Les indicateurs retenus pour la vérification des tendances parallèles sont : le taux d'immobilisations incorporelles, la productivité des salariés⁷, la production totale et le résultat comptable net. Si la production totale traduit le volume d'activité de l'entreprise, le résultat comptable est plutôt un indicateur de la santé financière de l'entreprise puisqu'il indique les ressources restantes à l'entreprise une fois les charges retranchées du chiffre d'affaires.

Les graphiques 5.2 à 5.5 montrent l'évolution de ces quatre indicateurs séparément pour les groupes traité et témoin sur la période 2007-2013. Sur les deux années précédant le début du traitement, les variables évoluent différemment dans les deux groupes : le taux d'immobilisations incorporelles chute en 2008 pour les

6. Les tendances parallèles sont en général vérifiées sur plusieurs années, mais nous faisons le choix de nous restreindre à deux années supplémentaires, puisque chaque appariement nous fait perdre des observations et réduit d'autant notre capacité à détecter des évolutions significativement différentes ou non.

7. Nous définissons la productivité des salariés comme le rapport entre la valeur ajoutée et l'effectif de l'entreprise.

partenaires traités alors qu'il augmente légèrement pour les témoins, dont le taux d'immobilisations incorporelles est par ailleurs plus élevé sur toute la période que pour les traités (figure 5.2), tout comme la productivité des salariés, qui augmente également sensiblement en 2008 pour les témoins, alors qu'elle reste relativement stable pour les traités, avant de diminuer entre 2008 et 2009 (figure 5.3). Les différences sont moins marquées pour ce qui est de la production totale (figure 5.4), contrairement à l'évolution du résultat comptable qui a fortement chuté pour les entreprises traitées en 2008, qui semblent avoir été beaucoup plus affectées par la crise que les entreprises témoins (figure 5.5). L'analyse graphique de l'évolution des différents indicateurs ne semble pas, *a priori*, valider l'hypothèse des tendances parallèles.

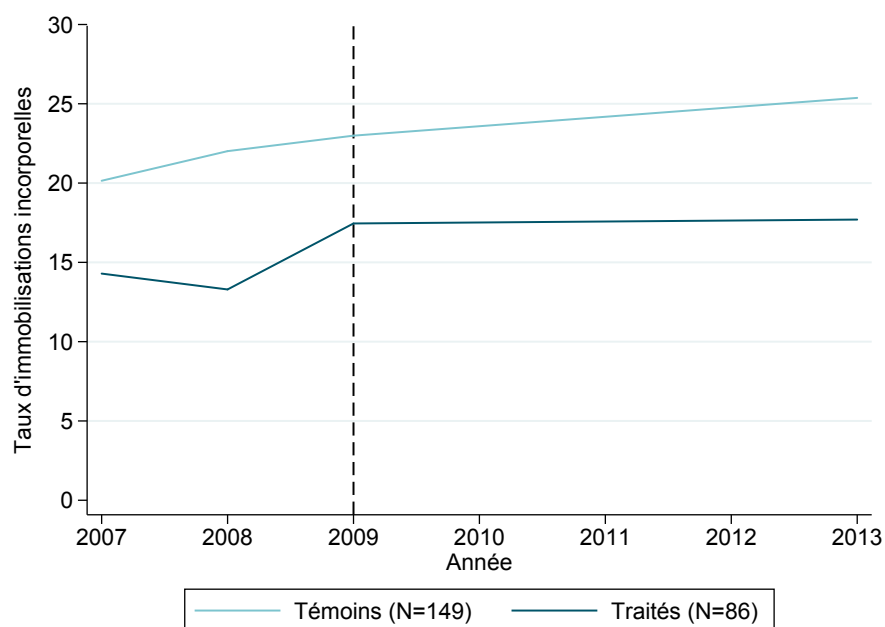
Afin de juger si ces divergences sont statistiquement significatives, nous avons effectué un test « placebo ». Nous estimons ainsi l'effet de bénéficier d'une aide de l'ADEME sur les années 2007 et 2008 à l'aide d'une régression, en interagissant la variable de traitement avec des indicatrices temporelles⁸. Si les entreprises des groupes traité et témoin évoluent de la même manière avant l'entrée en traitement, la différence entre les deux groupes ne devrait pas être significative. Nous effectuons ce test avec les quatre indicateurs précédents comme variables dépendantes, auxquels nous ajoutons également le chiffre d'affaires, la valeur ajoutée, l'excédent brut d'exploitation et les dividendes⁹. Les résultats de ce test sont reportés dans le tableau D.1 en Annexe D. Le nombre d'observations est en effet trop faible pour nous permettre d'estimer précisément nos coefficients.

L'hypothèse de tendances parallèles étant cruciale pour justifier de l'utilisation de la méthode des doubles différences, nous faisons le choix également de ne pas présenter les résultats des estimations de l'effet des aides de l'ADEME sur les indicateurs de performances des entreprises, d'emploi et de dépenses en R&D. Ces tableaux de régressions sont reportés à l'Annexe F et ne doivent pas être interprétés autrement que comme des corrélations.

8. Cette régression s'inspire de l'équation 5.2, avec $y = 2007, 2008, 2013$ et $z = 2007, 2008, 2013$. 2009 est l'année de référence.

9. S'intéresser à l'évolution des dividendes n'est toutefois pas pertinent pour les petites et moyennes entreprises, qui représentent une bonne partie des entreprises aidées par le PIA-ADEME.

FIGURE 5.2 – Evolution du taux d'immobilisations incorporelles pour la période 2007-2013 (en pourcentage)- Base ADEME-FARE-RD



LECTURE : En 2007, le taux d'immobilisations incorporelles est de 14,9 % pour les partenaires du groupe traité et de 20 % pour ceux du groupe témoin.

NOTE : Les immobilisations incorporelles sont un *proxy* pour les dépenses de R&D. Elles comprennent, entre autres, les frais de R&D et les concessions, brevets et licences.

SOURCES : Base ADEME ; FARE 2007, 2008, 2009 et 2013.

FIGURE 5.3 – Evolution de la productivité des salariés pour la période 2007-2013 - Base ADEME-FARE-RD

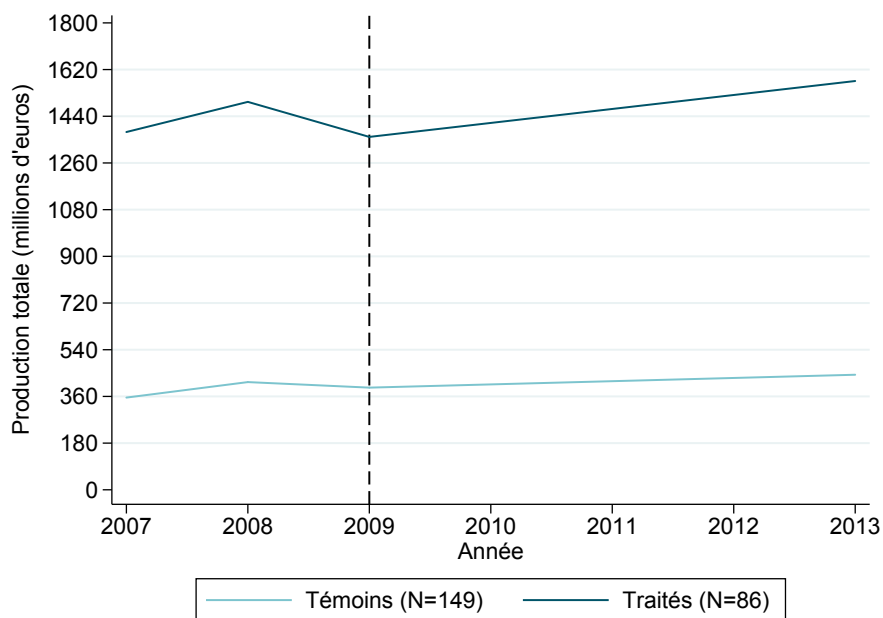


LECTURE : En 2007, chaque salarié des entreprises traitées produit en moyenne 72 000 euros de valeur ajoutée, contre 79 000 euros pour les salariés des entreprises témoins.

NOTE : La productivité des salariés est définie comme le rapport entre la valeur ajoutée de l'entreprise et les effectifs de l'entreprise.

SOURCES : Base ADEME ; FARE 2007, 2008, 2009 et 2013.

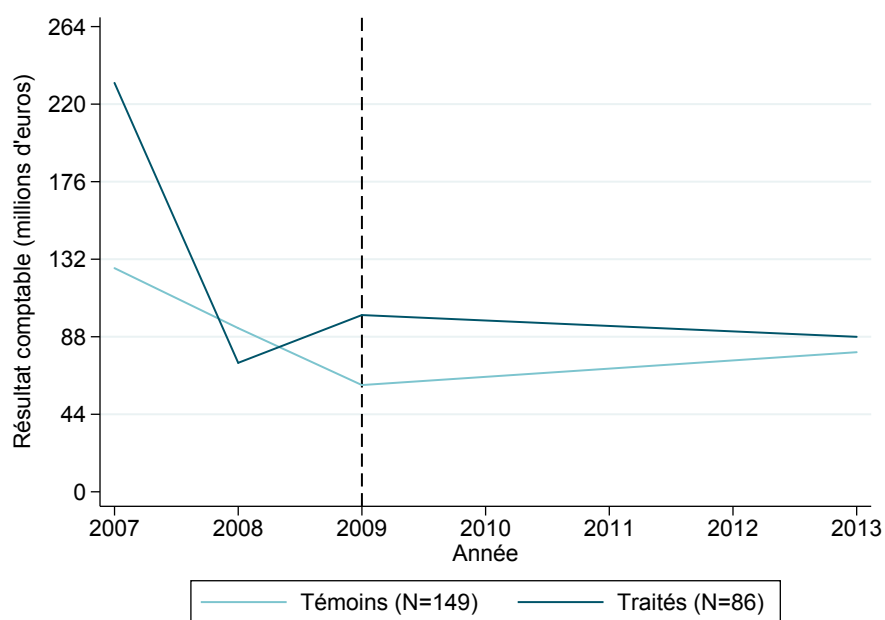
FIGURE 5.4 – Evolution de la production totale pour la période 2007-2013 - Base ADEME-FARE-RD



LECTURE : En 2007, la production totale est de 360 millions d'euros en moyenne pour les entreprises du groupe témoin contre près de 1 440 millions d'euros dans le groupe traité.

SOURCES : Base ADEME ; FARE 2007, 2008, 2009 et 2013.

FIGURE 5.5 – Evolution du résultat comptable pour la période 2007-2013 - Base ADEME-FARE-RD



LECTURE : En 2007, le résultat comptable est de 230 millions d'euros en moyenne pour les entreprises du groupe traité contre 130 millions d'euros pour les entreprises du groupe témoin.

SOURCES : Base ADEME; FARE 2007, 2008, 2009 et 2013.

5.2.2 L'analyse sur la base ADEME-FARE-DADS

La taille de l'échantillon ADEME-FARE-RD étant trop faible, d'une part, pour déterminer si les groupes traité et témoin évoluent de manière significativement différente avant la réception de l'aide, et donc valider ou invalider l'utilisation de la méthode des doubles différences ; et, d'autre part, pour conclure quant à l'effet du PIA-ADEME sur les différents indicateurs retenus, nous menons dans un second temps l'analyse sur la base ADEME-FARE-DADS. Le nombre d'observations étant bien plus conséquent, nous sommes en mesure d'améliorer la comparabilité des groupes en recourant aux méthodes d'appariement. Nous mettons en œuvre la procédure d'appariement dans un premier temps, puis vérifions ensuite l'évolution des deux groupes sur la période antérieure au traitement.

5.2.2.1 Mise en œuvre de la procédure d'appariement et vérification des tendances parallèles

Nous testons dans un premier temps l'hypothèse des tendances parallèles sur l'échantillon ADEME-FARE-DADS dans son ensemble. L'évolution des groupes traité et témoin étant encore sensiblement différente sur la période pré-traitement (voir Annexe E pour les graphiques), nous recourons aux méthodes d'appariement afin d'améliorer la comparabilité des deux groupes du point de vue des caractéristiques observables.

Nous calculons ainsi la probabilité de bénéficier du PIA-ADEME sur les caractéristiques de l'année 2009 en estimant un modèle *probit*¹⁰, dans lequel la variable dépendante est une variable dichotomique valant zéro pour les partenaires témoins et un pour les traités. L'équation estimée se présente comme suit :

$$Trait_i = \gamma + \delta_1 X_i + \delta_2 X_{i,t-1} + \mu_i \quad (5.3)$$

10. Le modèle *probit* est un modèle statistique adapté à l'estimation d'équation dont la variable expliquée est binaire, comme c'est le cas pour la variable de traitement ici considérée qui n'a que deux modalités : être partenaire ou non d'un projet aidé. Lorsque la variable expliquée est binaire, l'estimation d'un modèle linéaire est inappropriée et conduit à estimer des probabilités qui peuvent être inférieures à zéro ou supérieures à un. Le modèle *probit* fait des hypothèses plus fortes, notamment sur la distribution des termes d'erreurs, mais permet de contraindre les probabilités prédites à être dans l'intervalle [0;1]. Les coefficients sont estimés par la méthode du maximum de vraisemblance.

où

- $Trait_i$ est une indicatrice de traitement (vaut 1 pour les partenaires bénéficiaires) ;
- X_i est un ensemble de variables susceptibles d'influencer la probabilité de recevoir une aide ;
- $X_{i,t-1}$ est l'ensemble des valeurs prises par les variables X_i l'année précédente ;
- μ_i est le terme d'erreur qui regroupe les variables inobservables.

Les variables explicatives incluses (X_i), susceptibles d'influencer la probabilité de recevoir une aide, sont : l'âge de l'entreprise ; le montant des subventions d'investissement (*proxy* pour le fait de bénéficier d'autres aides que le PIA-ADEME) ; la masse salariale et la valeur ajoutée (qui permettent de contrôler de la taille de l'entreprise et de son activité) ; les effectifs de R&D, les stocks et les investissements en immobilisations incorporelles (pour contrôler de l'activité de R&D) ; ainsi qu'un indicateur d'ouverture commerciale, calculé comme la part du chiffre d'affaires à l'export dans le total des ventes. Enfin, en s'inspirant de González et Pazó (2008) et de Duguet (2010), nous incluons également les valeurs prises par ces variables explicatives en 2008 ($X_{i,t-1}$), sauf pour les subventions d'exploitations pour lesquelles l'information n'est disponible qu'en 2007. Les résultats de l'estimation du *probit* sont reportés dans le tableau 5.3.

Initialement, il était également prévu d'intégrer comme variables explicatives le coût initial du projet et le montant d'aide demandé. Ces variables, *a priori* susceptibles d'influencer la probabilité de bénéficier d'une aide, sont renseignées pour les projets bénéficiaires d'aide (données LOCO). Pour les entreprises n'ayant pas bénéficié d'une aide du PIA-ADEME, cette information a été récupérée par extraction à partir des dossiers de candidature, dans lesquelles elles sont renseignées. Toutefois, ces montants ne sont pas toujours renseignés et, d'autre part, une comparaison des valeurs indiquées avec celles consignées dans LOCO pour les partenaires bénéficiaires a montré de fortes évolutions des montants renseignés entre le moment du dépôt de dossier et celui de la contractualisation (voir chapitre 3). Il avait été envisagé d'inclure également la note de présélection des projets, attribuée par l'ADEME pour chaque dossier déposé en fonction de sa priorité d'instruction, dans le *probit*

pour le matching. Cette variable n'a finalement pas été retenue car les principes de notation changent sur la période de notre analyse, les notes étant généralement données par rapport aux autres projets soumis la même année. Enfin, nous avons considéré la possibilité d'inclure le secteur d'activité comme contrôle dans l'équation du *probit*. Toutefois, la classification des secteurs d'activité selon la nomenclature NAF n'est pas très pertinente pour les partenaires aidés par le PIA-ADEME, puisque leurs activités de R&D se situent souvent au croisement de plusieurs secteurs d'activité.

TABLEAU 5.3 – Résultats du *probit* - Base ADEME-FARE-DADS

Variable	Coefficients estimés
Age de l'entreprise	0.005** (0.002)
Part des exports/ventes	0.003** (0.002)
Taux d'investissements incorporels	-0.001*** (0.000)
Salaires bruts versés (millions d'euros)	0.006** (0.002)
Effectifs de R&D	0.001 (0.001)
Valeur ajoutée HT (millions d'euros)	0.000 (0.001)
Subventions d'investissement (millions d'euros)	0.004 (0.033)
Taux d'immobilisations incorporelles	0.001 (0.005)
Salaires bruts versés (millions d'euros) 2008	0.002 (0.003)
Effectifs de R&D 2008	0.000 (0.000)
Valeur ajoutée HT (millions d'euros) 2008	-0.003*** (0.001)
Taux d'immobilisations incorporelles 2008	-0.002 (0.005)
Subventions d'investissement (millions d'euros) 2007	-0.010 (0.045)
Nb. Observations	575
R ²	0.0506

LECTURE : L'âge de l'entreprise influence positivement la probabilité de bénéficier d'une aide, et cet effet est significatif à 5 %.

NOTES : Les coefficients reportés dans ce tableau sont issus de l'estimation d'un modèle *probit* avec en variable dépendante la probabilité d'être traité. Lorsqu'aucune année n'est indiquée, l'année considérée est 2009. Les étoiles indiquent la significativité du coefficient. * : significatif au seuil de 10 % ; ** : significatif au seuil de 5 % ; *** : significatif au seuil de 1 %. Les erreurs-types sont reportées entre parenthèses.

SOURCES : Base ADEME ; FARE et DADS 2007, 2008 et 2009.

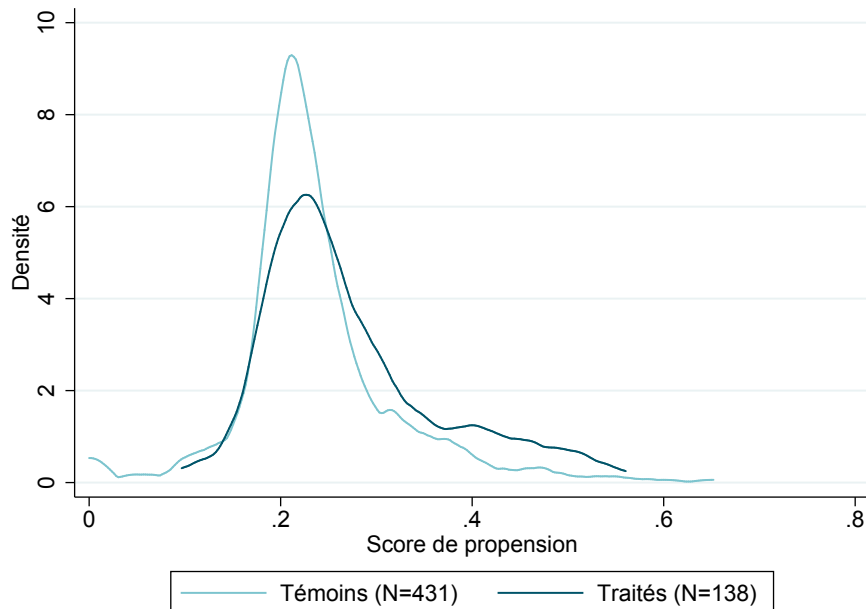
Parmi les variables influençant la probabilité de bénéficier d'une aide du PIA-ADEME, on retrouve l'âge de l'entreprise, l'indicateur d'ouverture commerciale, le taux d'investissements en immobilisations incorporelles ou encore la masse salariale. Le pouvoir prédictif de notre modèle est néanmoins relativement faible : nous sommes en mesure d'expliquer seulement 5,06 % de la variance de la probabilité de bénéficier d'une aide¹¹. Ces estimations impliquent que l'hypothèse d'indépendance conditionnelle, selon laquelle conditionnellement aux caractéristiques observables, l'attribution du traitement peut être considérée comme aléatoire, est difficile à valider dans notre contexte. Bien que la méthode des doubles différences permette d'« annuler » l'effet des inobservables fixes dans le temps, le fait que le R^2 des estimations *probit* soit si faible suggère que les inobservables jouent un rôle prépondérant. Nous sommes en effet limités par la qualité de nos données et ne pouvons observer des variables telles que les dépenses de R&D rapportées aux ventes, ou encore le fait de bénéficier d'une aide l'année antérieure, qui peuvent avoir un effet déterminant sur la probabilité de bénéficier d'une aide, comme ont pu le montrer les études de Duguet (2010), de Czarnitzki et al. (2011) ou encore de González et Pazó (2008).

Nous poursuivons toutefois l'exercice et, une fois les coefficients estimés, nous prédisons la probabilité d'être traité pour chaque partenaire et traçons la distribution de cette probabilité pour les groupes traité et témoin (figure 5.8). Les probabilités pour les entreprises du groupe témoin sont plus dispersées que celles du groupe traité : les premières varient de zéro à près de 0,65, quand les secondes se situent plutôt entre 0,10 et 0,55. Comme évoqué dans la revue de la littérature méthodologique (chapitre 4), une des deux hypothèses à valider pour justifier l'utilisation des méthodes d'appariement est celle du support commun, qui stipule que pour toute valeur du score de propension nous puissions observer à la fois des traités et des témoins. Nous restreignons donc notre échantillon au support commun en excluant les partenaires traités pour lesquels la probabilité d'être traité est supérieure à la probabilité maximum des partenaires témoins, puisque cela signifie qu'aucune entreprise comparable témoin n'existe pour ces observations traitées. Cela conduit à

11. À titre d'exemple, Jaffe et Le (2015) trouvent un R^2 qui se situe entre 0,3 et 0,38 et Czarnitzki et al. (2011) un R^2 de 0,22.

retirer six partenaires traités de notre échantillon.

FIGURE 5.6 – Distribution de la probabilité de bénéficier d'une aide du PIA-ADEME - Base ADEME-FARE-DADS



LECTURE : La distribution de la probabilité de recevoir une aide se concentre autour de 0,22 pour les groupes traité et témoin.

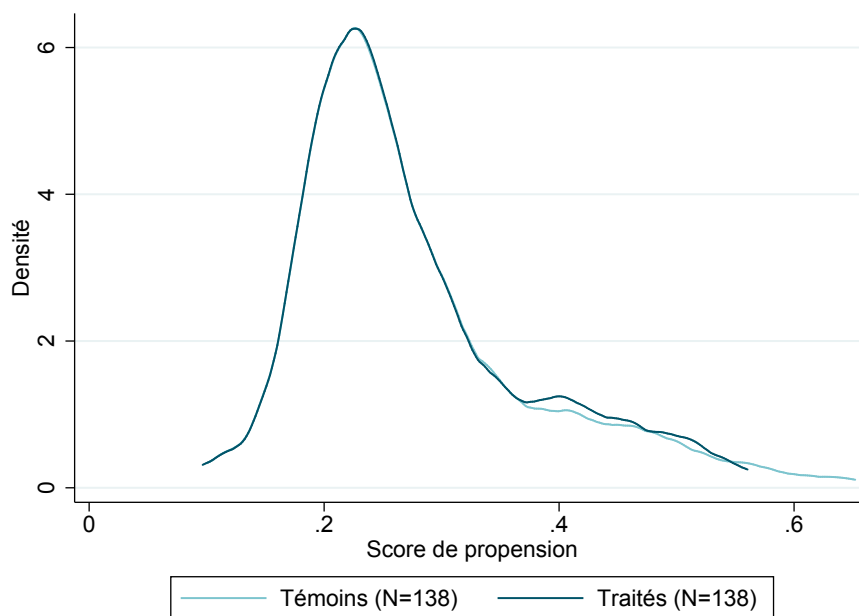
SOURCE : Base ADEME ; DADS et FARE 2007, 2008, 2009 et 2013.

Nous appliquons ensuite deux séries de restrictions pour constituer notre échantillon d'analyse. La première consiste à ne garder, pour chaque observation traitée, que l'observation témoin qui a le score de propension le plus proche (*one-to-one matching*). Une fois qu'une observation témoin a été utilisée, elle ne peut pas être réutilisée, et ce afin que la taille de l'échantillon ne soit pas trop petite (procédure sans remplacement). L'appariement « un pour un » a l'avantage d'être simple à mettre en œuvre mais n'utilise qu'une partie de l'information disponible puisque le contrefactuel pour le groupe traité est constitué uniquement des observations témoins ayant le score de propension le plus proche. Le fait d'utiliser également une procédure sans remplacement conduit en outre à réduire la qualité de l'appariement. C'est pourquoi nous appliquons également une deuxième série de restrictions pour constituer notre échantillon d'analyse en utilisant le *kernel matching*, qui

utilise l'ensemble des observations témoins. Chaque témoin se voit attribuer une pondération moyenne en fonction de sa distance à chaque traité, calculée comme la différence entre les scores de propension. L'estimation de l'effet du PIA-ADEME se fera ensuite en utilisant ces pondérations dans les régressions de doubles différences. L'avantage du *kernel matching* est qu'en utilisant plus d'informations, on réduit la variance de nos estimateurs et donc leurs écarts-types, mais on utilise potentiellement des partenaires témoins qui sont de mauvais jumeaux. C'est pourquoi nous suivons les recommandations de Caliendo et Kopeinig (2005) et utilisons une définition plus stricte du support commun : en plus de supprimer de notre échantillon les partenaires traités ayant une probabilité d'être traitée supérieure au maximum des témoins, nous retirons également les partenaires témoins qui auraient une probabilité d'être traitée inférieure au minimum des traités.

Les distributions des probabilités de recevoir une aide du PIA-ADEME pour les échantillons issus du *one-to-one matching* et du *kernel matching* reflètent bien l'arbitrage qu'il y a entre les deux méthodes. Les distributions des groupes traité et témoin sont beaucoup plus homogènes avec l'échantillon du *one-to-one matching* puisqu'on ne conserve que les témoins ayant les scores de propension les plus proches, mais cela implique de ne retenir qu'un petit nombre d'observations (138 traités et témoins). Le *kernel matching* donne un échantillon avec des probabilités plus disparates mais un nombre d'observations plus important, ce qui contribuera à améliorer la précision de nos estimations.

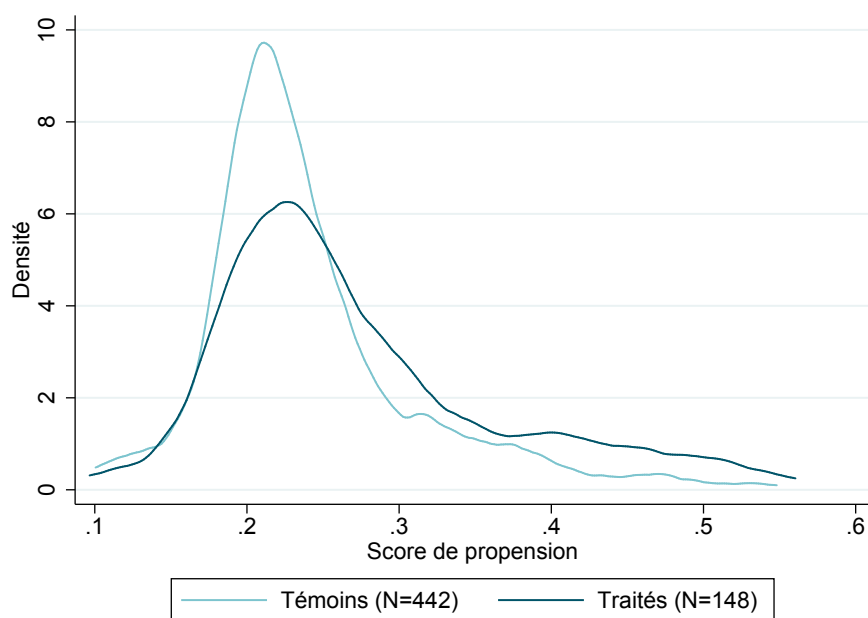
FIGURE 5.7 – Distribution de la probabilité de bénéficier d'une aide du PIA-ADEME - One-to-one matching - Base ADEME-FARE-DADS



LECTURE : La distribution de la probabilité de recevoir une aide se concentre autour de 0,22 pour les groupes traité et témoin.

SOURCE : Base ADEME ; DADS et FARE 2007, 2008, 2009 et 2013.

FIGURE 5.8 – Distribution de la probabilité de bénéficier d'une aide du PIA-ADEME - Kernel matching - Base ADEME-FARE-DADS



LECTURE : La distribution de la probabilité de recevoir une aide se concentre autour de 0,22 pour les groupes traité et témoin.

SOURCE : Base ADEME ; DADS et FARE 2007, 2008, 2009 et 2013.

De la même manière que pour la base ADEME-FARE-RD nous tentons de valider ou invalider l'hypothèse des tendances parallèles à la fois graphiquement et statistiquement. Tous les graphiques sont reportés à l'Annexe E. Le constat qui en ressort est que si la comparabilité des évolutions entre les entreprises traitées et témoins semble être légèrement améliorée par rapport à celles sur la base ADEME-FARE-RD, les graphiques ne semblent toujours pas valider l'hypothèse identifiante. Elle n'est pas non plus supportée par le test « placebo » qui révèle des différences significatives entre les deux groupes (tableaux 5.4 et 5.5). Les entreprises traitées enregistrent en 2007 un excédent brut d'exploitation inférieur de 8,8 millions d'euros à celui des entreprises témoins, ainsi qu'une productivité des salariés inférieure de 14 000 euros, différences toutes deux significatives à 10 %.

TABLEAU 5.4 – Vérification statistique de l'hypothèse des tendances parallèles - Base ADEME-FARE-DADS (*one-to-one matching*)

Variable dépendante	Année 2007 (Erreur-type)	Année 2008 (Erreur-type)	Année 2013 (Erreur-type)	Obs.
Chiffre d'affaires net (millions d'euros)	33.03 (98.93)	5.51 (98.93)	-29.89 (98.93)	1096
Résultat comptable (millions d'euros)	2.82 (59.55)	1.31 (60.06)	-1.50 (59.55)	1096
Taux d'immobilisations incorporelles	1.13 (4.00)	0.85 (4.00)	0.68 (3.99)	1096
Effectifs de R&D	15.17 (47.69)	-3.76 (47.69)	32.80 (47.69)	1096
Productivité des salariés	-15.36 (13.33)	-1.46 (13.31)	1.67 (13.32)	1096
Production totale (millions d'euros)	56.50 (93.45)	3.08 (93.45)	-21.26 (93.45)	1096
Valeur ajoutée HT (millions d'euros)	-1.12 (21.89)	-0.56 (21.89)	-4.42 (21.89)	1096
Excédent brut d'exploitation (millions d'euros)	-4.45 (8.19)	-0.97 (8.19)	-0.37 (8.19)	1096
Dividendes (milliers d'euros)	9118.87 (52812.06)	5230.39 (52812.06)	3399.69 (52812.06)	1096

LECTURE : Les entreprises traitées ont, en 2007, un taux d'immobilisations incorporelles supérieur de 1,13 points de pourcentage à celui des entreprises témoins en 2009, toutes choses égales par ailleurs. Cette différence est non significative.

NOTES : Les coefficients reportés dans ce tableau sont issus de l'estimation de régressions linéaires avec en variable dépendante les indicateurs de performance, et en variables explicatives des effets fixes pour les années ainsi que les termes d'interaction entre le traitement et l'année. 2009 est l'année de référence. Les étoiles indiquent la significativité du coefficient. * : significatif au seuil de 10 % ; ** : significatif au seuil de 5 % ; *** : significatif au seuil de 1 %.

SOURCES : Base ADEME ; FARE et DADS 2007, 2008, 2009 et 2013.

TABLEAU 5.5 – Vérification statistique de l'hypothèse des tendances parallèles - Base ADEME-FARE-DADS (*kernel matching*)

Variable	Année 2007 (Erreur-type)	Année 2008 (Erreur-type)	Année 2013 (Erreur-type)	Obs.
Chiffre d'affaires net (millions d'euros)	21.76 (62.99)	-0.90 (62.99)	-1.75 (62.99)	2212
Résultat comptable (millions d'euros)	-3.85 (9.27)	-2.96 (9.34)	-4.05 (9.27)	2212
Taux d'immobilisations incorporelles	0.80 (2.83)	-0.11 (2.83)	0.39 (2.83)	2212
Effectifs de R&D	9.07 (32.61)	-6.99 (32.61)	19.45 (32.61)	2212
Productivité des salariés	-14.15* (8.23)	-6.21 (8.23)	-5.25 (8.24)	2212
Production totale (millions d'euros)	26.72 (59.21)	-0.86 (59.21)	4.99 (59.21)	2212
Valeur ajoutée HT (millions d'euros)	-9.81 (14.97)	-3.50 (14.97)	-3.42 (14.97)	2212
Excédent brut d'exploitation (millions d'euros)	-8.84* (4.83)	-1.75 (4.83)	-2.49 (4.83)	2212
Dividendes (milliers d'euros)	1451.69 (6587.40)	1491.85 (6587.40)	2717.24 (6587.40)	2212

LECTURE : En 2007, les entreprises traitées ont un taux d'immobilisations incorporelles supérieur de 0,8 points de pourcentage à celui des entreprises témoins. Cette différence est non significative.

NOTES : Les coefficients reportés dans ce tableau sont issus de l'estimation de régressions linéaires avec en variable dépendante les indicateurs de performance, et en variables explicatives des effets fixes pour les années ainsi que les termes d'interaction entre le traitement et l'année. Les coefficients sont pondérés par le poids calculés lors du *kernel matching*. 2009 est l'année de référence. Les étoiles indiquent la significativité du coefficient. * : significatif au seuil de 10 % ; ** : significatif au seuil de 5 % ; *** : significatif au seuil de 1 %.

SOURCES : Base ADEME ; FARE et DADS 2007, 2008, 2009 et 2013.

À l'issue de ces différents tests, il en ressort que les données dont nous disposons ne nous permettent pas de valider les hypothèses nécessaires à l'identification d'un effet causal des aides du PIA-ADEME, à savoir l'hypothèse d'indépendance conditionnelle et celle des tendances parallèles. Si l'utilisation des techniques d'appariement nous permettent d'améliorer, par construction, la comparabilité des deux groupes, elles ne suffisent pas pour conclure à une évolution similaire des groupes traité et témoin sur la période antérieure au traitement. Ceci est dû au fait que les entreprises sont appariées sur des caractéristiques observables qui n'expliquent qu'une infime partie de la probabilité de recevoir une aide, et que le rôle joué par les inobservables est non négligeable, venant ainsi invalider l'hypothèse d'indépendance conditionnelle. Les résultats de l'estimation de l'équation des doubles différences ne peuvent ainsi pas être interprétés comme des effets causaux du PIA-ADEME, mais doivent plutôt être lus comme de simples corrélations. Nous reportons ces résultats à l'Annexe F.

5.2.3 Discussion sur la puissance statistique

Tout au long de cette étude, le faible nombre d'observations a été évoqué comme une des raisons pour lesquelles nous ne pouvions conclure quant à la significativité statistique de nos résultats. En effet, plus le nombre d'observations est faible, plus les coefficients des estimations ou des tests statistiques sont estimés avec du « bruit », c'est-à-dire avec beaucoup d'imprécision. Pour l'évaluation finale, une des questions à se poser afin de bien calibrer le protocole d'évaluation est celle du nombre d'observations minimum nécessaire pour détecter un effet. Nous présentons dans cette partie des calculs de puissance statistique, en prenant pour exemple les variables « effectifs de R&D » et « taux d'immobilisations incorporelles ». Notre capacité à détecter un effet dépend, entre autres, du nombre d'observations, et de la moyenne et de l'écart-type de la variable dans le groupe témoin ¹².

12. Les tests de puissances statistiques réalisés ici sont des tests statistiques classiques et consistent à comparer deux moyennes : μ_0 , la moyenne observée chez le groupe témoin, et μ_1 la moyenne supposée du groupe traité. La différence entre μ_0 et μ_1 est l'effet attendu du traitement. Nous testons ensuite l'hypothèse nulle $H_0 : \mu_0 = \mu_1$ (absence d'effet du traitement) contre l'hypothèse alternative $H_1 : \mu_0 \neq \mu_1$. En utilisant un seuil de significativité standard de 5 % et une puissance du test de 80 %, nous déterminons le nombre d'observations N nécessaire pour rejeter H_0 . Le résultat de ce

Le tableau 5.6 présente les résultats de ces calculs de puissance pour la variable « effectifs de R&D ». Afin de détecter un effet du traitement de +2 emplois en R&D par rapport à une moyenne de 39 emplois dans le groupe témoin, il faudrait un échantillon de 68 296 traités et de 140 000 témoins. La taille de l'échantillon nécessaire pour détecter un effet significatif sur cette variable est particulièrement élevée en raison de la très grande hétérogénéité des observations contenues dans notre échantillon, reflétée dans l'écart-type accompagnant la moyenne qui est de 134,8.

Le constat est moins sévère pour la variable « taux d'immobilisations incorporelles » (tableau 5.7). Pour observer une augmentation de 2 points de pourcentage du taux d'immobilisations incorporelles par rapport à une moyenne de 20,8 %, 841 traités et 1 682 témoins sont nécessaires. L'écart-type associé à la moyenne est en effet beaucoup plus faible pour cette variable (25,3), ce qui nous permet de détecter un effet de cette ampleur avec des tailles d'échantillons plus raisonnables.

Ces calculs soulignent ainsi l'importance de la comparabilité des entreprises des groupes traités et témoins pour être en mesure de détecter des effets significatifs sans avoir besoin d'un échantillon trop large, dont la taille est de toute manière déterminée par le nombre de partenaires bénéficiaires et non bénéficiaires du PIA-ADEME.

test peut ainsi s'exprimer comme suit : N est le nombre d'observations nécessaire pour détecter un effet de taille $\mu_1 - \mu_0$ significatif au seuil de 5 %.

TABLEAU 5.6 – Nombre d'observations nécessaire pour détecter un effet significatif - Variable « Effectifs de R&D »

Nombre d'obs. traitées requis ...	Nombre d'obs. témoins requis ...	Pour détecter un effet de ...
68 296	140 000	2 emplois supp.
17 075	34 149	4 emplois supp.
7 589	15 178	6 emplois supp.
4 270	8 539	8 emplois supp.
2 733	5 465	10 emplois supp.
1 898	3 796	12 emplois supp.
1 395	2 789	14 emplois supp.
1 068	2 136	16 emplois supp.

LECTURE : Pour observer une augmentation de 2 emplois en R&D supplémentaires par rapport à la moyenne des entreprises témoins de 2009, il faudrait un échantillon de 68 296 traités et de 140 000 témoins.

NOTES : La moyenne dans le groupe témoin en 2009 est de 39 emplois avec un écart-type de 134,8. Nous utilisons un niveau de significativité de 5 % et un critère de puissance statistique de 80 % (critères standards).

SOURCES : Base ADEME et DADS 2009.

TABLEAU 5.7 – Nombre d'observations nécessaire pour détecter un effet significatif - Variable « Taux d'immobilisations incorporelles »

Nombre d'obs. traitées requis ...	Nombre d'obs. témoins requis ...	Pour détecter un effet de ...
1 891	3 782	1 point de % supp.
841	1 682	2 points de % supp.
474	947	3 points de % supp.
304	607	4 points de % supp.
211	422	5 points de % supp.
155	310	6 points de % supp.

LECTURE : Pour observer une augmentation de 1 point de pourcentage supplémentaire du taux d'immobilisations incorporelles par rapport à la moyenne des entreprises témoins de 2009, il faudrait un échantillon de 1 891 traités et de 3 782 témoins.

NOTES : La moyenne dans le groupe témoin en 2009 est de 20,8 % avec un écart-type de 25,3. Nous utilisons un niveau de significativité de 5 % et un critère de puissance statistique de 80 % (critères standards).

SOURCES : Base ADEME et FARE 2009.

CHAPITRE 6

BILAN ET RECOMMANDATIONS

L'objectif de ce chapitre est de formuler des recommandations à destination de l'ADEME et du CGI en vue de l'évaluation finale, évaluation dont les conclusions doivent être remises à la Commission européenne en 2020. L'évaluation intermédiaire conduite par l'IPP s'est avérée utile à plusieurs égards, puisqu'elle a permis de mettre en lumière plusieurs difficultés, notamment concernant les données, dont certaines d'entre elles peuvent encore être corrigées pour l'évaluation finale. Elle a également permis de tester la faisabilité et la pertinence de plusieurs méthodes d'évaluation, et de nous éclairer sur ce qui est envisageable pour l'évaluation finale. Les recommandations formulées dans ce chapitre concernent les données transmises par l'ADEME et le CGI, l'utilisation de sources de données externes ainsi que la méthode d'identification d'un effet causal des aides. Pour chacun de ces points, nous procédons en établissant un bilan du travail réalisé et des difficultés rencontrées puis formulons ensuite nos recommandations.

6.1 Les données transmises par l'ADEME et le CGI

6.1.1 Travail réalisé et problèmes rencontrés avec les données ADEME-CGI

Le nombre de fichiers transmis, leur structure et les difficultés de prise en main ou d'extraction ont été évoquées dans le chapitre 3. Nous revenons ici sur les types

de problèmes associés aux fichiers transmis.

Les données transmises par le CGI

Tout d'abord, il convient de préciser que si les données de gestion sur le PIA-ADEME décrites dans ce rapport et mobilisées dans l'analyse sont celles issues de fichiers transmis par l'ADEME, trois fichiers de données nous ont également été transmis par le CGI et n'ont été utilisés qu'à des fins de statistiques descriptives (voir Annexe A). Les données du CGI étant issues des *reportings* des opérateurs du PIA, celles-ci ne différaient pas en termes de contenu de celles fournies par l'ADEME mais plutôt dans leur forme, et contenaient des informations agrégées au niveau de l'ensemble du PIA. Une première base de données regroupe des informations au niveau du partenaire*projet pour les projets contractualisés, synthétisant ainsi des informations présentes dans les fichiers LOCO de l'ADEME. Un second fichier présente les montants et les origines des cofinancements des projets contractualisés dans le cadre du PIA-ADEME. Enfin, un troisième fichier donne, pour chaque partenaire d'un projet bénéficiant d'aides du PIA, le montant total de ces aides. Cette base de données comporte donc des informations qui ne sont pas présentes dans les données ADEME, mais pour autant incomplètes : si le montant total d'aides perçu au titre du PIA nous semble être une variable de contrôle importante pour l'analyse des effets des aides du seul PIA-ADEME, il aurait toutefois fallu connaître la date des différents financements PIA, et notamment savoir si ceux-ci avaient lieu avant ou après 2013. Pour autant, l'information n'aurait pas été exploitable en l'état : l'appariement des données ADEME et CGI, *via* une table de correspondance entre les identifiants des projets (le numéro d'accord cadre pour l'ADEME et le nom du projet dans SISE), n'a pas pu aboutir en raison de divergences, en particulier concernant les dates auxquelles les données issues du CGI ou de l'ADEME ont été arrêtées.

Les données transmises par l'ADEME

La première difficulté dans la prise en main des données transmises par l'ADEME a été le nombre de fichiers et leurs différentes structures, le niveau d'observation étant parfois le projet, le projet*partenaire, le partenaire uniquement ou le par-

tenaire*type de financement, partenaire*dates d'échéances. L'appariement de ces fichiers initiaux a d'autre part fait émerger de nombreux problèmes et incohérences entre fichiers.

Les incohérences entre fichiers étaient principalement dues à des questions de saisie (par exemple, un décalage temporel entre la contractualisation, inscrite dans DIA, et la saisie du projet dans LOCO) et à des dates de « photographie » différentes entre fichiers, les différents fichiers n'ayant pas été tous arrêtés à la même date. Elles résultaient aussi d'une restriction, dans certains fichiers, aux projets relevant d'appels à projet (AAP) ou appels à manifestation d'intérêt (AMI) appartenant au champ de l'analyse, quand les projets dits « hors champ », comme les AAP réservés aux collectivités territoriales, étaient aussi inclus dans d'autres.

Une autre difficulté a résulté du caractère liminaire du dictionnaire de variables, qui avait été transmis en amont pour la sélection des variables d'intérêt pour l'analyse. Les acronymes et la duplication d'observations ou de variables liées à la procédure de financement des projets n'étaient pas facilement compréhensibles par des personnes connaissant mal la gestion de projets du PIA-ADEME, et rendaient moins évidente la sélection des variables pertinentes pour l'analyse.

Les fichiers bruts transmis par l'ADEME (Annexes 5)

Si l'extraction des fichiers bruts des Annexes remplies par les candidats à une aide du PIA-ADEME est une démarche originale, car elle permet de reconstituer un groupe d'entreprises témoins en théorie plus proche que si l'on se contente des entreprises réalisant de la R&D, elle s'est avérée compliquée par plusieurs aspects. Tout d'abord, les dossiers transmis ne contenaient pas les annexes pour l'ensemble des projets soumis et rejetés. Cela a eu pour conséquence principale de restreindre le groupe de partenaires de projets non retenus pouvant potentiellement servir à la constitution du groupe témoin.

Ensuite, une partie des données ainsi extraites n'a pas pu être exploitée. En effet, les numéros Siret étaient souvent mal ou non renseignés puisque, pour une part non négligeable des candidats, il s'agit d'entreprises en cours de création ne disposant pas encore de numéro de Siret. Ceci a rendu impossible l'appariement de

ces observations avec les autres fichiers.

Enfin, la qualité des variables extraites étant relativement médiocre, les informations sur le coût et l'aide demandée par partenaire n'ont pu être utilisées dans le calcul du score de propension notamment, à cause d'un pourcentage trop élevé de valeurs manquantes.

Les difficultés posées dans l'appariement par les données de l'ADEME

Nous avons vu que l'appariement des données ADEME avec des bases de données externes, y compris des fichiers exhaustifs, menait à une perte importante d'observations. Une explication serait que les partenaires de projets contractualisés sont souvent jeunes ou se créent lors de la contractualisation. Par ailleurs les entreprises des secteurs innovants visés pourraient connaître des probabilités de cessation d'activité plus importantes que les entreprises d'autres secteurs. Enfin, les changements de Siren pourraient aussi constituer une explication : ceux-ci concerneraient un très grand nombre de partenaires (près de 80 %). En effet, l'outil de gestion conserve le Siren inactif dans LOCO, mais seul le Siren actif est disponible lors des extractions. Ceci explique une partie des pertes d'observations lors de l'appariement entre les données ADEME et des données externes.

6.1.2 Recommandations sur les données ADEME-CGI

Au regard de ces difficultés dans l'exploitation des données, plusieurs recommandations peuvent ainsi être formulées quant aux données transmises, dans le but de faciliter le travail de l'équipe en charge de l'évaluation finale et de disposer des informations nécessaires à la mise en œuvre de la méthodologie.

Recommandation n°1 : bien définir le champ de l'étude et fournir des éléments de contexte

- En amont de la mise en œuvre de l'évaluation, il pourrait être envisagé que l'ADEME définisse, conjointement avec l'équipe d'évaluation sélectionnée, le champ de l'évaluation, afin de déterminer les types d'aides, les AAP ou les

AMI qui paraissent trop spécifiques pour être inclus dans l'étude, comme les AAP réservés aux collectivités territoriales, ou encore les projets financés sous forme de prises de participation. Idéalement, l'ensemble des bases de données extraites et des éventuels fichiers bruts transmis aux chercheurs devrait être restreint au champ de l'étude préalablement défini, afin de ne fournir que les observations pertinentes pour l'analyse et de faciliter les appariements entre les différents fichiers.

- Par ailleurs, si des éléments de contexte ont été transmis à l'équipe de l'IPP, ils se sont avérés incomplets, et l'information disponible publiquement demeure souvent partielle. Comprendre la façon dont ont été accordées les aides ainsi que les modes de sélection des projets et les évolutions de ces derniers au cours du temps est indispensable pour les évaluateurs. L'expertise des acteurs en charge de la politique à évaluer est en effet essentielle pour saisir les enjeux et les limites d'une stratégie d'identification. Ces éléments ont pu être reconstitués par l'IPP, à la suite d'une réunion prévue à cet effet avec l'ADEME et le CGI. Pour l'évaluation finale, il serait ainsi souhaitable qu'une telle réunion avec l'équipe de recherche soit prévue dès le début de l'évaluation et fasse éventuellement l'objet d'un document écrit.

Recommandation n°2 : constituer une base unique et documentée à partir des fichiers de gestion et s'assurer de la cohérence entre les variables

- Afin de gagner un temps précieux, il est impératif que l'ADEME transmette à l'équipe de recherche en charge de l'évaluation finale une unique base de données qui puisse être directement exploitable, et non une multitude de fichiers aux informations redondantes comme ce fut le cas pour l'évaluation intermédiaire. Ceci implique que l'ADEME poursuive ses efforts de mise en cohérence des données, qui est réalisé en amont des remontées d'informations mensuelles pour le CGI. Le format de cette base de données pourra par ailleurs s'inspirer de celui de la base constituée par l'IPP au niveau du partenaire, chaque partenaire étant identifié par son numéro de Siret. Il sera ainsi plus aisé d'apparier les données de l'ADEME avec d'autres sources de données.

- Dans la mesure du possible, il est essentiel que l'ADEME conserve une trace des changements de Siren des entreprises partenaires de projets contractualisés. L'exploitation de cette information, connue des gestionnaires pour le versement des aides, permettrait un meilleur appariement des données du PIA-ADEME avec les données externes.
- L'application de filtres excluant les projets qui ne rentrent pas dans le champ de l'analyse devra être effectuée par l'ADEME.
- Enfin, pour que l'équipe d'évaluation s'approprie plus facilement les données de gestion de l'ADEME, la base de données doit être accompagnée d'un dictionnaire des variables afin de préciser la signification des variables et des différentes modalités qu'elles peuvent prendre. D'autre part, l'information sur les détails des procédures et sur leur saisie informatique est nécessaire pour comprendre la nature et la structure des données ; celle-ci doit accompagner la description des bases et des variables. En plus de cette documentation écrite, une réunion entre l'ADEME et l'équipe d'évaluation pourrait être prévue au lancement de l'évaluation afin que la base de données soit présentée, comme cela avait été fait pour la présente étude.

Recommandation n°3 : reconstituer les changements de Siren des partenaires non-contractualisés

- Si le nouvel outil de gestion et de suivi des affaires de l'ADEME pourra éventuellement permettre de garder une trace datée des changements d'identifiants des partenaires contractualisés, un moment de l'étude devra être consacré à retracer les changements de Siren des entreprises du groupe témoin sur la période d'observation retenue. Cela pourra être fait à l'aide de la base de données SIRENE, désormais en accès libre depuis le 1^{er} janvier 2017. Cette base en accès libre n'est toutefois pas disponible sur les années antérieures à 2017. L'historique est néanmoins disponible sur le Centre d'accès sécurisé à distance (CASD) et pourra être demandé lors de l'évaluation finale.

Recommandation n°4 : uniformiser et verrouiller le format des annexes pour en faciliter l'extraction automatique

- La base de données constituée à partir des annexes 5 et qui rassemble, pour l'ensemble des entreprises partenaires de projets non retenus, les informations sur leur Siret, le coût et le montant d'aide demandé, pourrait être réalisée en interne par l'ADEME, pour ensuite être transmise à l'équipe en charge de l'évaluation.
- Dans le cas où cela ne pourrait être fait par l'ADEME, plusieurs recommandations peuvent être formulées pour faciliter et réduire le travail d'extraction des annexes par l'équipe en charge de l'évaluation finale :
 - Plutôt que de transmettre de nombreux dossiers contenant eux-mêmes des sous-dossiers, il serait utile que les annexes soient toutes rassemblées au sein d'un seul et unique dossier qui sera ensuite transmis à l'équipe en charge de l'évaluation finale.
 - Pour limiter le travail de lecture automatique des fichiers, les annexes doivent toutes avoir le même format, ou à défaut être rassemblées dans différents dossiers selon le type de format du document que les entreprises ont eu à remplir.
 - Ces documents doivent impérativement être en format Excel. Les fichiers PDF ne peuvent être lus automatiquement (sans avoir recours à la reconnaissance d'images).
 - Les champs nécessaires à l'analyse, et en particulier le Siret, doivent être verrouillés par l'ADEME, de sorte que le candidat ne puisse les laisser vierge et ne puisse renseigner un identifiant qui n'ait pas le bon format (Siret à 13 chiffres au lieu de 14, par exemple) ou un montant qui ne soit pas exprimé dans la bonne unité. L'unité doit en effet être homogène pour tous les projets afin d'éviter des retraitements et de permettre des comparaisons statistiques directes.

Recommandation n°5 : dater l'obtention des différents PIA dans la Requête 2 des données SISE

- Afin de pouvoir contrôler du fait de recevoir des aides PIA de la part d'opérateurs autres que l'ADEME, il est important de connaître la date à laquelle ces aides ont été perçues. Il est donc nécessaire que cette information soit consignée dans le système d'information SISE du CGI. Idéalement, il serait utile de pouvoir disposer de cette information depuis la création du PIA. Cette information étant renseignée au niveau du partenaire, elle pourrait alors être appariée avec les autres bases de données à partir du numéro de Siret.

6.2 Les données externes

6.2.1 Travail réalisé et problèmes rencontrés avec les données externes

Les données transmises par l'ADEME ont permis d'identifier les entreprises partenaires de projets contractualisés ou non, et de les classer dans les groupes « témoin » ou « traité » selon la date de contractualisation. Elles ne permettaient pas de conduire en tant que telle l'évaluation, celle-ci reposant sur des variables d'intérêt non présentes dans ces données : des variables indiquant la performance des entreprises, des variables décrivant les moyens mis en œuvre pour la R&D (effectifs, dépenses, etc.) et des variables décrivant les *outputs* de la R&D, c'est-à-dire l'innovation, communément mesurée par les dépôts de brevets.

Pour obtenir ces variables d'intérêt, comme décrit dans le chapitre 3, les données externes qui ont été mobilisées sont les données DADS, FARE et l'Enquête R&D. Comme cela a été détaillé dans ce chapitre, deux difficultés ont considérablement réduit notre échantillon d'entreprises, et ont conduit à mobiliser des données moins précises sur nos variables d'intérêt. La première difficulté est le caractère non-exhaustif de l'Enquête R&D. Cela induit par ailleurs une non-représentativité de l'échantillon apparié, l'Enquête R&D étant exhaustive pour les entreprises ayant d'importantes dépenses de R&D. Au vu du faible nombre d'observations résultant

de l'appariement et de sa faible représentativité, il a été décidé de procéder à un appariement des données ADEME et FARE avec les données DADS. Des *proxys* des variables d'intérêt ont pu être construits (voir tableau 5.2), au prix d'une perte de précision sur les activités de R&D. Bien que la perte d'observations soit moindre, le fait que notre étude porte sur des secteurs d'activités novateurs et en plein essor implique que les créations et cessions d'entreprises sont potentiellement plus fréquentes, ce qui expliquerait le nombre de Siren non présents dans nos bases les deux années qui nous intéressent.

Il convient de noter que, conformément à la décision de la Commission européenne sur le plan d'évaluation notifié, l'accès à d'autres bases de données externes avait été demandé pour cette étude. Il s'agit des données GECIR, CLAP et LIFI. L'accès aux données PATSTAT et Jeunes entreprises innovantes (JEI) n'a pas été demandé.

- Les données CLAP, au niveau postes, n'apportaient pas d'information supplémentaire à celles fournies par les données DADS Postes qui sont des données beaucoup plus précises. Les données CLAP avaient toutefois été demandées par mesure de précaution, dans le cas où les services producteurs de données et le Comité du secret statistique n'auraient pas donné leur accord pour l'utilisation des DADS.
- Les données GECIR sont les données de gestion sur le crédit impôt recherche (CIR). Elles donnent, pour les entreprises demandant le CIR, les variables nécessaires pour son calcul, comme les dépenses internes de R&D, ainsi que le montant du CIR versé. Ces données, en cours de reconfiguration à la Direction générale des finances publiques, n'ont pas pu être accessibles durant l'évaluation intermédiaire. Ces données, désormais quasi exhaustives pour les entreprises effectuant des activités de R&D au vu des réformes récentes du CIR (Dortet-Bernadet et Sicsic, 2015), constituent une alternative intéressante aux données de l'Enquête R&D pour l'évaluation finale.
- Les données LIFI n'ont pas été mobilisées pour l'étude. Celles-ci, dont l'accès est payant car reposant en partie sur les données Diane collectées par le Bureau van Dijk (2 500 euros par chercheur), auraient permis de préciser les

contours des entreprises aidées, afin d'agrèger les montants d'aide reçus et les variables d'intérêt au niveau du groupe et non des filiales. En effet, de nombreuses entreprises candidates au PIA-ADEME sont des filiales de grands groupes effectuant de la R&D dans le domaine environnemental¹. L'utilisation de cette base devra toutefois être envisagée pour l'évaluation finale puisqu'une des attentes de la Commission Européenne est que l'évaluation soit menée au niveau de la maison-mère.

- Les données PATSTAT sur les brevets déposés, dont l'accès est payant (6 500 euros), n'ont finalement pas été demandées : aucun effet des aides du PIA-ADEME sur les *inputs* de R&D n'ayant pu être identifié, il n'était par conséquent pas justifié de s'intéresser aux effets sur les *outputs* de R&D dans le cadre de cette évaluation intermédiaire.
- Les données Jeunes entreprises innovantes (JEI) étaient proposées dans le plan d'évaluation notifié à la Commission européenne mais elles n'ont pas été demandées pour cette étude. Elles auraient notamment pu permettre de compléter les données de l'Enquête R&D sur les petites entreprises grâce aux informations sur l'emploi en R&D, et de contrôler par le montant des exonérations de cotisations sociales octroyées dans le cadre de ce dispositif, mais l'essentiel de cette information est déjà disponible dans GECIR. En effet, outre les informations sur l'emploi en R&D, GECIR nous indique si l'entreprise bénéficie du dispositif JEI.

6.2.2 Recommandations sur les données externes

Recommandation n°6 : mieux prendre en compte le délai lié aux procédures d'accès aux données dans le calendrier de l'étude

- L'accès aux données externes *via* le Comité du secret statistique est une procédure longue. Celle-ci comprend la constitution du dossier, les contacts avec les services producteurs des différentes données, une attente de plus de deux mois entre le dépôt du dossier complété et la réunion du comité du secret (qui

1. Les données dont nous disposons ne nous permettent toutefois pas de quantifier la part d'entreprises candidates au PIA-ADEME qui sont des filiales de grands groupes.

a lieu tous les trois mois environ), la signature d'engagements et d'autorisations des membres de l'équipe et des services producteurs, une déclaration CNIL pour les données dites « ménage » (comme les DADS Postes), et un délai de deux mois entre la déclaration CNIL et l'autorisation d'accès aux données « ménage ». Ces délais peuvent en outre être allongés lors de la période estivale. Ceci doit être pris en compte dans le calendrier de l'évaluation finale, afin que la phase de constitution des bases appariées et des analyses à proprement parler ne soit pas trop courte. Entre trois et six mois peuvent ainsi s'écouler entre le moment de la constitution du dossier et celui de l'accès effectif aux données.

Recommandation n°7 : mener une réflexion en amont sur les données externes utiles à l'analyse

- Parce que les procédures à mener auprès du Comité du secret peuvent s'avérer relativement longues, il est indispensable de mener en amont une réflexion sur les sources de données externes qu'il est nécessaire de mobiliser pour l'évaluation. Une telle réflexion mériterait d'être menée avec l'équipe de recherche sélectionnée au démarrage de l'étude.
- L'accumulation de sources de données est par ailleurs coûteuse en temps, car la prise en main de chacune de ces sources n'est jamais immédiate. Si cette prise en main dépend fortement de l'expérience de l'équipe d'évaluation pour le traitement de ces données, un travail de « nettoyage » des données est dans tous les cas nécessaire. En effet, les formats des bases de données ne sont en général pas identiques d'une année sur l'autre : par exemple les variables peuvent changer de nom, ou ne sont pas calculées de la même manière et ne sont donc pas directement comparables. Nous estimons le temps de travail sur ces données afin de les rendre exploitables à un mois et demi environ.

Recommandation n°8 : utiliser la base SIRENE pour mieux documenter les pertes d'observations entre les différentes bases de données

- Que ce soit au moment de l'appariement entre deux bases différentes (par exemple ADEME et FARE) ou au moment de l'ajout d'une base similaire mais d'une année différente (par exemple l'ajout de FARE 2013 à la base ADEME-FARE 2009), on perd un nombre relativement important d'observations. Cela pourrait être dû au caractère innovant des entreprises aidées par le PIA-ADEME, pour lesquelles les créations et les destructions seraient plus fréquentes. Une exploitation de la base SIRENE, qui répertorie tous les établissements et les entreprises sur le territoire français, pourrait permettre de mieux documenter et de quantifier les raisons expliquant ces pertes d'observations lors des appariements, du moins du point de vue de la création d'entreprises (les cessations d'activités y sont moins bien renseignées).
- Pour documenter les pertes d'observations lors des appariements, les informations de la base SIRENE pourraient éventuellement être complétées par l'utilisation des variables « État statistique de l'entreprise » et « Date de l'état statistique » disponibles dans FARE, qui renseignent sur le statut actif ou inactif de l'entreprise. Si ces variables sont plutôt bien renseignées, leur qualité est assez médiocre car le statut est souvent mis à jour avec un décalage temporel.

Recommandation n°9 : exploiter les bases GECIR et/ou « Jeunes entreprises innovantes » plutôt que l'Enquête R&D pour les variables liées aux activités de R&D

- Si encore trop peu d'entreprises ont demandé le CIR sur la période d'observation de l'évaluation intermédiaire (faisant des données GECIR des données non exhaustives sur les entreprises ayant des activités de R&D)², l'exploitation des données de la base GECIR semble constituer une piste prometteuse pour l'évaluation finale. En effet, en 2020, un plus grand nombre d'entreprises

2. N'ayant pas pu avoir accès aux fichiers GECIR dans le cadre de ce projet, nous ne sommes pas en mesure de quantifier la part d'entreprises bénéficiant du PIA-ADEME qui sont également présentes dans ces bases.

auront demandé le CIR, notamment parmi les petites entreprises (Dortet-Bernadet et Sicsic, 2015). Cette base contient les principales variables d'intérêt : les dépenses de R&D (dépenses de personnel en R&D, dépenses de fonctionnement, dépenses liées aux brevets, *etc.*)³, les autres aides à la R&D perçues par l'entreprise comme le dispositif « jeunes entreprises innovantes » ou la réduction d'impôt en faveur des PME, *etc.* GECIR pourrait ainsi être utilisée à la place de l'Enquête R&D. L'appariement avec cette source de données devrait conduire à perdre moins d'observations au moment de l'évaluation finale.

- La base de données « Jeunes entreprises innovantes » (JEI) pourra éventuellement être demandée pour compléter les informations sur les emplois en R&D pour les petites entreprises si celles-ci ne sont pas déjà présentes dans la base GECIR. Elle pourra également permettre de contrôler du montant exact d'exonération de cotisations sociales pour les entreprises ayant bénéficié de ce dispositif. Néanmoins, comme cette base ne concerne qu'un peu moins de 4 000 entreprises (contre près de 22 000 pour la base GECIR), et que le gain d'informations qu'elle apporte est marginal, cette base pourrait être demandée dans un deuxième temps, s'il s'avère que les très petites entreprises sont mal couvertes par la base GECIR⁴.

Recommandation n°10 : utiliser la base PATSTAT si des effets sont observés sur les *inputs* de R&D

- Pour l'évaluation finale, si un effet est observé sur les *inputs* de R&D, il sera alors intéressant d'exploiter la base de données PATSTAT. Celle-ci décrit les dépôts et citations de brevets au niveau européen. En effet, on a vu (chapitre 4) que le nombre de brevets déposés constitue une bonne mesure de

3. Les dépenses éligibles au crédit d'impôt recherche couvrent l'essentiel des dépenses de R&D. Pour le calcul du CIR en 2016, sont uniquement exclues les dépenses liées au personnel de soutien (administration), les frais relatifs au dessin, modèle et marque de fabrique ainsi que les primes et les cotisations d'assurance.

4. Il est en effet possible d'ajouter des bases de données supplémentaires à un projet de recherche sans devoir repasser devant le Comité du secret : l'ajout de sources fait l'objet d'une consultation électronique du Comité, et cette procédure est assez rapide, bien que ces consultations soient à dates fixes. Il sera d'ailleurs plus facile de justifier de l'utilisation de la base JEI une fois établi que la base GECIR est incomplète pour les jeunes entreprises.

l'innovation au niveau de l'entreprise. Si cette base est payante, elle présente l'avantage d'être rapidement accessible.

- En revanche, elle implique un important travail pour pouvoir être appariée avec les données de l'ADEME sur le numéro Siren. En effet, la base PATSTAT ne contient pas l'identifiant Siren des entreprises mais uniquement leur raison sociale. Il faut donc dans un premier temps harmoniser les raisons sociales de PATSTAT avec celles du répertoire SIRENE, ce qui nécessite entre autres de coder un algorithme de reconnaissance de fautes d'orthographe. Ensuite, les observations de PATSTAT peuvent être appariées avec les données du répertoire SIRENE sur la raison sociale. Un troisième temps doit être consacré à l'identification des bons et des mauvais « matchs », afin de contrôler la qualité de l'algorithme. Bozio et al. (2016), qui ont eu à faire ce travail dans le cadre de leur travaux d'évaluation des effets du CIR, estiment une durée de travail de 3 à 6 mois⁵.

6.3 Méthodes économétriques

6.3.1 La méthodologie actuelle et ses limites

Pour l'évaluation intermédiaire, notre choix s'est rapidement porté vers la méthode des doubles différences, avec un groupe traité composé des entreprises ayant bénéficié d'une aide de l'ADEME et un groupe de contrôle retenu constitué des entreprises partenaires de projets déposés mais n'ayant pas été retenus pour bénéficier d'une aide. Cette méthodologie était une de celles envisagées dans le plan d'évaluation notifié et validé par la Commission européenne.

Outre le fait que cette méthode semblait pertinente au regard des données disponibles, elle semblait aussi particulièrement innovante, au regard de la littérature existante. En effet, rares sont les dispositifs pour lesquels on connaît les entreprises candidates mais non retenues : les travaux existants considèrent à défaut le plus

5. Si les auteurs souhaitaient initialement rendre public cet algorithme de « sirenisation » de la base PATSTAT, l'autorisation ne leur a pas été accordée. Ce travail devra donc être refait lors de l'évaluation finale.

souvent, pour le groupe témoin, le *pool* des entreprises nationales effectuant de la R&D (si l'information est disponible), parmi lesquelles un échantillon est constitué après une procédure de *matching* avec les entreprises traitées les plus comparables. Or, l'information sur la candidature des entreprises aux aides évaluées permet *a priori* de capturer un grand nombre d'inobservables sur l'activité innovante de l'entreprise, son domaine d'innovation (ici, la transition énergétique), *etc.*

Cependant, comme on l'a rappelé, la mise en œuvre de cette méthode d'évaluation suppose que l'hypothèse de tendances parallèles soit nécessairement vérifiée. Cette hypothèse n'a pu être validée dans le cadre de l'évaluation intermédiaire, même lorsque nous n'utilisons pas les données de l'Enquête R&D afin de disposer d'un plus grand nombre d'observations et de mettre en place une procédure de *matching* pour améliorer la comparabilité du groupe de traitement et du groupe de contrôle.

Cette difficulté majeure pourrait être renforcée par l'évolution du taux de sélection des financements PIA-ADEME à partir de 2014. En effet, hors appels à projet IPME, la sélection des projets se joue désormais en grande partie en amont du dépôt de dossier, lors de l'accompagnement par l'ADEME des candidatures, qui opère une préselection importante. Ainsi, aujourd'hui, une très large majorité des dossiers candidats sont retenus, ce qui décroît mécaniquement le nombre de partenaires témoins. Ces constats doivent amener une nouvelle réflexion sur la stratégie d'identification pour l'évaluation finale.

6.3.2 Recommandations sur les méthodes économétriques

Recommandation n°11 : tester d'autres groupes témoins pour la méthode des doubles différences

- La disponibilité et le caractère exhaustif des données GECIR au moment de l'évaluation finale pourra éventuellement permettre de conserver la stratégie d'identification proposée pour l'évaluation intermédiaire, si cela permet la constitution de groupes témoin et traité plus importants et donc de sélectionner, parmi ces groupes, des sous-groupes réellement comparables.

- Toutefois, quelles que soient les nouvelles pistes d'appariement de données, l'évolution des règles de sélection à partir de 2014 pour les projets hors IPME, entraînant une baisse du nombre d'entreprises rejetées au cours du temps signalée par l'ADEME, risque d'impliquer un changement de la définition du groupe témoin pour l'évaluation finale. Plusieurs possibilités peuvent être envisagées, comme la constitution d'un groupe témoin à partir d'entreprises effectuant de la R&D – mais pas nécessairement candidates à des aides de l'ADEME – identifiables dans les données GECIR ou dans l'Enquête R&D. La mise en œuvre d'un appariement sur score de propension permettrait ensuite d'en sélectionner un sous-groupe qui soit comparable au groupe traité. Cette stratégie, proposée initialement dans le plan d'évaluation soumis à la Commission européenne, a pour limite principale le rôle joué par des variables inobservables qui déterminent l'activité innovante des entreprises ainsi que leurs performances⁶.
- Les projets IPME pourraient être évalués à part, étant donné que les règles de sélection n'ont pas été modifiées pour ces projets spécifiques. Ceci permettrait de conserver un groupe témoin constitué des partenaires de projets rejetés, si les tailles d'échantillon le permettent.

Recommandation n°12 : utiliser une période de traitement variant en fonction de la date de réception de l'aide

- Plutôt que de retenir une période de traitement allant du début de la période d'observation à la fin (période 2009-2013 pour cette étude), il est également possible de retenir une période de traitement dont la date de début varie pour chaque observation. La date t d'entrée en traitement serait ainsi « glissante » et dépendrait de la date de dépôt de dossier du partenaire.
- Si cette option n'a pu être mise en œuvre faute de temps, cette définition de la période de traitement aurait l'avantage de permettre de mieux décomposer l'effet des aides du PIA-ADEME année après année, plutôt que de considé-

6. Cette stratégie n'a pas été testée dans le cadre de l'évaluation intermédiaire car le nombre d'entreprises témoins était suffisant pour pouvoir mener l'évaluation.

rer l'effet moyen sur la période. L'information utilisée serait également plus riche puisque les données sur les années intermédiaires seraient également utilisées.

- Puisque la contrainte d'être présent en 2009 et en 2013 dans les données serait levée, cela permettrait enfin d'avoir un échantillon d'analyse plus représentatif des entreprises aidées par le PIA-ADEME s'il s'avérait que la raison pour laquelle certaines entreprises ne se retrouvent pas lors des appariements entre les bases de début et de fin de période était liée à une probabilité de cessation d'activité plus élevée que la moyenne. En revanche, cela rendrait difficile la vérification des tendances parallèles pour les entreprises créées au moment de la contractualisation.

Recommandation n°13 : analyser l'hétérogénéité des effets des aides du PIA-ADEME

- Si pour l'évaluation intermédiaire l'analyse de l'hétérogénéité des effets n'a pu être menée, elle pourra être conduite pour l'évaluation finale puisque la taille de l'échantillon, et par conséquent la puissance statistique, devraient être plus importantes, à condition que les hypothèses nécessaires pour l'analyse d'effets soient vérifiées. Dans ce cas, il sera nécessaire de réfléchir aux différentes dimensions d'hétérogénéité qui sont pertinentes pour l'étude. Par exemple, des analyses différenciées selon la taille d'entreprise, selon le type d'aide (aide mixte ou uniquement subventions, appel à projet « Initiative PME » ou non), selon la durée de traitement ou encore, selon le programme PIA pourront être menées.

Recommandation n°14 : mettre en place un système de notation des projets permettant l'évaluation par régressions sur discontinuité

- Les problèmes de comparabilité des groupes témoin et traité pourraient être fortement minimisés si une autre stratégie d'identification peut être justifiée et appliquée. La mise en place, par l'ADEME, d'un système cohérent de notation des projets déposés permettrait de mettre en place une stratégie d'iden-

tification utilisant les régressions sur discontinuités (*Regression Discontinuity Design*). En effet, si un tel système existait, il y aurait alors un seuil au-dessous duquel les projets ne seraient pas subventionnés par l'ADEME, ce qui rendrait possible l'identification *causale* d'un effet des aides autour de la discontinuité ainsi créée, l'hypothèse identifiante étant que des projets situés directement de part et d'autre du seuil sont quasiment similaires. Cela permettrait en outre de rendre le processus de sélection des projets plus transparent.

6.4 Synthèse des recommandations

Le tableau suivant synthétise les différentes recommandations, que nous classons selon deux catégories : d'une part les éléments que nous jugeons essentiels à traiter en vue de l'évaluation finale, et d'autre part des méthodologies alternatives qui pourront être testées. Il nous paraît en effet prioritaire de concentrer les efforts sur la constitution d'un groupe témoin pertinent (*point 1*) et de maximiser la taille de l'échantillon global (*point 2*). Des méthodes alternatives pourront dans un second temps être testées, comme faire varier la période de traitement en fonction de la date d'obtention de l'aide (*point 3*) ou mettre en place un système de notation permettant la mise en œuvre des régressions sur discontinuités (*point 4*). Nous précisons à chaque fois les différents problèmes survenus lors de l'évaluation intermédiaire, les solutions envisageables, ainsi que les problèmes pouvant subsister ou survenir lors de l'évaluation finale.

Éléments nécessaires à l'évaluation	Problèmes survenus au cours de l'évaluation	Solutions envisageables pour l'évaluation finale	Problèmes pouvant subsister/survenir pour l'évaluation finale
<p>1. Constitution d'un groupe témoin pertinent</p> <p><i>Essentiel pour construire un contrefactuel crédible</i></p> <p>1.1 Définir un groupe témoin de taille suffisante</p>	<p><i>Les éléments suivants ont contribué à limiter le nombre d'observations parmi lesquelles nous pouvions "piocher" les témoins, ainsi que la qualité des témoins retenus :</i></p> <ul style="list-style-type: none"> - Le nombre d'annexes 5 transmis ne couvrait pas l'ensemble des partenaires de projets rejetés : seuls les partenaires de 561 projets sur 1009 ont été transmis, dont 192 concernent des projets validés. - Le pourcentage non négligeable de Siren manquants ou non renseignés a conduit à retirer de l'échantillon environ 13,5 % de témoins - Ne disposant que des numéros de Siren renseignés au moment du dépôt de dossier, nous n'avons pu apparier les observations témoins pour lesquelles le Siren a changé au cours de la période, cas relativement fréquent. - La baisse du nombre de partenaires rejetés à partir de 2014, sauf pour les IPME, limite notre capacité à utiliser les partenaires rejetés comme groupe témoin. Ce problème sera accentué lors de l'évaluation finale. 	<ul style="list-style-type: none"> - Conserver l'ensemble des annexes des partenaires de projets déposés (retenus ou non). Le nouvel outil de gestion CRM devrait permettre de régler ce problème - Verrouiller le champ Siren rempli par les entreprises et imposer un format de 14 chiffres (Recommandation n°4). - Reconstituer les changements de Siren des partenaires non contractualisés en utilisant la base SIRENE accessible via le CASD (Recommandation n°3) - Traiter à part les IPME, pour lesquels le groupe témoin pourra toujours être constitué des partenaires de projets rejetés - Pour les projets hors IPME, si le nombre de partenaires de projets rejetés s'avère trop petit, constituer un groupe témoin à partir des entreprises réalisant de la R&D, identifiables à partir de l'Enquête R&D ou des données GECIR (Recommandation n°11) 	<ul style="list-style-type: none"> - "Piocher" les entreprises témoin parmi des entreprises effectuant de la R&D, similaires aux entreprises traitées sur un certain nombre de caractéristiques, n'exclue pas le rôle de caractéristiques inobservables qui peuvent avoir un effet et sur la probabilité d'être traité, et sur les variables d'intérêt (activités de R&D, performances).

<p>1.2 Assurer la comparabilité du groupe témoin</p>	<p>- Le manque d'observables, notamment sur les activités de R&D des entreprises, sur les coûts et le montant d'aide demandé, et sur les autres aides à la R&D perçues par les partenaires, a conduit à choisir comme témoin des entreprises relativement peu comparables aux traités, même à l'issue de la procédure d'appariement.</p>	<p>- Verrouiller les champs du coût et de l'aide demandée remplis par les entreprises et imposer un format uniforme (en milliers d'euros par exemple) (Recommandation n°4) - Utiliser GECIR (et éventuellement JEI) pour les informations sur les activités de R&D (Recommandation n°9) - Dater la réception des autres aides PIA dans SISE (Recommandation n°5)</p>	
<p>2. Maximiser la taille de l'échantillon global <i>Essentiel pour garantir une certaine puissance statistique</i></p>	<p><i>Les éléments suivants ont contribué à limiter la taille de l'échantillon d'analyse final :</i></p> <p>- L'Enquête R&D étant exhaustive uniquement sur les grandes entreprises, l'appariement de cette base avec la base ADEME fait perdre près de deux tiers des observations, et nous a conduit à utiliser les bases FARE et DADS, à la place de l'Enquête R&D, pour obtenir des proxys pour les variables de R&D et augmenter la taille de notre échantillon</p> <p>- Les créations et disparitions d'entreprises étant plus fréquentes sur la population d'entreprises aidées par le PIA-ADEME, l'appariement avec plusieurs années de données conduit à une perte d'observations</p> <p>- Les partenaires traités pour lesquelles le Siren a changé au cours du temps n'ont pas pu être appariés avec les bases externes</p>	<p>- Utiliser la base GECIR qui devrait être quasi exhaustive sur la période concernée par l'évaluation finale, éventuellement complétée de JEI (Recommandation n°9)</p> <p>- Faire varier la période de traitement en fonction de la date réelle de début de traitement (permet de récupérer les entreprises créées après 2009) et de fin de traitement (permet de récupérer les entreprises ayant cessé leur activité avant 2013) (Recommandation n°12)</p> <p>- Conserver l'historique des changements de Siren dans l'outil de gestion de l'ADEME, ou reconstituer ces changements à partir des données SIRENE (Recommandation n°3)</p>	<p>- Impossible de vérifier l'hypothèse des tendances parallèles pour les entreprises créées au moment de l'octroi de l'aide</p>

Méthodologies alternatives	Problèmes survenus au cours de l'évaluation	Solutions envisageables pour l'évaluation finale	Problèmes pouvant subsister/survenir pour l'évaluation finale
<p>3. Faire varier la période de traitement en fonction de la date d'obtention de l'aide</p> <p><i>Option intéressante car permet d'utiliser plus d'information (en utilisant toutes les années et pas uniquement 2009 et 2013), de récupérer les entreprises créées après 2009, et de mieux décomposer l'effet du PIA-ADEME année après année.</i></p>	<ul style="list-style-type: none"> - Cette méthodologie n'a pas été mise en œuvre pour les raisons suivantes : le temps consacré à la préparation des données ADEME ainsi que le délai lié aux procédures d'accès aux données ont réduit le temps disponible pour la préparation des données externes. Seules les années 2007 et 2008 (pour les tendances parallèles) et 2009 et 2013 ont pu être traitées. - Cette méthode revient à décomposer l'effet du PIA-ADEME sur des sous-groupes en fonction de l'année de traitement. Des problèmes de puissance statistique étant déjà présents pour l'analyse sur l'ensemble de l'échantillon, ils auraient été d'autant plus importants pour une analyse par sous-groupes. 	<ul style="list-style-type: none"> - Transmettre une base ADEME unique et documentée (Recommandations n°1 & 2) - Prévoir plus de temps pour le traitement des données externes et mieux prendre en compte les délais liés à l'obtention des données (Recommandation n°6) - Améliorer la taille de l'échantillon global pour que les sous-groupes soient plus importants (voir points 1 et 2 du tableau) 	<ul style="list-style-type: none"> - Perte d'observations éventuelle au moment des appariements avec les années intermédiaires - Impossible de vérifier l'hypothèse des tendances parallèles pour les entreprises créées au moment de l'octroi de l'aide
<p>4. Utiliser la méthode de régression sur discontinuité (RDD)</p>	<ul style="list-style-type: none"> - Cette méthode nécessite de disposer d'un système de notation et d'un seuil au-dessus duquel une entreprise bénéficie du PIA-ADEME. Une telle note n'existe pas actuellement. 	<ul style="list-style-type: none"> - Mettre en place un système de notation pour l'octroi des aides du PIA-ADEME (Recommandation n°14) 	<ul style="list-style-type: none"> - Risque que le nombre d'observations soit trop faible (l'effet n'est estimé que sur les observations situées de part et d'autre de la discontinuité)

ANNEXE A. LES DONNÉES DU COMMISSARIAT GÉNÉRAL À L'INVESTISSEMENT

Les données transmises par le Commissariat général à l'investissement (CGI) consistent en différentes extractions faites à partir du Système d'Information pour le Suivi de l'Evaluation (SISE) des Programmes d'Investissement d'Avenir (PIA). Les différents opérateurs des PIA doivent faire remonter les informations relatives aux projets financés. Ces informations sont ensuite mises en forme par le système d'informations du CGI. Nous présentons dans cette annexe les trois fichiers qui nous ont été transmis dans le cadre de cette étude et qui ont été utilisés à des fins de statistiques descriptives dans le chapitre 1.

- **Requête 1** : La première base transmise est une base obtenue à la suite d'une requête faite au niveau du projet*partenaires pour le PIA-ADEME uniquement. Elle présente pour chacun des projets l'ensemble des entreprises partenaires et donne des informations relatives à l'action dans laquelle s'inscrit le projet (« Véhicule du futur » ou « Démonstrateurs de la transition écologique ») et au statut du projet (sélectionné, contractualisé ou abandonné). Pour chacun des partenaires, le type de partenaire (grande entreprise, petite et moyenne entreprise, etc.) et le secteur d'activité d'après la Nomenclature d'activités française (NAF) sont renseignés. Cette base contient également des variables de localisation géographique des projets financés. Enfin, les montants de subventions et d'aides remboursables sont reportés, et la distinction est faite entre les montants autorisés, contractualisés et décaissés. Une fois

restreinte au champ de notre étude (c'est-à-dire une fois exclues les observations correspondant aux appels à projets réservés aux collectivités territoriales ainsi que les prises de participation), nous dénombrons 670 partenaires de 272 projets, soit un total de 940 observations⁷.

- **Requête 2** : La seconde base est au niveau du partenaire et contient des informations sur les aides perçues dans le cadre de l'ensemble des PIA gérés par le CGI, et non pas uniquement les partenaires des projets qui ont pour opérateur l'ADEME. Cette base permet de disposer d'informations sur le total des montants perçus au titre des PIA, que ce soit sous forme de subventions, d'avances remboursables, de prises de participation, de prêts, de dotations à des fonds de garantie ou encore de dotations non-consommables. Au total, ce sont 3411 partenaires qui sont présents dans cette base, identifiables grâce à leur numéro de Siret. Cette base de données aurait été particulièrement intéressante afin de contrôler du montant d'aide reçu, autre que celui du PIA-ADEME, dans nos estimations. Mais ne connaissant pas la date à laquelle ces différentes aides ont été reçues, cette base n'a pu être exploitée.
- **Requête 3** : Enfin, la troisième base nous renseigne sur le montant des cofinancements de 385 projets qui ont pour opérateur l'ADEME et précise s'il s'agit d'un cofinanceur privé, public ou d'une collectivité territoriale. Le montant est au niveau du projet ; nous ne disposons pas du montant des cofinancements reçu au niveau du partenaire.

À partir de ces trois fichiers, deux bases ont été créées à des fins de statistiques descriptives. La première base est une base au niveau du partenaire, identifiable grâce à son numéro de Siret, et permet de savoir à combien de projets financés par le PIA un même partenaire participe et quels sont les montants totaux perçus. Cette base est obtenue en appariant la Requête 1 avec la Requête 2 sur le numéro Siret, et contient 670 partenaires uniques. La seconde base est au niveau du projet et résulte de l'appariement entre les Requête 1 et 3 de SISE. Elle nous renseigne sur le nombre de partenaires pour chacun des 273 projets soumis ainsi que sur les

7. Comme certains partenaires participent à plusieurs projets, il est normal que le nombre d'observations total soit supérieur au nombre de partenaires.

cofinancements obtenus pour chaque projet.

ANNEXE B. LES EMPLOIS R&D DANS LA NOMENCLATURE D'ACTIVITÉS FRANÇAISE

Chaque poste des déclarations annuelles de données sociales (DADS) est classifié selon la nomenclature d'activités françaises (NAF). Nous avons considéré qu'un poste correspondait à un emploi en R&D s'il correspondait à une des catégories ci-dessous :

- Ingénieurs et cadres d'étude :
 - Catégorie 383a : Ingénieurs et cadres d'étude, recherche et développement en électricité, électronique
 - Catégorie 383b : Ingénieurs et cadres de fabrication en matériel électrique, électronique
 - Catégorie 384a : Ingénieurs et cadres d'étude, recherche et développement en mécanique et travail des métaux
 - Catégorie 384b : Ingénieurs et cadres de fabrication en mécanique et travail des métaux
 - Catégorie 385a : Ingénieurs et cadres d'étude, recherche et développement des industries de transformation (agroalimentaire, chimie, métallurgie, matériaux lourds)
 - Catégorie 385b : Ingénieurs et cadres de fabrication des industries de transformation (agroalimentaire, chimie, métallurgie, matériaux lourds)
 - Catégorie 386b : Ingénieurs et cadres d'étude, recherche et développement de la distribution d'énergie, eau

- Catégorie 386c : Ingénieurs et cadres d'étude, recherche et développement des autres industries (imprimerie, matériaux souples, ameublement et bois)
- Catégorie 388a : Ingénieurs et cadres d'étude, recherche et développement en informatique
- Technicien de recherche-développement :
 - Catégorie 473b : Technicien de recherche-développement et des méthodes de fabrication en électricité, électromécanique et électronique
 - Catégorie 474b : Technicien de recherche-développement et des méthodes de fabrication en construction mécanique et travail des métaux
 - Catégorie 475a : Technicien de recherche-développement et des méthodes de production des industries de transformation
 - Catégorie 479a : Technicien des laboratoires de recherche publique ou de l'enseignement

ANNEXE C. CARACTÉRISTIQUES DES ENTREPRISES TRAITÉES ET TÉMOINS DE LA BASE ADEME-FARE-DADS EN 2009

Le tableau C.1 reporte les résultats des tests de comparaison de moyennes entre les groupes traité et témoin. Les entreprises des deux groupes de notre échantillon sont sensiblement différentes sur les caractéristiques de 2009. Les entreprises traitées ont une production supérieure de 94 millions d'euros, et elles réalisent un plus gros chiffre d'affaires avec 122 millions d'euros de plus que les entreprises témoins). Elles sont également de taille plus importante, comme en témoigne l'effectif moyen de salariés qui avoisine les 840 employés contre un peu moins de 320 pour les témoins. Leurs activités de R&D semblent également plus développées puisqu'elles comptent en moyenne 108 salariés en R&D contre seulement 32 dans le groupe témoin.

TABLEAU C.1 – Comparaison des moyennes des groupes traité et témoin en 2009 - Base ADEME-DADS-FARE

Nom de la variable	Moyenne témoins	Moyenne traitées	Différence	Obs.
Capital social (millions d'euros)	126.3	96.2	-30.2	613
Total de l'actif (millions d'euros)	555.8	694.7	138.9	613
Total du passif (millions d'euros)	481.0	566.6	85.6	613
Rentabilité des actifs	0.0	0.0	-0.0	608
Chiffre d'affaires net (millions d'euros)	149.9	272.3	122.4**	613
Chiffre d'affaires à l'export (millions d'euros)	64.9	116.3	51.4	613
Valeur ajoutée HT (millions d'euros)	38.3	72.8	34.5***	613
Excédent brut d'exploitation (millions d'euros)	6.1	3.1	-2.9	613
Résultat comptable (millions d'euros)	19.9	8.3	-11.6	613
Production totale (millions d'euros)	131.3	225.5	94.2**	613
Taux d'immobilisations incorporelles	20.8	18.7	-2.1	605
Taux investissement incorporel	697.7	40.7	-657.0	577
Total des immobilisations (millions d'euros)	71.9	123.7	51.8*	613
Salaires bruts versés (millions d'euros)	12.5	32.2	19.7***	613
Effectif moyen salariés	322.9	838.2	515.3***	613
Effectifs de R&D	32.2	108.0	75.8***	613
Productivité des salariés	67.9	66.3	-1.6	610

LECTURE : En 2009, les entreprises du groupe traités ont en moyenne un capital social de 96,2 millions d'euros, contre 126,3 millions pour celles du groupe témoin. Cette différence n'est toutefois pas significative.

NOTE : Les étoiles indiquent la significativité de la différence de moyenne entre les deux groupes.

* : significatif au seuil de 10 % ; ** : significatif au seuil de 5 % ; *** : significatif au seuil de 1 %.

SOURCES : Base ADEME ; FARE et DADS 2009.

ANNEXE D. VÉRIFICATION STATISTIQUE DES TENDANCES PARALLÈLES SUR LA BASE ADEME-RD-FARE

Les résultats présentés dans le tableau ci-dessous sont les résultats du test « placebo » afin de vérifier si les tendances différentes observées graphiquement sont statistiquement significatives. Les coefficients sont ainsi issus de la régression de la variable dépendantes sur une indicatrice de traitement, les effets fixes temporels, ainsi que sur l'interaction du traitement et des indicatrices temporelles.

TABLEAU D.1 – Vérification statistique de l'hypothèse des tendances parallèles - Base ADEME-FARE-RD

Variable dépendante	Année 2007 (Erreur-type)	Année 2008 (Erreur-type)	Année 2013 (Erreur-type)	Obs.
Chiffre d'affaires net (millions d'euros)	-17.69 (666.49)	64.44 (666.49)	155.52 (666.49)	940
Résultat comptable (millions d'euros)	65.39 (128.55)	-59.42 (128.87)	-31.00 (128.55)	940
Taux d'immobilisations incorporelles	-0.32 (4.57)	-3.18 (4.56)	-2.14 (4.56)	940
Productivité des salariés	6.77 (16.31)	-11.67 (16.32)	3.75 (16.30)	940
Production totale (millions d'euros)	57.22 (675.51)	113.70 (675.51)	165.77 (675.51)	940
Valeur ajoutée HT (millions d'euros)	47.74 (229.97)	5.09 (229.97)	53.95 (229.97)	940
Excédent brut d'exploitation (millions d'euros)	48.25 (109.80)	6.59 (109.80)	67.01 (109.80)	940
Dividendes (milliers d'euros)	-31074.19 (91999.80)	-28386.13 (91999.80)	2082.78 (91999.80)	940

LECTURE : En 2007, les entreprises traitées ont un taux d'immobilisations incorporelles inférieur de 0,32 points de pourcentage à celui des entreprises témoin. Cette différence n'est pas significative.

NOTES : Les coefficients reportés dans ce tableau sont issus de l'estimation de régressions linéaires avec en variables dépendantes les indicateurs de performances, et en variables explicatives des effets fixes pour les années ainsi que les termes d'interaction entre le traitement et l'année. 2009 est l'année de référence.

SOURCES : Base ADEME ; FARE 2007, 2008, 2009 et 2013.

ANNEXE E. VÉRIFICATION GRAPHIQUE DES TENDANCES PARALLÈLES SUR LA BASE ADEME-FARE-DADS

Les graphiques présentés dans cette annexes représentent les trajectoires des entreprises des groupes traité et témoin sur la période pré-traitement et post-traitement sur les variables suivantes : le taux d'immbilisations incorporelles, la productivité des salariés, la production totale, le résultat comptable, et l'effectif de R&D. Les graphiques de la partie 6.4 ont été réalisés sur la base ADEME-FARE-DADS complète, ceux des parties 6.4 et 6.4 sur les bases ADEME-FARE-DADS après l'appariement *one-to-one* et *kernel* respectivement.

Les tendances parallèles sur l'échantillon ADEME-FARE-DADS

FIGURE E.1 – Evolution du taux d'immobilisations incorporelles pour la période 2007-2013 (en pourcentage)- Base ADEME-FARE-DADS

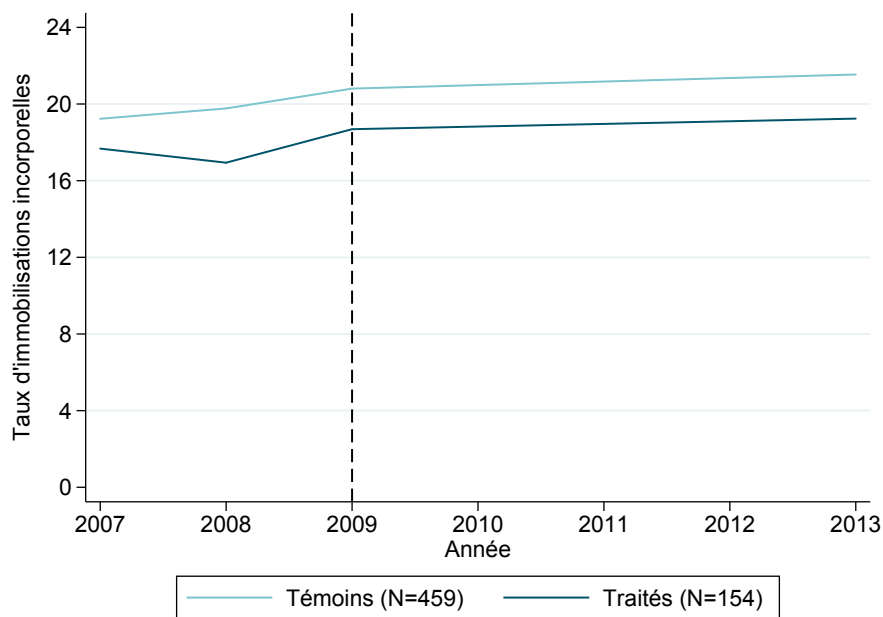


FIGURE E.2 – Evolution de la productivité des salariés pour la période 2007-2013 - Base ADEME-FARE-DADS

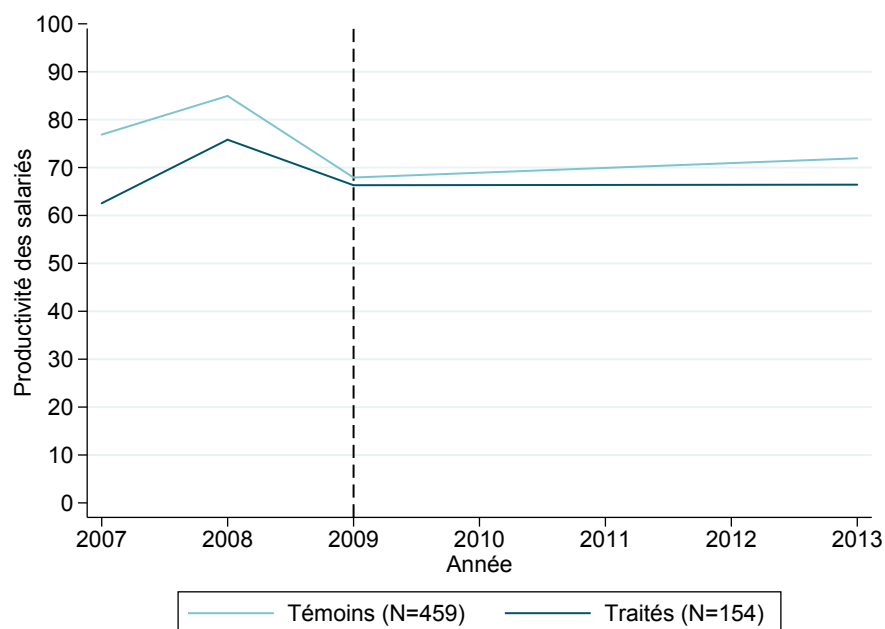


FIGURE E.3 – Evolution de la production totale pour la période 2007-2013 - Base ADEME-FARE-DADS

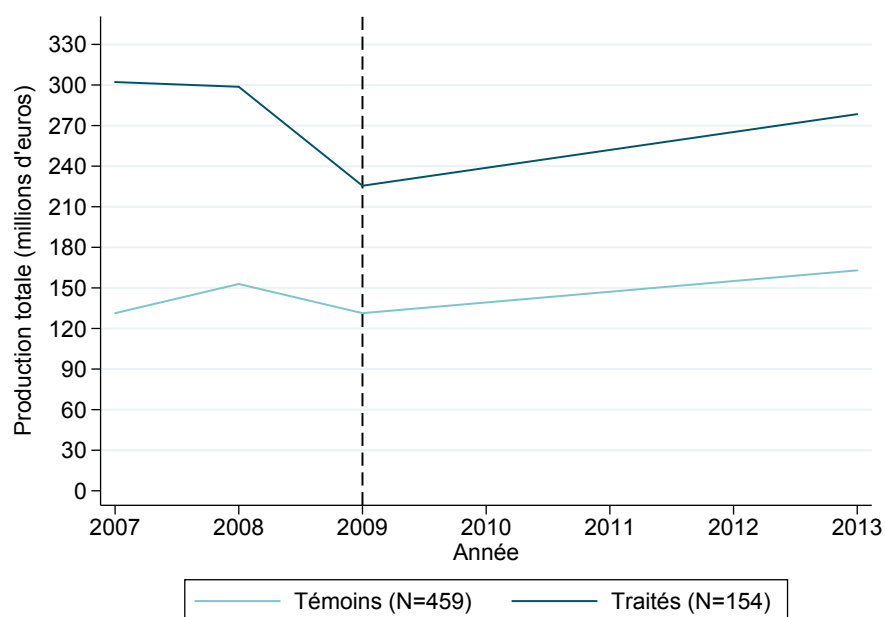


FIGURE E.4 – Evolution du résultat comptable pour la période 2007-2013 - Base ADEME-FARE-DADS

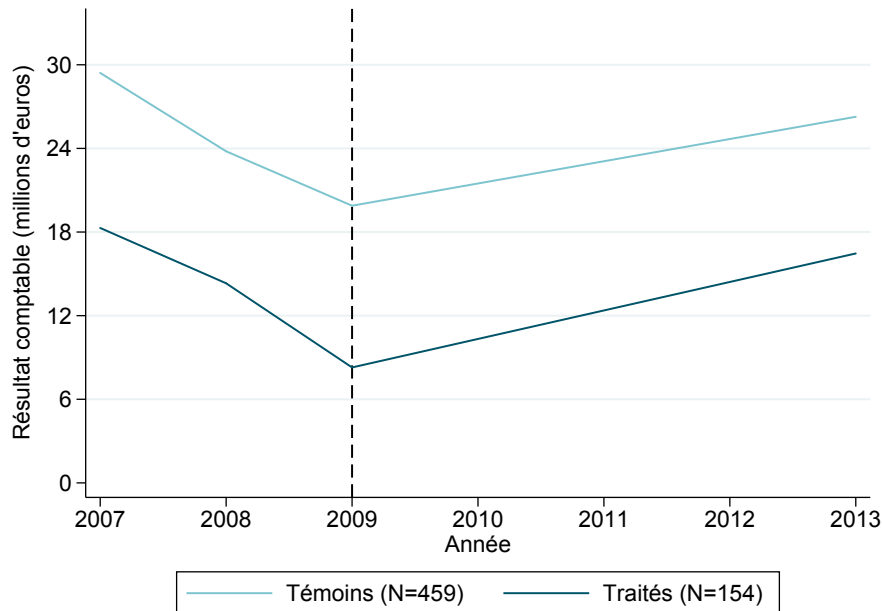


FIGURE E.5 – Evolution des effectifs en R&D pour la période 2007-2013 - Base ADEME-FARE-DADS



Les tendances parallèles sur l'échantillon ADEME-FARE-DADS - *one-to-one matching*

FIGURE E.6 – Evolution du taux d'immobilisations incorporelles pour la période 2007-2013 (en pourcentage)- Base ADEME-FARE-DADS (*one-to-one matching*)

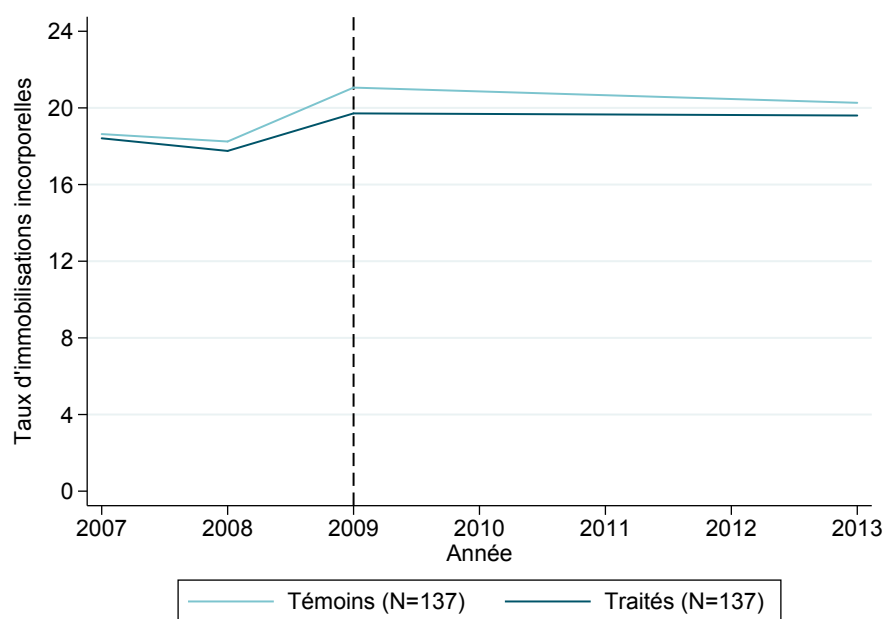


FIGURE E.7 – Evolution de la productivité des salariés pour la période 2007-2013 - Base ADEME-FARE-DADS (*one-to-one matching*)

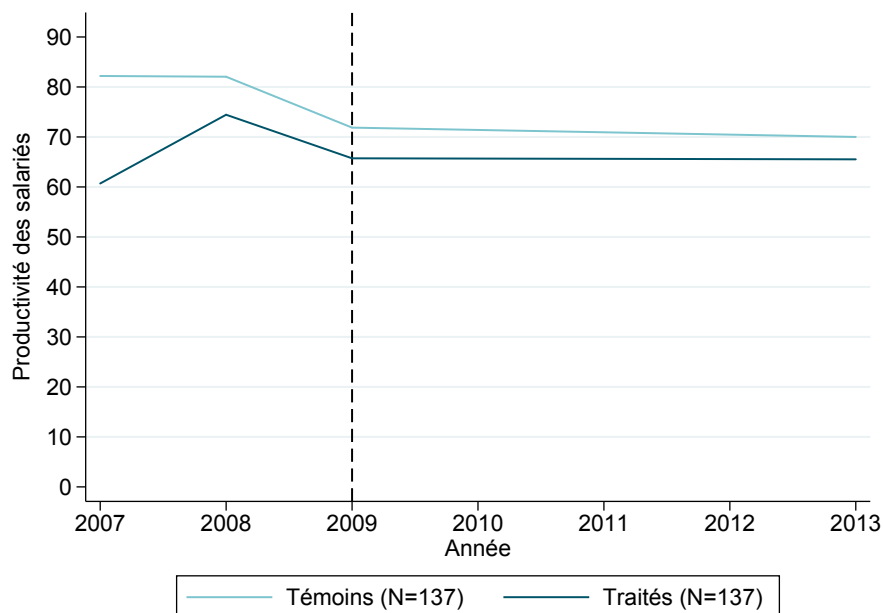


FIGURE E.8 – Evolution de la production totale pour la période 2007-2013 - Base ADEME-FARE-DADS (*one-to-one matching*)

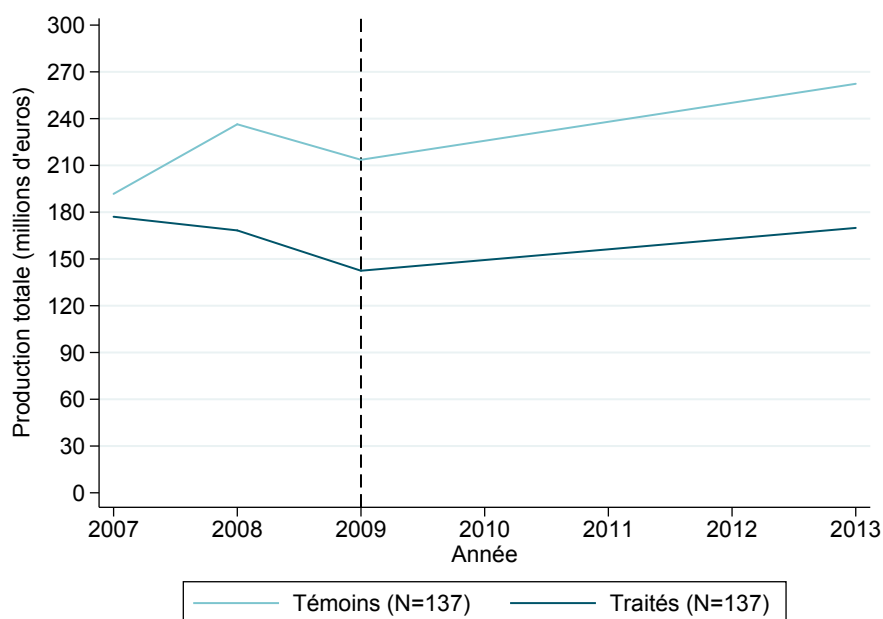


FIGURE E.9 – Evolution du résultat comptable pour la période 2007-2013 - Base ADEME-FARE-DADS (*one-to-one matching*)

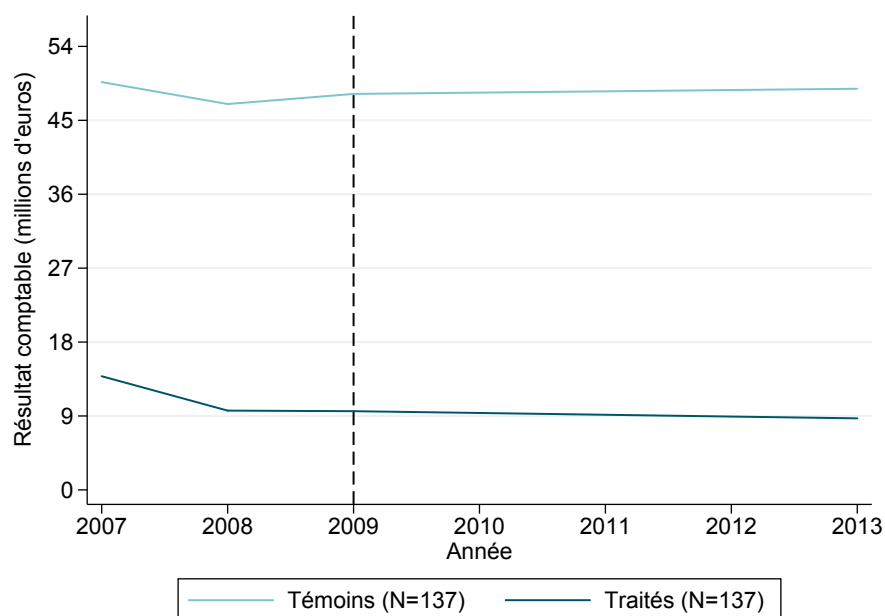


FIGURE E.10 – Evolution des effectifs en R&D pour la période 2007-2013 - Base ADEME-FARE-DADS (*one-to-one matching*)



Les tendances parallèles sur l'échantillon ADEME-FARE-DADS - *kernel matching*

FIGURE E.11 – Evolution du taux d'immobilisations incorporelles pour la période 2007-2013 (en pourcentage)- Base ADEME-FARE-DADS (*kernel matching*)

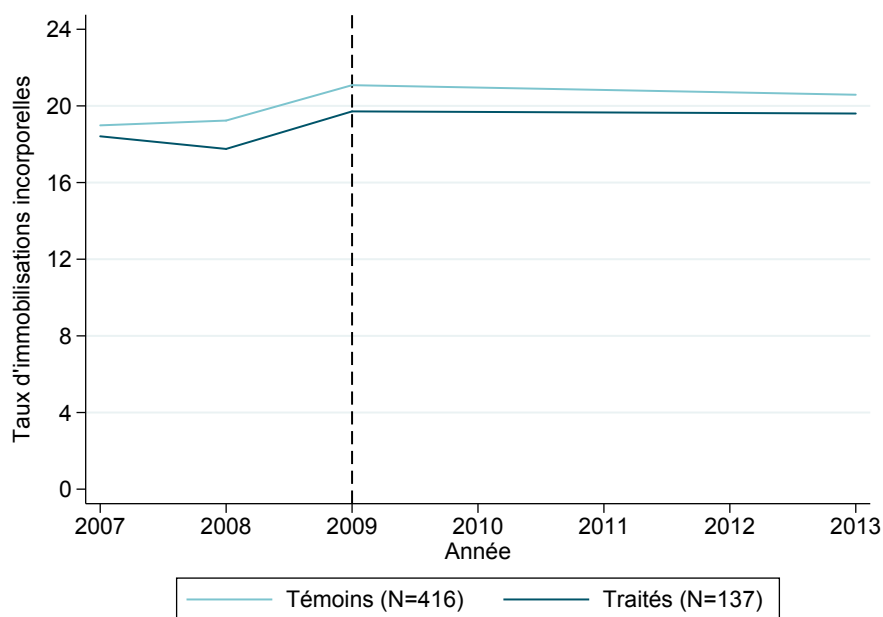


FIGURE E.12 – Evolution de la productivité des salariés pour la période 2007-2013 - Base ADEME-FARE-DADS (*kernel matching*)

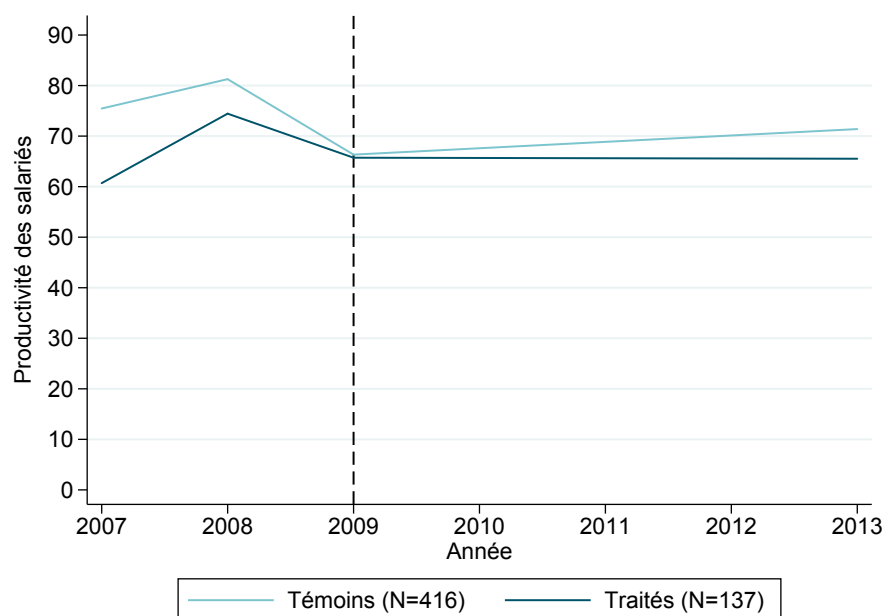


FIGURE E.13 – Evolution de la production totale pour la période 2007-2013 - Base ADEME-FARE-DADS (*kernel matching*)



FIGURE E.14 – Evolution du résultat comptable pour la période 2007-2013 - Base ADEME-FARE-DADS (*kernel matching*)

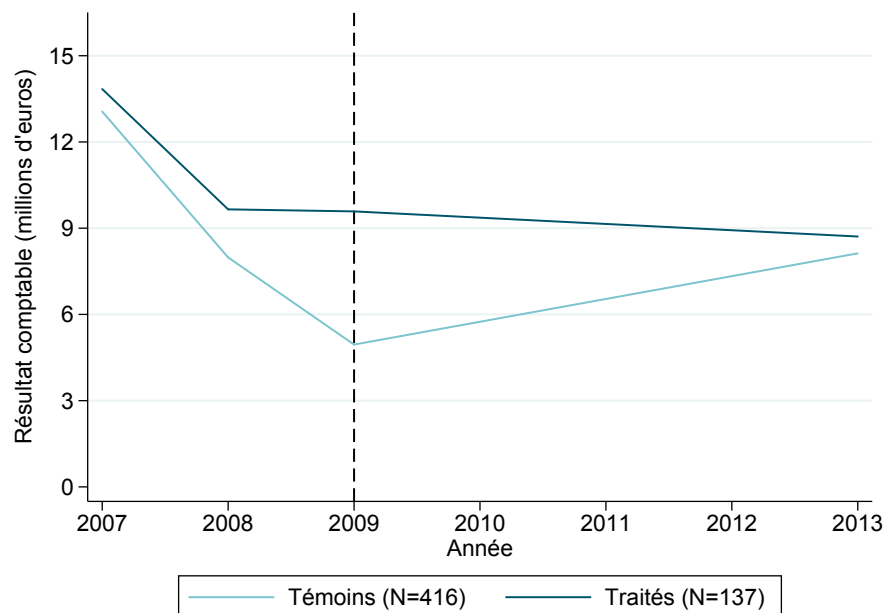


FIGURE E.15 – Evolution des effectifs en R&D pour la période 2007-2013 - Base ADEME-FARE-DADS (*kernel matching*)



ANNEXE F. RÉSULTATS DE L'ESTIMATION DES EFFETS DES AIDES DU PIA-ADEME

Les tableaux ci-dessous présentent les résultats de l'estimation économétrique des effets du PIA-ADEME sur les indicateurs de performances économiques, d'emploi et de dépenses de R&D présentés dans le tableau 3.6 sur les différentes bases d'analyses (estimation de l'équation 5.1). Comme nous l'avons montré au chapitre 5, l'hypothèse de tendances parallèles, cruciale pour interpréter de manière causale l'effet du traitement, n'a pu être validée. Les résultats suivants doivent être interprétés comme des corrélations entre le fait de bénéficier d'une aide du PIA-ADEME et les différents indicateurs, et non comme des effets causaux.

Résultats de l'estimation sur la base ADEME-FARE-RD

TABLEAU F.1 – Résultats des régressions - ADEME-FARE-RD

Variable dépendante	Coefficient	Erreur-type	Obs.
Rentabilité des actifs	-0.03***	0.01	470
Ratio budget R&D sur les ventes	0.92	1.48	470
Effectifs en R&D	36.21	30.48	470
Nombre de chercheurs	-29.59	23.83	470
Productivité des salariés	8.79*	4.86	470
Nombre de brevets déposés	-18.13	15.26	470
Taux d'investissements incorporels	-963.85	1056.49	470
Part des exports/ventes	11.24**	4.68	470
Production totale (millions d'euros)	12.70	36.19	470
Marge commerciale (millions d'euros)	-0.11	8.24	470
Valeur ajoutée HT (millions d'euros)	-35.17	26.86	470
Chiffre d'affaires net (millions d'euros)	-8.89	38.72	470

LECTURE : Sur la période *post* traitement, les entreprises traitées ont environ 36 emplois en R&D équivalent temps plein de plus que celles du groupe témoin. Cette différence n'est toutefois pas significative.

NOTES : Les coefficients reportés dans ce tableau sont issus de l'estimation de l'équation 5.1. * : significatif au seuil de 10 % ; ** : significatif au seuil de 5 % ; *** : significatif au seuil de 1 %. Les erreurs-types sont clusterisées au niveau du secteur d'activité.

SOURCES : Base ADEME ; Enquête R&D et FARE 2009 et 2013.

Résultats de l'estimation sur la base ADEME-FARE-DADS à l'issue du *one-to-one matching*

TABLEAU F.2 – Résultats des régressions - Base ADEME-DADS-FARE (*one-to-one matching*)

Variable dépendante	Coefficient	Erreur-type	Obs.
Rentabilité des actifs	-0.03**	0.01	548
Effectifs de R&D	17.03	18.04	548
Productivité des salariés	-0.66	9.52	548
Taux investissement incorporel	43.35	79.51	548
Part des exports/ventes	1.83	3.67	548
Production totale (millions d'euros)	-6.81	6.86	548
Marge commerciale (millions d'euros)	-2.32**	1.16	548
Valeur ajoutée HT (millions d'euros)	1.37	4.50	548
Chiffre d'affaires net (millions d'euros)	6.70	7.46	548

LECTURE : Sur la période *post* traitement, les entreprises traitées ont environ 17 emplois en R&D de plus que celles du groupe témoin. Cette différence n'est toutefois pas significative.

NOTES : Les coefficients reportés dans ce tableau sont issus de l'estimation de l'équation 5.1. * : significatif au seuil de 10 %; ** : significatif au seuil de 5 %; *** : significatif au seuil de 1 %. Les erreurs-types sont clusterisées au niveau du secteur d'activité.

SOURCES : Base ADEME; DADS et FARE 2009 et 2013.

Résultats de l'estimation sur la base ADEME-FARE-DADS à l'issue du *kernel matching*

TABLEAU F.3 – Résultats des régressions - Base ADEME-DADS-FARE (*kernel matching*)

Variable dépendante	Coefficient	Erreur-type	Obs.
Rentabilité des actifs	-0.04***	0.01	1106
Effectifs de R&D	25.49	17.46	1106
Productivité des salariés	4.37	3.45	1106
Taux investissement incorporel	61.93	73.53	1106
Part des exports/ventes	4.72	4.31	1106
Production totale (millions d'euros)	-0.21	6.59	1106
Marge commerciale (millions d'euros)	-1.31*	0.74	1106
Valeur ajoutée HT (millions d'euros)	-2.96	4.65	1106
Chiffre d'affaires net (millions d'euros)	-1.27	6.10	1106

LECTURE : Sur la période *post* traitement, les entreprises traitées ont environ 25,5 emplois en R&D de plus que celles du groupe témoin. Cette différence n'est toutefois pas significative.

NOTES : Les coefficients reportés dans ce tableau sont issus de l'estimation de l'équation 5.1. * : significatif au seuil de 10 % ; ** : significatif au seuil de 5 % ; *** : significatif au seuil de 1 %. Les erreurs-types sont clusterisées au niveau du secteur d'activité.

SOURCES : Base ADEME; DADS et FARE 2009 et 2013.

RÉFÉRENCES

- Abramovsky, L., Kremp, E., López, A., Schmidt, T., et Simpson, H. (2005). “Understanding cooperative R&D activity : Evidence from four European countries”. *Institute for Fiscal Studies Working Paper*.
- Acemoglu, D., Aghion, P., Bursztyn, L., et Hémous, D. (2012). “The environment and directed technical change”. *The American Economic Review*, 102(1), p. 131–166.
- ADEME (2013). « Investissements d’Avenir à l’Ademe, accélérer l’innovation verte ».
- ADEME (2015). « Développement d’une méthodologie d’évaluation ex post du programme Investissements d’Avenir opéré par l’Ademe ».
- Aghion, P., Dechezleprêtre, A., Hémous, D., Martin, R., et Van Reenen, J. (2016). “Carbon taxes, path dependency, and directed technological change : Evidence from the auto industry”. *Journal of Political Economy*, 124(1).
- Angrist, J. et Pischke, J. (2008). *Mostly harmless econometrics : An empiricist’s companion*. Princeton university press.
- Baron, J., Y., M., et Pohlmann, T. (2011). “R&D coordination in standard setting organizations : The role of consortia”. *7th International Conference on SITT*.
- Baumol, W. J. (2002). *The Free-Market Innovation Machine : Analyzing the Growth Miracle of Capitalism*. Princeton University Press.
- Béllego, C. et Dortet-Bernadet, V. (2014). « L’impact de la participation aux pôles de compétitivité sur les PME et les ETI ». *Economie et Statistiques*, 471.

- Bloom, N., Griffith, R., et Van Reenen, J. (2002). “Do R&D tax credits work? Evidence from a panel of countries 1979-1997”. *Journal of Public Economics*, 85, p. 1–31.
- Blundell, R. et Costa Dias, M. (2002). “Alternative approaches to evaluation in empirical microeconomics”. *Portuguese Economic Journal*, 1(2), p. 91–115.
- Bozio, A., Irac, D., et Py, L. (2016). “Impact of research tax credit on R&D and innovation : evidence from the 2008 French reform”. *Working paper*.
- Branstetter, L. et Sakakibara, M. (1998). “Japanese research consortia : a microeconomic analysis of industrial policy”. *Journal of Industrial Economics*, 46(2), p. 207–233.
- Bronzini, R. et Iachini, E. (2014). “Are incentives for R&D effective ? Evidence from a regression discontinuity approach”. *American Economic Journal*, 6(2), p. 100–134.
- Bronzini, R. et Piselli, P. (2016). “The impact of R&D subsidies on firm innovation”. *Research Policy*, 45(2), p. 442–457.
- Caliendo, M. et Kopeinig, S. (2005). “Some practical guidance for the implementation of propensity score matching”. *IZA Discussion paper series*, 1588.
- Cameron, A. et Trivedi, P. (2005). *Microeconometrics : methods and applications*. Cambridge university press.
- Cappelen, A., Raknerud, A., et Rybalka, M. (2012). “The effects of R&D tax credits on patenting and innovations”. *Research Policy*, 41(2), p. 334–345.
- Carayannis, E. et Provan, M. (2008). “Measuring firm innovativeness : towards a composite innovation index built on firm innovative posture, propensity and performance attributes”. *Innovation and Regional Development*, 1(1).
- Cerulli, G. (2010). “Modelling and Measuring the Effect of Public Subsidies on Business R&D : A Critical Review of the Econometric Literature ”. *Economic Record*.

- Chaminade, C. et Edquist, C. (2010). “Rationales for public policy intervention in the innovation process : A systems of innovation approach”. In Elgar, E., editor, *The Theory and Practice of Innovation Policy : An International Research Handbook*, p. 95–114. Cheltenham.
- Cohen, W. et Levinthal, D. (1989). “Innovation and Learning : The Two Faces of R&D”. *The Economic Journal*, 99(397), p. 569–596.
- Commissariat général à l’investissement (2015). « Rapport d’activité ».
- Commission européenne (2014a). “ Common methodology for state aide evaluation”. *Commission Staff Working Document*.
- Commission européenne (2014b). « Encadrement des aides d’État à la recherche, au développement et à l’innovation ». *Communication de la Commission*, (2014/C 198/01).
- Cour des comptes (2015). « Le programme d’investissements d’avenir : une démarche exceptionnelle, des dérives à corriger ». *Rapport public thématique*.
- Czarnitzki, D., Ebersberger, B., et Fier, A. (2007). “The relationship between R&D collaboration, subsidies and R&D performance : Empirical evidence from Finland and Germany”. *Journal of Applied Econometrics*, 22(7), p. 1347–1366.
- Czarnitzki, D. et Fier, A. (2003). “Publicly Funded R&D Collaborations and Patent Outcome in Germany”. ZEW Discussion Papers 03-24, ZEW - Zentrum für Europäische Wirtschaftsforschung / Center for European Economic Research.
- Czarnitzki, D., Hanel, P., et Rosa, J. (2011). “Evaluating the impact of R&D tax credit on innovation : a microeconomic study on Canadian firms”. *Research Policy*, 40(2), p. 217–229.
- Dasgupta, P. et Stiglitz, J. (1989). “Industrial Structure and the Nature of Innovative Activity”. *The Economic Journal*, 90(358).
- D’Aspremont, C. et Jacquemin, A. (1988). “ Cooperative and noncooperative R&D in duopoly with spillovers”. *The American Economic Review*, 78(5), p. 1133–1137.

- David, P. A., Hall, B. H., et Toole, A. A. (2000). "Is public R&D a complement or substitute for private R&D? A review of the econometric evidence". *Research Policy*, 29, p. 497–529.
- De Marchi, V. (2012). "Environmental innovation and R&D cooperation : Empirical evidence from Spanish manufacturing firms". *Research Policy*, 41, p. 614–623.
- Dechezleprêtre, A., Martin, R., et Mohnen, J. (2014). "Knowledge spillovers from clean and dirty technologies". *CEP Discussion papers*, 1300.
- Dortet-Bernadet, V. et Sicsic, M. (2015). « Effet des aides publiques sur l'emploi en R&D dans les petites entreprises ». *Direction des études et synthèse économique*, Document de travail(2015/11).
- Duguet, E. (2010). "The effect of the R&D tax credit on the private funding of R&D : An econometric evaluation on French firm level data". *Mimeo Discussion Paper*.
- Einiö, E. (2014). "R&D Subsidies and Company Performance : Evidence from Geographic Variation in Government Funding Based on the ERDF Population-Density Rule". *The Review of Economics and Statistics*, 96(3), p. 710–728.
- Fagerberg, J. (1994). "Technology and international differences in growth rates". *Journal of Economic Literature*, 32(3), p. 1147–1175.
- France Stratégie (2016). « Programme d'investissements d'avenir : rapport du comité d'examen à mi-parcours ».
- Galia, F. et Legros, D. (2004). "Complementarities between obstacles to innovation : evidence from France". *Research Policy*, 33(8).
- Galia, F., Mancini, S., et Morandini, V. (2012). "Obstacles to innovation : what hampers innovation in France and Italy". *Druid Society*, 2.
- García-Quevedo, J. (2004). "Do Public Subsidies Complement Business R&D? A Meta-Analysis of the Econometric Evidence". *Kyklos*, 57(1), p. 87–102.

- Givord, P. (2010). « Méthodes économétriques pour l'évaluation des politiques publiques ». *Insee Documents de travail*, (G2010/08).
- González, X. et Pazó, C. (2008). “Do public subsidies stimulate private R&D spending? ”. *Research Policy*, 37, p. 371–389.
- Griffith, R. (2000). “How Important is Business R&D for Economic Growth and Should the Government Subsidize it? ”. *IFS Briefing Notes*, 12.
- Griliches, Z. (1992). “The search for R&D spillovers”. *The Scandinavian journal of economics*, 94.
- Griliches, Z. (1998). Patent Statistics as Economic Indicators : A Survey. In “R&D and Productivity : The Econometric Evidence”, NBER Chapters, p. 287–343. National Bureau of Economic Research, Inc.
- Griliches, Z. et Regev, H. (1998). “An econometric evaluation of high-tech policy in Israel”. ATP-Conference in Washington DC.
- Guellec, D. et van Pottelsberghe, B. (1997). « Le soutien des pouvoirs publics stimule-t-il la R&D privée ? ». *Revue économique de l'OCDE*, (29).
- Hagedoorn, J. et Cloudt, M. (2003). “Measuring innovative performance : is there an advantage in using multiple indicators?”. *Research Policy*, 32(8), p. 1365–1379.
- Hall, B. (1996). “The private and social returns to R&D” . In Smith, B. et Barfield, C., editors, *Technology, R&D and the Economy*. Brookings Institution and American Enterprise Institute.
- Hall, B. (2002). “The Financing of Research and Development”. *Oxford Review of Economic Policy*, 18.
- Hall, B. et Van Reenen, J. (2000). “How effective are fiscal incentives for R&D? A review of the evidence”. *Research Policy*, 29.
- Hall, B., M. J. et Mohnen, P. (2010). “Measuring the returns to R&D ”. *United Nations University Working Paper Series*, 006.

- Hamberg, D. (1968). “ R&D : Essays on the Economics of Research and Development ”. *Journal of economic issues*, 2(2).
- Heckman, J., Ichimura, H., Smith, J., et Todd, P. (1998). “Characterizing selection bias using experimental data”. *Econometrica*, 66(5), p. 1017–1098.
- Higgins, R. et Link, A. (1981). “Federal support of technological growth in industry : some evidence of crowding out”. *Transactions on Engineering Management*, 28, p. 86–88.
- Hünermund, P. et Czarnitzki, D. (2015). “Estimating the Local Average Treatment Effect of R&D Subsidies in a Virtual Common Pot”. Annual Conference : Economic Development - Theory and Policy 112869, German Economic Association.
- Howe, J. et McFetridge, D. (1976). “The determinants of R&D expenditures”. *Canadian Journal of Economics*, 9, p. 57–71.
- Hyytinen, A. et Toivanen, O. (2005). “Do financial constraints hold back innovation and growth? Evidence on the role of public policy”. *Research Policy*, 34(9).
- Institut des Hautes Etudes pour la Science et la Technologie (2013). « Quelles politiques de soutien à l'innovation ? ». Dossier Innovation, Repère.
- Irwin, D. A. et Klenow, P. J. (1996). “High-tech R&D subsidies Estimating the effects of Sematech”. *Journal of International Economics*, 40(3-4), p. 323–344.
- Jaffe, A. B. et Le, T. (2015). “The Impact of R&D Subsidy on Innovation : a Study of New Zealand Firms”. NBER Working Papers 21479, National Bureau of Economic Research, Inc.
- Juppé, A. et Rocard, M. (2010). « Investir pour l'avenir, Priorités stratégiques d'investissement et emprunt national ».
- Kilponen, J. et Santavirta, T. (2007). “When do R&D subsidies boost innovation ? Revisiting the inverted U-shape”. Research Discussion Papers 10/2007, Bank of Finland.

- Klette, T. J. et Møen, J. (1999). “From Growth Theory to Technology Policy - Coordination Problems in Theory and Practice”. *Nordic Journal of Political Economy*, 25, p. 53–74.
- Klette, T. J. et Møen, J. (2011). “R&D investment responses to R&D subsidies : A theoretical analysis and a microeconomic study”. Discussion Papers 2011/15, Department of Business and Management Science, Norwegian School of Economics.
- Klette, T. J., Moen, J., et Griliches, Z. (2000). “Do subsidies to commercial R&D reduce market failures? Microeconomic evaluation studies”. *Research Policy*, 29(4-5), p. 471–495.
- Lach, S. (2000). “Do R&D subsidies stimulate or displace private R&D? Evidence from Israel”. *NBER Working Paper*, 7943.
- Lerner, J. (1999). “The Government as Venture Capitalist : The Long-Run Impact of the SBIR Program”. *The Journal of Business*, 72(3), p. 285–318.
- Levet, J. et Mathieu, C. (2013). « Evaluation *ex post* du programme d’investissement d’avenir : un dispositif à construire ». *Rapport du Commissariat général à l’investissement*.
- Leyden, D. et Link, A. (1991). “Why are government and private R&D complements?”. *Applied Economics*, 23, p. 1673–1681.
- Lhuillery, S., Marino, M., et Parrotta, P. (2013). « Evaluation de l’impact des aides directes et indirectes à la R&D en France. *Rapport pour le Ministère de l’Enseignement Supérieur et de la Recherche*.
- Lichtenberg, F. R. (1984). “The Relationship between Federal Contract R&D and Company R&D”. *American Economic Review*, 74(2), p. 73–78.
- Lichtenberg, F. R. (1988). “The Private R&D Investment Response to Federal Design and Technical Competitions”. *American Economic Review*, 78(3), p. 550–59.

- Link, A. (1982). "An Analysis of the Composition of R&D Spending". *Southern Economic Journal*, 49(2), p. 342–349.
- Lokshin, B. et Mohnen, P. (2013a). "Do R&D tax incentives lead to higher wages for R&D workers? Evidence from the Netherlands". *Research Policy*, 42, p. 823–830.
- Lokshin, B. et Mohnen, P. (2013b). "How effective are level-based R&D tax credits? Evidence from the Netherlands". *Applied Economics*, 44(12), p. 1527–1538.
- Martínez-Ros, E. et Corchuelo, B. (2009). "the effects of fiscal incentives for R&D in Spain". *Universidad Carlos III de Madrid Working Papers*, 09.
- Mauroux, A. (2012). « Le crédit d'impôt dédié au développement durable : une évaluation économétrique ». *Document de travail Direction des études et synthèses économiques*, G2012/11.
- Mohnen, P. et Dagenais, M. (2002). "Towards an innovation intensity index : the case of CIS 1 in Denmark and Ireland". In *Innovation and Firm Performance*, p. 3–30. Springer.
- Mohnen, P. et Röller, L.-H. (2000). "Complementarities in innovation policy". *WZB Working Paper*, (18).
- Mohnen, P. et Rosa, J. (2000). « Les obstacles à l'innovation dans les industries de services au Canada ». *CIRANO Working Paper*, 14.
- Mulkay, B. et Mairesse, J. (2013). "The R&D tax credit in France : Assessment and ex-ante evaluation of the 2008 reform". *NBER Working Paper*, 19073.
- Nadiri, I. (1993). "Innovation and technological spillovers". Economic research report 93-31, C.V. Starr center for applied economics, NYU.
- Negassi, S. et Hellier, J. (2011). « L'efficacité des aides publiques à l'innovation des entreprises en France ». *Conférence de l'Association Internationale de Management Stratégique*.
- OCDE (2002). « Manuel de Frascati : Méthode type proposée pour les enquêtes sur la recherche et le développement expérimental ».

- OCDE (2014). « Examen de l'OCDE des politiques d'innovation : France ».
- Pinget, A., Bocquet, R., et Mothe, C. (2015). “Barriers to environmental innovation in SMEs : Empirical evidence for French firms”. *Management*, 18.
- Popp, D. (2002). “Induced innovation and energy prices”. *American Economic Review*, 92(1), p. 160–180.
- Popp, D., Newell, R., et Jaffe, A. (2010). “Energy, the Environment, and Technological Change”. In Halland, B. et Rosenberg, N., editors, *Handbook of the Economics of Innovation Vol. II*, NBER Chapters, p. 873–938. Academic Press, Burlington.
- Pottelsberghe, B. V. et Capron, H. (1997). “Public support to business R&D : a survey and some new quantitative evidence”. In *Policy Evaluation in Innovation and Technology : Towards Best Practices*, chapter 10. OECD.
- Rennings, K. (2000). “Redifining innovation - Eco innovation research and the contribution from ecological economics”. *Ecological Economics*, 32, p. 319–332.
- Romer, P. (1990). “Endogenous technological change”. *Journal of political economy*, 98(5).
- Sala-i-Martin, X. (1990). “Lecture notes on economic growth : Introduction to the litterature and Neoclassical models”. *NBER Working Paper*, 3563.
- Scott, J. (1984). “Firm versus industry variability in R&D intensity”. In Griliches, Z., editor, *R&D, Patents and Productivity*. Univeristy of Chicago Press.
- Segarra, A., Teruel, M., et Garcia, J. (2008). “Barriers to innovation and public policy in Catalonia”. *International entrepreneurship and management journal*, 4.
- Sveikauskas, L. (2007). “R&D and Productivity Growth : A Review of the Literature”. *U.S. Bureau of Labor Statistics Working Paper*, (408).
- Tiwari, R. et Buse, S. (2007). “Barriers to innovation in SMEs : Can the internationalization of R&D mitigate their effects? ”. *Hamburg University of Technology Working Paper*, 50.

Wallsten, S. (2000). "The effects of government-industry R&D programs on private R&D : the case of the Small Business Innovation Research program". *RAND Journal of Economics*, 31(1), p. 82–100.

Zahra, S. et George, G. (2002). "Absorptive capacity : A review, reconceptualization, and extension". *Academy of Management Review*, 27(2), p. 185–203.

LISTE DES TABLEAUX

1.1	Opérateurs du PIA et crédits prévus initialement (en millions d'euros)	21
1.2	Modalités d'attribution et de remboursement - Programme « Véhicule du futur »	26
1.3	Modalités d'attribution et de remboursement - Programme « Démonstrateurs de la transition écologique et énergétique »	26
1.4	Profil des partenaires des projets bénéficiaires	28
1.5	Répartition des types de financement des projets PIA-ADEME par programme (en millions d'euros)	28
3.1	Description des fichiers LOCO	57
3.2	Nombre de valeurs manquantes pour les partenaires rejetés	60
3.3	Comparaison de valeurs entre les Annexes et DIA Projets pour les partenaires rejetés	60
3.4	Comparaison des entreprises présentes en 2009 et 2013 dans l'Enquête R&D à celles présentes seulement une année	66
3.5	Comparaison des partenaires ADEME-FARE présents dans l'Enquête R&D en 2009 et 2013 aux autres	71
3.6	Variables d'intérêt disponibles dans les bases constituées	73
4.1	Récapitulatif des méthodes d'évaluations disponibles et de leur applicabilité	94
5.1	Effectifs de partenaires traités et témoins	98
5.2	Les variables de contrôle utilisées	102

5.3	Résultats du <i>probit</i> - Base ADEME-FARE-DADS	112
5.4	Vérification statistique de l'hypothèse des tendances parallèles - Base ADEME-FARE-DADS (<i>one-to-one matching</i>)	118
5.5	Vérification statistique de l'hypothèse des tendances parallèles - Base ADEME-FARE-DADS (<i>kernel matching</i>)	119
5.6	Nombre d'observations nécessaire pour détecter un effet significatif - Variable « Effectifs de R&D »	122
5.7	Nombre d'observations nécessaire pour détecter un effet significatif - Variable « Taux d'immobilisations incorporelles »	122
C.1	Comparaison des moyennes des groupes traité et témoin en 2009 - Base ADEME-DADS-FARE	152
D.1	Vérification statistique de l'hypothèse des tendances parallèles - Base ADEME-FARE-RD	154
F.1	Résultats des régressions - ADEME-FARE-RD	166
F.2	Résultats des régressions - Base ADEME-DADS-FARE (<i>one-to-one mat- ching</i>)	167
F.3	Résultats des régressions - Base ADEME-DADS-FARE (<i>kernel mat- ching</i>)	168

LISTE DES FIGURES

1.1	Distributions des subventions et des avances remboursables par déciles de montants autorisés	29
3.1	Constitution de la base ADEME	62
3.2	Appariement des bases ADEME, FARE et R&D 2009 et 2013	70
3.3	Appariement des bases ADEME, FARE et DADS 2009 et 2013	72
4.1	Illustration de la méthode des doubles différences	90
4.2	Illustration de la méthode de la RDD - L'étude de Bronzini et Piselli (2016)	92
5.1	Temporalité de l'étude	96
5.2	Evolution du taux d'immobilisations incorporelles pour la période 2007-2013 (en pourcentage)- Base ADEME-FARE-RD	106
5.3	Evolution de la productivité des salariés pour la période 2007-2013 - Base ADEME-FARE-RD	107
5.4	Evolution de la production totale pour la période 2007-2013 - Base ADEME-FARE-RD	108
5.5	Evolution du résultat comptable pour la période 2007-2013 - Base ADEME-FARE-RD	109
5.6	Distribution de la probabilité de bénéficier d'une aide du PIA-ADEME - Base ADEME-FARE-DADS	114
5.7	Distribution de la probabilité de bénéficier d'une aide du PIA-ADEME - <i>One-to-one matching</i> - Base ADEME-FARE-DADS	116

5.8	Distribution de la probabilité de bénéficier d'une aide du PIA-ADEME - <i>Kernel matching</i> - Base ADEME-FARE-DADS	117
E.1	Evolution du taux d'immobilisations incorporelles pour la période 2007-2013 (en pourcentage)- Base ADEME-FARE-DADS	156
E.2	Evolution de la productivité des salariés pour la période 2007-2013 - Base ADEME-FARE-DADS	157
E.3	Evolution de la production totale pour la période 2007-2013 - Base ADEME-FARE-DADS	157
E.4	Evolution du résultat comptable pour la période 2007-2013 - Base ADEME-FARE-DADS	158
E.5	Evolution des effectifs en R&D pour la période 2007-2013 - Base ADEME-FARE-DADS	158
E.6	Evolution du taux d'immobilisations incorporelles pour la période 2007-2013 (en pourcentage)- Base ADEME-FARE-DADS (<i>one-to-one matching</i>)	159
E.7	Evolution de la productivité des salariés pour la période 2007-2013 - Base ADEME-FARE-DADS (<i>one-to-one matching</i>)	160
E.8	Evolution de la production totale pour la période 2007-2013 - Base ADEME-FARE-DADS (<i>one-to-one matching</i>)	160
E.9	Evolution du résultat comptable pour la période 2007-2013 - Base ADEME-FARE-DADS (<i>one-to-one matching</i>)	161
E.10	Evolution des effectifs en R&D pour la période 2007-2013 - Base ADEME-FARE-DADS (<i>one-to-one matching</i>)	161
E.11	Evolution du taux d'immobilisations incorporelles pour la période 2007-2013 (en pourcentage)- Base ADEME-FARE-DADS (<i>kernel mat- ching</i>)	162
E.12	Evolution de la productivité des salariés pour la période 2007-2013 - Base ADEME-FARE-DADS (<i>kernel matching</i>)	163
E.13	Evolution de la production totale pour la période 2007-2013 - Base ADEME-FARE-DADS (<i>kernel matching</i>)	163

E.14 Evolution du résultat comptable pour la période 2007-2013 - Base ADEME-FARE-DADS (<i>kernel matching</i>)	164
E.15 Evolution des effectifs en R&D pour la période 2007-2013 - Base ADEME-FARE-DADS (<i>kernel matching</i>)	164



L'Institut des politiques publiques (IPP) est développé dans le cadre d'un partenariat scientifique entre PSE-Ecole d'économie de Paris (PSE) et le Centre de Recherche en Economie et Statistique (CREST). 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-Ecole d'économie de Paris regroupe plus de 120 chercheurs, 200 doctorants et 300 étudiants, et constitue un pôle français en science économique de renommée mondiale. PSE a pour objectif premier de fédérer, animer et assurer le rayonnement de ses chercheurs, tout en proposant des formations généralistes et spécialisées à la pointe de la discipline, du M1 au doctorat. La fondation vise également à tisser des liens pérennes entre les différents univers « consommateurs » de savoirs économiques : les acteurs académiques, institutionnels et privés. www.parisschoolofeconomics.eu



PARIS SCHOOL OF ECONOMICS
ÉCOLE D'ÉCONOMIE DE PARIS

Le CREST est le centre de recherche du GENES (Groupe des Ecoles Nationales d'Économie et Statistiques) qui est devenu le 1^{er} janvier 2011 un établissement public à caractère scientifique, culturel et professionnel (EPSCP), sous la tutelle technique de l'INSEE (ministère de l'Économie, des Finances et de l'Industrie). Le GENES regroupe quatre établissements : le CREST, le CEPE, l'ENSAE et l'ENSAI. Il a vocation à conduire des travaux de recherche, des missions d'étude ou d'expertise et des actions de diffusion. Il est en outre habilité à développer des dispositifs d'accès aux données, notamment de la statistique publique. www.crest.fr

