

Retombées du Plan France Très Haut Débit sur les
entreprises: quels effets sur les usages numériques,
la performance, et l'innovation?

Rapport au 07/06/2022

CHLOE DUVIVIER
Inrae, UMR Territoires

LAURENT BERGE
Université de Bordeaux,
GREThA, UMR CNRS 5113

FLORIAN LEON
Ferd

CLAIRE BUSSIÈRE
Inrae, UMR Territoires

Table des matières

Résumé.....	1
Chapitre 1. Introduction générale.....	3
Objectifs de recherche.....	3
Couverture géographique, sectorielle et temporelle du rapport.....	4
Structure du rapport.....	6
Chapitre 2. Evolution de la couverture internet fixe.....	8
1. Evolution de la couverture en France (2013-2018).....	9
2. Géographie du déploiement en France (2013-2018).....	11
Chapitre 3. Méthodologie : identifier l'impact du THD.....	18
1. Stratégie d'identification.....	18
1.1. Définition des groupes de traités.....	18
1.2. Définition des groupes de contrôle.....	20
1.3. Menaces sur l'identification: sources d'endogénéité du traitement.....	21
1.4. Les solutions méthodologiques envisagées selon le type de données.....	23
2. L'identification dans le cas des données de panel.....	23
2.1. Mise en œuvre de la DD standard.....	24
2.2. Problèmes spécifiques à la DD avec traitement échelonné.....	25
3. L'identification dans le cas des données en coupe.....	31
3.1. Le difficile recours aux variables instrumentales.....	31
3.2. La méthode de l'appariement.....	32
Chapitre 4. Usages numériques.....	33
1. Revue de littérature.....	33
1.1. Définition des usages numériques.....	33
1.2. Impact de l'accès à internet sur les usages numériques.....	34
1.3. Un effet conditionnel des infrastructures THD ?.....	35
2. Données.....	39
2.1. Usages numériques des entreprises.....	39
2.2. Accès des entreprises au THD.....	42
2.3. Caractéristiques internes et externes des entreprises.....	47
3. Statistiques descriptives : accès au THD et usages numériques en France.....	48
4. Méthodologie.....	57
4.1. Appariement par le score de propension.....	57
4.2. Estimation de l'effet du traitement.....	60
4.3. Adaptation de la méthode d'appariement aux données d'enquête.....	60

4.4.	Etapas de l'estimation.....	61
5.	Résultats	64
5.1.	Appariement	64
5.2.	Impact moyen du programme.....	68
5.3.	Hétérogénéité de l'impact du programme	73
6.	Conclusion.....	83
	Annexes au chapitre 4	85
Chapitre 5.	Performance des entreprises et emploi	90
1.	Revue de littérature	90
1.1.	De la difficile comparaison d'études hétérogènes.....	90
1.2.	Effet sur l'emploi	92
1.3.	Effet sur les performances économiques régionales et individuelles.....	93
1.4.	Questions en suspens.....	94
2.	Méthodologie	96
2.1.	Données et échantillon.....	96
2.2.	Variables	97
2.3.	Méthode économétrique	103
3.	Résultats	105
3.1.	Estimation de l'effet global du THD et de la fibre	105
3.2.	Tests de robustesse	109
3.3.	Analyse par type d'entreprises.....	111
3.4.	Analyse selon les caractéristiques des communes.....	121
3.5.	Analyse au niveau entreprise.....	126
4.	Conclusion	128
	Annexe au chapitre 5.....	130
	A.5-A / Tableaux et figures additionnels	130
	A.5-B / Analyse au niveau entreprise : Précisions méthodologiques et résultats	133
Chapitre 6.	Performance d'innovation	144
1.	Cadre théorique et revue de littérature	144
1.1.	L'innovation: définitions.....	144
1.2.	Très haut débit et innovation: quelques mécanismes	145
1.3.	Mesures de l'innovation.....	153
2.	Analyse de la production de brevets.....	157
2.1.	Construction de données.....	157
2.2.	Stratégie empirique.....	161
2.3.	Statistiques descriptives.....	162

2.4.	Résultats	171
3.	Analyse de l'enquête CIS	179
3.1.	Présentation et construction de la base de données CIS	179
3.2.	Statistiques descriptives des données CIS	184
3.3.	Méthode économétrique	191
3.4.	Résultats principaux	200
3.5.	Hétérogénéité des résultats en fonction des caractéristiques des entreprises	209
4.	Conclusion.....	216
	Annexe au chapitre 6 : construction des variables	218
	Références	223

Résumé

Le projet vise à évaluer les retombées du Plan France Très Haut Débit sur les entreprises situées dans les zones peu denses en France. Plus précisément, il s'agit d'examiner l'impact du déploiement du très haut débit (>30Mbps) et de la Fibre (>100 Mbps) sur le comportement et les performances des entreprises. Deux dimensions du comportement des entreprises sont considérées: leurs usages numériques et leur processus d'innovation. La performance des entreprises est envisagée à travers l'analyse d'indicateurs standards de performance économique. En termes de géographie, on s'intéresse à l'impact du déploiement sur les entreprises situées dans les communes peu denses (zone d'initiative publique). Les effets sont évalués en mobilisant des méthodes quasi-expérimentales d'évaluation de politiques publiques.

Le chapitre sur les usages numériques examine l'effet de la fibre sur un ensemble de 10 pratiques différentes, incluant des usages organisationnels (ordinateurs, internet, site web, médias sociaux) et plus avancés (cloud computing, PGI, GRC, outils collaboratifs) ainsi que le commerce en ligne. Les résultats indiquent que l'accès à la fibre favorise l'usage de la plupart des outils organisationnels et avancés (ordinateurs, internet, site web, cloud computing, PGI, GRC). En revanche, la fibre n'est pas associée à une utilisation accrue du commerce en ligne, des médias sociaux et des outils collaboratifs. Par ailleurs, les retombées du Plan varient selon les caractéristiques des entreprises, avec des effets supérieurs sur les petites entreprises (fortement dépendantes des réseaux de télécommunication étant données leurs ressources limitées) et les multi-établissements (qui ont davantage d'incitations à adopter certains outils numériques en raison de leurs coûts de coordination supérieurs). L'impact de la fibre est également conditionnel à l'environnement externe des entreprises, avec des effets bénéfiques dans les zones péri-urbaines mais quasi-inexistants dans les espaces ruraux. Enfin, les retombées du déploiement ne sont pas immédiates, avec des effets mesurables environ trois ans après l'arrivée de la fibre.

Le chapitre sur la performance des entreprises considère l'effet du THD et de la fibre sur trois indicateurs standards de performance économique: la valeur ajoutée, les effectifs et la productivité. En raison de difficultés méthodologiques, l'analyse principale est conduite à l'échelle des communes (une analyse supplémentaire est fournie au niveau des entreprises mais est à interpréter avec précaution). Au niveau communal, nos résultats montrent que le THD et la fibre ont eu des effets positifs, rapides et croissants dans le temps sur la valeur ajoutée. Ces deux technologies ont également bénéficié à l'emploi bien que les effets soient plus tardifs à se

matérialiser (effets significatifs au bout de 4 à 5 ans). Enfin, on n'observe aucun effet du THD ou de la fibre sur la productivité au niveau communal. Par ailleurs, les retombées varient selon les caractéristiques des entreprises présentes dans les communes. En particulier, le THD a stimulé la création de valeur en priorité pour les grandes entreprises et les entreprises multi-établissements (même si toutes les entreprises en ont bénéficié). En revanche, l'effet sur l'emploi a été plus fort pour les PME. Les secteurs de l'industrie, de l'hôtellerie-restauration, et des TIC ont également bénéficié davantage du THD. La fibre semble avoir bénéficié quant à elle essentiellement aux PME, aux entreprises mono-établissement et à celles du secteur industriel. Les retombées du Plan sont également variables selon les caractéristiques des communes, avec des retombées observables rapidement pour toutes les communes, à l'exception des communes isolées (pour lesquelles seul le THD a un effet mais qui est tardif à se matérialiser). Enfin, seules les communes ayant déjà accès à du bon haut débit ont réellement bénéficié de l'arrivée de ces nouvelles technologies.

Le chapitre sur l'innovation estime l'impact du THD en considérant un ensemble de mesures basées sur les brevets et l'enquête sur les pratiques d'innovation menée par l'INSEE (enquête CIS). Ces divers indicateurs permettent de prendre en compte différents aspects de l'innovation. L'examen des données de brevets indique un impact non significatif ou négatif du déploiement à court terme. Ces résultats sont toutefois à interpréter avec prudence puisque les données de brevets ne sont exhaustives que jusqu'en 2016, et donc, l'analyse ne capte que les effets de très court terme des premiers déploiements de fibre. L'analyse des données CIS indique qu'à l'exception des entreprises multi-établissements, l'accès à la fibre n'a pas influencé la production d'innovation des entreprises, ni même leurs schémas collaboratifs. Les données CIS ont toutefois des caractéristiques particulières (perception subjective de l'innovation, mesure extensive des innovations, données disponibles jusqu'en 2018 seulement) qui peuvent expliquer l'absence d'effet. Pour les entreprises multi-établissements, l'accès à la fibre engendre toutefois une hausse des innovations de procédé, et en particulier des innovations organisationnelles, marketing et de support en 2018.

Chapitre 1. Introduction générale

Objectifs de recherche

Cette recherche examine les retombées du très haut débit¹ (THD) sur le comportement et les performances des entreprises, en considérant trois dimensions : leurs usages numériques, leur performance, et leur processus d'innovation.

Au-delà de l'estimation de l'impact moyen du THD sur les entreprises, la recherche vise à déterminer si certains types d'entreprises bénéficient davantage du Plan FTTHD que d'autres. Lorsque cela est possible, on examine ainsi si l'impact du Plan FTTHD est hétérogène en fonction des *caractéristiques internes* des entreprises, et notamment de leur taille, de leur structure organisationnelle (mono/multi-établissements) et de leur secteur d'activité.

La recherche vise également à déterminer si le déploiement du THD génère des effets particulièrement positifs dans certains contextes particuliers. Nos travaux antérieurs (Duvivier, 2019 ; Duvivier et al., 2018 ; 2021 ; Duvivier et Bussière, 2022) indiquent en effet que si le THD n'est pas suffisant à lui seul pour impulser le développement économique des espaces, il peut cependant constituer un réel levier de développement lorsque certaines conditions sont réunies. L'enjeu de la recherche est donc de comprendre dans quels *contextes territoriaux* et selon *quelles modalités de déploiement* les bénéfices du THD sont maximisés. En particulier, il semble primordial d'identifier les types d'espaces (ex : périurbain, pôles ruraux, espaces ruraux) dans lesquels le déploiement génère des effets plus limités, de sorte à mettre en place des mesures complémentaires. De même, comprendre comment le design de la politique numérique peut en affecter les retombées semble pertinent en termes de recommandations de politiques publiques. Lorsque l'échantillon mobilisé le permet, on analyse ainsi si l'impact du THD est conditionnel au débit déployée (THD vs fibre), au niveau de couverture², et au type de transition technologique à l'œuvre^{3,4}.

¹ Dans ce rapport, on emploie le terme générique de THD bien que l'analyse empirique vise à évaluer les retombées de l'arrivée du très haut débit (>30 Mbps) et de la fibre (>100 Mbps).

² Plus précisément, on cherche à déterminer si l'impact du Plan FTTHD est plus fort lorsque la part des locaux couverts au sein d'une commune est supérieure.

³ Il s'agira notamment d'analyser si un déploiement progressif du THD (ex : passage du HD au THD puis à la fibre) génère des effets économiques différents d'un déploiement « radical » (ex : passage d'un débit faible à de la fibre optique).

⁴ Nous aurions également souhaité examiner si l'effet du déploiement est supérieur dans les espaces où des dispositifs d'accompagnement numérique à destination des entreprises existent (ex : subventions aux entreprises

Outre les questions relatives au contexte territorial et au design de la politique numérique, se pose également la question de la *temporalité de l'impact* du THD. Lorsque les données le permettent, la recherche tente ainsi d'évaluer à partir de quand on perçoit les impacts du THD et si ces derniers se renforcent au cours du temps (en raison d'effets d'apprentissage par exemple).

Du point de vue méthodologique, les effets sont évalués en mobilisant des méthodes quasi-expérimentales d'évaluation de politiques publiques. Plus précisément, nous recourons aux doubles-différences échelonnées dans les chapitres mobilisant des données de panel et à l'appariement lorsque l'analyse repose sur des données en coupe répétée.

Couverture géographique, sectorielle et temporelle du rapport

En termes de géographie, on s'intéresse à l'impact du THD sur les entreprises situées dans les communes peu denses, bénéficiant d'un déploiement de THD public (zones d'intervention publique, ci-après ZIPU)⁵.

Par ailleurs, le rapport considère ici l'ensemble des entreprises du secteur marchand à l'exception du secteur agricole et ce, pour plusieurs raisons. Tout d'abord, la plupart des bases de données mobilisées ici (ex : enquêtes-TIC, données FARE) n'incluent pas les activités non marchandes ni agricoles. En outre, et plus fondamentalement, la logique des activités agricoles est très spécifique, et donc, leur analyse nécessiterait une étude à part entière. Enfin, l'agriculture numérique repose en grande partie sur les objets connectés⁶ qui communiquent peu de données mais sur des distances importantes. Ainsi, pour les agriculteurs, l'enjeu est aussi (surtout ?) d'accéder à des réseaux longue portée (Lora, SigFox) qui ne sont pas concernés par le Plan FTTHD.

pour financer de l'accompagnement, coaching/formation par des experts, prêts bancaires, etc.). Malheureusement, malgré nos recherches, nous ne sommes pas parvenus à obtenir de données précises sur l'existence de tels dispositifs.

⁵ La liste des communes situées en zone d'intervention publique a légèrement évolué depuis 2011, année où le gouvernement français a lancé l'AMII auprès des opérateurs privés. Certains opérateurs privés ont notamment décidé de déployer des réseaux sur fonds propres dans des communes initialement classées comme ZIPU. Par exemple, Orange a décidé de déployer son propre réseau dans 8 communes initialement concernées par le RIP Yvelines Fibre. Dans ce projet, nous considérons comme appartenant à la ZIPU toute commune n'ayant pas fait l'objet d'une déclaration d'intérêt par les opérateurs privés en 2011.

⁶ Incluent par exemple les drones (permettant de surveiller la qualité du sol, l'élevage, l'épandage), les capteurs sensoriels (assurant le suivi de l'état de la production, des conditions climatiques, la présence d'insectes, les niveaux d'irrigation) ou encore les colliers connectés (pour suivre la santé des bovins, identifier les périodes de gestation, détecter d'éventuels troubles de l'alimentation).

En termes de temporalité, le rapport se compose de deux parties afin d'évaluer de manière séparée les retombées du THD pendant la période antérieure à la Covid 19 (chapitres 2 à 6) et pendant la crise sanitaire (rapport additionnel, qui sera remis fin juin 2022).

Concernant la période pré-Covid 19, on analyse les retombées des déploiements réalisés jusqu'en 2018 (chapitres 4 et 6) ou 2019 (chapitre 5). En effet, la disponibilité actuelle des données sur les entreprises⁷ ne nous permet pas de tenir compte des déploiements plus récents, dont le nombre s'est pourtant accéléré à partir de 2019. Malgré cela, l'évaluation des déploiements réalisés jusqu'en 2018 ou 2019 devrait fournir une estimation assez représentative des effets du Plan FTHD puisque les RIP les plus précoces⁸ sont situés dans des régions diversifiées en termes de conditions géographiques et socio-économiques, ce qui assure une bonne représentativité des différents types d'espaces peu denses dans l'analyse (cf. chapitre 2).

Afin d'évaluer les retombées du Plan FTHD sur les entreprises pendant la pandémie, nous sommes actuellement en train de réaliser une nouvelle analyse à partir des données de l'enquête flash Acemo Covid. Il s'agit d'examiner ici si le THD a permis aux entreprises de moins souffrir des restrictions mises en place pendant la crise sanitaire (ex : en facilitant la vente en ligne). Cette analyse supplémentaire prend donc en compte les déploiements les plus récents et porte spécifiquement sur la période débutant en avril 2020.

Plusieurs raisons expliquent qu'on traite séparément la période pré- et post-Covid 19 dans ce rapport. Si l'absence de données empêche de considérer la crise Covid dans les chapitres 4 à 6, des arguments économiques plus fondamentaux nous font douter de la pertinence de combiner les périodes pré- et post-Covid 19. Premièrement, la situation depuis 2020 est tellement particulière qu'on risque d'estimer des effets conjoncturels très atypiques dans le rapport focalisé sur la période Covid, bien différents des effets structurels du THD qu'on évalue dans les chapitres 4 à 6⁹. Deuxièmement, certaines variables dépendantes considérées dans les chapitres 4 à 6 ne sont pas des indicateurs pertinents à étudier pendant la pandémie. Par exemple, en raison du soutien important de l'Etat aux entreprises, on n'observe peu d'effets

⁷ Concernant les usages numériques, la dernière enquête TIC-Entreprise disponible est celle de 2019, qui renseigne sur les usages des entreprises en 2018 (ou, pour certains usages, sur les pratiques en début d'année 2019). Les dernières données FARE, mobilisées pour l'étude de la performance, portent quant à elle sur l'exercice 2019. Enfin, concernant l'innovation, la dernière enquête CIS porte sur l'année 2018 et les données sur les brevets ne sont exhaustives que jusqu'en 2016.

⁸ Ex : Bretagne, Oise, Seine et Marne, Loire, Nord-Pas-de-Calais, Alsace.

⁹ Par exemple, il est probable que certains secteurs bénéficiant beaucoup du THD (ex : le tourisme) soient ceux qui aient été touchés le plus durement par la crise.

nets de la crise sur la performance des entreprises à l'heure actuelle (l'année 2020 a d'ailleurs été caractérisée par un nombre de faillites relativement faible). Sur la performance des entreprises, les effets réels de la crise seront donc plus tardifs et ne peuvent être véritablement évalués aujourd'hui.

Si l'analyse menée dans le rapport additionnel (Covid) semble intéressante, il faut d'ores et déjà garder en mémoire deux éléments. Premièrement, bien qu'une première analyse des effets du THD pendant la crise sanitaire semble intéressante, trois ou quatre années de recul supplémentaires seraient nécessaires pour examiner si les effets observés sont durables dans le temps. Deuxièmement, nous examinons l'effet du THD sur les entreprises pendant la Covid alors que les effets de la crise concernent probablement surtout le résidentiel (achats de biens, prix immobiliers) et le télétravail, aspects que nous n'abordons pas dans notre rapport.

Structure du rapport

Le présent rapport (hors futur chapitre Covid) s'articule en trois volets de recherche, chacun étant consacré à l'analyse des effets du THD sur une des dimensions du comportement ou de la performance des entreprises (usages, performance, et innovation).

Puisque le THD génère des effets essentiellement si les entreprises se l'approprient, le premier volet est consacré à l'analyse de leurs pratiques en ligne. Il s'agit ici d'examiner si l'accès à la fibre modifie les pratiques en ligne des entreprises (en distinguant des usages organisationnels et complexes) et leur niveau d'intensité.

Toutefois, s'emparer des outils numériques ne constitue évidemment pas un gage de réussite économique, en particulier dans les espaces peu denses où une forte proportion d'entreprises en font une utilisation relativement peu efficace (Aronica et al., 2021). De ce fait, le deuxième volet examine l'effet du Plan FTHD sur les indicateurs usuels de performance économique des entreprises, à savoir la création de valeur ajoutée, l'emploi et la productivité.

Enfin, le troisième volet s'interroge sur l'impact du THD sur la propension à innover des entreprises. Afin de capturer différents aspects de l'innovation, on considère l'effet de la fibre sur un ensemble de mesures basées sur les brevets et l'enquête sur les pratiques d'innovation menée par l'INSEE (enquête CIS).

La suite du rapport est structurée de la manière suivante. Le chapitre 2 présente des statistiques descriptives sur l'évolution de la couverture internet fixe en France. Le chapitre 3 présente la méthodologie qui sera employée pour évaluer les retombées du Plan FTHD. Les

chapitres suivants sont consacrés à l'évaluation de l'impact du THD sur les usages numériques des entreprises (chapitre 4), leur performance (chapitre 5), et leur processus d'innovation (chapitre 6). Les chapitres 4 à 6 sont structurés de manière indépendante, incluant chacun une revue de littérature, une description des données et de la méthodologie, ainsi que l'analyse des résultats.

Chapitre 2. Evolution de la couverture internet fixe

Ce chapitre présente l'évolution de la couverture internet fixe des communes de France métropolitaine entre 2013 et 2018 (période correspondant aux analyses empiriques conduites dans les chapitres 4 à 6)¹⁰. On présente des données de couverture plus récentes (2019-2021) dans le rapport additionnel focalisé sur la crise sanitaire (qui sera remis fin juin 2022).

L'objectif de ce chapitre est double. D'une part, il s'agit de fournir un rapide état des lieux de la diffusion d'internet depuis la mise en place du Plan France Très Haut Débit. D'autre part, ces premières statistiques descriptives permettent de vérifier le nombre et la répartition géographique des communes par classes de débit, ce qui est essentiel avant de réfléchir à la stratégie empirique à mettre en œuvre.

Les données de couverture mobilisées ici sont décrites dans l'encadré 1¹¹. A partir des données de couverture disponibles, on distingue ici quatre groupes de communes en fonction du débit le plus élevé présent dans la commune¹² :

- Accès à moins de 8 Mbps,
- Accès à du « bon » haut-débit (entre 8 Mbps et 30 Mbps),
- Accès à du très haut débit (entre 30 Mbps et 100 Mbps),
- Accès à de la fibre (débit supérieur à 100 Mbps).

Cette classification en quatre groupes a été retenue car elle reflète les retombées des différents objectifs du plan : garantir à tous un accès au « bon » haut débit (>8 Mbps) d'ici 2020, au très haut débit (>30 Mbps) d'ici 2022, et à la fibre FttH (> 100 Mbps) d'ici 2025.

¹⁰ Pour chaque année, on se base sur la couverture au quatrième trimestre, sauf pour 2018 pour laquelle on se base sur la couverture au premier trimestre pour des raisons de disponibilité des données.

¹¹ L'encadré 2 décrit les données existantes pour la couverture internet mobile, que nous ne mobilisons pas ici en raison de leur mauvaise qualité.

¹² Par exemple, si 10% des locaux d'une commune sont éligibles à 100 Mbps et plus et 90% à un débit compris entre 30 et 100 Mbps, on classe cette commune comme ayant accès à 100 Mbps et plus.

1. Evolution de la couverture en France (2013-2018)

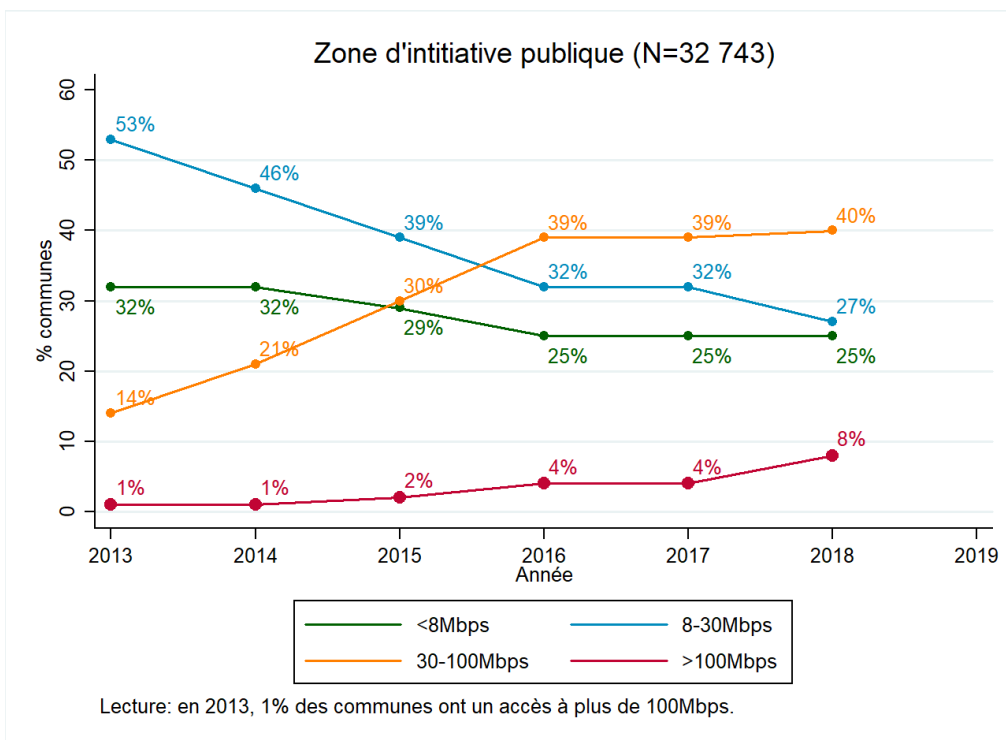
Les graphiques 2.1 et 2.2 représentent la répartition des communes selon le niveau de débit le plus élevé auquel elles ont accès.

Au sein de la zone d'initiative publique (graphique 2.1), le THD a fortement progressé sur la période, représentant ainsi la technologie la plus rapide dans 40% des communes en 2018 contre 14% en 2013. Le développement du THD coïncide avec une forte diminution de la part des communes ayant accès à du bon haut débit (8-30 Mbps) tandis que la part des communes ayant accès à moins de 8 Mbps a quant à elle seulement légèrement diminué sur la période¹³. Il semble donc que ce soit essentiellement les communes qui disposaient déjà d'un bon haut débit qui ont bénéficié de l'arrivée du THD. Enfin, la progression de la fibre s'est accélérée depuis 2017, avec 8% des communes ayant accès à plus de 100 Mbps en 2018 contre 4% seulement l'année précédente. Malgré cette évolution, la fibre demeure la technologie la moins répandue en 2018, comparée au THD (40% des communes), au bon HD (27%) et au débit plus faible ou à l'absence de débit (25%).

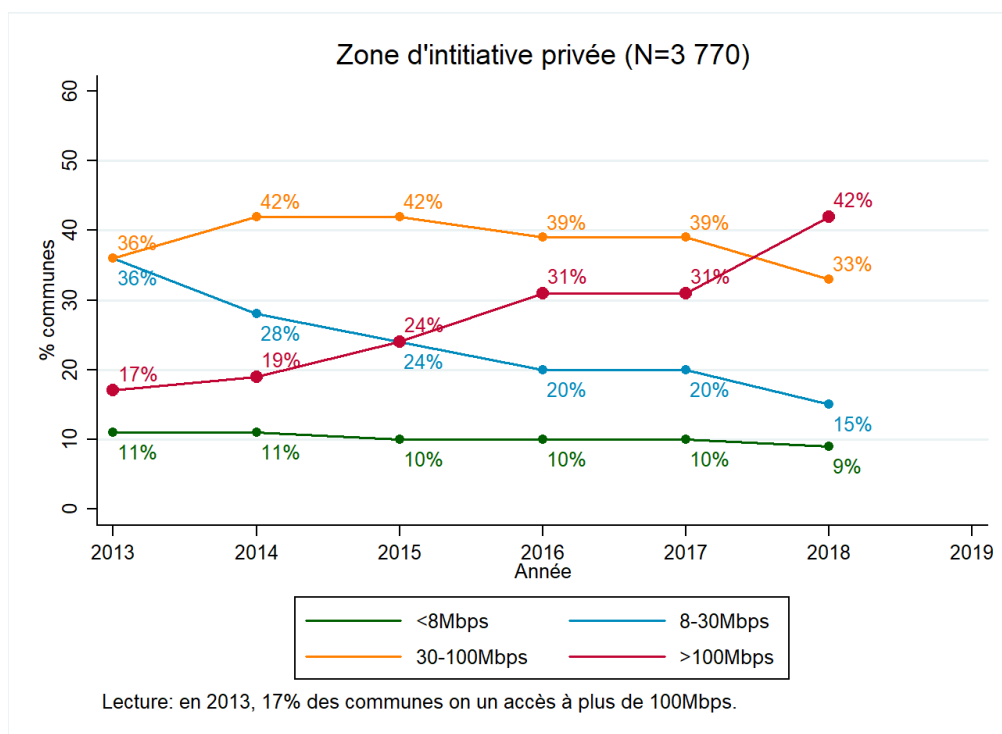
Les communes de la zone privée présentent une évolution différente, caractérisée par une forte hausse de la fibre (passant de 17 à 42%), qui est d'ailleurs la seule technologie qui progresse sur la période. En 2018, la fibre constitue la technologie la plus répandue au sein de la zone privée (42%), suivie du THD (33%) et du bon HD (15%).

La suite des statistiques descriptives concerne exclusivement les communes de la zone d'initiative publique qui font l'objet de notre recherche.

¹³ Rappelons toutefois que cette catégorie (<8Mbps) est relativement agrégée, ce qui peut masquer certaines dynamiques favorables (ex : hausse de la part des communes éligibles à plus de 3 Mbps).



Graphique 2.1. Répartition des communes selon le niveau de débit maximal (ZIPU)



Graphique 2.2. Répartition des communes selon le niveau de débit maximal (ZIPRI)

2. Géographie du déploiement en France (2013-2018)

Les cartes 2.1, 2.2 et 2.3 représentent les communes couvertes au sein de la zone d'initiative publique, selon le niveau de débit (bon HD ou plus, THD ou plus, Fibre) et la date d'arrivée de chaque technologie. Les cartes font également apparaître les frontières des RIP pour que les dynamiques locales soient plus apparentes.

Dès 2013, la majorité des communes (64%) dispose, au moins en partie, d'un accès à du bon HD (carte 2.1). Rappelons toutefois que les cartes, comme les graphiques précédents, indiquent seulement si les communes disposent d'un accès au débit considéré, quel que soit leur taux de couverture, ce qui risque de surestimer leur couverture effective¹⁴. Par ailleurs, si le nombre de communes éligibles au bon HD a augmenté entre 2013 et 2018, la plupart des communes (61%) qui ne disposaient pas d'un accès au bon haut débit en 2013 demeurent dans la même situation en 2018¹⁵. Ainsi, en 2013 comme en 2018, la couverture en bon haut débit est supérieure dans l'Ouest (notamment en Bretagne et dans les Pays de la Loire), en Rhône-Alpes et en Alsace, où la quasi-totalité des communes sont couvertes dès le début de la période. L'accès au bon HD a toutefois nettement progressé dans quelques rares RIP sur la période (ex : Haute-Saône). Plus globalement, en 2018, on observe une fracture opposant la partie Ouest du pays à la partie Est, relativement moins bien couverte. Globalement, malgré l'intérêt d'évaluer l'impact de l'accès au bon haut débit, sa faible variabilité sur la période nous empêche malheureusement d'en évaluer l'effet.

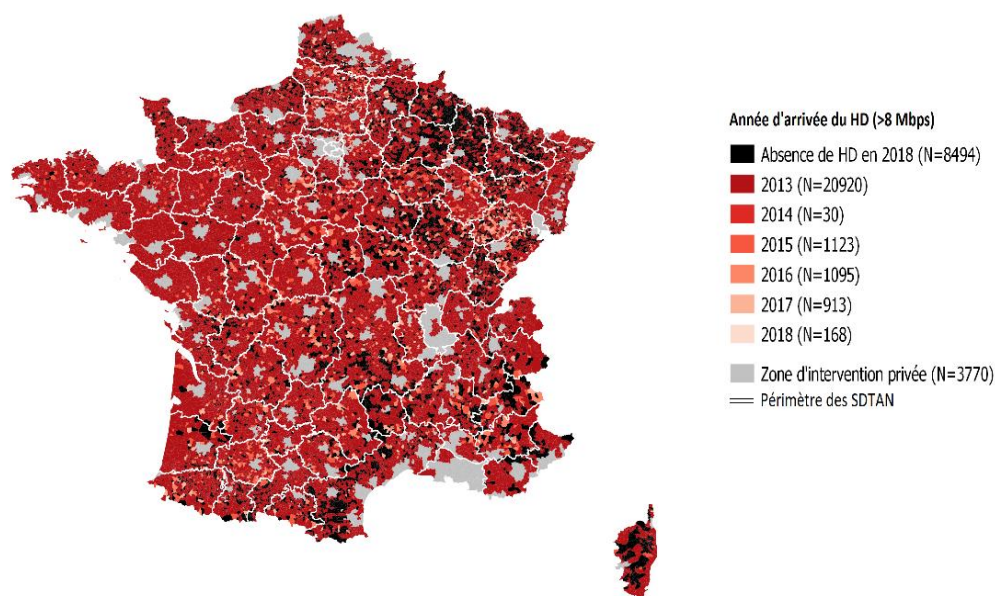
La carte de la couverture en THD fait apparaître davantage d'hétérogénéité entre la situation des communes (carte 2.2). Celles bénéficiant d'un déploiement plus précoce du THD sont concentrées sur la côte Ouest (Loire Atlantique, Maine et Loire, Gironde et Landes), la côte Est (Var), à proximité de la frontière Suisse (Ain, Haute-Savoie) et en Alsace. De plus, dans l'ensemble des RIP, les communes peu denses situées à proximité des agglomérations urbaines (zone d'initiative privée) bénéficient en règle générale d'un déploiement plus précoce. Par ailleurs, entre 2013 et 2018, le THD a nettement progressé dans certains RIP qui en étaient peu dotés en 2013 (ex : Indre, Loire et Cher, Mayenne, Oise, Aveyron, Gers).

Enfin, les différences de couverture en fibre sont très nettes entre les RIP (carte 2.3). Certains sont ainsi très en avance, à la fois parce que la mise en œuvre locale du Plan FTTH

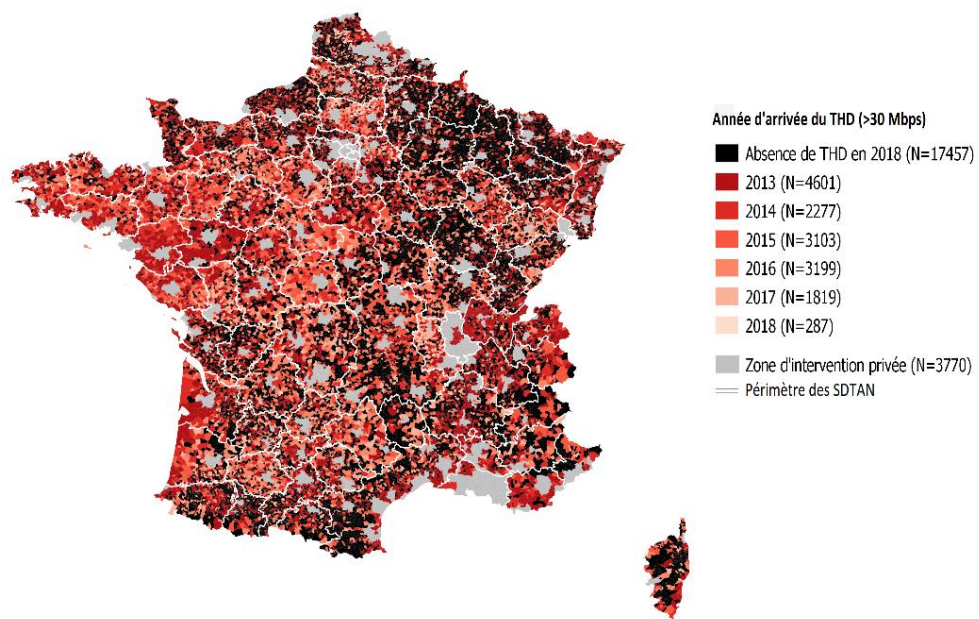
¹⁴ Par exemple, en 2013, le taux de couverture (% des locaux couverts) moyen des communes disposant d'un accès au bon HD est de 62%.

¹⁵ De plus, le taux de couverture moyen des communes a seulement légèrement augmenté, passant de 62% à 69%.

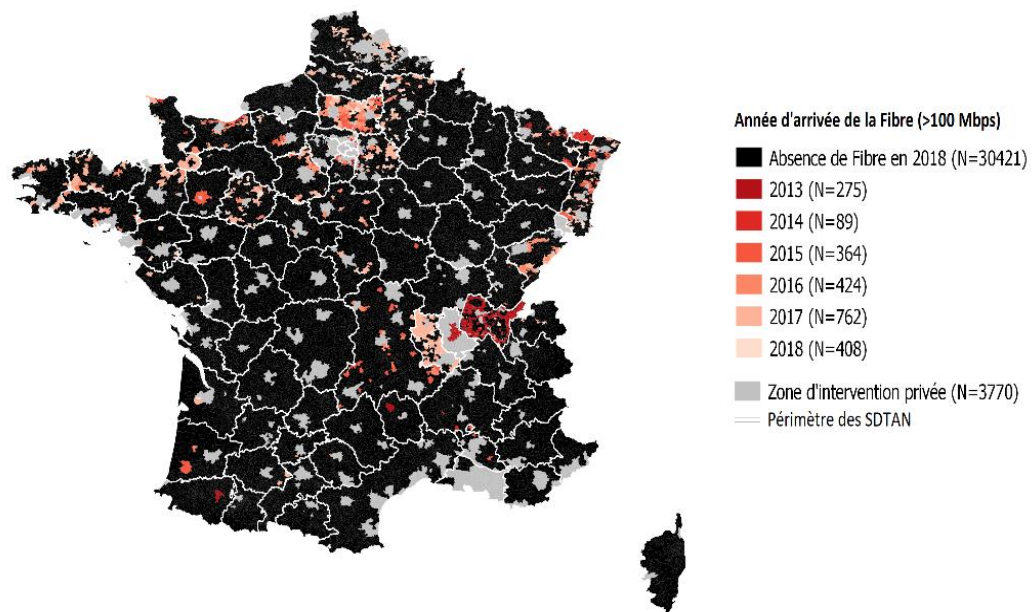
s'est faite de manière plus précoce et parce que les acteurs ont fait le choix de déployer de la fibre directement plutôt que de s'appuyer sur des technologies alternatives dans un premier temps. Trois RIP se démarquent de par leur déploiement précoce de la fibre : le Rhône, l'Ain, la Loire et l'Oise. A cela s'ajoute quelques RIP où des déploiements, bien que moins précoces et généralisés, ont déjà eu lieu (Sarthe, RIP bretons, Calvados, Manche, Nord-Pas-de-Calais, Alsace, Doubs, Auvergne).



Carte 2.1. Accès des communes à 8 Mbps et plus au sein de la ZIPU



Carte 2.2. Accès des communes à 30 Mpbs et plus au sein de la ZIPU



Carte 2.3. Accès des communes à 100 Mbps et plus au sein de la ZIPU

Encadré 2.1. Données de couverture internet fixe.

Les données de couverture internet fixe, publiées d'abord par l'Observatoire du THD de l'Agence du Numérique puis par l'ARCEP, renseignent sur la couverture de l'ensemble des communes françaises par classes de débit. Plus précisément :

Les données de l'Agence du Numérique renseignent sur la couverture des communes, du 4^{ème} trimestre 2013 au 1^{er} trimestre 2018. On connaît ainsi pour chaque commune la part de locaux (logements, entreprises et sites publics) selon six classes de débit:

- Non éligibles à internet,
- Éligibles à un débit inférieur à 3Mbps,
- Éligibles à un débit entre 3 et 8 Mbps,
- Éligibles au « bon » haut-débit (entre 8Mbps et 30 Mbps),
- Éligibles au très haut débit (entre 30Mbps et 100 Mbps),
- Éligibles à la fibre (débit supérieur à 100Mbps).

Ces données, qui étaient initialement publiées par classes de débit mais pour toutes les technologies confondues, sont ventilées, à partir du T2 2015, par classes de débit et par technologie (DSL, câble, fibre).

Données de l'ARCEP. Via son outil « Ma connexion internet », l'ARCEP publie en open-data des données trimestrielles de couverture à la maille infra-communale du T3 2019 au T4 2021. Les données sont ventilées par technologie et selon sept classes de débit :

- Non éligibles au haut débit,
- Éligibles à un débit entre 512 Kbps et 3 Mbps,
- Éligibles à un débit entre 3 et 8 Mbps,
- Éligibles à un débit entre 8 et 30 Mbps,
- Éligibles à un débit entre 30 et 100 Mbps,
- Éligibles à un débit entre 100 Mbps et 1 Gbps,
- Éligibles à un débit supérieur à 1 Gbps.

Contrairement aux données de l'Agence, qui ne renseignent que sur le déploiement des technologies filaires, l'Arcep fournit des données sur (i) toutes les technologies (Fibre, câble, DSL, THD Radio, 4g five, HD Radio, et Satellite), (ii) seulement les technologies filaires, et (iii) seulement les technologies terrestres.

Combiner les données de l'Agence du Numérique et de l'ARCEP permet donc d'avoir des informations sur la couverture communale du T4 2013 au T4 2021 (avec des données manquantes entre le T2 2018 et le T3 2019). Toutefois, cela pose certains problèmes puisque

les données de l'Agence et de l'ARCEP ne sont pas totalement comparables (celles-ci ne reposent pas sur le même référentiel de locaux).

Etant donné ces problèmes de comparabilité et la disponibilité des données sur les entreprises, dans ce rapport nous mobilisons exclusivement les données de l'Agence du Numérique, et donc, évaluons les retombées des réseaux déployés entre 2014 et 2018.

Enfin, bien que les données de l'Agence offrent des intérêts certains (couverture exhaustive des communes, information sur le taux d'éligibilité des locaux et les débits), il est important de garder en mémoire qu'elles ne renseignent que sur le déploiement des technologies filaires. De ce fait, on risque potentiellement de sous-estimer la couverture de communes bénéficiant, par exemple, de technologies radio (ce qui peut « contaminer les groupes de contrôle »).

Encadré 2.2. Données de couverture internet mobile.

Nous disposons de deux principales sources de données sur la couverture internet mobile. D'une part, l'ANFR publie des données géolocalisées sur les antennes 3G et 4G. Cependant, celles-ci ne permettent pas de reconstituer l'état de la couverture mobile dans le temps¹⁶. D'autre part, nous disposons également des données SIG sur la couverture 3G et 4G du territoire, publiées par l'ARCEP, qui sont directement issues des déclarations des opérateurs. Celles-ci permettent de calculer, pour chaque année depuis 2013, la part de la surface de chaque commune française couverte en 3G et en 4G, selon chacun des quatre grands opérateurs (Orange, SFR, Free, Bouygues). Connaître le nombre d'opérateurs présents permet aussi d'identifier les zones blanches et zones grises en 3G et en 4G.

¹⁶ Nos échanges avec l'ANFR nous ont confirmé que la date associée à chaque antenne correspond à la date de sa dernière mise à jour. Comme les antennes connaissent des mises à jour fréquentes (changement de technologie, etc.), il n'est pas possible de reconstituer l'historique pour chaque antenne, ce qui exclut toute possibilité de mobiliser ces données dans une analyse en panel.

Malheureusement, de simples statistiques descriptives nous ont confirmé que ces données surestiment grandement la couverture mobile réelle. A ce stade, il nous semble donc peu prudent de tenter de mesurer la couverture internet mobile au niveau national. Notons que l'absence de données fiables sur la couverture mobile en France n'empêche pas l'évaluation des retombées du déploiement de RIP, qui reposent essentiellement sur des technologies filaires. L'absence de données sur la couverture mobile, courant dans les travaux existants (Bertschek et al., 2015) est toutefois regrettable car un nombre croissant d'agents recourent à ce type de connexion, notamment dans les espaces ruraux (Manlove et Whitacre, 2019).

Chapitre 3. Méthodologie : identifier l'impact du THD

1. Stratégie d'identification

On souhaite estimer l'impact de deux traitements sur les entreprises :

- L'effet du très haut débit (> 30 Mbps),
- L'effet de la Fibre (> 100 Mbps).

Pour établir l'impact causal d'internet sur les entreprises, nous choisissons de mobiliser des méthodes quasi-expérimentales d'évaluation de politiques publiques (appariement et doubles différences). Celles-ci consistent à comparer des entreprises situées dans des communes ayant bénéficié de l'arrivée d'internet (« groupe traité ») à des entreprises similaires situées dans des communes comparables mais n'ayant pas encore bénéficié du déploiement (« groupe de contrôle »). La mise en œuvre de ces méthodes requièrent de définir dans un premier temps les groupes traités et de contrôle.

1.1. Définition des groupes de traités

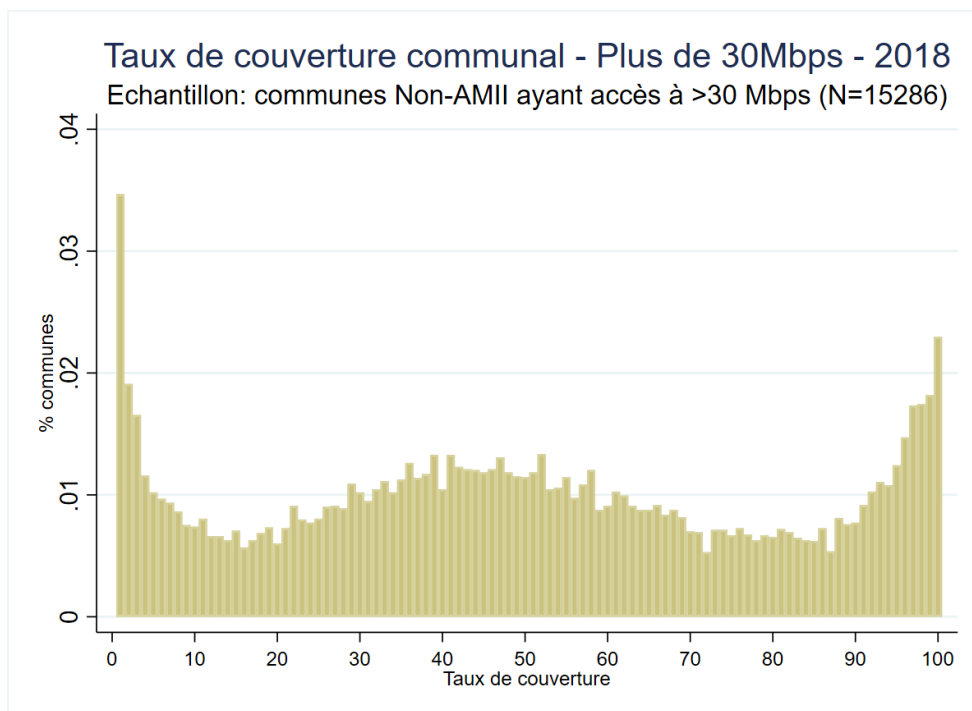
On définit les deux groupes de traités suivants :

- Traités THD ou plus : accès à un débit supérieur ou égal à 30 Mbps,
- Traités Fibre : accès à un débit supérieur à 100 Mbps¹⁷.

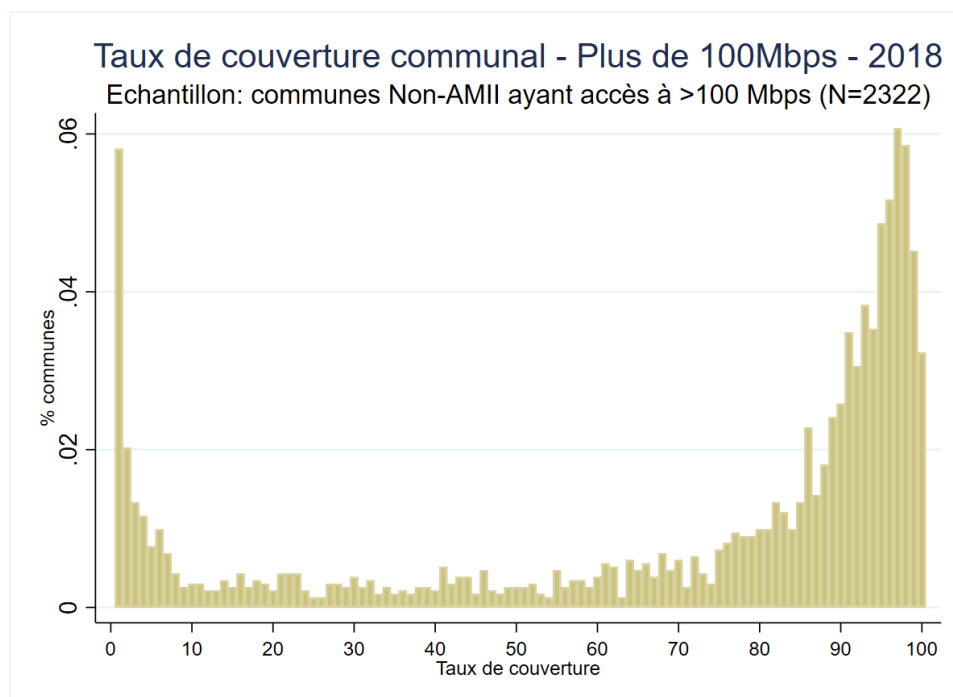
Par ailleurs, les spécificités du déploiement et, en particulier, le fait que certaines communes ont un taux de couverture très faible complexifie la constitution des groupes de traités. Les graphiques 3.1 et 3.2 représentent la distribution des taux de couverture des communes selon les deux classes de débit auxquelles on s'intéresse ici (> 30 Mbps, > 100 Mbps)¹⁸. Les graphiques sont basés sur les niveaux de couverture en 2018 (des observations similaires peuvent être faites si en retenant d'autres années).

¹⁷ Une autre option serait de définir les groupes de traités de manière « exclusive » : traités bon HD (8-30 Mbps), traités THD (30-100 Mbps), traités fibre (> 100 Mbps). Cependant, cette option risquerait, selon nous, de créer des effets de composition dont les conséquences méthodologiques sont difficilement prévisibles.

¹⁸ Seuls les taux de couverture des communes éligibles aux classes de débit données sont représentés.



Graphique 3.1 : Taux de couverture des communes ayant accès à plus de 30Mbps



Graphique 3.2 : Taux de couverture des communes ayant accès à plus de 100Mbps

Pour le THD comme pour la fibre, une proportion importante de communes a accès à la technologie de manière très partielle. Par exemple, près de 10% des communes ayant accès à la fibre ont moins de 3% de leurs locaux éligibles à 100Mbps ou plus.

L'existence de taux de couverture très faibles constitue un enjeu méthodologique. D'une part, comme l'impact d'internet est vraisemblablement négligeable dans les communes dont la part de locaux éligibles est réduite, on peut s'interroger sur la pertinence de considérer comme traitées des communes dont le taux de couverture est très faible. D'autre part, la grande variabilité des taux de couverture pour une technologie donnée rend nécessaire la prise en compte de cette hétérogénéité dans l'analyse.

Plusieurs options sont envisageables. La première, la plus simple mais aussi la plus arbitraire, consisterait à ne considérer comme traitées que les communes dont le taux de couverture est supérieur à un seuil minimum (ex : 10% des locaux couverts). Une deuxième solution reviendrait à estimer un effet du traitement séparément pour différents sous-échantillons, en fonction du taux de couverture des communes. Cette méthode est plus satisfaisante car elle permettrait d'examiner si l'effet du déploiement varie en fonction de son intensité. Enfin, une dernière option consisterait à estimer un effet moyen différent selon l'intensité du traitement, en mobilisant par exemple une méthode d'appariement par score de propension généralisé ou en introduisant une variable de traitement continue dans le modèle de doubles-différences.

1.2. Définition des groupes de contrôle

Les entreprises situées dans les communes non couvertes de la *zone d'initiative publique* constituent les unités de contrôle les plus évidentes, puisqu'elles sont situées dans un contexte territorial a priori très comparable à celles implantées dans les communes traitées de la zone publique (espaces peu denses).

On peut toutefois regrouper différemment les communes non couvertes de la zone d'initiative publique pour créer les groupes de contrôle. Plusieurs groupes de comparaison peuvent être envisagés pour les groupes de traités « THD ou plus » et « Fibre ».

Le groupe de traités « THD ou plus » peut être comparé à trois groupes de contrôle :

- GC1_{THD} : débit < 30 Mbps,
- GC2_{THD} : débit < 8Mbps,
- GC3_{THD} : débit entre 8 et 30 Mbps.

Chacun de ces groupes de contrôle a ses propres avantages et inconvénients. Le groupe GC1_{THD} permet de disposer d'un grand nombre d'unités de comparaison. Toutefois, comparer les entreprises ayant accès au THD à celles ayant soit un accès au bon HD, soit un accès à rien, peut rendre l'interprétation du coefficient estimé difficile (capte-t-on un effet THD ? Un effet HD ?). Pour résoudre cette difficulté, on peut alors comparer les traités « THD ou plus » au GC2_{THD} et GC3_{THD}. En particulier, le GC3_{THD} (débit entre 8 et 30 Mbps) permettrait d'estimer l'effet *additionnel* du THD par rapport à du bon HD.

De même, le groupe de traités « Fibre » peut être comparé à au moins trois groupes de contrôle :

- GC1_{Fibre} : débit < 100 Mbps,
- GC2_{Fibre} : débit < 30 Mbps,
- GC3_{Fibre} : débit entre 30 et 100 Mbps.

Ici encore, le GC1_{Fibre} peut rendre difficile l'analyse de l'impact estimé (capte-t-on un effet Fibre ? Un effet THD ? Un effet HD ?). Comparer le groupe de « traités Fibre » au groupe GC3_{Fibre} permettrait d'analyser si la fibre génère des gains additionnels par rapport au THD.

1.3. Menaces sur l'identification: sources d'endogénéité du traitement

Timing du déploiement au sein des RIP

Dans le cas de notre évaluation du programme Auvergne Très Haut Débit (Duvivier et al., 2018), des entretiens préliminaires avec des experts locaux nous ont indiqué que le déploiement se fait de manière prioritaire dans certaines communes, en fonction de leurs caractéristiques. Concrètement, celui-ci est programmé de sorte à réaliser un compromis entre rentabilité économique et nécessité de raccorder en priorité les moins connectés. Les communes couvertes prioritairement sont donc, d'une part, les communes les plus grandes et les plus accessibles, pour lesquelles le nombre d'abonnés potentiels est le plus élevé et le coût du raccordement le plus faible, et, d'autre part, les communes qui rencontrent les problèmes d'accès à internet les plus importants.

De ce fait, une simple comparaison entre les entreprises situées dans des communes bénéficiaires et celles situées dans les communes non encore bénéficiaires risquerait de biaiser l'impact estimé, en attribuant l'effet d'autres caractéristiques au Plan FTTHD.

Le problème peut être facilement solutionné si le calendrier du déploiement des RIP dépend

essentiellement de caractéristiques observables ou inobservables invariantes dans le temps. En revanche, le défi sera plus complexe à résoudre si les communes ayant accès au THD de manière plus précoce sont aussi celles pour lesquelles des événements non-observables empiriquement affectent la performance future. Par exemple, si l'arrivée du THD est directement liée à des investissements publics massifs, comme des infrastructures, qui ont pour conséquence d'augmenter l'emploi et générer de l'activité par un effet multiplicateur.

Est-ce que cette menace sur l'identification est plausible ? Bien que cela ne soit pas exclu, plusieurs éléments laissent à penser que les conséquences sur notre analyse empirique sont limitées.

Tout d'abord, même en admettant que le mécanisme précité affecte toutes les unités étudiées, il n'affecte sûrement pas toutes les mesures de performance ni tous les secteurs. En particulier, il est peu probable qu'un tel mécanisme affecte toutes les entreprises d'un territoire, ce qui fait que les analyses au niveau de l'entreprise permettent de se préserver de tels problèmes.

Ensuite, il est très peu probable que la décision de connexion au THD soit uniquement attaché à ces investissements futurs. Dans les RIP, ce sont les acteurs politiques qui décident de l'ordre de déploiement de la fibre et il semble donc plus plausible que la date de déploiement dépende de caractéristiques observables, telle que la taille des communes, le niveau d'accessibilité ou le débit préexistant, avec différentes priorités en fonction des objectifs des élus. Ainsi, même si au final les communes les plus denses sont connectées les premières, cela ne remet pas en question la causalité des estimations à partir du moment où les critères de décision sont observés (ex : population, revenu par habitant, taux d'accès à internet)¹⁹.

Enfin, il existe souvent un délai entre l'élaboration du calendrier et la date effective de déploiement. Cela implique que même s'il y avait de l'endogénéité au moment de la décision, le temps que le THD soit déployé, cela peut être pris en compte via la dynamique avant traitement et ainsi ne pas affecter la causalité des résultats.

Sorting des entreprises

La migration des entreprises constitue une autre source d'endogénéité potentielle. On pourrait en effet supposer que les entreprises les plus performantes, ou qui souhaitent précisément utiliser internet, s'implantent dans les communes bénéficiant d'un déploiement du

¹⁹ Les travaux existants (Canzian et al., 2018 ; Malgouyres et al., 2019) indiquent que la diffusion de la technologie s'explique essentiellement par des variables observables ou inobservables invariantes dans le temps (ciblage des communes les plus denses ; existence préalable d'infrastructures). Le timing du déploiement serait donc exogène une fois que l'hétérogénéité observée et inobservée fixe dans le temps sont contrôlées.

THD. Cette hypothèse est toutefois peu plausible.

En effet, notre analyse de la performance des entreprises, qui repose dans la plupart des cas sur l'estimation de doubles différences (cf. ci-après), ne concerne que des entreprises qui existaient déjà avant l'arrivée du THD. Or, comme les coûts de relocalisation sont très élevés, il est très peu probable que les entreprises existantes se réimplantent dans une autre commune uniquement pour bénéficier d'un accès au THD (DeStefano et al, 2014).

Pour conclure, pour qu'il existe un biais positif au déploiement sur la performance, il faut réunir un ensemble d'hypothèses fortes. De plus, comme l'évaluation porte sur un panel large d'indicateurs, quand bien même ce biais existerait sur une mesure, il est peu plausible qu'il porte sur toutes.

1.4. Les solutions méthodologiques envisagées selon le type de données

Dans la suite de ce chapitre, nous présentons les solutions méthodologiques envisagées pour tenir compte de l'hétérogénéité observée et inobservée entre les unités traitées et leurs contrefactuels. Notre évaluation s'intéresse à de nombreuses variables de résultats (usages, performance, innovation), qui sont observables à différentes échelles de temps. Pour la plupart d'entre elles, nous pouvons construire un panel où les unités sont observées de façon répétée au cours du temps et s'intègrent donc très bien dans l'analyse par doubles différences. L'identification d'un effet causal du Plan FTHD semble alors possible puisque la méthode permet de contrôler l'hétérogénéité fixe dans le temps (au moyen d'effets fixes) et l'hétérogénéité observée (au moyen d'un appariement préalable).

D'autres indicateurs (mesures qualitatives de l'innovation, usages numériques) ne sont toutefois observables qu'en coupe répétée, c'est à dire que l'on n'observe pas ou seulement peu de fois les mêmes entreprises dans le temps, ce qui empêche d'utiliser le cadre des doubles différences et complique la tâche de trouver des effets causaux. Il s'agira alors de mettre en œuvre un batterie d'analyses de sensibilité pour s'assurer de la robustesse des effets obtenus à différentes sources de biais potentiels.

Nous présentons d'abord la méthodologie envisagée dans le cas des données de panel (section 2) puis celle identifiée dans le cas des analyses en coupe répétée (section 3).

2. L'identification dans le cas des données de panel

Pour établir l'impact causal de l'effet du THD sur les entreprises, le cadre d'analyse principal que nous utiliserons est la méthode des différences-en-différences (DD). Cette méthode

quantifie l'effet du traitement, ici l'accès au THD, en comparant l'évolution des individus traités, avant et après le traitement, à celle des contrôles.

Nous détaillons ici les différentes mises en œuvre de la méthode des DD à notre objet d'analyse. En particulier, la mise en œuvre du THD se fait de façon progressive, c'est à dire que la période à laquelle le traitement est reçu est hétérogène. Ce cadre d'analyse a ses propres difficultés et potentielles solutions que nous discuterons.

2.1. Mise en œuvre de la DD standard

La relation économétrique que nous estimerons est représentée par l'équation suivante:

$$y_{it} = \alpha_i + \beta_t + \gamma \cdot \tau_{it} + \epsilon_{it} \quad (1)$$

avec y_{it} la variable dépendante qui nous intéresse (par exemple le profit), pour l'entreprise i et l'année t ; α_i et β_t sont deux effets fixes au niveau de l'entreprise et de l'année, capturant tout effet invariant au niveau de l'individu ou du temps; et τ_{it} est une variable binaire représentant le traitement et vaut 1 si l'unité i a accès au traitement en t , et 0 sinon. Le coefficient qui nous intéresse est γ qui représente l'effet causal du traitement.

Il peut être intéressant aussi de capturer l'effet dynamique du traitement. L'utilisation du THD requiert vraisemblablement un apprentissage, et l'on peut donc s'attendre à ce qu'il se passe plusieurs années avant que l'impact de l'accès ne se réalise. Afin de capturer l'effet dynamique du traitement, on utilisera également la représentation suivante:

$$y_{it} = \alpha_i + \beta_t + \sum_{d \neq -1} \gamma_d \cdot 1_{t-c_i=d} + \epsilon_{it} \quad (2)$$

avec γ_d l'effet du traitement d périodes après l'avoir reçu, c_i correspondant à la date d'accès au traitement de l'unité i , c'est à dire sa cohorte (pour les unités de contrôle, $c_i = +\infty$). Les coefficients γ_d représentent l'évolution de l'effet du traitement. En particulier on s'attend à ce que ces coefficients soient non statistiquement différents de 0 pour $d < 0$, c'est à dire pour les périodes antérieures au traitement, afin que l'hypothèse de tendances parallèles entre les contrôles et les traités soit valide²⁰. Si l'effet du traitement est dynamique, les coefficients devraient s'éloigner de 0 graduellement pour $d > 0$.

²⁰ L'hypothèse des tendances parallèles, centrale dans le cadre cette méthode, stipule qu'en l'absence de traitement, les traités auraient évolué de la même façon que les contrôles. Cette hypothèse peut être vérifiée empiriquement par une analyse graphique et l'estimation d'un modèle sur la période précédant le traitement.

2.2. Problèmes spécifiques à la DD avec traitement échelonné

Dans notre analyse, les communes « traitées » de la zone d'intervention publique ne bénéficient pas toutes du THD à la même période, ainsi le traitement se fait de façon échelonnée dans le temps. Cette caractéristique amène des problèmes méthodologiques spécifiques que nous détaillons à présent.

Le modèle économétrique décrit dans l'équation (1) porte sur l'effet total du traitement et contient des effets fixes individuels et annuels. Ceux-ci permettent de capturer les effets spécifiques et invariants au niveau de l'entreprise et de l'année. En particulier les effets fixes années sont importants car ils permettent de ne pas attribuer au traitement, de façon erronée, un choc temporel négatif (ou positif) affectant toutes les entreprises. Toutefois, le problème avec les effets fixes annuels est qu'ils capturent, par définition, l'effet moyen de l'année. Ainsi, en extrapolant, si au cours d'une année il n'y a que des individus traités, alors l'effet du traitement *pour cette année* est indiscernable de l'effet annuel commun à tous les individus. De plus on voit bien que dans le cas d'un traitement échelonné la composition des groupes traités varie chaque année²¹. Ce changement de composition implique que ce que capture l'effet fixe annuel est modifié au cours du temps (en particulier, si la taille des groupes augmente dans le temps, il devrait capturer de plus en plus l'effet du traitement *spécifique à l'année*). Dans le cas d'un effet du traitement homogène, c'est à dire identique pour chaque année, cela ne pose pas de problème d'inférence. Par contre dans le cas d'un effet du traitement hétérogène²², par exemple croissant au cours du temps, cela risque d'entraîner une atténuation de l'effet du traitement.

²¹ Selon la définition des groupes de traités, l'évolution de leur taille sera différente. A ce stade, nous devons encore choisir entre deux options. Une première option consiste à définir les groupes de traités de manière exclusive : groupe HD (8-30 Mbps), groupe THD (30-100 Mbps) et groupe Fibre (> 100 Mbps). Une deuxième option consiste à définir les groupes de traités en fonction d'un seuil minimum de débit : groupe HD ou plus (> 8 Mbps), groupe THD ou plus (> 30 Mbps), groupe Fibre (> 100 Mbps). Dans le premier cas, la taille des groupes de traitement peut diminuer ou augmenter au cours du temps. En revanche, dans le second cas, la fraction des unités traitées augmente chaque année (car aucun individu ne sort du traitement).

²² L'hétérogénéité peut prendre deux formes : (i) en fonction de la durée écoulée depuis le traitement ; (ii) en fonction de la cohorte.

Illustration du problème.

Pour donner un exemple du problème, générons les données suivantes: supposons que l'effet du traitement soit positif et progressif et que les individus soient traités à des périodes différentes. La variable dépendante est générée de la façon suivante:

$$y_{it} = \tau_{it} \cdot (1 + t - c_i) + \epsilon_{it}$$

avec τ_{it} un indicateur indiquant le statut du traitement, $(t - c_i)$ représentant le nombre de périodes depuis que le traitement a été reçu et ϵ un bruit blanc. Considérons un panel équilibré de 10 périodes où il y a 11 cohortes, une cohorte ne recevant jamais de traitement et les dix autres recevant le traitement de la période 1 à la période 10. Comme décrit par l'équation précédente, l'effet du traitement augmente au fil du temps. Cela est représenté dans la Figure 3.2 pour les cohortes traitées. On constate que la moyenne de l'effet du traitement est de 4.

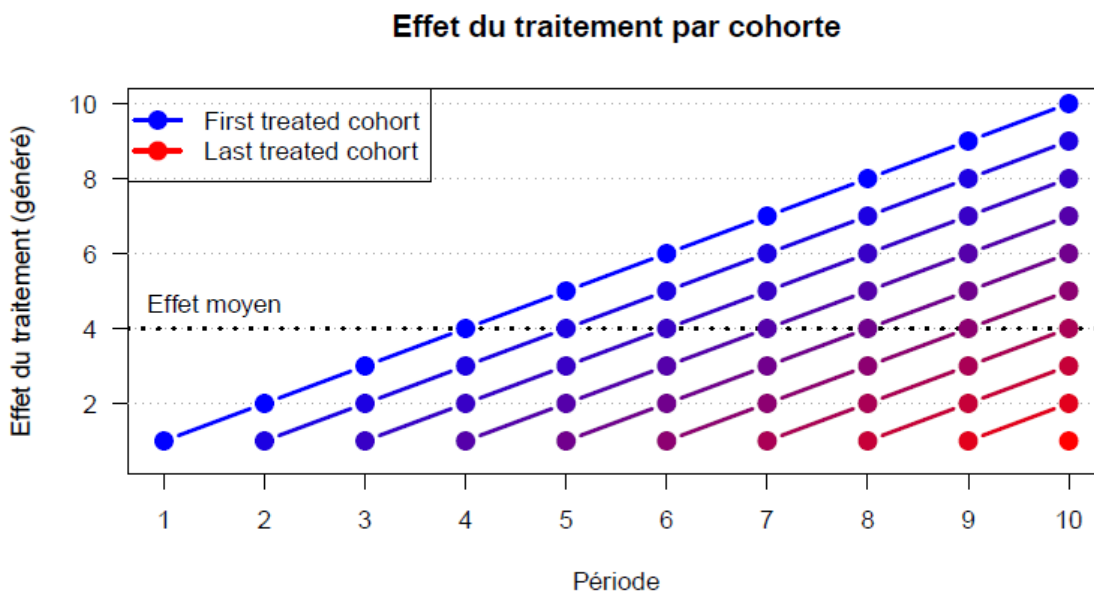


Figure 3.2: Effet moyen du traitement théorique par cohorte.

Maintenant estimons l'effet total du traitement par l'équation (1) tout d'abord sans individus de contrôle puis avec autant d'individus de contrôle que d'individus traités. Les résultats sont présentés dans le tableau 3.1. On constate tout d'abord que si l'on n'utilise pas d'individus de contrôle, l'effet total du traitement estimé est même négatif (colonne 1) ! Lorsque l'on introduit les contrôles (colonne 2), l'effet estimé est cette fois positif mais reste en deçà de moitié de la valeur moyenne de 4. Cet exemple très simple illustre donc les problèmes liés à un traitement échelonné, comme dans notre cas d'étude.

Dependent Variable:		y
With controls	no	yes
Model:	(1)	(2)
<i>Variables</i>		
treat × post	-0.0266 (0.2934)	2.148 *** (0.2479)
<i>Fixed-effects</i>		
Individual	Yes	Yes
Year	Yes	Yes
<i>Fit statistics</i>		
Observations	500	1000
R ²	0.766734	0.651971
Within R ²	2.931×10 ⁻⁵	0.120433

*One-way (Individual) standard-errors in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Tableau 3.1. Estimation de l'effet total du traitement, sans et avec individus de contrôle.

Solutions possibles

Ce problème de biais de l'estimateur lorsque le traitement est hétérogène a motivé de nombreux travaux de recherche. Récemment de Chaisemartin et D'Haultfoeuille (2020) ont proposé un estimateur non biaisé pour ce genre de données. Si l'on ne considère que des individus qui accèdent au traitement sans en sortir²³, définissons DID_t l'estimateur de différence-en-différence pour une cohorte qui est traitée en t :

$$DID_t \equiv \frac{1}{N_{0 \rightarrow 1,t}} \cdot \sum_{i \in G_{0 \rightarrow 1,t}} (y_{i,t} - y_{i,t-1}) - \frac{1}{N_{0 \rightarrow 0,t}} \cdot \sum_{i \in G_{0 \rightarrow 0,t}} (y_{i,t} - y_{i,t-1})$$

avec $G_{0 \rightarrow 1,t}$ l'ensemble des individus qui sont traités à la période t et $G_{0 \rightarrow 0,t}$ les individus n'ayant jamais été traités jusqu'en t ; et $N_{0 \rightarrow 1,t}$ (resp. $N_{0 \rightarrow 0,t}$) le cardinal de $G_{0 \rightarrow 1,t}$ (resp. $G_{0 \rightarrow 0,t}$). L'estimateur de Chaisemartin and D'Haultfoeuille (2020) a alors la forme suivante:

$$DID_M \equiv \sum_{t=2}^T \frac{N_{0 \rightarrow 1,t}}{N_S} DID_t$$

avec $N_S = \sum_{t=2}^T N_{0 \rightarrow 1,t}$ le nombre total d'individus traités (en excluant les individus toujours traités). DID_M est donc une moyenne pondérée de plusieurs estimateurs de différence-en-différence. Dans le théorème 3, les auteurs montrent que DID_M est un estimateur non biaisé de l'effet total du traitement. Afin d'estimer DID_M , il faut comparer les individus qui viennent juste d'accéder au traitement à ceux

²³ Cas où l'on définirait les groupes de traités en fonction d'un seuil minimum de débit : groupe HD ou plus (> 8 Mbps), groupe THD ou plus (> 30 Mbps), groupe Fibre (> 100 Mbps).

qui n'y ont jamais eu accès. Cela implique que les individus traités ne peuvent être utilisés qu'une seule fois, l'année où ils ont reçu le traitement (bien qu'ils puissent être utilisés plusieurs fois comme contrôle). L'inconvénient majeur de cette méthode est donc qu'il n'est pas possible d'estimer l'effet du traitement à plusieurs périodes. Cela est particulièrement problématique dans notre étude où l'on s'attend à ce que l'accès au très haut débit n'ait pas un effet immédiat, rendant l'emploi de cette méthode vacant²⁴.

Une autre méthode consiste à estimer l'effet du traitement total à partir des effets annuels, via l'équation (2). En effet, si l'hétérogénéité du traitement est seulement annuelle alors l'estimation d'effets annuels est sans biais, comme illustré dans la Figure 3.3 qui reporte les effets annuels du traitement.

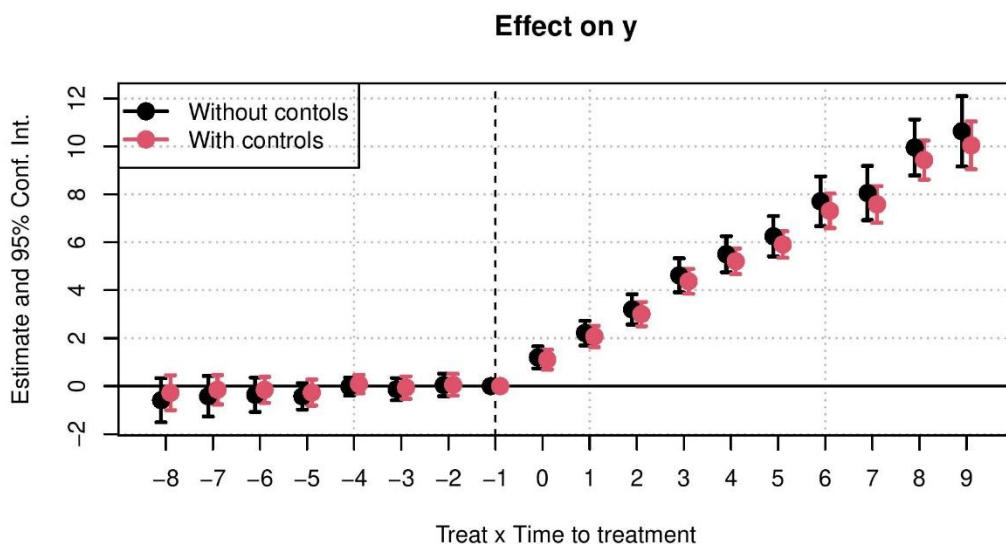


Figure 3.3: Estimateurs de l'effet annuel du traitement.

Mais dans le cas où l'effet du traitement est aussi hétérogène par cohorte traitée²⁵, les effets annuels peuvent également être biaisés. Sun et Abraham (2021) montrent que l'effet du traitement d'une année peut être contaminé par l'effet du traitement des autres années. Soit $CATT_{c,d}$ l'effet du traitement pour la cohorte c et la période relative d , la cohorte c étant définie par la période de première réception du traitement. Avec ces notations, la proposition 3 de Sun et Abraham (2021) énonce que:

²⁴ A noter que cette méthode peut être adaptée par exemple en regroupant les périodes (i.e. faire du « binning »); cela éloignerait néanmoins ce nouvel estimateur de leur méthode originale et de ses propriétés.

²⁵ Cela risque d'être le cas dans notre évaluation puisque les communes les plus grandes, pour lesquelles on s'attend à ce que l'impact du THD soit supérieur, ont tendance à bénéficier les premières du déploiement.

$$\gamma_d = \sum_c \omega_{c,d}^d CATT_{c,d} + \sum_{d' \neq d, d' > 0} \sum_c \omega_{c,d'}^d CATT_{c,d'}$$

c'est à dire que l'effet du traitement estimé par l'équation (2), γ_d , est une somme pondérée des effets du traitement à différentes périodes, et pas seulement en d . Les poids du deuxième terme, $\omega_{c,d'}^d$, ont la particularité d'annuler ce terme dans le cas où l'effet du traitement est homogène entre cohortes.

Les auteurs proposent alors une méthode alternative qui ne souffre pas de biais dans le contexte où les effets du traitement sont hétérogènes par cohortes. Cette méthode est la suivante:

Étape 1. Estimer le modèle suivant par les MCO:

$$\gamma_{it} = \alpha_i + \beta_t + \sum_{c \neq \infty} \sum_{d \neq -1} \delta_{c,d} \cdot 1_{c_i=c} \cdot 1_{t-d_i=d} + \epsilon_{it}$$

avec c_i la cohorte de l'individu i et $c_i = \infty$ si i n'a jamais été traité.

Étape 2. Calculer l'effet du traitement à la période d , \hat{v}_d , par :

$$\hat{v}_d \equiv \sum_c s_{c,d} \cdot \delta_{c,d}$$

avec $s_{c,d}$ la proportion des individus recevant le traitement en période d qui proviennent de la cohorte c .

Callaway et Sant'Anna (2021) offrent également une alternative: plutôt que d'effectuer une régression avec des effets fixes, ils proposent d'estimer de multiples estimations par DD simples, à chaque fois avec deux groupes et deux périodes, et ce pour toutes les cohortes c et toutes les périodes t . Soit $I_c = \{i | c_i = c\}$, l'estimateur proposé est :

$$CATT_{c,t} = \frac{1}{|I_c|} \sum_{i \in I_c} (y_{i,t} - y_{i,t-1}) - \frac{1}{|I_\infty|} \sum_{i \in I_\infty} (y_{i,t} - y_{i,t-1})$$

qui est l'estimateur par différence-en-différence comparant l'évolution de y pour la cohorte c à l'année t à l'évolution pour le groupe de contrôle. A noter que cette estimation de différence-en-différence peut être faite de multiple façons: avec ou sans appariement, en utilisant les individus jamais traités ou bien ceux non encore traités comme contrôles, avec des méthodes double robustes, etc. (voir le théorème 1 de Callaway et Sant'Anna). Une fois que tous les $CATT_{c,t}$ sont obtenus, il est possible de les agréger comme on le souhaite pour avoir, par exemple, l'effet total du traitement ou l'effet dynamique du traitement. Contrairement à la méthode de Sun et Abraham (2021) où il est possible d'obtenir des écarts-types analytiques, ici les écarts-types des estimateurs sont obtenus par bootstrapping ce qui est plus intense en termes de calculs et limite la capacité de traiter de grands volumes de données.

Pour illustrer ces estimateurs reprenons l'exemple précédent et ajoutons de l'hétérogénéité par cohorte. Maintenant y est généré de la façon suivante:

$$y_{it} = \tau_{it} \cdot [1 + (t - c_i) - c_i] + \varepsilon_{it}$$

où $(t - c_i)$ représente la période relative et c_i l'effet cohorte.

Après avoir soustrait le groupe toujours traité pour lequel ces estimations ne sont pas définies, nous estimons l'effet annuel du traitement selon trois méthodes: la méthode classique définie par l'équation (2), la méthode de Sun et Abraham (2021) et celle de Callaway et Sant'Anna (2021). Les résultats sont présentés à la figure 3.4.

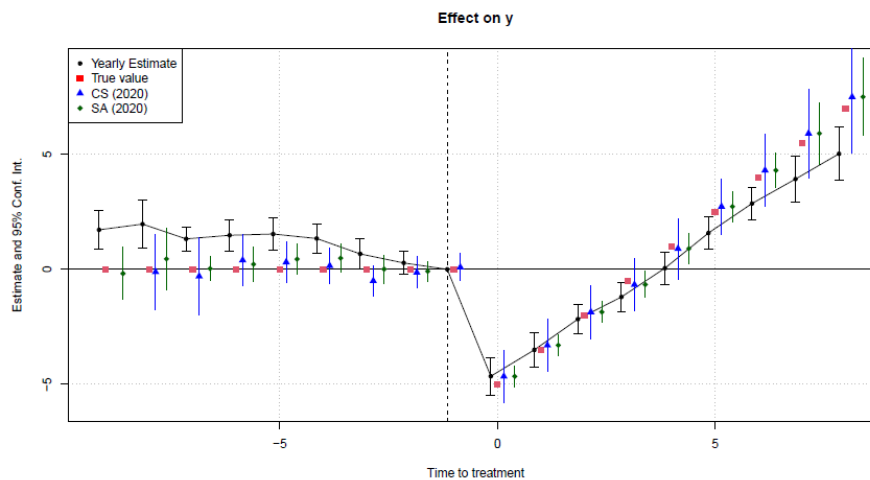


Figure 3.4: Estimateurs de l'effet annuel du traitement en présence d'hétérogénéité par cohorte.

On constate tout d'abord que la méthode classique dans ce cas trouve une présence de pré-tendance, comme tous les coefficients avant la période 0 sont statistiquement positifs, alors qu'il n'y a pas de pré-tendance par construction. Cette méthode conduirait ainsi à rejeter, à tort, l'hypothèse de comparabilité entre les groupe de traitement et ceux de contrôle (l'hypothèse de tendance parallèle). Ensuite, la méthode classique sous-estime l'effet du traitement pour les périodes relatives les plus avancées, jusqu'à 30% pour la dernière période.

Les méthodes SA et CS produisent des estimateurs très proches (à noter que les années de référence ne sont pas les mêmes). En particulier, les coefficients précédant le traitement ne sont pas statistiquement différents de 0, comme attendu, et les coefficients pour les périodes les plus avancées restent proches des valeurs génératives.

Conclusion : estimateur retenu pour l'analyse

Dans notre analyse, quand il sera possible de le faire, nous utiliserons la méthode de Sun et Abraham qui permet d'avoir des estimateurs non biaisés en présence d'hétérogénéité et qui est simple en termes de calculs. Cet estimateur permettra de calculer divers impacts : effets annuels par cohorte, effet total par cohorte, effet annuel total, etc.

3. L'identification dans le cas des données en coupe

Si la majorité des données mobilisées permettent une analyse en panel, l'enquête annuelle sur l'usage des TIC dans les entreprises (TIC Entreprises) et l'enquête Communautaire sur l'innovation (CIS) ne fournissent que des coupes répétées, ce qui empêche l'utilisation des doubles différences. L'identification d'un effet causal est alors délicate en raison de la non prise en compte de l'hétérogénéité inobservée variante et invariante dans le temps.

3.1. Le difficile recours aux variables instrumentales

Pour identifier l'effet causal d'internet, les travaux antérieurs en coupe les plus avancés recourent généralement à la méthode des variables instrumentales. Les instruments mobilisés sont basés sur un raisonnement économique du choix de déploiement et correspondent donc aux facteurs qui affectent les coûts et les revenus du déploiement d'internet. Parmi les instruments employés, on peut notamment citer la pente du terrain (Kolko, 2012; Ivus et Boland, 2015), le réseau téléphonique historique (Czernich, 2014 ; Ciapanna et Colonna, 2019), et la densité de population (Mack et al., 2011; Mack et Rey, 2014; Mack et Wentz, 2017; Van Gaasbeck, 2008)²⁶.

Recourir à ces instruments ne constitue toutefois pas, d'après nous, une solution satisfaisante dans le cadre de cette évaluation. Tout d'abord, ces variables apparaissent parfois faiblement corrélées avec l'accès à internet (instruments faibles), notamment lorsque l'on considère les technologies de nouvelles générations (Canzian et al., 2019). De plus, ces instruments, basés sur un raisonnement économique, semblent plus pertinents pour expliquer le déploiement d'opérateurs privés qu'un déploiement public. En effet, même si les variables de coûts et de revenu jouent vraisemblablement un rôle important dans l'élaboration du calendrier au sein d'un RIP (les collectivités ont besoin d'obtenir un retour sur investissement rapide pour autofinancer le programme), d'autres objectifs entrent en jeu. En particulier, le déploiement public dépend aussi de critères d'équité (ex : résoudre en priorité les problèmes de couverture les plus importants) et d'arbitrages politiques. De plus, chaque RIP peut avoir des stratégies de diffusion différentes ce qui rendrait l'analyse encore plus complexe.

Par ailleurs, on peut s'interroger sur le caractère exogène des instruments. Par exemple, même dans un modèle incluant beaucoup de contrôles, peut-on s'assurer que la pente

²⁶ Concernant la pente du terrain, l'idée est que les coûts de déploiement sont supérieurs, et donc l'accès à internet moins développé, dans les zones escarpées. D'autres auteurs ont utilisé les réseaux de télécommunication préexistant, en avançant que les coûts de déploiement sont moindres là où des infrastructures existent déjà. De plus, ces réseaux historiques ont parfois été créés par les opérateurs historiques pour des raisons extra-économiques (cf. Italie). Enfin, d'autres emploient la densité de population avec l'idée que les revenus espérés des opérateurs seront plus élevés, et donc l'accès à internet supérieur, dans les zones denses.

du terrain ou la densité de population n'affectent pas la performance des entreprises au-delà de leur effet transitant via internet ? De même, concernant les réseaux de télécommunication historiques, il est difficile de savoir si ceux-ci ne recouvrent pas d'autres réseaux (notamment de transport) qui auraient pu influencer le développement économique local à long-terme.

3.2. La méthode de l'appariement

Pour tenter d'identifier un effet causal avec des données en coupe répétée, nous envisageons de recourir à la méthode de l'appariement, qui permet de contrôler l'hétérogénéité observée entre les traités et les contrôles. Pour réduire le risque d'hétérogénéité inobservée, notre stratégie consiste à comparer les entreprises situées dans les communes traitées à celles situées dans les communes du même RIP et disposant d'un débit moindre.

Comparer les entreprises traitées à des contrefactuels appartenant au même RIP devrait réduire l'hétérogénéité inobservée. En effet, cela devrait permettre de contrôler certains facteurs inobservés invariant (ex : histoire, culture, géographie) et variant dans le temps (politiques publiques départementales et régionales).

Chapitre 4. Usages numériques

1. Revue de littérature

1.1. Définition des usages numériques

Les travaux existants ont examiné des usages numériques très divers. Si certains considèrent un indice synthétique des usages numériques, obtenu à partir d'une méthode de classification (Boquet et Brossard, 2007 ; Ruiz-Rodriguez et al., 2018), en règle générale les différents types d'usages sont analysés de manière séparée, puisque leurs facteurs explicatifs varient (DeStephano et al., 2017 ; Nicoletti et al., 2020).

Les travaux ont d'abord porté sur l'utilisation de matériel informatique, d'internet, d'emails (Chatzoglou et al., 2010 ; Galliano et Roux, 2006 ; 2008) puis sur les sites web, le commerce en ligne, la facturation électronique, l'e-administration ou les réseaux sociaux (Giotopoulos et al., 2017 ; Haller et Siedschlag, 2011). Plus récemment, les recherches ont examiné les déterminants de l'utilisation du cloud computing, de progiciel de gestion intégré, et de la gestion de la relation client (Alshamaila et al., 2013 ; Nicoletti et al., 2020).

Plutôt que d'examiner les différents types d'usages numériques séparément, certains auteurs regroupent les usages en fonction de leur niveau de complexité (Forman et al., 2005a ; 2005b ; Nicoletti et al., 2020), distinguant ainsi des applications standards (ex : utilisation d'emails, existence d'un site web, vente en ligne) et des outils plus complexes (ex : gestion de la relation client, cloud computing). D'autres distinguent les outils selon leur degré de nouveauté (Faquet et Malardé, 2020), opposant ainsi outils matures (ex : progiciels de gestion, traitement automatisé des factures) et émergents (cloud computing, big data, intelligence artificielle).

Par ailleurs, de manière intéressante, certains travaux distinguent les usages selon les besoins auxquels ils répondent. Par exemple, Forman et al. (2005b) différencient les usages favorisant la coordination des activités au sein d'un établissement (ex : intranet) et entre établissements (ex : extranet). Selon les auteurs, étant donné l'éloignement des entreprises dans le rural, celles-ci devraient être particulièrement incitées à adopter des applications favorisant la coordination entre établissements. De manière similaire, on peut également distinguer les outils favorisant le fonctionnement interne d'une entreprise (ex : systèmes intégrés de gestion) de ceux jouant un rôle dans leurs relations externes (ex : internet, commerce électronique). Dans

leur analyse des usages en milieu rural, Galliano et al. (2008) distinguent quant à eux les usages répondant à un besoin informationnel (ex : publicité en ligne, veille concurrentielle et technologique) de ceux répondant à un besoin de coordination externe (ex : extranet, visioconférence, outils de partage de ressources). Les entreprises rurales, qui font face à des coûts informationnels et de coordination supérieurs, devraient retirer un gain important de l'usage de ces deux types d'outils.

Enfin, certains auteurs distinguent l'adoption de l'intensité d'usage des outils numériques²⁷ (Bayo-Moriones et Lera-Lopez, 2007 ; Galliano et Roux, 2008 ; Haller et Siedschlag, 2011). Cette distinction est importante, notamment lorsque l'on considère des outils qui en sont à différents stades de diffusion (Battisti et Stoneman, 2003 ; 2005). En effet, si pour les technologies récentes on observe surtout des écarts en termes d'adoption, lorsque celles-ci se diffusent, les écarts entre entreprises se situent surtout au niveau de l'intensité d'usage des technologies.

1.2. Impact de l'accès à internet sur les usages numériques

Peu de travaux ont examiné l'effet de l'accès aux infrastructures internet sur les usages numériques des entreprises et, dans la plupart des cas, celui-ci ne constitue qu'une variable de contrôle parmi d'autres. Les résultats obtenus sont contrastés.

Par exemple, alors que l'accès à une connexion plus rapide est associé à un usage plus intensif d'internet et des emails dans les entreprises rurales en France (Galliano et Roux, 2008), la relation n'est pas significative en Bretagne et Midi-Pyrénées (Galliano et al., 2008). Selon les auteurs, l'absence d'effet significatif s'expliquerait par le fait que l'accès à une connexion rapide ne serait nécessaire que pour certains usages spécifiques (ex : visioconférence).

Des travaux plus récents indiquent toutefois qu'avec la hausse du volume de données échangées et la multiplication d'applications gourmandes en data, la vitesse de connexion devrait constituer un déterminant croissant des usages numériques (Ezell et al., 2009 ; Nicoletti et al., 2020). Ainsi, dans les pays Européens, l'accès à une connexion supérieure à 30 Mbps serait positivement associé à l'adoption d'outils numériques sophistiqués, tels que le cloud computing, les progiciels de gestion intégré, et la gestion de la relation client (Nicoletti et al.,

²⁷ L'intensité d'utilisation d'un outil est généralement mesurée par la part des employés utilisant cet outil (ex : ordinateur, emails) ou par la part du chiffre d'affaire de l'entreprise obtenu grâce à cet outil (ex : commerce en ligne).

2020). Au Luxembourg, l'accès au très haut débit et à la fibre favoriseraient également les pratiques en ligne des entreprises et le e-commerce mais n'auraient pas d'effet sur l'usage de l'e-administration qui, pour garantir une inclusion maximale des agents économiques, ne requiert pas l'utilisation d'applications sophistiquées (Ben-Aoun Peltier et Ben Youssef, Non publié).

Si ces travaux soulignent l'importance du très haut débit, les analyses portant spécifiquement sur les espaces ruraux, dont la capacité absorptive demeure faible, indiquent cependant que le seul déploiement de technologies ultra-rapides n'a que peu, voire pas d'effet sur les entreprises (Duvivier et al., En cours ; Townsend et al., 2013). La mise en œuvre de politiques de médiation numérique, visant à développer les compétences, la confiance et la compréhension des bénéfices associés à ces technologies, semble fortement conditionner l'impact du déploiement dans ces espaces.

1.3. Un effet conditionnel des infrastructures THD ?

Le déploiement d'infrastructures produit vraisemblablement des effets variables selon les entreprises et le contexte spatial dans lequel il est mis en œuvre. L'estimation de l'effet moyen du déploiement risque ainsi de masquer des dynamiques très hétérogènes selon les facteurs internes et externes aux entreprises. Comprendre dans quels contextes et pour quels types d'entreprises le déploiement numérique générera le plus (ou le moins) d'impact est primordial pour identifier les obstacles limitant les retombées du déploiement, et donc, les mesures à mettre en œuvre pour y remédier.

Facteurs internes à l'entreprise

L'accès à internet devrait favoriser davantage l'usage numérique des entreprises disposant d'un **niveau de capital humain** élevé, puisqu'elles en perçoivent plus clairement les bénéfices, en tirent davantage partie grâce à une utilisation plus efficiente, et n'ont pas à réaliser de dépenses élevées (formation, embauches) pour s'en emparer. L'effet du capital humain est loin d'être négligeable puisque, par exemple, une PME sur trois déclare que le manque de connaissances suffisantes limite son utilisation du cloud computing (Eurostat, 2016)²⁸. Plus que

²⁸ Eurostat (2016), "Cloud computing - statistics on the use by enterprises", Statistics Explained (http://ec.europa.eu/eurostat/statistics-explained/index.php/Cloud_computing_-_statistics_on_the_use_by_enterprises).

le niveau de formation général de la main d'œuvre, ce sont principalement ses connaissances dans le domaine des TIC qui déterminent l'usage numérique d'une entreprise (Giotopoulos 2017 ; Nicoletti et al., 2020). Pour les petites entreprises, qui n'ont généralement pas les moyens d'embaucher des spécialistes du domaine, le recours ponctuel à des consultants externes constitue un facteur clé des usages numériques (Thong, 2001). De plus, les entreprises plus innovantes, qui ont une capacité absorptive supérieure, devraient également davantage utiliser les outils numériques (Giotopoulos 2017 ; Battisti et Stoneman, 2005).

Un meilleur accès à internet agira également différemment sur l'adoption d'outils numériques selon la **taille** des entreprises, bien que l'effet soit controversé. Pour certains, les entreprises plus grandes disposent de ressources (financières, humaines) supérieures et font face à des coûts de coordination et de surveillance plus élevés, ce qui les conduit à adopter davantage les outils numériques, et notamment les plus sophistiqués (Giotopoulos 2017; Nicoletti et al., 2020). Pour d'autres, les petites entreprises, plus flexibles, ont davantage de facilités à s'adapter rapidement aux évolutions technologiques et sont davantage encouragées à adopter certains outils (ex : cloud computing) afin de réduire leurs investissements (Alshamaila et al., 2013).

Au-delà de la taille des entreprises, c'est aussi leur **structure organisationnelle** qui importe. Les entreprises multi-établissements ou appartenant à un groupe, qui font face à des coûts de coordination supérieurs, mobilisent davantage les outils numériques (Bayo-Moriones et Lera-Lopez, 2007 ; Forman et al., 2005b ; Galliano et Roux, 2006). Cela est particulièrement vrai pour les usages favorisant la coordination (intranet, extranet, outils de partage de travail et de ressources, etc.).

Par ailleurs, l'accès à internet pourrait engendrer des effets variables selon « **l'âge** » des entreprises puisque leurs besoins évoluent au cours du temps. Par exemple, les entreprises récentes, qui cherchent principalement à accroître leurs ventes, devraient adopter en priorité des outils de gestion de la relation client, tandis que les entreprises plus développées mobiliseront davantage des outils visant à optimiser le fonctionnement des processus de l'entreprise (ex : progiciel de gestion intégré) (Nicoletti et al., 2020).

Les effets du déploiement sont aussi extrêmement variables en fonction du **secteur industriel** puisque les usages numériques, notamment sophistiqués, sont indispensables dans certains secteurs (ex : tertiaire supérieur) et bien moins dans d'autres (ex : services à la personne, action sociale). Pour Forman et al. (2005a ; 2005b), le secteur industriel serait d'ailleurs un des déterminants principaux des écarts d'usages numériques entre les entreprises

urbaines et rurales. Ainsi, selon les auteurs, le moindre usage des outils numériques dans les espaces ruraux s'expliquerait intégralement par les différences de composition industrielle entre zones urbaines et rurales.

Enfin, un meilleur accès à internet devrait accroître davantage les usages numériques des **entreprises exportatrices**, dont les besoins de communication sont supérieurs et qui, plus exposées à la concurrence internationale, se doivent d'adopter ces outils pour maintenir un niveau de productivité élevé (Alshamaila et al., 2013 ; Galliano et Roux, 2008 ; Haller et Siedschlag, 2011).

Facteurs externes à l'entreprise

Le déploiement d'infrastructures internet devrait également impacter différemment l'usage numérique des entreprises selon l'environnement économique dans lequel elles opèrent. Dans les espaces où la **concurrence** est élevée, le déploiement d'internet devrait fortement inciter les entreprises à mobiliser des outils numériques pour accroître leur performance et chances de survie. Les résultats empiriques sont cependant contrastés, certains concluant à un effet positif de la concurrence sur l'usage de technologies (Nicoletti et al., 2020), d'autres à un effet limité voire non significatif (Alshamaila et al., 2013 ; Galliano et Roux, 2008).

Par ailleurs, un meilleur accès à internet favorise d'autant plus l'adoption d'outils numériques par une entreprise que le **nombre d'utilisateurs dans son environnement** (secteur industriel et/ou région) est élevé (Forman et al., 2005b ; Galliano et Roux, 2008 ; Haller et Siedschlag, 2011). En effet, plus le nombre d'entreprises utilisant ces outils est élevé, plus les externalités informationnelles et de réseau sont fortes, et plus les entreprises sont incitées à développer leurs usages numériques. D'une part, la diffusion d'informations permet de renseigner les entreprises non-utilisatrices sur l'existence, les modes d'utilisation et les bénéfices des outils numériques, réduisant ainsi les incertitudes liées à leur adoption, qui constituent un frein majeur aux usages numériques (Alshamaila et al., 2013). D'autre part, les bénéfices associés à l'usage d'outils numériques augmentent avec le nombre d'utilisateurs au sein d'un réseau (ex : une entreprise a peu d'intérêt à avoir une messagerie électronique ou à proposer de la vente en ligne si aucun de ses partenaires n'utilise les outils numériques). De plus, pour rester compétitive, une entreprise sera obligée d'utiliser les mêmes outils que ses concurrents.

Le déploiement d'internet risque enfin d'impacter différemment les usages numériques des entreprises selon le **degré de ruralité des territoires**. En plus des facteurs déjà évoqués (problème d'accès à une main d'œuvre qualifiée ou à des consultants externes, externalités informationnelles et de réseau inférieures, etc.), l'éloignement des fournisseurs de matériels et services informatiques risque de réduire encore davantage la capacité des entreprises rurales à utiliser des outils numériques. Cependant, si la capacité d'adoption est moindre dans le rural, les incitations à utiliser ces outils sont probablement supérieures (Forman et al., 2005b). En effet, l'éloignement des partenaires et des clients rend leur usage plus important encore pour les entreprises rurales, puisqu'elles font face à des coûts de recherche de l'information et de coordination supérieurs.

Les résultats empiriques existants sont cependant assez variables, probablement parce que les travaux contrôlent dans leurs modèles différents facteurs associés à la ruralité (ex : composition industrielle, niveau qualification, interactions entre entreprises, etc.). Pour la France, Galliano et Roux (2008) concluent que le fait d'être situé dans une zone rurale n'impacte pas la probabilité d'adopter des outils basiques (internet, email) mais réduit l'intensité de leur utilisation. Forman et al. (2005b) indiquent quant à eux que, une fois qu'on contrôle la plupart des caractéristiques des espaces urbains et ruraux, les entreprises rurales ont une probabilité supérieure d'utiliser des outils numériques basiques, et en particulier ceux répondant à un besoin de coordination entre établissements. En revanche, les entreprises rurales utilisent significativement moins les outils numériques complexes. Enfin, les travaux comparant les espaces ruraux sous influence des pôles et les espaces ruraux isolés observent peu, voire pas de différences entre les usages numériques des entreprises (Galliano et Roux, 2006 ; 2008 ; Galliano et al., 2008).

Conclusion : quels enseignements pour notre évaluation ?

Les résultats de la littérature laissent penser que l'effet moyen du plan FTTHD sur (certains) usages numériques des entreprises rurales risque d'être limité en raison, d'une part, de leurs caractéristiques internes et, d'autre part, du contexte économique. La concentration d'entreprises de petite taille, à faible niveau en capital humain, et appartenant souvent à des secteurs économiques pour lesquels les bénéfices des outils numériques sont moindres, risque de limiter le développement des usages, même dans le cas d'un bon accès à internet. L'environnement économique des espaces peu denses, caractérisé par un éloignement des

fournisseurs de services informatiques et par la faiblesse des externalités informationnelles et de réseau, risque de freiner encore davantage le développement des usages numériques.

Au-delà de l'estimation de l'effet moyen du Plan, il convient donc d'examiner plus précisément dans quelle mesure les retombées varient selon les caractéristiques internes et externes des entreprises. Dans le cas où le Plan aurait des effets moyens limités, l'estimation des effets conditionnels permettra d'examiner si cela provient simplement de l'absence d'utilité des outils numériques pour certaines entreprises (ex : certains secteurs économiques) ou si cela résulte de l'existence d'obstacles (manque de compétences, déficit informationnel, ressources financières limitées, etc.) freinant l'usage d'outils numériques par des entreprises qui en bénéficieraient.

2. Données

2.1. Usages numériques des entreprises

Les informations sur les usages numériques des entreprises sont issues de l'Enquête sur les technologies de l'information et de la communication dans les entreprises (enquête « TIC-Entreprises ») de l'Insee. Cette enquête, réalisée depuis 2007, renseigne sur l'usage des TIC dans les entreprises d'au moins 10 personnes occupées (hors fonction publique et hors champs agricole). A notre connaissance, il s'agit de la seule enquête réalisée de manière annuelle, à grande échelle et sur la France entière²⁹.

L'enquête fournit, d'une part, des informations sur les usages numériques des entreprises et, d'autre part, des données de cadrage (numéro Siren, effectif, code APE, chiffre d'affaire, adresse). Le numéro Siren permet ainsi d'apparier l'Enquête TIC-Entreprises avec d'autres sources de données afin de disposer d'informations supplémentaires sur les entreprises et la couverture internet disponible à leur lieu d'implantation.

²⁹ Cette enquête ne renseigne pas sur les TPE. Des informations sur les usages des TPE sont fournies dans l'Enquête sur l'usage de l'informatique et des technologies de la communication dans les entreprises de moins de 10 personnes occupées (enquête « TIC-TPE »). Toutefois, cette dernière n'est pas mobilisée ici pour deux raisons. Premièrement, l'enquête TIC-TPE a été seulement réalisée en 2012 et 2016, et donc, permet d'évaluer l'effet des déploiements ayant eu lieu jusqu'en 2015. Or, quasiment aucune commune de la ZIPU n'avait bénéficié alors d'un déploiement de la fibre. Deuxièmement, l'enquête « TIC-TPE » est moins riche que l'enquête TIC-Entreprises, puisqu'elle n'inclut pas d'information sur certains usages numériques (ex : pas d'information sur le recours à l'e-administration, le cloud computing, les PGI ou encore les outils collaboratifs).

Etant données les informations fournies dans l'enquête, nous considérerons les 10 pratiques numériques suivantes³⁰ :

- Utilisation d'ordinateur (adoption et intensité d'utilisation),
- Utilisation d'internet (adoption et intensité d'utilisation),
- Présence d'un site web,
- Recours à l'e-administration,
- Utilisation des médias sociaux,
- Utilisation du cloud computing,
- Utilisation d'un progiciel de gestion intégrée (PGI)³¹,
- Recours à une gestion de la relation client (GRC)³²,
- Utilisation d'outils collaboratifs³³,
- Recours au commerce électronique (achats et ventes) : adoption et intensité d'utilisation.

Le tableau 4.1 définit les variables d'usages numériques considérées dans notre analyse. Comme le tableau l'indique, si pour certains indicateurs nous disposons d'information pour chaque année, pour d'autres les informations sont plus irrégulières. Cela sera pris en compte dans la mise en œuvre de l'appariement.

A notre connaissance, peu de travaux considèrent une gamme aussi large d'usages numériques. Nous considérons en effet ici à la fois des pratiques organisationnelles (ex : ordinateur) et sophistiquées (ex : cloud computing), impliquant des outils utiles pour les relations externes (ex : commerce électronique) et internes (ex : PGI) de l'entreprise, et répondant pour certains à un besoin informationnel (ex : internet) et pour d'autres à un besoin de coordination (ex : outils collaboratifs).

³⁰ Nous avons inclus tous les usages numériques pour lesquels nous disposons au moins de deux années d'enquêtes. Quelques rares usages (ex : utilisation d'un intranet ; utilisation d'applications d'appel vocal via internet) ne sont donc pas retenus car nous ne disposons que d'une année d'observation.

³¹ « Progiciel permettant de gérer les processus d'une entreprise et de partager les informations entre les métiers via une base de données unique (SAP, PeopleSoft, Oracle, etc.). Il intègre les fonctions suivantes : planning, achats, ventes, marketing, relation client, finances et ressources humaines, etc. » (Enquête TIC-Entreprises, 2018).

³² « Progiciel permettant de gérer l'ensemble des relations avec les clients dans un même processus en regroupant la gestion des campagnes marketing, l'informatisation des forces de vente, le suivi de la relation client au quotidien, etc. » (Enquête TIC-Entreprises, 2018).

³³ Inclut l'utilisation d'outils de travail collaboratifs (ex : vidéoconférence, groupware) et/ou d'outils de conception collaborative (ex : conception assistée par ordinateur, maquette numérique partagée, gestion de cycle de vie du produit).

Tableau 4.1 : Indicateurs d'usages numériques

Variable	Définition	Années disponibles
Ordinateur : adoption	Muette égale à 1 si l'entreprise utilise au moins un ordinateur, 0 sinon.	2014 à 2018
Ordinateur : intensité	Part des salariés de l'entreprise utilisant un ordinateur.	2014, 2015 et 2017
Internet : adoption	Muette égale à 1 si l'entreprise dispose d'un accès à internet, 0 sinon.	2014 à 2019
Internet : intensité	Part des salariés de l'entreprise utilisant un appareil avec accès à internet.	2014 à 2019
Site web	Muette égale à 1 si l'entreprise dispose d'un site web, 0 sinon.	2014 à 2019
E-administration	Muette égale à 1 si l'entreprise échange avec les administrations publiques, 0 sinon. Ces échanges incluent l'utilisation d'internet pour obtenir de l'information ou des formulaires sur les sites web des administrations, déclarer électroniquement la TVA et/ou les cotisations sociales, faire une déclaration unique d'embauche par voie électronique, effectuer une déclaration électronique de formalités douanières à l'import et à l'export, accéder à une information sur des appels d'offre publics en ligne ou répondre à des appels d'offre publics en ligne.	2015 et 2019
Médias sociaux	Muette égale à 1 si l'entreprise dispose d'un profil utilisateur sur un réseau social (ex : Facebook, LinkedIn, Xing, Viadeo, Yammer, Google+, etc.), 0 sinon.	2015, 2016, 2017 et 2019
Cloud computing	Muette égale à 1 si l'entreprise achète des services de cloud computing, 0 sinon.	2014, 2016 et 2018
PGI	Muette égale à 1 si l'entreprise utilise un progiciel de gestion intégrée, 0 sinon.	2014, 2015, 2017 et 2019
GRC	Muette égale à 1 si l'entreprise utilise une application pour la gestion de la relation client, 0 sinon.	2014, 2015, 2017 et 2019
Outils collaboratifs	Muette égale à 1 si l'entreprise utilise des outils informatiques collaboratifs (vidéoconférence, groupware, etc.), 0 sinon.	2015 et 2017

Ventes web-EDI : adoption	Muette égale à 1 si l'entreprise a reçu des commandes passées via un site/une application web ou des messages de type EDI, 0 sinon.	2014 à 2019
Ventes web-EDI : intensité	Part du chiffre d'affaire de l'entreprise issue des commandes passées via un site/une application web ou des messages de type EDI.	2014 à 2019
Achats web-EDI	Muette égale à 1 si l'entreprise a passé des commandes via un site/une application web ou des messages de type EDI, 0 sinon.	2014, 2015, 2016 et 2017

2.2. Accès des entreprises au THD

Si l'enquête TIC-Entreprises ne renseigne pas sur la couverture internet disponible au lieu d'implantation de l'entreprise, il est toutefois possible d'obtenir cette information en appariant l'enquête avec les données de couverture internet de l'Agence du Numérique (cf. Encadré 1 chapitre 2) puisque nous connaissons le Siren (et donc le lieu d'implantation) des entreprises.

Pour calculer un indicateur de couverture internet pertinent pour chaque entreprise, il convient toutefois de tenir compte du fait que l'unité enquêtée est *l'entreprise* et non l'établissement. L'enquête renseigne donc sur les usages numériques de *l'ensemble* des établissements constituant l'entreprise (et pas seulement ceux du siège social ou de l'établissement enquêté). Prendre en compte uniquement la couverture internet existante dans la commune où est situé le siège social de l'entreprise risquerait alors de conduire à des erreurs de mesure importantes dans le cas des entreprises multi-établissements.

De ce fait, pour chaque entreprise, nous avons calculé un indicateur de couverture en prenant en compte l'accès à internet de l'ensemble de ses établissements. De plus, pour tenir compte du fait qu'au sein d'une entreprise certains établissements sont plus importants que d'autres, la couverture internet de l'entreprise i à l'année t est calculée comme la moyenne pondérée de la couverture de ses différents établissements :

$$Couverture_{i,t} = \sum_{k=1}^K Couverture_{k,i,t} \cdot Poids_{k,i,t} \quad (1)$$

Où $Couverture_{k,i,t}$ est la part de locaux couverts dans la commune où est implanté l'établissement k de l'entreprise i à l'année t . $Poids_{k,i,t}$ correspond à la part des effectifs de l'établissements k dans le total des effectifs de l'entreprise i ³⁴.

Par ailleurs, nous comparons les données de l'enquête TIC en t avec les données de couverture internet en $t-1$ pour plusieurs raisons. Premièrement, l'enquête TIC-Entreprises est réalisée en début d'année t tandis qu'on exploite les données de couverture au T4 de l'année t . Deuxièmement, une partie des questions de l'enquête TIC portent sur l'exercice précédent (ex : commerce en ligne). Enfin, il existe vraisemblablement un délai entre l'arrivée d'internet et la modification des usages numériques des entreprises. Comme la dernière enquête TIC-Entreprises est disponible pour l'année 2019³⁵ et que les données de couverture internet sont disponibles à partir de 2013, nous comparons donc les enquêtes TIC-Entreprises réalisées entre 2014 et 2019 avec les données de couverture de 2013 à 2018.

L'existence d'entreprises multi-établissements complique également la définition de l'échantillon. En effet, celles-ci ont souvent des établissements situés à la fois en zone d'initiative privée (ci-après, ZIPRI) et en ZIPU. Au moins trois options peuvent être envisagées pour sélectionner les entreprises opérant en ZIPU. Premièrement, on peut considérer qu'une entreprise appartient à la zone d'intervention publique dès lors qu'au moins un de ses établissements est situé dans une commune de la ZIPU. Toutefois, cette option risquerait de nous conduire à inclure de nombreuses entreprises situées majoritairement en zone d'initiative privée (ex : cas d'une entreprise dont neuf établissements seraient situés en ZIPRI et un seul en ZIPU). Une deuxième option serait d'inclure dans notre échantillon toutes les entreprises dont au moins $k\%$ des effectifs opèrent dans des établissements situés dans des communes en ZIPU. Toutefois, la définition du seuil k pose des difficultés (quel seuil retenir ?) et cette option n'élimine pas le risque d'inclure dans l'analyse des entreprises opérant en majorité dans la zone d'intervention privée. De ce fait, on choisit ici de considérer qu'une entreprise appartient à la ZIPU dès lors que son siège social est situé dans une commune en ZIPU. Cette option a l'avantage d'être simple et « restrictive », dans le sens où elle permet d'exclure les entreprises qui opéreraient majoritairement dans des communes de la ZIPRI.

³⁴ Les données d'effectifs sont issues de la base REE-Stock d'établissements.

³⁵ Vérification effectuée au 9 août 2021.

Le tableau 4.2 indique le nombre d'entreprises enquêtées, selon leur accès à différentes classes de débit³⁶. Pour rappel, nous ne considérons ici que les entreprises localisées en France métropolitaine et dont le siège social est situé dans les communes en ZIPU.

Comme on souhaite évaluer idéalement l'effet de deux « traitements » (THD et fibre), on distingue deux groupes de traités (GT^{THD} , GT^{Fibre}). Comme nous l'avons décrit au chapitre 3, chacun d'entre eux peut être comparé à différents groupes de contrôle. Premièrement, les entreprises disposant d'un accès à du THD (≥ 30 Mbps) peuvent être comparées à trois groupes de contrôle différents : celles disposant d'un accès inférieur à 30 Mbps (GC_1^{THD}), celles disposant d'un accès inférieur à 8 Mbps (GC_2^{THD}) ou encore celles disposant d'un accès compris entre 8 et 30 Mbps (GC_3^{THD})³⁷. Deuxièmement, les entreprises disposant d'un accès à la fibre (≥ 100 Mbps) peuvent être comparées à trois groupes de contrôles différents : celles disposant d'un accès inférieur à 100 Mbps (GC_1^{Fibre}), celles disposant d'un accès inférieur à 30 Mbps (GC_2^{Fibre}) ou encore celles disposant d'un accès compris entre 30 et 100 Mbps (GC_3^{Fibre}).

D'après le tableau, la majorité des entreprises enquêtées dispose déjà d'un accès au THD dès 2014. Cela tient au fait qu'un certain nombre de communes étaient déjà reliées alors au THD et que ces communes couvertes les premières sont en général plus grandes, et donc, regroupent davantage d'entreprises. D'un point de vue méthodologique, le nombre bien supérieur d'entreprises traitées par rapport au nombre d'entreprises de contrôle rend extrêmement complexe l'évaluation de l'effet du THD. En effet, les entreprises de contrôle sont trop peu nombreuses par rapport à celles traitées pour que la procédure d'appariement permette d'identifier des entreprises traitées et non traitées comparables.

³⁶ Tous les effectifs reportés sont pondérés.

³⁷ Pour rappel, chacun de ces groupes de contrôle a ses propres avantages et inconvénients. Le groupe GC_1^{THD} permet de disposer d'un grand nombre d'unités de comparaison. Toutefois, comparer les entreprises disposant du THD à des entreprises ayant pour certaines accès à du bon HD et pour d'autres pas d'accès à internet, peut rendre l'interprétation du coefficient estimé difficile (capte-t-on un effet THD ? Un effet HD ?). Pour résoudre cette difficulté, on peut alors comparer les traités « THD ou plus » au GC_2^{THD} et GC_3^{THD} . En particulier, le GC_3^{THD} (débit entre 8 et 30 Mbps) permet d'estimer l'effet *additionnel* du THD par rapport à du bon HD.

Tableau 4.2 : Nombre d'entreprises enquêtées selon le débit maximal dans la commune d'implantation (ZIPU uniquement)

	2014	2015	2016	2017	2018	2019	2014-2019
Total	64 634	62 470	63 175	64 691	62 219	61 930	379 120
Très haut débit							
- Plus de 30 Mbps (GT^{THD})	42 602	48 556	52 682	56 363	55 459	54 549	310 210
- Moins de 30 Mbps (GC_1^{THD})	22 032	13 913	10 493	8 328	6 760	7 382	68 910
- Moins de 8 Mbps (GC_2^{THD})	2 847	2 664	1 867	2 179	1 702	1 822	13 082
- 8-30 Mbps (GC_3^{THD})	19 185	11 249	8 626	6 150	5 058	5 560	55 828
Fibre							
- Plus de 100 Mbps (GT^{Fibre})	4 712	6 237	9 999	12 435	14 958	16 075	64 417
- Moins de 100 Mbps (GC_1^{Fibre})	59 922	56 232	53 176	52 256	47 262	45 855	314 703
- Moins de 30 Mbps (GC_2^{Fibre})	22 032	13 913	10 493	8 328	6 760	7 382	68 910
- 30-100 Mbps (GC_3^{Fibre})	37 890	42 319	42 683	43 927	40 501	38 473	245 794

Sources: Enquêtes TIC-Entreprises, Insee (2014-2019); Données de couverture internet de l'Agence du Numérique (2013-2018).

Champ: entreprises dont le siège social est situé dans une commune de la zone d'initiative publique.

En revanche, pour l'ensemble de la période, le nombre d'entreprises disposant d'un accès à la fibre est bien moindre que celui disposant d'un accès inférieur à 100 Mbps (GC_1^{Fibre}). Ainsi, même en 2019, les premières sont près de trois fois moins nombreuses que les secondes. De même, en 2019, les entreprises disposant d'un accès à la fibre sont plus de deux fois moins nombreuses que celles disposant d'un accès à 30-100 Mbps (GC_2^{Fibre}).

On se focalisera donc dans ce chapitre sur l'évaluation de l'impact de la fibre optique. Dans la plupart des analyses conduites, on considère qu'une entreprise est traitée dès lors qu'elle est située dans une commune disposant d'un accès à la fibre (quel que soit le pourcentage de locaux couverts). De plus, les entreprises traitées seront comparées à celles disposant d'un accès inférieur à 100 Mbps (GC_1^{Fibre}), puisque cela nous permet de conserver le plus grand nombre d'unités de contrôle. Des tests additionnels seront réalisés pour tenir compte de l'intensité de la couverture en fibre (part de locaux couverte).

Test des effets de « maillon faible » et de « maillon fort »

Suite aux échanges avec le comité scientifique, nous avons tenté d'examiner si on observait un effet de « maillon faible » et/ou de « maillon fort » au sein de l'échantillon d'entreprises multi-établissements. Selon l'hypothèse du maillon faible, le fait qu'un seul établissement n'ait pas accès à internet au sein d'une entreprise suffirait à limiter les usages numériques de l'ensemble de l'entreprise. Selon l'hypothèse du maillon fort, il suffirait que l'établissement stratégique (ex : services centraux, R&D) dispose d'un accès au THD pour que l'entreprise dans son ensemble utilise intensément les outils numériques.

Pour tester ces hypothèses, nous avons créé des indicateurs de couverture différents de celui décrit à l'équation (1) qui, pour rappel, mesure la couverture d'une entreprise comme la couverture moyenne de l'ensemble de ses établissements. Pour tester l'hypothèse du maillon faible, on considère que la couverture internet de l'entreprise i à l'année t est égale au niveau de couverture de son établissement le moins bien couvert :

$$Couverture_{i,t} = \min(Couverture_{k,i,t}) \quad (2)$$

Où $Couverture_{k,i,t}$ est la part de locaux couverts dans la commune où est implanté l'établissement k de l'entreprise i à l'année t . Ainsi, seules les entreprises multi-établissements dont tous les établissements ont accès au THD sont considérées comme traitées.

Pour tester l'hypothèse du maillon fort, on considère que la couverture internet de l'entreprise i à l'année t est égale au niveau de couverture de son siège social³⁸ :

$$Couverture_{i,t} = Couverture_{siège,i,t} \quad (3)$$

Où $Couverture_{siège,i,t}$ est la part de locaux couverts dans la commune où est implanté le siège social de l'entreprise i à l'année t .

Malheureusement, la création de ces indicateurs conduit à un nombre d'observations traitées trop faible pour réaliser des estimations. D'une part, ce test ne concerne que les entreprises multi-établissements, ce qui réduit déjà de manière importante notre échantillon (plus de 70% de notre échantillon est composé d'entreprises mono-établissements). D'autre part, de nombreuses entreprises considérées comme traitées selon l'indicateur calculé à l'équation (1) ont une couverture nulle selon l'équation (2) ou (3). Par exemple, seules 4839 entreprises multi-établissements ont un taux de couverture non nulle selon l'équation (2)³⁹. Les hypothèses du maillon faible et du maillon fort n'ont donc pas pu être examinées dans le cadre de ce rapport focalisé sur les ZIPU.

2.3. Caractéristiques internes et externes des entreprises

Comme évoqué ci-dessus, l'enquête TIC-Entreprises fournit des données de cadrage sur les entreprises enquêtées, et notamment leur numéro Siren. Ce dernier permet d'apparier l'Enquête TIC-Entreprises avec d'autres sources de données afin de disposer d'informations supplémentaires sur les caractéristiques internes et externes des entreprises. Ces caractéristiques seront utilisées dans la suite de l'analyse comme variables de contrôle et/ou afin d'évaluer si l'effet du THD est hétérogène en fonction du type d'entreprises. Ces variables, principalement issues du répertoire Sirene et du recensement de la population, sont décrites dans la Section 5 « Méthodologie ».

³⁸ Nous reconnaissons que le siège social d'une entreprise ne recoupe toutefois pas nécessairement ses activités stratégiques.

³⁹ Le problème du faible nombre d'observations est particulièrement important pour les indicateurs d'usages peu répandus et pour lesquels nous n'avons que peu d'années d'enquête (ex : cloud computing, outils collaboratifs). Ainsi, lorsqu'on calcule la couverture selon l'équation (2), seules 324 entreprises traitées déclarent recourir au cloud computing et 316 aux outils collaboratifs. Notons, de plus, qu'il s'agit ici de données pondérées, qui reposent donc sur un nombre très réduit d'unités effectivement enquêtées.

3. Statistiques descriptives : accès au THD et usages numériques en France

Cette section présente des statistiques descriptives sur les usages numériques des entreprises, en comparant celles situées dans une commune disposant d'un accès à la fibre à celles situées dans une commune disposant d'un débit moindre (<100Mbps).

L'objectif de cette section est double. D'une part, il s'agit d'étudier si on observe une corrélation entre l'accès à la fibre et l'adoption/l'intensité de ces usages. D'autre part, il s'agit de vérifier que tous les usages considérés peuvent bien être étudiés dans notre analyse statistique (ex : nombre d'observations suffisant ; hétérogénéité suffisante entre les entreprises, i.e. les usages ne doivent être ni unanimement partagés ni trop exceptionnels).

Les tableaux 4.3 à 4.16 synthétisent les usages numériques des entreprises selon le niveau de débit maximal dans leur commune d'implantation.

Utilisation d'ordinateurs et d'internet : des différences en termes d'intensité et non d'adoption

Concernant l'utilisation d'ordinateurs et d'internet (tableaux 4.3 à 4.6), les écarts entre entreprises se situent surtout en termes d'intensité d'usage et non d'adoption. Ces technologies relativement anciennes sont en effet largement répandues puisque la quasi-totalité des entreprises disposent au moins d'un ordinateur et d'un accès à internet dès 2014 (tableaux 4.3 et 4.5).

En revanche, la part de salariés qui dispose d'un ordinateur ou qui utilise internet est significativement plus élevée (10 à 15 points de pourcentage en plus) dans les entreprises situées dans une commune disposant d'un accès à la fibre (tableaux 4.4 et 4.6).

Tableau 4.3 : Utilisation d'au moins un ordinateur dans l'entreprise

	2014	2015	2016	2017	2018	2019	2014-2019
Nb. entreprises							
- Moins de 100 Mbps	59 827	56 036	53 036	52 212	47 232	45 855	314 199
- Accès à la Fibre (>100 Mbps)	4 712	6 237	9 999	12 403	14 917	16 075	64 343
% entreprises							
- Moins de 100 Mbps	99,8%	99,7%	99,7%	99,9%	99,9%	100,0%	99,8%
- Accès à la Fibre (>100 Mbps)	100,0%	100,0%	100,0%	99,7%	99,7%	100,0%	99,9%

Sources: Enquêtes TIC-Entreprises, Insee (2014-2019); Données de couverture internet de l'Agence du Numérique (2013-2018).

Champ: entreprises dont le siège social est situé dans une commune de la zone d'initiative publique.

Lecture: en 2014, 4 712 entreprises situées dans les communes disposant d'un accès à la fibre sont équipées d'au moins un ordinateur.

Tableau 4.4 : Part des salariés utilisant un ordinateur

	2014	2015	2016	2017	2018	2019	2014-2019
% entreprises							
- Moins de 100 Mbps	40,5	40,7	-	42,3	-	-	41,1
- Accès à la Fibre (>100 Mbps)	55,0	53,5	-	52,4	-	-	53,2
Test de différence de moyenne [†]							
Différence de moyenne	14.51***	12.80***	-	10.07***	-	-	12.08***
$P> t $	0,000	0,000	-	0,000	-	-	0,000

Sources: Enquêtes TIC-Entreprises, Insee (2014-2019); Données de couverture internet de l'Agence du Numérique (2013-2018).

Champ: entreprises dont le siège social est situé dans une commune de la zone d'initiative publique.

Lecture : en 2014, 55% des salariés des entreprises situées dans les communes disposant d'un accès à la fibre utilisent un ordinateur.

[†] Un test de Student a été réalisé. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Tableau 4.5. Utilisation d'internet dans l'entreprise

	2014	2015	2016	2017	2018	2019	2014-2019
Nb. entreprises							
- Moins de 100 Mbps	59 714	55 594	52 837	52 094	46 957	45 829	313 025
- Accès à la Fibre (>100 Mbps)	4 712	6 237	9 963	12 329	14 877	16 038	64 157
% entreprises							
- Moins de 100 Mbps	99,7%	98,9%	99,4%	99,7%	99,4%	99,9%	99,7%
- Accès à la Fibre (>100 Mbps)	100,0%	100,0%	99,6%	99,1%	99,5%	99,8%	99,7%

Sources: Enquêtes TIC-Entreprises, Insee (2014-2019); Données de couverture internet de l'Agence du Numérique (2013-2018).

Champ: entreprises dont le siège social est situé dans une commune de la zone d'initiative publique.

Lecture: en 2014, 4 712 entreprises situées dans les communes disposant d'un accès à la fibre utilisent internet. En 2014, 99,7% des entreprises situées dans les communes ne disposant pas d'un accès à la fibre utilisent internet.

Tableau 4.6 : Part des salariés utilisant internet

	2014	2015	2016	2017	2018	2019	2014-2019
% entreprises							
- Moins de 100 Mbps	35,1	34,9	36,0	37,5	42,2	42,7	37,8
- Accès à la Fibre (>100 Mbps)	49,6	48,2	45,5	47,3	51,7	52,3	49,5
Test de différence de moyenne [†]							
Différence de moyenne	14.49***	13.29***	9.51***	9.76***	9.54***	9.52***	11.75***
$P> t $	0,000	0,000	0,000	0,000	0,000	0,000	0,000

Sources: Enquêtes TIC-Entreprises, Insee (2014-2019); Données de couverture internet de l'Agence du Numérique (2013-2018).

Champ: entreprises dont le siège social est situé dans une commune de la zone d'initiative publique.

Lecture: en 2014, 49,6% des salariés des entreprises situées dans les communes disposant d'un accès à la fibre utilisent internet.

[†] Un test de Student a été réalisé. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Des usages inégalement répandus

Si certains usages sont largement répandus au sein des entreprises, d'autres demeurent plus rares. Ainsi, dès 2014, plus de la moitié des entreprises disposent d'un site web et plus de 90% recourent à l'e-administration. L'usage des réseaux sociaux, de PGI, de GRC ou encore les achats web-EDI sont, quant à eux, relativement moins répandus (environ une entreprise sur trois ou sur deux sur la période). Enfin, le cloud computing, les outils collaboratifs et les ventes web et EDI font partis des usages les moins répandus, avec moins de 25% des entreprises y ayant recourt sur la période. Evidemment, si la nouveauté et la complexité de certains usages expliquent principalement leur faible utilisation (ex : cloud computing), le recourt relativement restreint à d'autres pratiques tient plutôt au fait qu'elles ne concernent, par définition, que certaines entreprises (ex : vente en ligne). Par ailleurs, de manière cohérente, une part croissante d'entreprises recourt à ces différents usages numériques entre 2014 et 2019.

La plupart des usages sont positivement corrélés à l'accès à la fibre

Les statistiques descriptives indiquent qu'il existe une corrélation positive entre l'accès à la fibre des entreprises et la plupart de leurs pratiques en ligne (site web, réseau social, cloud computing, PGI, ERP, outils collaboratifs, achats web-EDI). Seul le recourt à l'e-administration, qui concerne la quasi-totalité des entreprises traitées et non traitées, n'est pas corrélé à l'accès à la fibre⁴⁰. De plus, si la probabilité de déclarer des ventes web et EDI est corrélée à l'accès à la fibre certaines années, leur intensité (part du chiffre d'affaire issu des ventes web et EDI) n'est pas associée quant à elle à l'accès à la fibre. Enfin, pour la plupart des pratiques en ligne, les différences entre traités et non traités s'estompent légèrement au cours du temps, probablement en raison d'une amélioration de la couverture des non traités⁴¹.

Pour conclure, l'analyse descriptive révèle deux enseignements. Premièrement, au niveau méthodologique, une analyse statistique plus poussée ne sera possible que pour certains usages. En particulier, nous excluons de l'analyse à venir l'étude de l'adoption des technologies relativement anciennes (ordinateurs, internet), dont les taux d'adoption avoisinent les 100%. Pour celles-ci, l'étude consistera à examiner non pas les déterminants de leur adoption mais ceux de leur *intensité* d'utilisation. De plus, l'e-administration, dont l'usage est également

⁴⁰ Pour 2019, on observe même une corrélation négative entre l'accès à la fibre et le recourt à l'e-administration. Une explication potentielle est que les entreprises situées dans les communes non traitées, généralement plus reculées, recourent davantage à l'e-administration en raison du nombre plus faible de guichets publics « physiques ».

⁴¹ Il est probable qu'une part croissante de non traités dispose d'un accès au THD entre 2014 et 2018.

largement répandu, est aussi exclue de l'analyse. Au final, l'analyse économétrique portera donc sur les usages numériques suivants :

- Utilisation d'ordinateur (intensité d'utilisation),
- Utilisation d'internet (intensité d'utilisation),
- Présence d'un site web,
- Utilisation des médias sociaux,
- Utilisation du cloud computing,
- Utilisation d'un progiciel de gestion intégrée (PGI),
- Recours à une gestion de la relation client (GRC),
- Utilisation d'outils collaboratifs,
- Recours au commerce électronique (achats et ventes) : adoption et intensité d'utilisation.

Deuxièmement, l'analyse descriptive indique qu'il existe une corrélation positive entre l'accès à la fibre et la plupart des usages numériques des entreprises. L'enjeu de l'analyse économétrique est à présent de vérifier si cette corrélation provient d'un effet spécifique de la fibre (en tant que facteur facilitant les usages numériques) ou de variables omises, tel que la surreprésentation de secteurs utilisateurs d'outils numériques dans les communes dotées de fibre.

Tableau 4.7 : Présence d'un site web

	2014	2015	2016	2017	2018	2019	2014-2019
Nb. entreprises							
- Moins de 100 Mbps	34 822	35 556	33 847	32 919	30 692	30 899	198 733
- Accès à la Fibre (>100 Mbps)	3 268	4 624	7 284	9 032	10 950	12 344	47 503
% entreprises							
- Moins de 100 Mbps	58%	63%	64%	63%	65%	67%	63%
- Accès à la Fibre (>100 Mbps)	69%	74%	73%	73%	73%	77%	74%
Test de différence de moyenne [†]							
Différence de moyenne	0.112***	0.109***	0.092***	0.096***	0.083***	0.094***	0.106***
$P> t $	0,003	0,001	0,001	0,000	0,001	0,000	0,000

Sources: Enquêtes TIC-Entreprises, Insee (2014-2019); Données de couverture internet de l'Agence du Numérique (2013-2018).

Champ: entreprises dont le siège social est situé dans une commune de la zone d'initiative publique.

Lecture: en 2014, 58% des entreprises situées dans les communes disposant d'un accès à la fibre ont un site web.

[†] Un test du Chi-deux a été réalisé. * p<0.1, ** p<0.05, *** p<0.01.

Tableau 4.8 : Utilisation de l'e-administration

	2014	2015	2016	2017	2018	2019	2014-2019
Nb. entreprises							
- Moins de 100 Mbps	-	53 153	-	-	-	44 077	97 230
- Accès à la Fibre (>100 Mbps)	-	6 025	-	-	-	15 015	21 040
% entreprises							
- Moins de 100 Mbps	-	95%	-	-	-	96%	95%
- Accès à la Fibre (>100 Mbps)	-	97%	-	-	-	93%	94%
Test de différence de moyenne [†]							
Différence de moyenne		0,021				-0.027**	-0,009
<i>P> t </i>	-	0,122	-	-	-	0,042	0,356

Sources: Enquêtes TIC-Entreprises, Insee (2014-2019); Données de couverture internet de l'Agence du Numérique (2013-2018).

Champ: entreprises dont le siège social est situé dans une commune de la zone d'initiative publique.

Lecture: en 2015, 97% des entreprises situées dans les communes disposant d'un accès à la fibre recourent à l'e-administration.

[†] Un test du Chi-deux a été réalisé. * p<0.1, ** p<0.05, *** p<0.01.

Tableau 4.9 : Utilisation des réseaux sociaux

	2014	2015	2016	2017	2018	2019	2014-2019
Nb. entreprises							
- Moins de 100 Mbps	-	12 748	13 696	17 008	-	18 943	62 395
- Accès à la Fibre (>100 Mbps)	-	1 685	3 506	4 917	-	7 899	18 006
% entreprises							
- Moins de 100 Mbps	-	23%	26%	33%	-	41%	30%
- Accès à la Fibre (>100 Mbps)	-	27%	35%	40%	-	49%	40%
Test de différence de moyenne [†]							
Différence de moyenne	-	0,043	0.093***	0.070***	-	0.078**	0.102***
<i>P> t </i>	-	0,120	0,000	0,006	-	0,002	0,000

Sources: Enquêtes TIC-Entreprises, Insee (2014-2019); Données de couverture internet de l'Agence du Numérique (2013-2018).

Champ: entreprises dont le siège social est situé dans une commune de la zone d'initiative publique.

Lecture: en 2015, 27% des entreprises situées dans les communes disposant d'un accès à la fibre ont un compte sur les réseaux sociaux.

[†] Un test du Chi-deux a été réalisé. * p<0.1, ** p<0.05, *** p<0.01.

Tableau 4.10 : Utilisation du cloud computing

	2014	2015	2016	2017	2018	2019	2014-2019
Nb. entreprises							
- Moins de 100 Mbps	4 409	-	5 913	-	5 699	-	16 021
- Accès à la Fibre (>100 Mbps)	618	-	1 837	-	2 572	-	5 027
% entreprises							
- Moins de 100 Mbps	7%	-	11%	-	12%	-	10%
- Accès à la Fibre (>100 Mbps)	13%	-	18%	-	17%	-	17%
Test de différence de moyenne [†]							
Différence de moyenne	0.058***	-	0.073***	-	0.051***	-	0.069***
$P> t $	0,008	-	0,000	-	0,003	-	0,000

Sources: Enquêtes TIC-Entreprises, Insee (2014-2019); Données de couverture internet de l'Agence du Numérique (2013-2018).

Champ: entreprises dont le siège social est situé dans une commune de la zone d'initiative publique.

Lecture: en 2014, 13% des entreprises situées dans les communes disposant d'un accès à la fibre recourent au cloud computing.

[†] Un test du Chi-deux a été réalisé. * p<0.1, ** p<0.05, *** p<0.01.

Tableau 4.11 : Utilisation d'un progiciel de gestion intégrée

	2014	2015	2016	2017	2018	2019	2014-2019
Nb. entreprises							
- Moins de 100 Mbps	18 805	19 783	-	17 315	-	20 218	76 121
- Accès à la Fibre (>100 Mbps)	2 333	3 256	-	5 817	-	8 602	20 008
% entreprises							
- Moins de 100 Mbps	31%	35%	-	33%	-	44%	36%
- Accès à la Fibre (>100 Mbps)	50%	52%	-	47%	-	54%	51%
Test de différence de moyenne [†]							
Différence de moyenne	0.181***	0.170***	-	0.136***	-	0.094***	0.152***
$P> t $	0,000	0,000	-	0,000	-	0,000	0,000

Sources: Enquêtes TIC-Entreprises, Insee (2014-2019); Données de couverture internet de l'Agence du Numérique (2013-2018).

Champ: entreprises dont le siège social est situé dans une commune de la zone d'initiative publique.

Lecture: en 2014, 50% des entreprises situées dans les communes disposant d'un accès à la fibre utilisent un PGI.

[†] Un test du Chi-deux a été réalisé. * p<0.1, ** p<0.05, *** p<0.01.

Tableau 4.12 : Gestion de la relation client

	2014	2015	2016	2017	2018	2019	2014-2019
Nb. entreprises							
- Moins de 100 Mbps	11 740	14 644	-	10 494	-	9 372	46 250
- Accès à la Fibre (>100 Mbps)	1 621	2 535	-	3 940	-	4 365	12 461
% entreprises							
- Moins de 100 Mbps	20%	26%	-	20%	-	20%	22%
- Accès à la Fibre (>100 Mbps)	34%	41%	-	32%	-	27%	32%
Test de différence de moyenne [†]							
Différence de moyenne	0.148***	0.146***	-	0.116***	-	0.067***	0.100***
<i>P> t </i>	0,000	0,000	-	0,000	-	0,002	0,000

Sources: Enquêtes TIC-Entreprises, Insee (2014-2019); Données de couverture internet de l'Agence du Numérique (2013-2018).

Champ: entreprises dont le siège social est situé dans une commune de la zone d'initiative publique.

Lecture: en 2014, 34% des entreprises situées dans les communes disposant d'un accès à la fibre recourent à la GRC.

[†] Un test du Chi-deux a été réalisé. * p<0.1, ** p<0.05, *** p<0.01.

Tableau 4.13 : Utilisation d'outils collaboratifs

	2014	2015	2016	2017	2018	2019	2014-2019
Nb. entreprises							
- Moins de 100 Mbps	-	6 638	-	7 006	-	-	13 645
- Accès à la Fibre (>100 Mbps)	-	1 379	-	2 814	-	-	4 193
% entreprises							
- Moins de 100 Mbps	-	12%	-	13%	-	-	13%
- Accès à la Fibre (>100 Mbps)	-	22%	-	23%	-	-	22%
Test de différence de moyenne [†]							
Différence de moyenne	-	0.103***	-	0.092***	-	-	0.099***
<i>P> t </i>	-	0,000	-	0,000	-	-	0,000

Sources: Enquêtes TIC-Entreprises, Insee (2014-2019); Données de couverture internet de l'Agence du Numérique (2013-2018).

Champ: entreprises dont le siège social est situé dans une commune de la zone d'initiative publique.

Lecture: en 2015, 22% des entreprises situées dans les communes disposant d'un accès à la fibre utilisent des outils collaboratifs.

[†] Un test du Chi-deux a été réalisé. * p<0.1, ** p<0.05, *** p<0.01.

Tableau 4.14 : Commerce électronique : ventes web ou EDI (adoption)

	2014	2015	2016	2017	2018	2019	2014-2019
Nb. entreprises							
- Moins de 100 Mbps	7 957	11 474	10 748	10 255	8 555	8 086	57 074
- Accès à la Fibre (>100 Mbps)	846	1 756	2 438	2 774	2 999	3 217	14 030
% entreprises							
- Moins de 100 Mbps	13%	20%	20%	20%	18%	18%	18%
- Accès à la Fibre (>100 Mbps)	18%	28%	24%	22%	20%	20%	22%
Test de différence de moyenne [†]							
Différence de moyenne	0,047*	0,077***	0,042*	0,027	0,019	0,024	0,036***
$P> t $	0,063	0,006	0,064	0,184	0,294	0,189	0,000

Sources: Enquêtes TIC-Entreprises, Insee (2014-2019); Données de couverture internet de l'Agence du Numérique (2013-2018).

Champ: entreprises dont le siège social est situé dans une commune de la zone d'initiative publique.

Lecture: en 2014, 18% des entreprises situées dans les communes disposant d'un accès à la fibre déclarent réaliser des ventes web ou EDI.

[†] Un test du Chi-deux a été réalisé. * p<0.1, ** p<0.05, *** p<0.01.

Tableau 4.15 : Commerce électronique : % du chiffre d'affaire issue des ventes web ou EDI

	2014	2015	2016	2017	2018	2019	2014-2019
(i) Toutes entreprises (réalisant ou non des ventes web et/ou EDI)							
% moyen du chiffre d'affaire							
- Moins de 100 Mbps	3%	4%	5%	4%	5%	5%	4%
- Accès à la Fibre (>100 Mbps)	3%	7%	5%	5%	5%	5%	5%
Test de différence de moyenne							
Différence de moyenne	0,001	0,024**	0,007	0,014**	0,005	0,005	0,011***
$P> t $	0,797	0,02	0,434	0,049	0,451	0,448	0,001
(ii) Entreprises réalisant des ventes web et/ou EDI							
% moyen du chiffre d'affaire							
- Moins de 100 Mbps	20%	21%	23%	20%	25%	26%	22%
- Accès à la Fibre (>100 Mbps)	16%	23%	22%	24%	26%	26%	24%
Test de différence de moyenne [†]							
Différence de moyenne	-0,045	0,027	-0,010	0,040	0,002	-0,004	0,014
$P> t $	0,105	0,384	0,763	0,149	0,941	0,898	0,265

Sources: Enquêtes TIC-Entreprises, Insee (2014-2019); Données de couverture internet de l'Agence du Numérique (2013-2018).

Champ: entreprises dont le siège social est situé dans une commune de la zone d'initiative publique.

Lecture: en 2014, les entreprises qui réalisent des ventes web ou EDI et qui sont situées dans les communes couvertes en fibre ont en moyenne un chiffre d'affaire provenant à 16% des ventes web et/ou EDI.

[†] Un test de Student a été réalisé. * p<0.1, ** p<0.05, *** p<0.01.

Tableau 4.16 : Commerce électronique : achats web ou EDI

	2014	2015	2016	2017	2018	2019	2014-2019
Nb. entreprises							
- Moins de 100 Mbps	13 561	20 805	26 666	24 336	-	-	85 367
- Accès à la Fibre (>100 Mbps)	1 476	3 048	5 632	6 814	-	-	16 970
% entreprises							
- Moins de 100 Mbps	23%	37%	50%	47%	-	-	39%
- Accès à la Fibre (>100 Mbps)	31%	49%	56%	55%	-	-	51%
Test de différence de moyenne [†]							
Différence de moyenne	0.087**	0.119***	0.062**	0.082***	-	-	0.123***
<i>P> t </i>	0,012	0,000	0,027	0,002	-	-	0,000

Sources: Enquêtes TIC-Entreprises, Insee (2014-2019); Données de couverture internet de l'Agence du Numérique (2013-2018).

Champ: entreprises dont le siège social est situé dans une commune de la zone d'initiative publique.

Lecture: en 2014, en 2014, 31% des entreprises situées dans les communes disposant d'un accès à la fibre déclarent réaliser des achats en ligne.

[†] Un test du Chi-deux a été réalisé. * p<0.1, ** p<0.05, *** p<0.01.

4. Méthodologie

4.1. Appariement par le score de propension

L'analyse descriptive indique une corrélation positive entre l'accès des entreprises à la fibre et leurs usages numériques. Plusieurs facteurs omis pourraient toutefois être à l'origine de cette corrélation. Ainsi, la fibre est généralement déployée en priorité dans les communes les plus peuplées et les plus accessibles qui concentrent vraisemblablement certains types d'entreprises davantage utilisatrices d'outils numériques (ex : surreprésentation d'entreprises du tertiaire supérieur). On peut aussi imaginer que les RIP qui déploient la fibre de manière plus précoce mettent également en place des dispositifs d'accompagnement numérique à destination des entreprises (coaching/formation par des experts, prêts bancaires pour l'équipement des entreprises, etc.)⁴².

Pour contrôler ces facteurs, nous recourrons à la méthode de l'appariement par le score de propension. Celle-ci permet de sélectionner au sein de notre échantillon les entreprises bénéficiaires et non bénéficiaires comparables, i.e. partageant des caractéristiques *observables* similaires⁴³.

Pour mettre en œuvre la méthode, on distingue deux groupes d'entreprises. Tout d'abord, les entreprises traitées sont celles situées dans une commune disposant d'un accès à la fibre, quel que soit le taux de locaux couverts. On décide de ne pas imposer de seuil minimal de couverture pour considérer qu'une entreprise est traitée. Si cela peut conduire à considérer comme traitées des entreprises situées dans des communes où une part très faible de locaux dispose effectivement d'un accès à la fibre, nous préférons ne pas imposer de seuil arbitraire (ex : 10% de locaux couverts ? 20% ?) pour définir le traitement. Il faut toutefois garder en tête que ce choix peut conduire à sous-estimer l'effet du traitement. Des tests additionnels seront donc proposés dans la section « Résultats » pour vérifier que le biais d'atténuation n'impacte

⁴² Le risque de « sorting » des entreprises est quant à lui relativement limité. Comme les coûts de relocalisation sont très élevés, il est en effet très peu probable que les entreprises changent de commune dans le but bénéficier d'un accès au THD (DeStefano et al, 2014).

⁴³ A l'instar des modèles de régression classiques (ex : moindres carrés ordinaires), la méthode de l'appariement repose donc sur l'hypothèse de sélection sur les observables (i.e. conditionnellement aux caractéristiques observables, les entreprises traitées et de contrôle ne diffèrent pas de manière systématique en termes de caractéristiques inobservables). Toutefois, l'appariement offre le double avantage d'écarter de l'analyse les entreprises qui sont hors du support commun et de ne pas faire d'hypothèse sur la forme fonctionnelle du modèle et la distribution des termes d'erreurs.

pas les estimations dans une trop grande mesure. Par ailleurs, les entreprises de contrôle sont celles situées dans une commune où aucun local n'a accès à la fibre.

Pour réaliser l'appariement, on sélectionne un ensemble de variables suspectées d'être corrélées avec le déploiement de la fibre et les usages numériques des entreprises. Les variables d'appariement sont composées d'indicateurs au niveau communal et au niveau entreprise ainsi que de muettes départements et années⁴⁴. Concernant les variables au niveau communal, comme les communes les plus grandes ont une probabilité plus élevée d'être couvertes (en raison d'un nombre espéré d'abonnés supérieurs), on contrôle pour la population totale et le nombre total d'établissements. De même, comme les communes plus accessibles devraient bénéficier de la fibre plus précocement (en raison de coûts de déploiement inférieurs), on contrôle pour la topographie des communes et la proximité à la ville moyenne la plus proche. Enfin, on s'attend à ce que les déploiements de fibre, qui s'appuient sur le réseau déjà existant, soient effectués prioritairement dans les communes déjà relativement bien couvertes en THD⁴⁵. De ce fait, on contrôle également pour une muette indiquant si la commune souffre d'un mauvais accès au THD (moins de 25% des locaux couverts en THD). Ces variables sont mesurées en 2013, avant la détermination du calendrier de déploiement au sein des RIP.

On introduit également des variables au niveau entreprise parmi les variables d'appariement. Evidemment, au sien d'un RIP, le calendrier de déploiement est basé sur les caractéristiques des communes et il est extrêmement peu probable que celles des entreprises aient joué un quelconque impact. Toutefois, en ciblant prioritairement certaines communes, le calendrier de déploiement risque de conduire certains types d'entreprises à bénéficier de la fibre de manière plus précoce (ex : si les entreprises les plus grandes ou appartenant à certains secteurs sont surreprésentées dans les communes les plus peuplées et accessibles). Pour en tenir compte, on introduit parmi les variables d'appariement des indicatrices renseignant la taille des entreprises, leur structure organisationnelle (mono ou multi-établissement), leur âge et leur

⁴⁴ Pour rappel, il est très peu fréquent qu'une même entreprise soit enquêtée plusieurs années dans notre échantillon. Pour l'analyse économétrique, nous avons décidé de regrouper les différentes vagues d'enquêtes (2014-2019) entre elles afin d'augmenter la taille des échantillons et donc la précision des estimations.

⁴⁵ Notons toutefois que certains RIP (ex : Auvergne) ciblent prioritairement les communes souffrant des plus gros problèmes de connexion (Duvivier et al., 2018). Cependant, le ciblage prioritaire de ces communes se traduit vraisemblablement, au moins dans un premier temps, par des déploiements de technologies alternatives (montée des débits sur le réseau ADSL, satellite, Wimax, 4G) plutôt que par de la fibre.

secteurs d'activité. Contrairement aux variables communales, qui sont mesurées en 2013, ces variables sont mesurées l'année du traitement⁴⁶.

Les variables d'appariement sont décrites dans le tableau A.4.1 en annexe. Au total, 9 variables d'appariement sont retenues. Dit autrement, nous souhaitons que les entreprises disposant de la Fibre et leurs contrefactuelles aient des valeurs similaires pour ces 9 variables. Se pose alors le problème bien connu de la « malédiction de la dimension », selon lequel il est impossible de trouver, pour chaque entreprise « traitée », une entreprise contrefactuelle qui partagent exactement les mêmes caractéristiques⁴⁷.

Puisqu'un appariement exact n'est pas envisageable, nous recourons alors à un appariement par le score de propension (Rosenbaum et Rubin, 1983). Ainsi, pour identifier les contrefactuelles, nous ne sélectionne pas les entreprises qui partagent exactement les mêmes caractéristiques que celles disposant de la Fibre (appariement exact) mais les entreprises qui ont des valeurs similaires du score de propension⁴⁸.

Pour estimer le score de propension, nous introduisons les 9 variables d'appariement comme variables indépendantes dans un modèle logistique. Les entreprises sont ensuite appariées selon la méthode du plus proche voisin avec remplacement.

Par ailleurs, comme discuté dans la Section 3.1., toutes les variables d'usage ne sont pas renseignées systématiquement dans chaque vague d'enquête. Ainsi, si nous disposons d'informations annuelles sur toute la période pour certains usages (ex : utilisation d'internet, site web), pour d'autres (ex : cloud computing, outils collaboratifs) nous ne disposons d'informations que pour certaines années. Pour nous assurer de la bonne comparabilité des groupes traités et de contrôle, la méthode de l'appariement est donc réalisée plusieurs fois de sorte à tenir compte de la disponibilité des données pour chaque variable d'usage⁴⁹.

⁴⁶ Il n'est pas possible de mesurer les caractéristiques de l'ensemble des entreprises en 2013 puisque certaines d'entre elles sont créées ultérieurement.

⁴⁷ Nous utilisons 9 variables d'appariement dont plusieurs sont continues. Cela conduit à des millions de combinaisons de caractéristiques possibles (même en introduisant des « calipers » pour les variables continues ou en transformant les variables continues en variables catégorielles).

⁴⁸ Le score de propension désigne la probabilité pour chaque entreprise d'être située dans une commune bénéficiant de la fibre à l'année t .

⁴⁹ Par exemple, pour les variables pour lesquelles on dispose d'informations annuelles entre 2014 et 2019 (ex : utilisation d'internet, site web, ventes web), on réalise l'appariement en incluant dans notre échantillon toutes les entreprises enquêtées de 2014 à 2019. En revanche, pour la variable « % de salariés utilisant un ordinateur », qui

4.2. Estimation de l'effet du traitement

Une fois l'appariement réalisé, l'effet moyen du traitement peut être calculé de deux manières différentes. Premièrement, il peut être simplement obtenu en calculant la différence de moyenne des usages numériques entre le groupes de traités et le groupe de contrôle appariés. Deuxièmement, il peut également être estimé en recourant à un modèle de régression, dans lequel les variables d'usages sont successivement régressées sur l'indicatrice de traitement (et, possiblement, des variables de contrôle additionnelles). Cette deuxième option offre au moins trois avantages. D'une part, cela permet de tenir compte de la structure des données mobilisées, et en particulier d'appliquer les pondérations issues des enquêtes « TIC-Entreprises » (cf. Section 5.3.). D'autre part, le modèle de régression permet d'introduire des variables de contrôle et ainsi d'améliorer la performance des estimations, notamment lorsque de légères différences de caractéristiques persistent entre les traités et les contrôles à la suite de la procédure d'appariement (Elder et Jepsen, 2014 ; Lenis et al., 2019). Enfin, l'estimation d'un modèle de régression permet de tester plus formellement si l'effet du traitement varie significativement entre différents groupes d'entreprises (cf. Section 5.4.).

Pour les raisons évoquées ci-dessus, l'impact de la fibre est obtenu ici en estimant un modèle de régression pour les entreprises traitées et de contrôle appariées. Dans le cas des variables d'usages binaires (ex : site web, GRC, cloud computing, etc.), on recourt à un modèle logit tandis que pour les variables d'usages en proportion (ex : % de salariés utilisant un ordinateur, % du chiffre d'affaire issu des ventes web et EDI) on recourt à un modèle tobit.

4.3. Adaptation de la méthode d'appariement aux données d'enquête

Bien que la méthode de l'appariement par le score de propension soit très répandue, peu d'indications ont été fournies sur la façon de la mettre en œuvre lorsque qu'on utilise des données d'enquête. De ce fait, la plupart des travaux appliquant cette méthode à des données d'enquête ne tiennent pas compte de leur structure souvent complexe (pondération, stratification, etc.). Pourtant, ces pratiques, et notamment la non-application des pondérations risque de compromettre la validité externe de l'analyse, puisque les effets ainsi estimés sont

n'est disponible que dans les enquêtes 2014, 2015 et 2017, nous réalisons l'appariement en incluant seulement les entreprises concernées par les vagues d'enquêtes correspondantes.

valables pour l'échantillon mais non représentatifs de la population cible dans son ensemble (DuGoff et al., 2014).

Quelques travaux fournissent toutefois des recommandations sur la façon de mettre en œuvre la méthode de l'appariement dans le cadre de données d'enquête (Austin et al., 2018 ; DuGoff et al., 2014 ; Lenis et al., 2019 ; Ridgeway et al., 2015 ; Zanutto, 2006). En particulier, ces papiers soulèvent plusieurs questions. D'une part, faut-il utiliser les poids de l'enquête lors de l'estimation du score de propension ? D'autre part, faut-il utiliser les poids de l'enquête dans la suite de l'analyse (lors du test d'égalité des co-variables entre traités et contrôles et lors de l'estimation de l'effet du traitement) ?

Concernant la première question, la majorité des travaux indiquent qu'il n'est pas nécessaire de prendre en compte les poids de l'enquête lors de l'estimation du score de propension. DuGoff et al. (2014) indiquent cependant qu'ajouter la variable de pondération de l'enquête parmi les variables d'appariement permet généralement d'en améliorer la qualité. Nous choisissons donc de retenir la méthode de DuGoff et al. lors de l'estimation du score de propension.

Par ailleurs, à notre connaissance seuls Lenis et al. (2019) ont examiné s'il était préférable d'appliquer la pondération au moment du test d'égalité des co-variables entre traités et contrôles. De ce fait, nous suivons leur recommandation selon laquelle il convient d'appliquer les poids de l'enquête lors de ce test.

Enfin, les travaux concluent de manière unanime qu'il est nécessaire de prendre en compte les poids de l'enquête au moment de l'estimation de l'effet moyen du traitement. De plus, dans le cas où l'appariement génère des pondérations, il convient de recalculer un poids final, égal au produit du poids de l'enquête et du poids issu de la méthode de l'appariement (DuGoff et al., 2014 ; Ridgeway et al., 2015).

4.4. Etapes de l'estimation

Pour estimer l'effet de la fibre sur les usages numériques des entreprises, on procède en deux étapes. Tout d'abord, on estime l'effet moyen du programme sur l'ensemble des entreprises traitées. Pour cela, on applique la méthode exposée ci-dessus (appariement par score de propension adapté aux données d'enquête puis estimation de l'effet du traitement sur l'échantillon apparié) sur l'ensemble de l'échantillon.

Deuxièmement, on teste si l'effet moyen du traitement varie selon les caractéristiques internes et externes des entreprises, ainsi que selon les modalités du déploiement de la fibre.

Concernant les caractéristiques internes, nous examinons si l'effet de la fibre varie selon l'âge, la taille, la structure organisationnelle, et le secteur des entreprises.

Concernant les caractéristiques externes, nous analysons si l'effet varie selon le degré de ruralité des territoires et selon la présence de fournisseurs d'équipements et de services informatiques. Pour tenir compte du degré de ruralité des territoires, nous utilisons le zonage en aires urbaines de l'Insee⁵⁰. Contrairement aux classifications plus classiques (ex : grille de densité communale), le zonage en aires urbaines permet de tenir compte à la fois de la taille des communes et de leur intégration avec les espaces urbains voisins, ce qui est particulièrement pertinent ici puisqu'on s'attend à ce que l'effet de la Fibre varie selon la distance aux aires urbaines. Suivant Détang-Dessendre et Piguët (2016), nous regroupons certaines catégories du zonage en aires urbaines afin de classer les communes en quatre types d'espaces⁵¹:

- Les grands pôles urbains (> 10 000 emplois)
- Les zones péri-urbaines (communes dont au moins 40% de leurs résidents travaillent dans les grands pôles urbains voisins)
- Les pôles moyens et petits (1 500 - 10 000 emplois)
- Les communes rurales (pour conserver un nombre suffisant d'observations dans chaque catégorie, nous regroupons les communes rurales sous influence des pôles et les communes rurales isolées).

Comme Détang-Dessendre et Piguët (2016), on considère ici que les zones urbaines incluent les grands pôles urbains et les zones péri-urbaines tandis que l'espace rural est composé des pôles moyens et petits ainsi que des communes rurales. Si l'on croise le zonage en aires urbaines avec la liste des ZIPU, il apparaît que certaines communes urbaines sont situées dans les RIP. Ce résultat a priori surprenant, qui concerne toutefois un nombre très minoritaire de communes⁵², s'explique par le fait que la liste des ZIPU est issue des investissements

⁵⁰ Le zonage de l'Insee est réalisé en deux étapes : (i) identification des centres urbains (selon le nombre d'emplois), et (ii) définition de la sphère d'influence de chaque centre urbain (une commune est considérée comme appartenant à une aire d'influence urbaine lorsqu'au moins 40% de sa population active résidente travaille dans le(s) centre(s) urbain(s) voisin(s)).

⁵¹ Voir Tableau A.4.2 pour plus de détails.

⁵² 4,3% des communes de la ZIPU appartiennent à un grand pôle urbain, 44,1% à l'espace péri-urbain, 4% aux petits et moyens pôles et 47,5% à l'espace rural. A l'inverse, les communes hors ZIPU appartiennent à 49,4% à un grand pôle urbain, à 46,6% au péri-urbain et seulement à 1% aux petits et moyens pôles et à 3% à l'espace rural.

prévisionnels des opérateurs privés, eux-mêmes basés essentiellement sur un critère de densité. Ainsi, certaines communes urbaines selon le zonage en aires urbaines peuvent ne pas constituer des destinations suffisamment rentables pour des opérateurs privés. Dans ce la suite du chapitre, les termes « urbain », « périurbain » et « rural » renvoient donc à la définition du zonage en aires urbaines tandis que les termes ZIPU ou « peu denses » renvoient aux zones dans lesquelles les opérateurs privés ne souhaitent pas investir.

Enfin, nous examinons aussi si l'effet du traitement varie selon les modalités du déploiement de la fibre : l'intensité de la couverture (*i.e.* la part de locaux couverts en fibre dans la commune)⁵³ et le nombre d'années depuis lequel la fibre est déployée. Ces variables sont décrites dans le tableau A.4.2⁵⁴.

Pour tester si le traitement est hétérogène en fonction des caractéristiques internes et externes des entreprises, on introduit des variables interactives dans le modèle de deuxième étape (modèle consistant à estimer l'effet du traitement sur l'échantillon apparié). Si l'hétérogénéité du traitement est généralement examinée en scindant l'échantillon total en plusieurs sous-échantillons, nous préférons recourir ici aux variables interactives afin de tester plus formellement si l'effet de la fibre varie significativement entre les différents groupes d'entreprises. Par ailleurs, pour tester si l'impact du Plan varie en fonction des modalités de déploiement, on scinde la variable de traitement pour distinguer différents groupes de traités, en fonction du taux de couverture communal ou du nombre d'années écoulées depuis l'arrivée de la fibre dans la commune. A noter, pour tester si l'effet du Plan est supérieur dans les communes bénéficiant de la fibre depuis plus longtemps, on restreint l'échantillon aux entreprises mono-établissements. En effet, il serait trop complexe de calculer une durée d'exposition à la fibre pour les entreprises multi-établissements dont l'indice de couverture est une moyenne pondérée de la couverture de leurs différents établissements.

⁵³ Pour tester si l'effet du traitement varie selon son intensité, une autre option aurait été de recourir à la méthode du score de propensity généralisé (Hirano et Imbens, 2004). Toutefois, à notre connaissance, cette méthode n'a pas été adaptée aux données d'enquête.

⁵⁴ Comparé aux variables d'appariement, nous avons parfois dû regrouper les modalités de certaines variables afin de conserver un nombre suffisants d'observations dans chaque catégorie, notamment pour les usages numériques encore peu répandus. Cela est notamment le cas pour le secteur économique, pour lequel nous ne pouvons alors distinguer que trois secteurs : Manufacture-construction ; Commerce-Transport-Hébergement ; Services.

5. Résultats

5.1. Appariement

Le tableau 4.17 examine les caractéristiques des entreprises traitées et non traitées. La colonne (1) compare les entreprises traitées à l'ensemble des entreprises non traitées tandis que la colonne (2) compare uniquement les traitées et non traitées similaires (i.e. appariées).

Avant appariement, les tests de différence de moyenne indiquent clairement que les entreprises traitées, qui bénéficient du déploiement de la fibre en premier, sont significativement différentes des autres. Comme attendu, ces entreprises sont situées dans des communes plus grandes, plus accessibles (plus proches d'une ville moyenne et moins souvent en zone de montagne) et qui disposaient déjà d'un meilleur accès au THD (30-100 Mbps) initialement.

Bien que le calendrier de déploiement soit déterminé sur la base de caractéristiques communales, le ciblage prioritaire des communes plus grandes et plus accessibles a logiquement conduit certains types d'entreprises à bénéficier de la fibre de manière plus précoce. Ainsi, les entreprises plus grandes (plus de 50 salariés), avec plusieurs établissements et opérant dans les secteurs du commerce/transport/hébergement et du tertiaire supérieur bénéficient plus souvent de la fibre que les autres. Ce résultat n'est guère surprenant puisque ce type d'entreprises est davantage présent dans les pôles de l'espace rural.

Le tableau 4.18 présente les résultats de l'estimation des scores de propension obtenus avec un modèle logit. L'estimation du modèle est satisfaisante avec un pseudo-R² de 0,36 et 83% des observations correctement classées. Les résultats obtenus sont cohérents⁵⁵ : les entreprises situées dans les communes les plus peuplées et les plus accessibles ont une probabilité supérieure de bénéficier les premières de la fibre. A l'inverse, celles implantées dans une commune souffrant initialement d'un faible accès au THD ont une probabilité plus faible de bénéficier de la fibre. Par ailleurs, comparées aux entreprises de 250 salariés et plus, les entreprises plus petites sont situées dans des communes qui bénéficient moins souvent de la fibre. De plus, comparées aux entreprises du secteur des « autres services », celles du tertiaire supérieur sont plus souvent situées dans des communes bénéficiant prioritairement de la fibre

⁵⁵ Les rapports de cote sont reportés. Pour rappel, un rapport de cote supérieur (inférieur) à 1 indique que la variable a un effet positif (négatif) sur la probabilité de bénéficier de la fibre.

tandis que celles du secteur manufacturier sont davantage implantées dans des communes non dotées de fibre.

L'estimation du modèle logit permet de calculer les scores de propension. Les entreprises traitées et non traitées sont ensuite appariées selon la méthode du voisin le plus proche (avec remplacement) et les entreprises hors-support commun sont éliminées⁵⁶ (selon le critère du minimum-maximum). La distribution des scores de propension avant et après l'appariement (figure 4.1) indiquent qu'ils se chevauchent de manière satisfaisante après appariement. De plus, une fois l'échantillon restreint aux entreprises appariées, les traitées et non-traitées ont des caractéristiques similaires dans l'ensemble (colonne 2 du tableau 4.17). En effet, les traitées et non traitées appariées sont identiques en tout point à l'exception que les entreprises traitées sont significativement plus souvent des entreprises mono-établissements. Cette différence demeure significative malgré plusieurs tentatives de modification de la procédure d'appariement (appariement par les 3 et les 5 plus proches voisins ; caliper ; appariement exact sur la variable « mono-établissement »). Pour tenir compte de cette différence entre traités et contrôles, nous introduisons une muette « mono-établissement » dans le modèle destiné à estimer l'impact de la fibre (cf. section 6.2). Notons toutefois que la surreprésentation de mono-établissements parmi les traités pourrait conduire à un biais d'atténuation (l'hypothèse, qui sera confirmée par la suite, étant que les entreprises mono-établissements ont moins d'intérêt à utiliser les outils numériques).

⁵⁶ L'élimination des entreprises hors-support conduit à éliminer 1,6% des traitées.

Tableau 4.17: Comparaison des caractéristiques des entreprises traitées et non traitées

	(1)			(2)		
	Avant appariement			Après appariement		
	Traitées	Contrôles	Différence [†]	Traitées	Contrôles	Différence [†]
Variables communales						
Population	8,52	4,34	4.18***	7,80	9,12	-1,31
Stock établissements	1,37	1,73	-0.36***	1,37	1,40	-0,03
Prox. Ville moyenne	0,32	0,16	0.16***	0,31	0,33	-0,02
Montagne	0,12	0,15	-0.03***	0,12	0,12	0,00
Mauvais accès THD	0,55	0,63	-0.07***	0,56	0,55	0,01
Variables au niveau entreprise						
Taille						
0-9 salariés	0,11	0,15	-0.04***	0,11	0,09	0,02
10-49 salariés	0,64	0,73	-0.09***	0,64	0,64	0,00
50-249 salariés	0,19	0,11	0.09***	0,19	0,21	-0,02
>250 salariés	0,05	0,01	0.04***	0,05	0,06	-0,01
Monoétablissement	0,46	0,77	-0.32***	0,46	0,29	0.16***
Age						
0-5 ans	0,11	0,11	0,01	0,11	0,15	-0,04
6-9 ans	0,09	0,11	-0.02***	0,09	0,08	0,01
>10 ans	0,80	0,78	0,02	0,80	0,77	0,03
Secteur						
Manufacture	0,26	0,28	-0.02**	0,27	0,25	0,02
Construction	0,13	0,21	-0.09***	0,12	0,16	-0,04
Comm., transp., héb.	0,44	0,39	0.04***	0,44	0,43	0,01
Tertiaire Sup.	0,11	0,05	0.06***	0,11	0,09	0,01
Autres services	0,07	0,06	0,01	0,06	0,07	-0,01
<i>N (pondéré)[‡]</i>	64 416	314 703	-	63 372	38 036	-
<i># hors support</i>	-	-	-	1 044	-	-

Note : [†]Pour les variables continues, un test de Student a été effectué tandis qu'un test du Chi-deux a été réalisé pour les variables qualitatives. * p<0.1, ** p<0.05, *** p<0.01.

[‡]Pour rappel, pour réaliser les tests de différence de moyenne nous appliquons les poids de l'enquête.

Tableau 4.18 : Estimation du score de propension (logit)

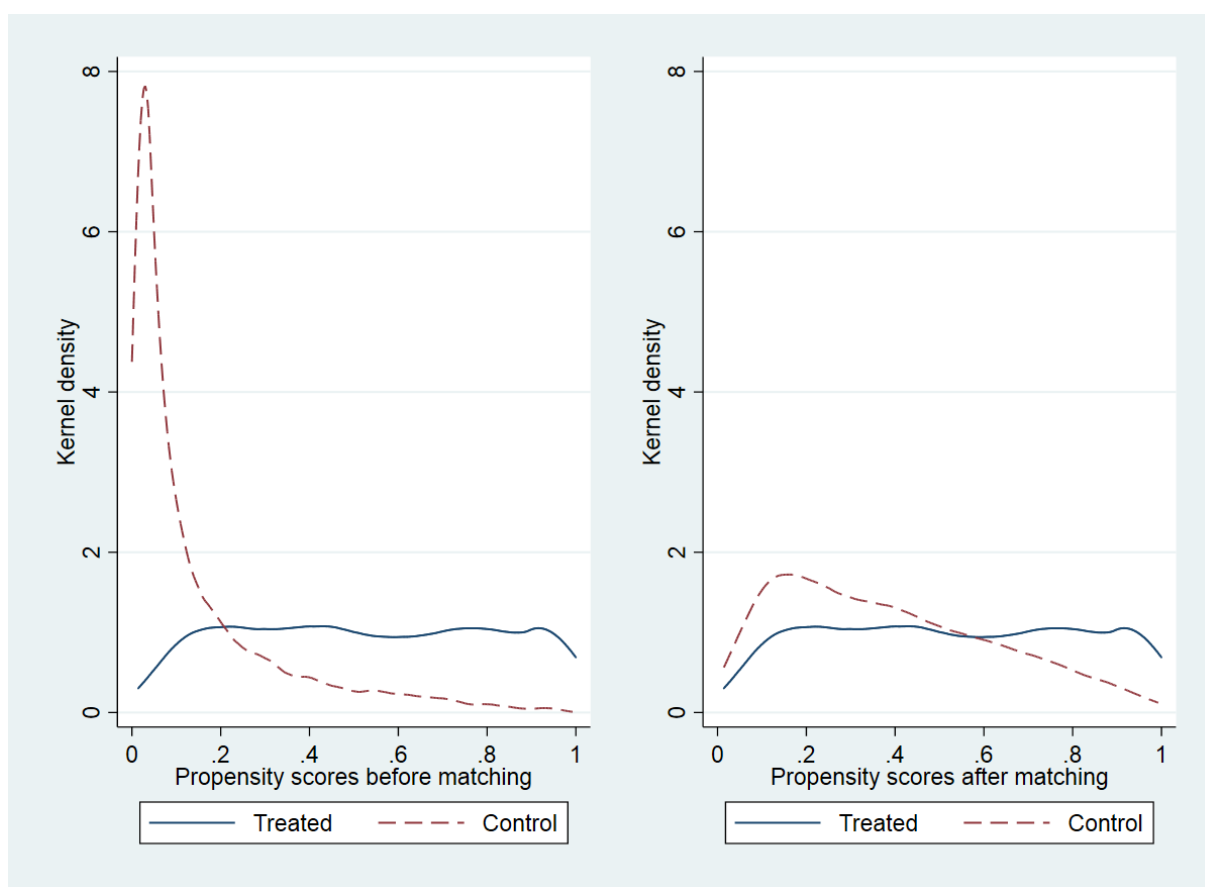
Variables au niveau communal	
Population	1.058*** (0.010)
Stock établissements	1.021 (0.039)
Stock établissements ²	0.997 (0.003)
Prox. ville moyenne	1.436*** (0.096)
Montagne	0.761*** (0.078)
Mauvais accès THD	0.770*** (0.038)
Variables au niveau entreprise	
<i>Taille (réf. >250)</i>	
0-9 salariés	0.277*** (0.034)
10-49 salariés	0.259*** (0.022)
50-249 salariés	0.412*** (0.030)
Mono-établissement	0.133*** (0.006)
<i>Age (réf: > 10 ans)</i>	
0-5 ans	1.250** (0.109)
6-9 ans	1.004 (0.083)
<i>Secteur (réf: autres services)</i>	
Manufacture	0.664*** (0.070)
Construction	0.848 (0.103)
Comm., trnsp., héb.	1.140 (0.117)
Tertiaire Sup.	1.948*** (0.242)
Pondération enquête	0.990*** (0.003)
Muettes départements	Oui
Muettes années	Oui
N (non pondéré)	18 078
% correctement classés	83.30%
Pseudo R2	0.357

Note : Rapports de cote reportés.

Ecart types entre parenthèses.

* p<0.10, ** p<0.05, *** p<0.01.

Figure 4.1 : Distribution des scores de propension avant et après appariement



5.2. Impact moyen du programme

Pour faciliter la lecture des résultats, on présente l'impact du programme en regroupant les usages numériques en trois catégories : (i) les pratiques organisationnelles (tableau 4.19), (ii) les usages plus avancés (tableau 4.20), et (iii) le commerce électronique (tableau 4.21). Les pratiques organisationnelles incluent l'intensité d'utilisation d'ordinateurs et d'internet, l'existence d'un site web, et l'usage des médias sociaux. Les pratiques plus complexes incluent le cloud computing, les PGI, la GRC, et les outils collaboratifs. Le commerce électronique regroupe les ventes web et EDI (adoption et intensité) et les achats web et EDI (adoption).

Pour chaque usage, on propose quatre estimations : (i) le modèle de base, dans lequel on régresse simplement la variable binaire de traitement sur la variable d'usage numérique correspondante (colonne 1) ; (ii) le « modèle de base augmenté », dans lequel on introduit en plus une variable muette « mono-établissement » pour contrôler les différences de structures organisationnelles entre traités et non traités (colonne 2); (iii) le modèle de base augmenté estimé en incluant uniquement les entreprises disposant de THD (30-100 Mbps) dans le groupe

de contrôle⁵⁷ (colonne 3); (iv) le modèle de base augmenté en excluant du groupe de traités les entreprises situées dans une commune où moins de 10% des locaux sont couverts en fibre⁵⁸ (colonne 4).

Bien que peu de statistiques post-estimations soient disponibles dans le cadre de l'analyse de données d'enquête, nous avons réalisé pour chaque estimation logistique le test de Hosmer et Lemeshow. Dans chacun des cas, le test indique qu'on ne peut pas rejeter l'hypothèse nulle selon laquelle les valeurs prédites et observées sont similaires pour les différents déciles (résultats non reportés, disponibles sur demande).

D'après nos estimations, le plan FTHD encourage certaines pratiques en ligne dans les espaces peu denses. Plus précisément, les entreprises situées dans des communes ayant accès à la fibre font à la fois davantage usage d'outils organisationnels (utilisation d'internet, d'ordinateurs, site web) et complexes (cloud computing, PGI, GRC). Les effets ne sont pas négligeables. D'après notre spécification préférée (colonne 2)⁵⁹, l'accès à la fibre augmenterait la part des salariés utilisant un ordinateur de 8 points de pourcentage et celle utilisant internet de 11 points de pourcentage. L'accès à la fibre augmenterait également la probabilité d'utiliser un PGI ou le cloud computing ou de 1,3 et celle de recourir à la GRC de 1,5.

Si l'effet de la fibre sur les usages avancés n'est pas surprenant (ces pratiques nécessitent un débit très rapide)⁶⁰, son impact sur les usages organisationnels était moins attendu. Plusieurs raisons peuvent expliquer que la fibre stimule l'usage d'outils organisationnels, accessibles même avec des débits moins rapides. Premièrement, en favorisant l'usage d'outils avancés, l'accès à la fibre accroît mécaniquement l'usage d'outils organisationnels (on ne peut accéder au cloud sans utiliser un ordinateur et internet). Deuxièmement, comme on l'a déjà évoqué

⁵⁷ Pour rappel, l'intérêt de comparer les entreprises ayant accès à la fibre à celles ayant accès au THD (et non à toutes celles disposant d'un accès inférieur à 100 Mbps comme dans le modèle de base) est qu'on estime l'effet *additionnel* de la fibre par rapport au THD. De plus, l'exclusion du groupe de contrôle des entreprises situées dans les communes disposant d'un accès inférieur à 30 Mbps permet d'éliminer dès le départ des entreprises situées dans des contextes territoriaux vraisemblablement très différents (communes très peu denses, peu accessibles, etc.).

⁵⁸ Certaines entreprises sont en effet situées dans des communes où une part très faible des locaux est éligible à la fibre, ce qui pourrait sous-estimer l'impact du traitement.

⁵⁹ On préfère le modèle augmenté (colonne 2) au modèle de base (colonne 1) puisqu'il permet de contrôler les différences de structures organisationnelles entre traités et non traités. Conformément à notre hypothèse, les entreprises avec un établissement unique utilisent significativement moins les outils numériques que les multi-établissements. Toutes les analyses suivantes (section 6.3) sont réalisées à partir de ce modèle augmenté.

⁶⁰ L'accès au cloud computing nécessite un débit (de plus en plus) rapide. Si l'usage de PGI et de GRC ne nécessite pas a priori d'accéder à un débit très rapide, l'accès à ces ressources se fait de manière croissante sur le cloud (Nicoletti et al., 2020), ce qui exige alors l'accès à des débits très rapides.

plusieurs fois, l'impact estimé ici (colonne 2) capte à la fois l'effet de la fibre, du THD, et du bon HD⁶¹. On peut donc se demander si, pour les usages organisationnels, l'impact estimé ici reflète réellement un effet pur de la fibre ou plutôt l'effet de l'accès à des classes de débit inférieures. Pour le savoir, on restreint le groupe de contrôle aux entreprises ayant accès à 30-100 Mbps (colonne 3), de sorte à ne capter que l'effet *additionnel* de la fibre. Si, pour la plupart des usages, modifier la définition du groupe de contrôle ne change rien, cela engendre néanmoins une diminution de l'impact du traitement pour l'utilisation d'ordinateurs et d'internet. Ce résultat indique que ces deux pratiques de base semblent donc dépendre non seulement de l'accès à la fibre mais aussi (surtout ?) de l'accès à des débits moindres, que l'on capte partiellement dans notre modèle de référence (colonne 2).

Par ailleurs, changer la définition du groupe de traitement (colonne 4) n'engendre pas de modification notable des résultats. En d'autres termes, inclure au sein du groupe de traités des entreprises situées dans des communes ayant un taux de couverture en fibre faible ne semble pas sous-estimer l'effet moyen du traitement. La question de l'effet de l'intensité de la couverture sera examinée davantage dans la section 6.3.

Enfin, si la fibre a un effet significatif sur une majorité d'usages, elle n'a pas d'impact sur l'utilisation des médias sociaux, des outils collaboratifs et le commerce en ligne. L'absence d'effet sur la probabilité d'usage des outils collaboratifs est particulièrement surprenante puisque ce type de pratique (ex : visioconférence) requiert un débit rapide et symétrique⁶².

⁶¹ En effet, l'impact est estimé en comparant les entreprises ayant accès à la fibre à celles ayant accès à moins de 100 Mbps. Le groupe de contrôle inclut donc à la fois des entreprises ayant accès à du THD mais aussi des entreprises ayant accès à des débits très faibles.

⁶² Une explication potentielle est que seules les enquêtes-TIC de 2015 et 2017 renseignent sur l'usage des outils collaboratifs, et donc, pour cet usage, on ne peut mesurer que l'effet de la couverture fibre en 2014 et 2016 qui était alors encore très limitée.

Tableau 4.19 : Impact moyen de la fibre sur les pratiques organisationnels

	(1) Baseline	(2) Mono-étab.	(3) GC_3^{Fibre} (30- 100)	(4) GT : couv. >1%
Ordinateur : intensité				
Traitement	6.610** (2.725)	8.378*** (2.689)	4.974* (2.764)	4.127 (3.018)
Mono-établissement	-	-12.991*** (2.521)	-11.101*** (2.724)	-14.826*** (3.084)
N (pondéré)	47 766	47 766	47 478	43 836
Internet : intensité				
Traitement	9.163*** (1.848)	10.648*** (1.803)	7.530*** (1.856)	6.437*** (2.180)
Mono-établissement	-	-8.980*** (1.611)	-9.535*** (1.678)	-8.783*** (1.936)
N (pondéré)	129 559	129 559	130 412	123 398
Site web				
Traitement	1.230* (0.143)	1.299** (0.141)	1.393*** (0.147)	1.369*** (0.157)
Mono-établissement	-	0.724*** (0.076)	0.712*** (0.074)	0.726*** (0.082)
N (pondéré)	129 559	129 559	130 412	123 398
Médias sociaux				
Traitement	1.071 (0.099)	1.106 (0.102)	1.019 (0.102)	1.139 (0.124)
Mono-établissement	-	0.663*** (0.060)	0.639*** (0.061)	0.666*** (0.071)
N (pondéré)	88 987	88 987	87 331	84 591

Note : Rapports de cote reportés pour les modèles logit (site web ; médias sociaux). Ecarts types entre parenthèses.
* p<0.10, ** p<0.05, *** p<0.01. Pour rappel, les estimations sont pondérées par un poids final, égal au produit des poids de l'enquête et de l'appariement.

Tableau 4.20 : Impact moyen de la fibre sur les pratiques plus avancées

	(1) Baseline	(2) Mono-étab.	(3) GC_3^{Fibre} (30- 100)	(4) GT : couv. >1%
Cloud computing				
Traitement	1.245 (0.196)	1.303* (0.200)	1.373** (0.207)	1.717*** (0.260)
Mono-établissement	-	0.753* (0.111)	0.752** (0.108)	0.676*** (0.099)
N (pondéré)	61 165	61 165	59 319	54 482
PGI				
Traitement	1.220 (0.155)	1.304** (0.161)	1.425*** (0.178)	1.263* (0.163)
Mono-établissement	-	0.688*** (0.081)	0.618*** (0.072)	0.709*** (0.085)
N (pondéré)	80 081	80 081	79 128	74 543
GRC				
Traitement	1.380*** (0.166)	1.482*** (0.180)	1.474*** (0.185)	1.404*** (0.182)
Mono-établissement	-	0.667*** (0.079)	0.516*** (0.059)	0.563*** (0.066)
N (pondéré)	80 081	80 081	79 128	74 543
Outils collaboratifs				
Traitement	0.833 (0.133)	0.906 (0.143)	1.205 (0.177)	1.252 (0.212)
Mono-établissement	-	0.448*** (0.073)	0.464*** (0.075)	0.388*** (0.066)
N (pondéré)	35 049	35 049	37 428	32 828

Note : Rapports de cote reportés pour les modèles logit (cloud computing, PGI, GRC, outils collaboratifs). Ecarts types entre parenthèses. * p<0.10, ** p<0.05, *** p<0.01. Pour rappel, les estimations sont pondérées par un poids final, égal au produit des poids de l'enquête et de l'appariement.

Tableau 4.21 : Impact moyen de la fibre sur le commerce électronique

	(1) Baseline	(2) Mono-étab.	(3) GC_3^{Fibre} (30- 100)	(4) GT : couv. >1%
Ventes web-EDI : adoption				
Traitement	0.916 (0.098)	0.961 (0.100)	1.019 (0.102)	1.127 (0.128)
Mono-établissement	-	0.740*** (0.071)	0.727*** (0.068)	0.789** (0.079)
N (pondéré)	129 559	129 559	130 412	123 398
Ventes web-EDI : intensité				
Traitement	-2.318 (2.953)	-0.949 (2.859)	-0.261 (2.911)	4.441 (3.049)
Mono-établissement	-	-9.160*** (2.629)	-9.037*** (2.777)	-7.864*** (2.731)
N (pondéré)	129 559	129 559	130 412	123 398
Achats web-EDI				
Traitement	1.242 (0.214)	1.262 (0.210)	1.169 (0.138)	1.239* (0.150)
Mono-établissement	-	0.853 (0.117)	0.905 (0.116)	0.894 (0.103)
N (pondéré)	68 975	68 975	67 623	62 462

Note : Rapports de cote reportés pour les modèles logit (ventes web-EDI : adoption ; achats web-EDI). Ecarts types entre parenthèses. * p<0.10, ** p<0.05, *** p<0.01. Pour rappel, les estimations sont pondérées par un poids final, égal au produit des poids de l'enquête et de l'appariement.

5.3. Hétérogénéité de l'impact du programme

Hétérogénéité selon les caractéristiques des entreprises

On examine à présent si l'impact de la fibre varie selon les caractéristiques suivantes des entreprises : l'âge, la structure organisationnelle, la taille, et le secteur d'activité.

Les résultats sont reportés dans les tableaux 4.22 (usages organisationnels), 4.23 (usages avancés), et 4.24 (commerce électronique). Afin de faciliter la lecture des résultats, les tableaux présentés ici reportent seulement l'effet du traitement estimé pour chaque groupe d'entreprises. Les résultats complets (incluant les variables interactives et additives) sont disponibles sur demande.

Parmi toutes les caractéristiques des entreprises, c'est la taille et la structure organisationnelle qui conditionnent le plus l'effet du traitement. Tout d'abord, concernant la taille : si la présence de fibre accroît l'usage des outils numériques dans les petites entreprises (< 50 salariés), l'impact est moins souvent significatif pour les entreprises intermédiaires (50-249 salariés) et quasiment inexistant pour les grandes entreprises (> 250 salariés). Cela ne

signifie pas que les entreprises les plus grandes font moins usages des outils numériques, bien au contraire : les coefficients associés aux variables de taille introduites additivement dans le modèle indiquent que les grandes entreprises utilisent significativement plus que les autres les outils numériques (à l'exception, logiquement, du cloud computing)⁶³. En revanche, nos résultats indiquent que les pratiques numériques des grandes entreprises sont *peu affectées par la disponibilité de la fibre dans leur commune*. Même en l'absence de fibre dans leur commune, les grandes entreprises ont en effet les moyens de se procurer leur propre accès internet. A l'inverse, l'usage numériques des entreprises plus petites, dont les ressources financières sont plus limitées, apparaît fortement dépendant des réseaux publics existants au niveau communal. Les petites entreprises constituant la grande majorité des entreprises, le Plan FTTHD devrait donc vraisemblablement concourir à accroître les pratiques numériques d'un grand nombre.

Par ailleurs, si la disponibilité de la fibre accroît significativement les pratiques numériques des entreprises multi-établissements, l'effet est beaucoup plus limité pour les mono-établissements. Logiquement, l'arrivée de la fibre engendre essentiellement un impact pour les entreprises multi-établissements, qui font face à des coûts de coordination supérieurs, et donc, ont davantage d'intérêt à s'emparer des outils numériques (et notamment de ceux, tels que les PGI, qui visent à optimiser le fonctionnement interne d'une entreprise).

L'effet du traitement conditionnellement à l'âge est moins évident. En effet, la couverture fibre semble stimuler davantage à la fois les usages numériques des entreprises les plus jeunes (< 5 ans) et ceux des plus anciennes (> 10 ans). Cependant, de manière intéressante, la disponibilité de la fibre favorise l'émergence *d'usages différents* selon l'âge des entreprises. Ainsi, la fibre accroît significativement plus le recours à la GRC et au cloud computing dans les entreprises les plus récentes⁶⁴, qui ont le plus à gagner de l'usage de ces outils (les jeunes entreprises bénéficient en effet particulièrement de l'adoption de la GRC, qui leur permet d'accroître leurs ventes, et du cloud computing, qui leur permet de réduire leurs investissements).

Enfin, les résultats par secteurs indiquent que la couverture en fibre stimule significativement les usages numériques des entreprises de la manufacture-construction et du commerce-transport-hébergement. En revanche, cela n'a quasiment pas d'impact sur celles

⁶³ Les coefficients associés aux variables additives ne sont pas reportés dans les tableaux. Les résultats complets sont disponibles sur demande.

⁶⁴ Des tests de Wald ont été réalisés pour tester l'égalité des coefficients entre chaque catégorie d'entreprise. Les résultats sont disponibles sur demande.

opérant dans le secteur des services. Malheureusement, le manque d'observations nous a contraint à créer des catégories sectorielles très agrégées, regroupant des activités fortement hétérogènes en termes de pratiques numériques. En particulier, l'absence d'impact de la fibre pour les entreprises du tertiaire masque sans aucun doute une forte hétérogénéité entre les entreprises opérant dans des domaines qui ont beaucoup à gagner des outils numériques (ex : « Activités informatiques et services d'information », « Edition, audiovisuel et diffusion ») et celles de secteurs moins numérisés (ex : « Activités de services administratifs et de soutien »).

Tableau 4.22 : Effet de la fibre selon les caractéristiques des entreprises
– usages organisationnels

	(1) Ordinateur : intensité	(2) Internet : intensité	(3) Site web	(4) Médias sociaux
<i>Effet selon l'âge</i>				
0-5 ans	12.555 (8.342)	12.810** (6.160)	0.863 (0.436)	1.041 (0.284)
6-9 ans	5.324 (7.873)	7.848 (5.442)	1.364 (0.364)	0.658 (0.190)
≥ 10 ans	7.068*** (2.531)	9.962*** (1.593)	1.383*** (0.146)	1.189 (0.126)
<i>Effet selon la structure organisationnelle</i>				
Mono-établissement	5.174 (3.896)	3.789* (2.108)	1.103 (0.158)	1.041 (0.144)
Multi-établissement	9.909*** (3.494)	14.721*** (2.518)	1.454** (0.226)	1.151 (0.141)
<i>Effet selon la taille</i>				
< 50 salariés	9.044*** (3.499)	10.745*** (2.401)	1.309** (0.179)	1.071 (0.124)
50-249 salariés	8.126** (3.671)	10.856*** (2.254)	1.415* (0.295)	1.363** (0.209)
≥ 250 salariés	3.270 (3.633)	8.894** (3.570)	0.796 (0.269)	0.743 (0.162)
<i>Effet selon le secteur</i>				
Manufacture-Const.	8.370** (3.360)	8.197*** (2.486)	1.204 (0.240)	1.266* (0.169)
Comm.-Transp.-Héb.	9.350*** (3.099)	12.589*** (2.165)	1.368** (0.198)	1.063 (0.154)
Services	7.155 (6.707)	7.475 (5.154)	1.364 (0.327)	1.036 (0.244)

Note : Rapports de cote reportés pour les modèles logit (site web ; médias sociaux). Ecart types entre parenthèses. * p<0.10, ** p<0.05, *** p<0.01. Pour rappel, les estimations sont pondérées par un poids final, égal au produit des poids de l'enquête et de l'appariement. Le tableau reporte l'effet moyen du traitement recalculé pour chaque catégorie d'entreprises. Chaque modèle inclut des variables de contrôle (mono-établissement et, le cas échéant, variables muettes pour l'âge, la taille, et le secteur) ; les coefficients associés à ces variables ne sont pas reportés dans les tableaux.

Tableau 4.23 : Effet de la fibre selon les caractéristiques des entreprises – usages avancés

	(1) Cloud computing	(2) PGI	(3) GRC	(4) Outils collaboratifs
<i>Effet selon l'âge</i>				
0-5 ans	3.080* (2.097)	1.412 (0.756)	3.579*** (1.721)	0.643 (0.324)
6-9 ans	1.341 (0.566)	1.720 (0.613)	1.004 (0.398)	0.850 (0.475)
≥ 10 ans	1.192 (0.201)	1.201 (0.144)	1.373*** (0.166)	0.962 (0.167)
<i>Effet selon la structure organisationnelle</i>				
Mono-établissement	1.125 (0.270)	1.034 (0.193)	0.852 (0.168)	0.684 (0.187)
Multi-établissement	1.390* (0.270)	1.485** (0.240)	1.931*** (0.288)	0.994 (0.188)
<i>Effet selon la taille</i>				
< 50 salariés	1.423* (0.302)	1.353** (0.208)	1.448** (0.236)	0.887 (0.216)
50-249 salariés	1.145 (0.326)	1.138 (0.215)	1.511** (0.281)	0.882 (0.233)
≥ 250 salariés	1.088 (0.312)	1.359 (0.449)	1.581* (0.411)	0.884 (0.285)
<i>Effet selon le secteur</i>				
Manufacture-Const.	0.877 (0.181)	1.201 (0.288)	1.507* (0.317)	0.638* (0.152)
Comm.-Transp.-Héb.	1.917** (0.600)	1.663*** (0.291)	1.672*** (0.282)	1.103 (0.263)
Services	1.292 (0.433)	0.875 (0.230)	1.046 (0.287)	1.296 (0.535)

Note : Rapports de cote reportés pour les modèles logit (cloud computing, PGI, GRC, outils collaboratifs). Ecart types entre parenthèses. * p<0.10, ** p<0.05, *** p<0.01. Pour rappel, les estimations sont pondérées par un poids final, égal au produit des poids de l'enquête et de l'appariement. Le tableau reporte l'effet moyen du traitement recalculé pour chaque catégorie d'entreprises. Chaque modèle inclut des variables de contrôle (mono-établissement et, le cas échéant, variables muettes pour l'âge, la taille, et le secteur) ; les coefficients associés à ces variables ne sont pas reportés dans les tableaux.

Tableau 4.24 : Effet de la fibre selon les caractéristiques des entreprises
– commerce électronique

	(1) Ventes web- EDI : adoption	(2) Ventes web- EDI : intensité	(3) Achats web-EDI
<i>Effet selon l'âge</i>			
0-5 ans	1.440 (0.684)	17.541 (12.644)	2.753 (1.802)
6-9 ans	0.632 (0.210)	-10.186 (11.287)	1.020 (0.302)
≥ 10 ans	0.931 (0.088)	-3.041 (2.556)	1.080 (0.111)
<i>Effet selon la structure organisationnelle</i>			
Mono-établissement	0.812 (0.113)	-4.002 (3.807)	1.434** (0.219)
Multi-établissement	1.045 (0.145)	0.639 (3.866)	1.185 (0.283)
<i>Effet selon la taille</i>			
< 50 salariés	1.038 (0.149)	0.268 (3.712)	1.355 (0.279)
50-249 salariés	0.758* (0.111)	-4.093 (3.946)	1.007 (0.171)
≥ 250 salariés	1.091 (0.254)	1.092 (5.650)	1.048 (0.211)
<i>Effet selon le secteur</i>			
Manufacture-Const.	1.027 (0.196)	-3.626 (5.312)	1.879* (0.664)
Comm.-Transp.-Héb.	0.947 (0.125)	2.162 (3.731)	0.813 (0.115)
Services	0.842 (0.258)	-3.486 (7.282)	1.684** (0.383)

Note : Rapports de cote reportés pour les modèles logit (ventes web-EDI : adoption ; achats web-EDI). Ecart types entre parenthèses. * p<0.10, ** p<0.05, *** p<0.01. Pour rappel, les estimations sont pondérées par un poids final, égal au produit des poids de l'enquête et de l'appariement. Le tableau reporte l'effet moyen du traitement recalculé pour chaque catégorie d'entreprises. Chaque modèle inclut des variables de contrôle (mono-établissement et, le cas échéant, variables muettes pour l'âge, la taille, et le secteur) ; les coefficients associés à ces variables ne sont pas reportés dans les tableaux.

Hétérogénéité selon l'environnement externe des entreprises

On examine à présent si l'impact de la fibre varie selon l'environnement externe des entreprises, et notamment en fonction du degré de ruralité du territoire et de la présence de fournisseurs en équipements et services informatiques.

Les résultats sont reportés dans les tableaux 4.25 (usages organisationnels), 4.26 (usages avancés), et 4.27 (commerce électronique). Comme précédemment, les tableaux reportent seulement l'effet du traitement estimé pour chaque groupe d'entreprises. Les résultats complets (incluant les variables interactives et additives) sont disponibles sur demande.

L'impact du Plan FTHD dépend du degré de ruralité du territoire dans lequel se situent les entreprises. Plus précisément, la fibre est essentiellement associée à une hausse des usages numériques dans les espaces péri-urbains. À l'inverse, le Plan FTHD n'a que peu d'effet dans les grands pôles urbains et les petits et moyens pôles et aucun effet dans le rural. Ces résultats sont conformes à ceux de Mack (2015) et de Duvivier et al. (2021) selon lesquels le THD a davantage d'impact dans les espaces intermédiaires, caractérisés à la fois par une capacité d'appropriation des outils élevée (niveau de capital humain et externalités informationnelles relativement élevés⁶⁵) et des incitations fortes à les utiliser (en raison de l'éloignement des partenaires et des clients, et des coûts d'information et de coordination supérieurs). À l'inverse, dans les espaces ruraux, même si les entreprises auraient beaucoup à gagner à s'emparer des outils numériques (Forman et al., 2005b), le Plan FTHD n'a pas d'impact, probablement en raison du niveau de capital humain plus réduit et de la faible pénétration des outils numériques (externalités informationnelles et de réseau plus faibles).

Par ailleurs, l'impact de la fibre n'est pas plus faible sur les entreprises situées dans des bassins de vie dépourvus de fournisseurs en équipements et services informatiques. Des résultats similaires sont obtenus lorsque l'on considère séparément les fournisseurs d'équipements et de services informatiques. Ce résultat contredit donc l'idée selon laquelle l'absence de fournisseurs d'équipements et de services informatiques limite l'effet des politiques de déploiement dans les espaces peu denses (cf. section 2.3).

⁶⁵ Nous ne disposons pas d'indicateur pertinent pour contrôler ces facteurs dans le modèle d'appariement.

Tableau 4.25 : Effet de la fibre selon les caractéristiques des communes

– usages organisationnels

	(1) Ordinateur : intensité	(2) Internet : intensité	(3) Site web	(4) Médias sociaux
<i>Effet selon le degré de ruralité</i>				
Grands poles urbains	10.567 (6.886)	10.940** (4.795)	1.100 (0.271)	1.090 (0.216)
Périurbain	7.406** (3.490)	10.315*** (2.271)	1.397** (0.232)	1.222 (0.186)
Petits et moyens pôles	0.954 (4.963)	12.119*** (3.263)	1.727*** (0.354)	1.026 (0.195)
Rural	6.280 (4.044)	6.120* (3.469)	1.061 (0.243)	0.958 (0.219)
<i>Effet selon la présence d'équipements et de services informatiques</i>				
Présence	11.108* (6.174)	15.510*** (3.906)	1.392 (0.320)	1.065 (0.178)
Absence	6.180** (2.513)	7.391*** (1.768)	1.237* (0.140)	1.137 (0.128)

Note : Rapports de cote reportés pour les modèles logit (site web ; médias sociaux). Ecart types entre parenthèses. * p<0.10, ** p<0.05, *** p<0.01. Pour rappel, les estimations sont pondérées par un poids final, égal au produit des poids de l'enquête et de l'appariement. Le tableau reporte l'effet moyen du traitement recalculé pour chaque catégorie d'entreprises. Chaque modèle inclut des variables de contrôle (mono-établissement et, le cas échéant, variables muettes pour le type d'espace ou la présence d'équipements et services informatiques) ; les coefficients associés à ces variables ne sont pas reportés dans les tableaux.

Tableau 4.26 : Effet de la fibre selon les caractéristiques des communes – usages avancés

	(1) Cloud computing	(2) PGI	(3) GRC	(4) Outils collaboratifs
<i>Effet selon le degré de ruralité</i>				
Grands poles urbains	1.214 (0.306)	2.908** (1.375)	1.799 (0.856)	0.797 (0.237)
Périurbain	1.695** (0.424)	1.750*** (0.316)	1.627** (0.312)	0.963 (0.282)
Petits et moyens pôles	1.232 (0.504)	1.217 (0.311)	0.806 (0.210)	0.783 (0.231)
Rural	0.883 (0.362)	1.106 (0.342)	1.616 (0.531)	1.048 (0.443)
<i>Effet selon la présence d'équipements et de services informatiques</i>				
Présence	1.226 (0.287)	2.548** (1.194)	1.510 (0.698)	1.159 (0.314)
Absence	1.323 (0.276)	1.485*** (0.200)	1.407** (0.202)	0.811 (0.155)

Note : Rapports de cote reportés pour les modèles logit (cloud computing, PGI, GRC, outils collaboratifs). Ecart types entre parenthèses. * p<0.10, ** p<0.05, *** p<0.01. Pour rappel, les estimations sont pondérées par un poids final, égal au produit des poids de l'enquête et de l'appariement. Le tableau reporte l'effet moyen du traitement recalculé pour chaque catégorie d'entreprises. Chaque modèle inclut des variables de contrôle (mono-établissement et, le cas échéant, variables muettes pour le type d'espace ou la présence d'équipements et services informatiques) ; les coefficients associés à ces variables ne sont pas reportés dans les tableaux.

Tableau 4.27 : Effet de la fibre selon les caractéristiques des communes
– commerce électronique

	(1) Ventes web- EDI : adoption	(2) Ventes web- EDI : intensité	(3) Achats web-EDI
<i>Effet selon le degré de ruralité</i>			
Grands pôles urbains	0.939 (0.227)	0.927 (6.361)	1.685 (0.729)
Périurbain	1.057 (0.161)	2.334 (4.367)	1.268 (0.196)
Petits et moyens pôles	0.956 (0.171)	-3.062 (4.559)	0.847 (0.174)
Rural	0.944 (0.210)	-6.925 (6.836)	1.105 (0.280)
<i>Effet selon la présence d'équipements et de services informatiques</i>			
Présence	0.969 (0.217)	-0.374 (5.866)	1.551 (0.608)
Absence	0.971 (0.101)	-0.865 (3.009)	1.147 (0.129)

Note : Rapports de cote reportés pour les modèles logit (ventes web-EDI : adoption ; achats web-EDI). Ecart types entre parenthèses. * p<0.10, ** p<0.05, *** p<0.01. Pour rappel, les estimations sont pondérées par un poids final, égal au produit des poids de l'enquête et de l'appariement. Le tableau reporte l'effet moyen du traitement recalculé pour chaque catégorie d'entreprises. Chaque modèle inclut des variables de contrôle (mono-établissement et, le cas échéant, variables muettes pour le type d'espace ou la présence d'équipements et services informatiques) ; les coefficients associés à ces variables ne sont pas reportés dans les tableaux.

Hétérogénéité selon les modalités du déploiement

On examine à présent si l'impact varie selon les modalités de déploiement, et plus particulièrement selon le taux de couverture communal et le nombre d'années écoulées depuis l'arrivée de la fibre.

Les résultats sont reportés dans les tableaux 4.28 (usages organisationnels), 4.29 (usages avancés), et 4.30 (commerce électronique). Contrairement à ce qu'on pourrait penser, l'impact du Plan n'est pas plus élevé dans les communes qui ont un niveau de couverture supérieur. Une explication possible est que les entreprises, en partie concentrées dans les zones d'activités, sont probablement les premières à bénéficier de l'arrivée de la fibre dans une commune. Ainsi, si une part élevée de locaux couverts est désirable au niveau des ménages (pour garantir notamment un accès universel aux services publics en ligne), l'impact du Plan sur les usages numériques des entreprises varie peu une fois qu'un niveau minimal de locaux est couvert⁶⁶.

⁶⁶ Pour donner un ordre d'idée de ce niveau minimal, on estime ici que l'effet diffère peu en-dessous et au-dessus du deuxième quartile de couverture, qui s'élève à 32% de locaux couverts dans notre échantillon. Rappelons toutefois qu'il s'agit ici d'un seuil moyen et que, pour les pôles ruraux regroupant un certain nombre d'entreprises, le seuil est probablement plus élevé.

En revanche, l'effet du Plan dépend fortement du nombre d'années écoulées depuis l'arrivée de la fibre. En effet, pour les usages numériques affectés par la fibre⁶⁷, l'effet est uniquement significatif pour les entreprises situées dans les communes qui bénéficient de la fibre depuis trois ans au moins. L'impact croissant du Plan au cours du temps s'explique vraisemblablement par l'existence de délais dans l'appropriation du THD par les entreprises (prise de décision, externalités informationnelles et de réseau notamment).

Tableau 4.28 : Effet de la fibre selon les modalités du déploiement – usages organisationnels

	(1) Ordinateur : intensité	(2) Internet : intensité	(3) Site web	(4) Médias sociaux
<i>Effet selon l'intensité du déploiement</i>				
≤ Q1	5.679* (3.290)	8.336*** (2.221)	1.545*** (0.227)	1.051 (0.131)
> Q1 & ≤ Q2	11.611*** (3.715)	13.990*** (2.403)	1.297* (0.188)	1.229 (0.159)
> Q2 & ≤ Q3	13.076*** (3.714)	11.513*** (2.377)	1.103 (0.153)	0.955 (0.127)
> Q3	3.004 (3.430)	8.620*** (2.224)	1.282* (0.177)	1.209 (0.163)
<i>Effet selon le temps écoulé depuis l'arrivée de la fibre</i>				
Trois ans ou plus	10.407 (7.360)	3.589 (2.410)	1.280* (0.191)	1.162 (0.205)
Deux ans	-4.357 (4.802)	-0.150 (2.646)	1.094 (0.182)	0.963 (0.183)
Un an	0.683 (5.010)	0.684 (2.651)	1.147 (0.184)	0.955 (0.189)

Note : Rapports de cote reportés pour les modèles logit (site web ; médias sociaux). Ecart types entre parenthèses. * p<0.10, ** p<0.05, *** p<0.01. Pour rappel, les estimations sont pondérées par un poids final, égal au produit des poids de l'enquête et de l'appariement. Le tableau reporte l'effet moyen du traitement recalculé pour chaque catégorie d'entreprises.

⁶⁷ Pour rappel, ce test est réalisé uniquement pour les entreprises mono-établissements, ce qui explique que le traitement a un effet significatif sur un nombre relativement restreint d'usages.

Tableau 4.29 : Effet de la fibre selon les modalités du déploiement – usages avancés

	(1) Cloud computing	(2) PGI	(3) GRC	(4) Outils collaboratifs
<i>Effet selon l'intensité du déploiement</i>				
≤ Q1	1.483** (0.286)	1.476** (0.226)	1.468*** (0.213)	0.763 (0.149)
> Q1 & ≤ Q2	1.516** (0.301)	1.237 (0.195)	1.467** (0.224)	0.869 (0.178)
> Q2 & ≤ Q3	1.343 (0.277)	1.230 (0.197)	1.628*** (0.264)	1.095 (0.241)
> Q3	0.806 (0.182)	1.270 (0.204)	1.373* (0.235)	0.963 (0.239)
<i>Effet selon le temps écoulé depuis l'arrivée de la fibre</i>				
Trois ans ou plus	1.912** (0.600)	1.998*** (0.474)	1.681** (0.377)	0.481 (0.242)
Deux ans	1.429 (0.565)	1.293 (0.291)	0.949 (0.224)	1.014 (0.362)
Un an	1.585 (0.485)	1.725** (0.375)	1.129 (0.254)	0.564 (0.230)

Note : Rapports de cote reportés pour les modèles logit (cloud computing, PGI, GRC, outils collaboratifs). Ecart types entre parenthèses. * p<0.10, ** p<0.05, *** p<0.01. Pour rappel, les estimations sont pondérées par un poids final, égal au produit des poids de l'enquête et de l'appariement. Le tableau reporte l'effet moyen du traitement recalculé pour chaque catégorie d'entreprises.

Tableau 4.30 : Effet de la fibre selon les modalités du déploiement – commerce électronique

	(1) Ventes web- EDI : adoption	(2) Ventes web- EDI : intensité	(3) Achats web-EDI
<i>Effet selon l'intensité du déploiement</i>			
≤ Q1	1.124 (0.143)	2.507 (3.518)	1.304 (0.263)
> Q1 & ≤ Q2	0.965 (0.129)	-1.096 (3.804)	1.520** (0.317)
> Q2 & ≤ Q3	0.828 (0.112)	-4.218 (3.728)	1.062 (0.197)
> Q3	0.901 (0.124)	-1.930 (3.891)	1.178 (0.210)
<i>Effet selon le temps écoulé depuis l'arrivée de la fibre</i>			
Trois ans ou plus	1.116 (0.183)	2.716 (4.920)	1.242 (0.299)
Deux ans	1.059 (0.194)	3.231 (5.522)	1.383 (0.304)
Un an	1.124 (0.201)	-0.178 (5.210)	0.944 (0.182)

Note : Rapports de cote reportés pour les modèles logit (ventes web-EDI : adoption ; achats web-EDI). Ecart types entre parenthèses. * p<0.10, ** p<0.05, *** p<0.01. Pour rappel, les estimations sont pondérées par un poids final, égal au produit des poids de l'enquête et de l'appariement. Le tableau reporte l'effet moyen du traitement recalculé pour chaque catégorie d'entreprises.

6. Conclusion

Ce chapitre évalue l'impact du THD sur les usages numériques des entreprises. En raison du nombre d'observations disponibles dans les enquêtes-TIC, la présente analyse se focalise sur l'impact de la fibre. On considère ici 10 usages numériques différents, incluant des pratiques organisationnelles (utilisation d'ordinateurs et d'internet, existence d'un site web, usage des médias sociaux) et plus avancées (cloud computing, PGI, GRC, outils collaboratifs) ainsi que le commerce en ligne. Comme nous disposons de données en coupe répétée, nous employons une méthode d'appariement par score de propension pour identifier l'effet de la fibre. Il convient donc de garder en mémoire que les estimations reposent sur l'hypothèse de sélection sur les observables.

Nos résultats indiquent que le déploiement de la fibre dans le cadre du Plan FTTHD encourage les entreprises situées dans les espaces peu denses à utiliser davantage certains outils numériques. De manière intéressante, la fibre stimule l'usage d'outils organisationnels (utilisation d'internet, d'ordinateurs, site web) mais aussi sophistiqués (cloud computing, PGI, GRC), ces derniers étant encore relativement peu répandus dans les espaces peu denses. En revanche, la couverture en fibre n'est pas associée à une probabilité plus élevée d'usages des médias sociaux, des outils collaboratifs et du commerce en ligne.

Au-delà de l'impact moyen du Plan FTTHD, nos résultats indiquent que la fibre a des effets très hétérogènes en fonction des caractéristiques internes des entreprises, de leur environnement externe, et des modalités du déploiement. Ainsi, le déploiement de RIP accroît essentiellement les pratiques numériques des petites entreprises, dont les ressources financières limitées ne permettent pas de financer leur propre accès à internet lorsqu'elles sont situées dans des communes non desservies. Le déploiement de la fibre est également essentiellement bénéfique aux entreprises multi-établissements, qui ont davantage intérêt à adopter des outils numériques en raison de leurs coûts de coordination supérieurs.

Par ailleurs, l'environnement externe conditionne également fortement les retombées du déploiement. En particulier, la fibre bénéficie principalement aux entreprises situées dans les zones péri-urbaines, qui ont à la fois des incitations fortes à utiliser les outils numériques et la capacité suffisante de se les approprier. Les retombées sont en revanche quasi-inexistantes pour les entreprises situées dans les espaces ruraux. Par ailleurs, bien que certains travaux concluent que l'éloignement des fournisseurs d'équipements et de services informatiques risquent de

limiter l'impact des politiques de déploiement dans les espaces peu denses, nos résultats n'indiquent pas que ce mécanisme réduit l'efficacité du Plan FTTHD.

Enfin, les retombées du Plan ne sont pas immédiates, avec des effets significatifs qui apparaissent trois ans environ après l'arrivée de la fibre dans la commune. En revanche, au-delà d'un certain seuil de locaux couverts (estimés ici à 32% en moyenne), le déploiement additionnel de fibre optique ne semble pas stimuler davantage l'usage d'outils numériques par les entreprises.

Ce résultat invite alors à s'interroger sur les modalités du déploiement du Plan France Très Haut Débit. Premièrement, est-ce qu'une couverture systématique en fibre optique est réellement nécessaire (option choisie par certains RIP) ? Deuxièmement, même dans le cas où l'on retiendrait l'idée d'une couverture totale à long terme, est-ce qu'un déploiement plus rapide et plus ciblé (ex : zones d'activité au sein d'une commune) dans un premier temps ne serait pas préférable ?

Concernant la première question, si l'usage numérique des entreprises évolue peu une fois qu'un niveau minimal de locaux est couvert, parvenir à un taux de couverture très élevé semble toutefois primordial pour les ménages. Dit autrement, il est donc difficile de conclure sur la pertinence d'une couverture systématique à partir de cette étude qui est focalisée sur les usages numériques des entreprises.

Concernant la deuxième question, nos résultats suggèrent en effet qu'un déploiement plus ciblé et plus rapide permettrait aux entreprises de s'approprier plus rapidement les outils numériques. Si ce type de déploiement apparaît donc plus efficace d'un point de vue économique, se pose toutefois la question de son acceptabilité sociale, en particulier au sein de communes peu denses où un tel déploiement conduirait à des écarts de débit importants.

Annexes au chapitre 4

Tableau A.4.1 : Définition des variables d'appariement

Variable	Définition	Source
<i>Variables communales</i>		
Population	Nombre d'habitants (1000 hab.)	Recensement de la population (2013)
Stock établissements	Stock d'établissements	Répertoire des entreprises et des établissements (2013), Insee
Prox. ville moyenne	Variable muette égale à 1 si l'entreprise est implantée dans une commune située à moins de 20 minutes d'une ville de 20 000 habitants, 0 sinon	Recensement de la population (2013), Insee. Calculs réalisés sur Odomatrix*
Montagne	Variable muette égale à 1 si la commune est située en zone de montagne ou de moyenne montagne, 0 sinon	Cemagref
Mauvais accès THD	Variable muette égale à 1 si moins de 25% des locaux ont accès au THD en 2013, 0 sinon	Observatoire du THD, Agence du Numérique
<i>Variables au niveau entreprise</i>		
Taille		
- 0-9 salariés	Variable muette égale à 1 si l'entreprise a entre 0 et 9 salariés, 0 sinon	Répertoire Sirene (2014 ; 2019)
- 10-49 salariés	Variable muette égale à 1 si l'entreprise a entre 10 et 49 salariés, 0 sinon	Répertoire Sirene (2014 ; 2019)
- 50-249 salariés	Variable muette égale à 1 si l'entreprise a entre 50 et 249 salariés, 0 sinon	Répertoire Sirene (2014 ; 2019)
- >250 salariés	Variable muette égale à 1 si l'entreprise a plus de 250 salariés, 0 sinon	Répertoire Sirene (2014 ; 2019)
Mono-établissement	Variable muette égale à 1 si l'entreprise a un seul établissement, 0 sinon	Répertoire Sirene (2014 ; 2019)
Age		
- 0-5 ans	Variable muette égale à 1 si l'entreprise a été créée il y a moins de 5 ans, 0 sinon	Répertoire Sirene (2014 ; 2019)
- 6-9 ans	Variable muette égale à 1 si l'entreprise a été créée il y a entre 6 et 9 ans, 0 sinon	Répertoire Sirene (2014 ; 2019)

-	>10 ans	Variable muette égale à 1 si l'entreprise a été créée il y a plus de 10 ans, 0 sinon	Répertoire Sirene (2014 ; 2019)
Secteur			
-	Manufacture	Variable muette égale à 1 si l'entreprise appartient au secteur manufacturier**, 0 sinon	Répertoire Sirene (2014 ; 2019)
-	Construction	Variable muette égale à 1 si l'entreprise appartient au secteur de la construction**, 0 sinon	Répertoire Sirene (2014 ; 2019)
-	Comm., transp., héb.	Variable muette égale à 1 si l'entreprise appartient au secteur du commerce, transport ou hébergement**, 0 sinon	Répertoire Sirene (2014 ; 2019)
-	Tertiaire Sup.	Variable muette égale à 1 si l'entreprise appartient au tertiaire supérieur**, 0 sinon	Répertoire Sirene (2014 ; 2019)
-	Autres services	Variable muette égale à 1 si l'entreprise appartient au secteur des autres services, ** 0 sinon	Répertoire Sirene (2014 ; 2019)

Note : *Odomatrix, INRA UMR1041 CESEAR, d'après IGN Route 500®, BD ALTI 500®, RGC®

** Pour définir les secteurs, nous utilisons ici la nomenclature de synthèse agrégée pour l'activité en NAF rév.2.

Nous retenons ici le niveau d'agrégation en 10 ou 38 postes selon les secteurs.

- Secteur manufacturier : « Industrie manufacturière, industries extractives et autres » (A10=BE),
- Construction : « Construction » (A10=FZ),
- Commerce-transport-hébergement : « Commerce de gros et de détail, transports, hébergement et restauration » (A10=GI),
- Tertiaire supérieur : « Information et communication » (A10 = JZ), « Activités financières et d'assurance » (A10 = KZ), « Activités immobilières » (A10 = LZ), « Activités juridiques, comptables, de gestion, d'architecture, d'ingénierie, de contrôle et d'analyses techniques » (A38 = MA), « Recherche-développement scientifique » (A38 = MB), « Autres activités spécialisées, scientifiques et techniques » (A38 = MC),
- Autres services : « Administration publique, enseignement, santé humaine et action sociale » (A10 = OQ), « Autres activités de services » (A10 = RU), « Activités de services administratifs et de soutien » (A38 =NZ).

Tableau A.4.2 : Variables pour tester l'hétérogénéité de l'impact du programme

Variable	Définition	Source
<i>Variables au niveau entreprise</i>		
Taille		
- < 50 salariés	Variable muette égale à 1 si l'entreprise a moins de 50 salariés, 0 sinon	Répertoire Sirene (2014 ; 2019)
- 50-249 salariés	Variable muette égale à 1 si l'entreprise a entre 50 et 249 salariés, 0 sinon	Répertoire Sirene (2014 ; 2019)
- >250 salariés	Variable muette égale à 1 si l'entreprise a plus de 250 salariés, 0 sinon	Répertoire Sirene (2014 ; 2019)
Structure organisationnelle		
- Multi-établissement	Variable muette égale à 1 si l'entreprise a un plusieurs établissement, 0 sinon	Répertoire Sirene (2014 ; 2019)
- Mono-établissement	Variable muette égale à 1 si l'entreprise a un seul établissement, 0 sinon	Répertoire Sirene (2014 ; 2019)
Age		
- 0-5 ans	Variable muette égale à 1 si l'entreprise a été créée il y a moins de 5 ans, 0 sinon	Répertoire Sirene (2014 ; 2019)
- 6-9 ans	Variable muette égale à 1 si l'entreprise a été créée il y entre 6 et 9 ans, 0 sinon	Répertoire Sirene (2014 ; 2019)
- >10 ans	Variable muette égale à 1 si l'entreprise a été créée il y a plus de 10 ans, 0 sinon	Répertoire Sirene (2014 ; 2019)
Secteur		
- Manufacture-Const.	Variable muette égale à 1 si l'entreprise appartient au secteur manufacturier ou de la construction, 0 sinon	Répertoire Sirene (2014 ; 2019)
- Comm., transp., héb.	Variable muette égale à 1 si l'entreprise appartient au secteur du commerce, transport ou hébergement, 0 sinon	Répertoire Sirene (2014 ; 2019)
- Services	Variable muette égale à 1 si l'entreprise appartient au secteur tertiaire, 0 sinon	Répertoire Sirene (2014 ; 2019)
<i>Environnement externe de l'entreprise</i>		
Degré de ruralité		
- Grands poles urbains	Variable muette égale à 1 si l'entreprise est située dans un grand pôle (catégorie 111 du zonage en aires urbaines), 0 sinon	Zonage en aires urbaines, Insee (2010)

- Périurbain	Variable muette égale à 1 si l'entreprise est située dans une commune appartenant à la couronne d'un grand pôle ou dans une commune multipolarisée des grandes aires urbaines (catégories 112 et 120 du zonage en aires urbaines), 0 sinon	Zonage en aires urbaines, Insee (2010)
- Petits et moyens pôles	Variable muette égale à 1 si l'entreprise est située dans un petit ou moyen pôle (catégories 211 et 221 du zonage en aires urbaines), 0 sinon	Zonage en aires urbaines, Insee (2010)
- Rural	Variable muette égale à 1 si l'entreprise est située dans la couronne d'un petit ou moyen pôle, une autre commune multipolarisée ou une commune isolée (catégories 212, 222, 300 ou 400 du zonage en aires urbaines), 0 sinon	Zonage en aires urbaines, Insee (2010)
Présence d'équipements et de services informatiques		
- Présence	Variable muette égale à 1 si l'entreprise est située dans un bassin de vie où il existe au moins un établissement spécialisé dans les équipements et/ou services informatiques*, 0 sinon	Répertoire Sirene (2014 ; 2019)
- Absence	Variable muette égale à 1 si l'entreprise est située dans un bassin de vie où il existe au moins un établissement spécialisé dans les équipements et/ou services informatiques*, 0 sinon	Répertoire Sirene (2014 ; 2019)

Variables caractérisant les modalités du déploiement

Effet selon l'intensité du déploiement

≤ Q1	Variable de traitement égale à 1 si l'entreprise est située dans une commune dont le taux de couverture est inférieur au premier quartile, 0 sinon	Observatoire du THD, Agence du Numérique (2013-2018)
> Q1 & ≤ Q2	Variable de traitement égale à 1 si l'entreprise est située dans une commune dont le taux de couverture est supérieur au premier quartile et inférieur au	Observatoire du THD, Agence du Numérique (2013-2018)

	taux de couverture médian, 0 sinon	
> Q2 & ≤ Q3	Variable de traitement égale à 1 si l'entreprise est située dans une commune dont le taux de couverture est supérieur au taux de couverture médian et inférieur au troisième quartile, 0 sinon	Observatoire du THD, Agence du Numérique (2013-2018)
> Q3	Variable de traitement égale à 1 si l'entreprise est située dans une commune dont le taux de couverture est supérieur au troisième quartile, 0 sinon	Observatoire du THD, Agence du Numérique (2013-2018)
Effet selon le temps écoulé depuis l'arrivée de la fibre		
Trois ans ou plus	Variable de traitement égale à 1 si l'entreprise est située dans une commune où la fibre a été déployée depuis trois ans ou plus, 0 sinon	Observatoire du THD, Agence du Numérique (2013-2018)
Deux ans	Variable de traitement égale à 1 si l'entreprise est située dans une commune où la fibre a été déployée depuis deux ans, 0 sinon	Observatoire du THD, Agence du Numérique (2013-2018)
Un an	Variable de traitement égale à 1 si l'entreprise est située dans une commune où la fibre a été déployée depuis un an, 0 sinon	Observatoire du THD, Agence du Numérique (2013-2018)

* Le secteur de l'équipement et des services informatiques est défini en se basant sur la nomenclature de synthèse agrégée pour l'activité en NAF rév.2 et inclut les classes suivantes :

- 4651Z Commerce de gros d'ordinateurs, d'équipements informatiques périphériques et de logiciels
- 4652Z Commerce de gros de composants et d'équipements électroniques et de télécommunication
- 4741Z Commerce de détail d'ordinateurs, d'unités périphériques et de logiciels en magasin spécialisé
- 4742Z Commerce de détail de matériels de télécommunication en magasin spécialisé
- 6201Z Programmation informatique
- 6202A Conseil en systèmes et logiciels informatiques
- 6202B Tierce maintenance de systèmes et d'applications informatiques
- 6203Z Gestion d'installations informatiques
- 6209Z Autres activités informatiques
- 7733Z Location et location-bail de machines de bureau et de matériel informatique 9511Z Réparation d'ordinateurs et d'équipements périphériques
- 9512Z Réparation d'équipements de communication

Chapitre 5. Performance des entreprises et emploi

1. Revue de littérature

1.1. De la difficile comparaison d'études hétérogènes

Les recherches sur l'effet du très haut débit (THD) sur l'activité économique sont difficilement comparables en raison de différences de méthodologie portant notamment sur l'unité d'analyse (entreprise ou zone géographique), la mesure de la performance (variable expliquée) et la mesure du THD⁶⁸ (variable de traitement).

Unité d'analyse

Les études nationales évaluant l'effet du THD sur l'activité économique recouvrent deux grands types de travaux.⁶⁹ D'une part, de nombreuses études ont analysé l'effet de l'arrivée d'internet sur le développement économique local au niveau méso-économique (de la commune à la région) . D'autre part, des études ont analysé l'effet du THD sur les performances des entreprises. Or, ces deux niveaux d'analyses induisent des résultats parfois différents comme cela est discuté ci-dessous.

Variables expliquées

La variable la plus fréquemment utilisée dans les différentes études porte sur l'évolution de l'emploi. Les travaux distinguent dans la mesure du possible les effets selon le niveau de qualification des travailleurs.

Les études portant sur le développement des entreprises analysent en outre l'effet sur la productivité des entreprises. Néanmoins, d'autres mesures sont parfois incluses pour mesurer la performance des entreprises (chiffre d'affaire, valeur ajoutée, profitabilité). Certaines études sur le développement régional portent sur des mesures proches comme le salaire médian (Kolko, 2012 ; Whithacre et al., 2014 ; Briglauer et al., 2019) ou le nombre d'entreprises actives (Mack et Rey, 2014 ; Hasbi, 2020 ; Canzian et al., 2019).

⁶⁸ Pour simplifier, il est fait référence au très haut débit dans la suite du texte même si la revue inclut des mesures allant du bon haut débit (HD) à la fibre. Ces différentes technologies / débits ne sont distingués que lorsque la question du débit importe. Pour les autres points de réflexion, il est fait référence au THD de manière générique.

⁶⁹ Nous ignorons les travaux macroéconomiques qui comparent différents pays.

Mesure du très haut débit : Quel débit ? Adoption versus accès ?

Au-delà de la variable expliquée, la définition du THD est rarement harmonisée. Il est parfois difficile, sinon impossible, de savoir le débit auquel il est fait référence pour identifier le (très) haut débit. Seule la technologie est connue et parfois le débit maximum. Le terme « haut débit » peut donc recouvrir des réalités très différentes. Par exemple, Akerman et al. (2015) définissent le haut débit par un débit supérieur à 256 Kbps alors que certaines études font référence à des débits supérieurs à 20 Mbps (Ford, 2018 ; Canzian et al., 2019) voire allant jusqu'à 100 Mbps ou plus (Bai, 2017). Cette distinction recouvre en partie la forte croissance du débit au cours des dernières années. Néanmoins, elle ne facilite pas la comparabilité des études.

En outre, les recherches se concentrent en général sur le débit final (après traitement) mais considèrent rarement le débit initial (avant traitement). Or, les effets d'une montée en débit sont conditionnels à ce point de départ. Afin de mieux comprendre l'effet réel de l'accès au THD et à la fibre, il est donc important de connaître le débit disponible par les entreprises avant la mise en place de la politique évaluée (absence d'internet, faible débit, haut débit, THD).

Un autre point méthodologique important est la distinction entre adoption et accès. Certaines études se concentrent sur l'adoption, le plus souvent à partir de données d'enquêtes sur les entreprises. L'avantage de cette mesure est qu'elle permet de comparer les entreprises ayant adopté le THD et celles qui n'ont pas choisi d'y recourir, même si elles y ont potentiellement accès. En raison d'un problème évident d'endogénéité dans le choix d'adoption, les articles recourent toutefois le plus souvent à des stratégies de variables instrumentales dont les instruments sont l'accès au THD. L'intuition est que l'absence d'accès est un facteur central de la non adoption, ce qui peut être questionnable (Townsend et al., 2013). Une autre hypothèse est que les conditions d'accès sont exogènes à la localisation de l'entreprise (absence de « sorting »). Une nouvelle fois, cette hypothèse d'exogénéité est questionnable car les entreprises servies par le THD ont sans doute accès à d'autres services dont ne sont pas pourvues les entreprises éloignées des centres économiques (infrastructures de transport par exemple).

En raison de la disponibilité des données, la plupart des travaux utilisent toutefois une mesure de l'accès à internet, et non de son adoption (Duvivier, 2019). Par construction, la plupart des études à un niveau méso-économique (commune, département, région) utilisent des mesures d'accès faute d'information précise sur les entreprises. La limite de cette approche est qu'elle considère que toutes les entreprises disposant d'un accès à internet l'adoptent ce qui, à nouveau, est loin d'être le cas, en particulier dans les espaces ruraux. Techniquement, ces

recherches estiment donc une intention d'être traité (« Intention-to-treat »). Cette approche a néanmoins un double avantage. Tout d'abord, il est possible d'inclure de nombreuses entreprises dans l'étude puisque l'information sur l'accès n'est pas contrainte par l'existence d'enquêtes sur l'usage numérique des entreprises. Ensuite, même les entreprises n'ayant pas adopté le THD peuvent bénéficier de l'accès grâce à des effets de débordement (« spillover effects »). Ces effets peuvent être positifs (effet d'agglomération) ou négatifs (concurrence). L'approche par l'accès permet donc d'avoir une vision plus générale de l'effet net du haut débit sur les entreprises dans une région desservie.

Dans la suite, nous reprenons les principaux travaux en organisant notre revue selon la variable de performance considérée.

1.2. Effet sur l'emploi

La plupart des études portant sur l'emploi ne se situe pas au niveau des effectifs des entreprises mais exploite l'évolution de l'emploi au niveau local (municipalités, départements, régions). Les travaux sur le haut débit montrent le plus souvent un effet positif sur l'emploi mais qui est très inégalitaire (Bertschek et al., 2015 ; France Stratégie, 2020). Seuls les plus qualifiés semblent réellement bénéficier du haut débit, conformément à l'hypothèse d'un biais technologique favorable aux plus qualifiés (Akerman et al., 2015; Michaels et al., 2013). Les moins qualifiés souffrent quant à eux, en règle générale, de l'arrivée du haut débit. Les travaux plus récents sur le très haut débit semblent néanmoins conclure à l'absence d'effet sur l'emploi au niveau local (Bai, 2017 [Etats-Unis]; Canzian et al., 2019 [Italie]; Ford, 2018 [Etats-Unis]). Dans le cas de la France, Hasbi (2020) met en évidence un effet de la fibre sur la réduction du taux de chômage. Néanmoins, il convient de noter que ce travail exploite des données anciennes (2010-15) eu égard aux récents développement de cette technologie en France.

Les évaluations du très haut débit sur les effectifs au niveau des entreprises peinent à confirmer ces premiers résultats.⁷⁰ Fabling et Grimes (2016) concluent que le THD a peu d'effet sur les effectifs (sauf pour les entreprises ayant investi au préalable pour accueillir cette technologie). Cette absence d'effet est confirmée dans le cas de l'Italie par Canzian et al. (2019). Néanmoins, ces auteurs montrent un effet positif sur d'autres variables d'activités comme la création d'activité et de valeur ajoutée. L'absence d'effet sur les effectifs peut s'expliquer par

⁷⁰ Si à un niveau agrégé l'évolution de l'emploi est essentielle pour les décideurs politiques, cette mesure est plus ambiguë pour approcher la performance des entreprises individuelles car l'effectif est autant une mesure de l'output (hausse de l'activité) que des facteurs de productions.

un délai temporel entre la hausse d'activités et les embauches. Ciapanna et Colonna (2019) qui étudient le même programme que Canzian et al. (2019) montrent quant à eux un effet différencié du THD sur l'emploi en fonction du degré de qualification. Seuls les plus qualifiés tirent un bénéfice à l'arrivée du très haut débit. A l'opposé, les moins qualifiés pâtissent de son arrivée. De plus, De Stefano et al. (2020) étudient l'adoption du *cloud computing* au Royaume-Uni et mettent en évidence un effet différencié selon le type d'entreprises. Si le *cloud computing* a un effet positif sur la croissance des effectifs et de l'activité (revenus), cet impact est essentiellement valable pour les entreprises récemment créées.

1.3. Effet sur les performances économiques régionales et individuelles

Si de nombreuses études observent un effet positif du très haut débit sur l'activité économique locale (Kandilov et Renkow, 2010 ; Mack et Rey, 2014 ; Hasbi, 2020 ; Ivus et Boland, 2015 ; Ford, 2018), d'autres travaux ne trouvent pas d'effet significatif (Kolko, 2012 ; Whitacre et al., 2014).

Ces résultats positifs au niveau méso-économique sont néanmoins difficilement comparables avec les travaux microéconomiques. La plupart des analyses de l'effet du haut débit à partir de données d'entreprises individuelles porte sur la productivité. Si Grimes et al. (2012) montrent un effet bénéfique de l'adoption du haut débit sur la productivité des entreprises en Nouvelle Zélande, les études ultérieures, ne parviennent pas à mettre en évidence un effet positif (Bertschek et al., 2013 [Allemagne]; Colombo et al., 2013 [Italie]; De Stefano et al., 2018 [Royaume-Uni]; Haller et Lyons, 2019 [Irlande]).

Il existe peu de travaux s'intéressant à l'accès au THD sur la productivité. Dans une étude récente sur la Nouvelle-Zélande, Fabling et Grimes (2016) concluent à un effet non significatif de l'adoption du THD sur la productivité. Ce résultat tend à souligner que l'effet ne semble pas conditionnel au pays considéré (Nouvelle Zélande). Une étude de Canzian et al. (2019) dans une région rurale d'Italie (Trentin) montre un effet positif d'un programme d'accès au THD (technologie ADSL+2 allant jusqu'à 20Mbps) sur la productivité des sociétés anonymes. De manière intéressante, ce travail met en évidence un effet décalé dans le temps (effet qui apparaît après un an). En outre, il s'agit d'une des rares études à considérer le niveau de débit initial (2 Mbps ou 7 Mbps). Les résultats montrent que le débit initial ne joue pas sur l'effet positif observé. Si cette étude est intéressante pour nous de par sa proximité avec le plan FTTHD, elle reste très particulière en raison de l'échantillon choisi (sociétés anonymes). Elle tend néanmoins à nuancer les effets nuls observés jusqu'alors pour le HD et le THD.

1.4. Questions en suspens

La littérature sur l'effet de l'internet haut débit sur les performances économiques met en évidence une scission entre les études selon l'unité d'analyse. Les études à un niveau agrégé (commune, région) mettent plus souvent en évidence un effet positif que les études microéconomiques (entreprise). Néanmoins, les travaux à la fois microéconomiques et méso-économiques peinent à mettre en évidence un effet clairement positif THD et la fibre. L'effet d'internet serait plus important pour des débits moindres. Néanmoins, à de rares exceptions (Canzian et al., 2019), la situation de départ en termes de débit n'est pas connue. Cette absence d'information suppose que toutes les entreprises bénéficient de manière égale de la montée en débit. Or, passer d'un faible débit, voire d'une absence d'internet, à du THD n'a probablement pas le même effet que passer d'un bon HD à du THD (même si Canzian et al., 2019, ne mettent pas en évidence un tel effet). Cela peut sans doute expliquer pourquoi les études les plus récentes portant sur des niveaux de débit plus élevés peinent à trouver des effets significatifs, même à un niveau agrégé.

Ensuite, l'effet moyen cache le plus souvent une hétérogénéité importante entre les entreprises. Si l'effet moyen du HD ou du THD est faible ou nul, cela ne signifie cependant pas qu'il est non significatif pour l'ensemble des entreprises. Plusieurs travaux ont cherché à identifier les conditions permettant de tirer pleinement profit de l'arrivée du haut débit. Akerman et al. (2015) ont montré que le haut débit avait des effets différenciés sur les travailleurs en fonction de leur niveau de qualification, confirmant des études antérieures sur l'hypothèse de « skill-biased technology » des technologies de l'information et de la communication (Michaels et al., 2013). Si cette technologie semble être un complément pour les plus qualifiés (hausse de l'emploi et de la productivité), elle agit en revanche comme un substitut pour les moins qualifiés. D'autres travaux sur la productivité ont tenté de prendre en compte le rôle de la qualification de la main d'œuvre sans parvenir à des conclusions claires. Dans une étude portant sur l'Italie, Ciapanna et Colonna (2019) confirment les résultats d'Akerman et al. (2015) en montrant que le haut débit améliore la productivité des plus qualifiés, et donc des entreprises ayant recours de manière intensive à ces travailleurs. En revanche, Grimes et al. (2012) ne trouvent pas de différence d'effet entre les entreprises en fonction de l'intensité en connaissance du secteur dans lequel elles opèrent.

Par ailleurs, l'organisation interne de l'entreprise semble également conditionner l'effet du THD sur les entreprises. Il existe une littérature assez riche sur les complémentarités entre les

investissements en nouvelles technologies et l'organisation interne des entreprises (Bloom et al., 2012; Draca et al., 2009). Ainsi, plusieurs études (Colombo et al., 2013 [Italie]; De Stefano et al., 2018 [Royaume-Uni]; Grimes et al., 2012 [Nouvelle-Zélande]) ont montré que les entreprises qui ont préparé l'arrivée du très haut débit en ont tiré un avantage.⁷¹ La structure interne de l'entreprise a souvent été négligée. En particulier, à notre connaissance, aucun travail n'a distingué les entreprises à établissement unique et les multi-établissements. Or, il est probable que ces deux groupes d'entreprises ne tirent pas le même profit de l'entrée du THD ou de la fibre. Dans le cas d'une entreprise mono-établissement, internet est surtout un moyen pour communiquer avec les partenaires externes de l'entreprise (fournisseurs, clients, etc.). Pour les entreprises multi-établissements, une meilleure connexion entre les différents établissements peut améliorer les performances internes de l'entreprise dans son ensemble en fluidifiant les échanges d'information. Autrement dit, les entreprises multi-établissements peuvent tirer un double bénéfice de l'entrée du THD/fibre en améliorant l'échange de données externe et interne.

Enfin, la plupart des travaux sur la performance négligent l'environnement dans lequel opère l'entreprise. En particulier, l'impact positif du déploiement du numérique est potentiellement conditionnel à l'existence d'autres aménités (infrastructures de transport par exemple). Peu de travaux ont étudié si l'effet du THD est conditionnel à l'environnement dans lequel opère l'entreprise. Canzian et al. (2019) montrent que le THD bénéficie aux (grandes) entreprises en zone rurale. Grimes et al. (2012) n'ont pas relevé de différence dans l'effet positif du HD selon la localisation de l'entreprise. Néanmoins, les études restent muettes sur les conditions précises favorisant les retombées du THD. Sur ce point, les travaux examinant l'impact du THD sur la création d'entreprises fournissent plus de réponses. Selon Mack (2015) et Duvivier et al. (2021), l'arrivée d'internet aurait un impact particulièrement bénéfique dans les « espaces intermédiaires » (i.e. petits et moyens pôles). En effet, ces espaces ont à la fois beaucoup à gagner de l'arrivée d'internet⁷² et, en même temps, offrent un contexte économique relativement favorable aux entreprises (niveau minimum d'activité et de main d'œuvre, etc.). D'autres travaux indiquent également que les bénéfices d'internet sont supérieurs dans les espaces ruraux dont le contexte économique est favorable avant l'arrivée d'internet, c'est-à-dire

⁷¹ Ces investissements initiaux recouvrent aussi bien l'embauche d'un personnel qualifié, que la formation du personnel existant ou l'adaptation du management interne (processus de décision, mécanismes d'incitation). D'autres caractéristiques sont parfois évoquées comme le secteur ou l'âge de l'entreprise (par exemple, De Stefano et al., 2018).

⁷² Les économies d'agglomération y étant limitées, internet peut réduire les coûts d'information et de coordination.

dans les espaces bien dotés en aménités naturelles, proche d'un pôle urbain, ou disposant d'un niveau minimum d'activités (Cumming et Johan, 2010; Duvivier et al., 2018 ; Kandilov et Renkow, 2010; Kim et Orazem, 2017; Mack, 2014-a; 2015).

2. Méthodologie

Nous présentons dans un premier temps les données et l'échantillon retenu pour l'analyse. Nous décrivons ensuite les variables et enfin nous présentons l'approche économétrique retenue. L'analyse principale se concentre sur une évaluation du plan FTHD au niveau communal. Nous discutons des résultats au niveau microéconomique en Annexe.

2.1. Données et échantillon

Ce chapitre évalue l'accès de l'internet haut débit (THD et fibre) sur les performances des entreprises françaises opérant dans des zones d'initiative publique. Les données financières sur les entreprises sont extraites du Fichier Approché des Résultats d'ESANE (FARE). Ces données contiennent des informations comptables pour l'ensemble des entreprises françaises. Les données reprennent les principaux éléments des bilans et comptes de résultats des entreprises. Ces informations sont combinées avec des données extraites du répertoire SIRENE. Ce dernier permet d'obtenir des informations générales sur les entreprises (fichier SIREN) mais également des informations sur chacun des établissements de l'entreprise (fichier SIRET).

Les données sur les entreprises sont combinées à la couverture internet par communes. Le niveau de couverture est fourni par l'Agence du numérique (2013-2018) et par l'Arcep (2019)⁷³ pour quatre niveaux de débit : 3, 8, 30 et 100 Mbps. Seuls les deux derniers niveaux de couverture nous intéressent dans ce rapport. Le chapitre 2 détaille ces données, ainsi que les évolutions récentes dans la couverture internet en France.

⁷³ La combinaison de ces deux sources de données pose certains problèmes. D'une part, les données de l'Arcep, issues de « Ma connexion internet », sont à considérer avec précaution puisqu'elles posent certains problèmes de fiabilité notamment pour les espaces peu denses (ex : problème concernant le référentiel des locaux employés). D'autre part, il existe des problèmes de comparabilité entre les données de couverture de l'Agence du Numérique et de l'Arcep. En croisant les données de l'Agence pour 2018 et celles de l'Arcep pour 2019, on observe en effet que certaines communes qui avaient accès à internet en 2018 n'y ont plus accès en 2019. Cela concerne 0.6% des communes pour la Fibre et 2.6% des communes pour le THD. Le taux d'erreur supérieur semble toutefois limité, et donc, nous choisissons de combiner ici les deux sources de données afin de tenir compte des déploiements plus récents, et plus nombreux, ayant eu lieu en 2019.

Les données de connexion sont fournies au niveau municipal, la fusion des deux bases est directe à partir des codes INSEE de chaque commune.

Dans la mesure où notre objectif est d'évaluer le programme France Très Haut Débit, nous ne retenons que les communes et entreprises localisées dans une commune de zone d'initiative publique soit un peu plus de 30 000 communes (voir tableau 5.2 ci-dessous).

En outre, les données de couverture étant absentes avant 2013, nous excluons les communes et entreprises localisées dans les communes qui auraient bénéficié du THD ou de la fibre avant cette date (« *déjà traitées en 2013* »). Comme l'illustrent le tableau 5.2⁷⁴, cela concerne un nombre limité de communes et d'entreprises pour la fibre (moins de 1% des communes, 2,5% des établissements) mais un nombre beaucoup plus important pour le THD qui est déjà bien diffusé en zone d'initiative publique en 2013 (15% des communes, la moitié des établissements).

2.2. Variables

Variables expliquées

Pour chaque commune, nous avons construit trois indicateurs d'activités : valeur ajoutée produite, emploi et productivité. Il convient de noter que ces mesures s'appuient uniquement sur des données du secteur privé marchand. Ainsi, la valeur ajoutée et l'emploi au niveau communal ignore la création de richesse produite par les activités non marchandes et l'emploi se réfère uniquement à l'emploi privé.

La valeur ajoutée est une approximation du PIB communal issu du secteur privé productif qui permet d'analyser la création de richesse au niveau municipal. L'effectif mesure le nombre d'emplois créés par le secteur privé dans la commune. Ces deux mesures sont au cœur des préoccupations des décideurs politiques locaux. La productivité moyenne permet de mesurer la richesse créée par travailleur. La mesure de productivité, si elle est souvent étudiée au niveau microéconomique (entreprise) ou macroéconomique (pays), l'est moins au niveau méso-économique (communes, régions). Or, à l'image des études portant sur les différences de productivité entre pays, elle permet en partie de renseigner sur la qualité des emplois créés en termes de rémunérations et donc *in fine* de niveaux de vie. En outre, les différences de productivité entre communes sont relativement importantes et expliquent en partie les différences entre pays (Acemoglu et Dell, 2010).

⁷⁴ Le tableau 5.2 ne recense que les communes/entreprises appartenant à la zone d'initiative publique.

L'emploi par commune est obtenu à partir du fichier des établissements (SIRET) dans le répertoire SIRENE. Pour chaque commune, nous avons l'ensemble des établissements ainsi que le nombre d'employés par établissement. L'emploi total dans la commune i l'année t est donc la somme des emplois des établissements (e) opérant sur cette commune :

$$EMP_{it} = \sum_{e \in i} EFF_{et} \quad (5.1)$$

La construction de la valeur ajoutée produite par le secteur privé par commune (« PIB privé municipal ») est moins directe à obtenir. En effet, la valeur ajoutée n'est disponible qu'au niveau des entreprises (et non des établissements). Pour les entreprises mono-établissement, il existe une correspondance directe entre les deux. Cela n'est cependant pas le cas pour les entreprises multi-établissements. Une entreprise multi-établissements a de multiples établissements dans des communes parfois différentes. Il convient donc de construire une mesure de la valeur ajoutée par établissement avant de pouvoir la calculer par commune. Pour ce faire, nous avons opéré en trois temps.

Premièrement, nous avons calculé le poids relatif de chaque établissement par année ($\theta_{e,t}$) pour chaque entreprise (E) à partir des données d'emplois par établissement^{75,76}:

$$\theta_{e,t} = \frac{EFF_{e,t}}{\sum_{k \in E,t} EFF_{k,t}} \quad (5.2)$$

Avec e l'établissement concerné et k l'ensemble des établissements de l'entreprise E . Deuxièmement, nous avons décomposé la valeur ajoutée de l'entreprise selon le poids de chaque établissement :

$$VA_{e,t} = \theta_{e,t} * VA_{E,t} \quad (5.3)$$

Enfin, nous avons sommé les valeurs ajoutées de l'ensemble des établissements de la commune i afin de calculer la valeur ajoutée au niveau municipal :

$$VA_{i,t} = \sum_{e \in i} VA_{e,t} \quad (5.4)$$

A partir des deux variables de valeur ajoutée et d'effectifs, nous construisons la productivité apparente du travail au niveau communal comme le ratio de la valeur ajoutée sur les effectifs.

$$PT_{i,t} = VA_{i,t} / EMP_{i,t} \quad (5.5)$$

⁷⁵ Il convient de noter que par construction $\theta_{et} = 1$ pour les entreprises mono-établissement.

⁷⁶ Lorsque les données d'emplois sont nulles pour l'ensemble des établissements, nous imputons un poids identique à chacun.

Le tableau 5.1 présente les principales statistiques descriptives pour les trois variables retenues pour l'analyse. L'unité d'analyse est commune - année. Le tableau 5.1 ne présente que les statistiques pour les observations utilisées pour l'analyse.

En moyenne, les communes traitées génèrent une valeur ajoutée (privée) totale d'environ 82 430 euros. La valeur ajoutée moyenne par commune peut paraître faible mais il convient de noter que les communes concernées sont très petites (90% ont moins de 2000 habitants, plus de la moitié moins de 500 habitants) avec parfois un nombre restreint d'établissements. En outre, la valeur ajoutée calculée ne porte que sur les activités productives des entreprises, contrairement à la mesure usuelle du PIB. Enfin, il convient de souligner que la distribution de la valeur ajoutée est très hétérogène. La médiane est beaucoup plus faible que la moyenne, cette dernière étant tirée par les plus grosses communes.

Tableau 5.1. Statistiques descriptives

Variables	Obs.	Moyenne	Erreur standard	Distribution						
				Min	10%	25%	Médiane	75%	90%	Max
Valeur ajoutée (milliers euros)	301196	82.43	310.72	-5836.4	0.535	2.274	9.067	39.671	167.59	13793.7
Effectif (nb)	301196	144.04	500.9	0	0	3	15	73	316	19372
Productivité (milliers euros)	260662	0.743	1.649	-20.693	0.365	0.461	0.578	0.779	1.171	380.21
Population (nombre)	301196	905.5	1697.2	1	103	187	397	910	2053	82427
Indice de diversification	301196	15.89	127.7	0.004	0.189	0.356	0.746	2.013	100	45311.6

Note : Le tableau représente la distribution des variables dépendantes et de contrôle. L'unité d'analyse est commune-année. Seules les observations incluent dans l'analyse sont retenues (exclusion des communes en zone d'initiative privée). La valeur ajoutée et la productivité sont exprimées en milliers d'euros. Les effectifs et la population sont exprimés en nombre.

Un constat très similaire s'observe pour l'emploi. La moitié des communes génèrent moins de 15 emplois dans le secteur productif (voire moins de 3 emplois dans un quart des communes).

Une simple analyse des différences pour les variables dépendantes (en Annexe) met en évidence que les communes traitées ont des niveaux de valeur ajoutée et d'effectifs plus importants. Ce résultat n'est guère surprenant dans la mesure où le très haut débit et la fibre ont été déployés en priorité des communes relativement grandes. Enfin, cette différence est prise en compte dans notre analyse statistique qui se concentre sur l'évolution intra-communale suite à l'arrivée du très haut débit et de la fibre.

Variable de traitement

Nous distinguons les communes en fonction de trois catégories pour les deux classes de débit (THD et fibre). Les communes étant déjà traitées en 2013, les communes traitées entre 2014 et 2019 et les communes non traitées à la fin de la période d'observation. Une commune est considérée comme traitée si elle a accès à du THD (30 Mbps ou plus) ou de la fibre (100 Mbps ou plus). Nous créons une muette égale à 1 dès que la commune est raccordée (sans tenir compte du pourcentage de locaux couverts, ce qui est considéré dans une analyse additionnelle). Ne pouvant pas connaître la date précise du traitement pour les communes étant déjà raccordées au THD ou à la fibre avant 2013, nous excluons ces communes de l'analyse.

Le tableau 5.2 présente la distribution des communes et le nombre d'établissements opérants en zone d'initiative publique en fonction des cohortes étudiées. Concernant le THD, près de 50% des communes en zone d'initiative publique ont été traitées entre 2014 et 2019. Il restait en 2019 35% des communes en zone d'initiative publique qui n'avaient pas accès au THD. Néanmoins, comme l'illustre les dernières colonnes du tableau 5.1, les communes non encore traitées sont de petites communes en termes de population (11% des habitants des communes des zones d'initiative publique) et d'activité économique (moins de 10% des établissements).

Un nombre plus faible de communes ont bénéficié d'un accès à la fibre au cours de la période 2014-2019. En effet, uniquement 16% des communes ont été traitées au cours de la période d'observation. Ainsi, la majorité des communes (82%) des zones d'initiative publique n'ont pas d'accès à la fibre à la fin de notre période d'étude. Sans surprise, les communes connectées à la fibre ont une densité de population et d'établissements plus importante que les communes non traitées.

Tableau 5.2 : Distribution des cohortes

Très haut débit (30 Mbps ou plus)							
	Communes		Etablissements		Population		
	Nombre	%	Nombre	%	Nombre	%	
Déjà traitées en 2013	4 629	15.00	1 028 857	52.30	13 100 000	46.07	
Traitées entre 2014 et 2019	15 392	49.87	753 655	38.31	11 997 196	42.19	
-2014	2 346	7.60	238 938	12.15	3 500 867	12.31	
-2015	3 099	10.04	178 363	9.07	2 764 307	9.72	
-2016	3 245	10.51	114 838	5.84	1 879 030	6.61	
-2017	1 830	5.93	58 200	2.96	1 032 391	3.63	
-2018	333	1.08	11 700	0.59	233 833	0.82	
-2019	4 539	14.71	151 616	7.71	2 586 768	9.10	
Jamais traitées	10 844	35.13	184 599	9.38	3 339 319	11.74	

Fibre (100 Mbps ou plus)							
	Communes		Etablissements		Population		
	Nombre	%	Nombre	%	Nombre	%	
Déjà traitées en 2013	276	0.89	56 386	2.87	747 521	2.63	
Traitées entre 2014 et 2019	5 160	16.72	559 448	28.44	8 328 023	29.25	
-2014	96	0.31	22 120	1.12	327 453	1.15	
-2015	369	1.20	60 179	3.06	848 646	2.98	
-2016	435	1.41	55 203	2.81	830 496	2.92	
-2017	806	2.61	79 024	4.02	1 279 216	4.49	
-2018	442	1.43	30 621	1.56	536 202	1.88	
-2019	3 012	9.76	312 301	15.88	4 506 010	15.82	
Jamais traitées	25 429	82.39	1 351 286	68.69	19 400 000	68.13	

Note : Le tableau reporte le nombre de communes, d'établissements et la population totale en fonction de la cohorte : traitée avant 2013, traitée entre 2014 et 2018 et non-traitée en 2019. Seules les communes et entreprises appartenant à une zone d'initiative publique sont incluses.

Variable de contrôle

Nous considérons plusieurs variables de contrôle au niveau communal. Nous incluons la population totale de la commune dans la mesure où nos variables dépendantes (emplois et valeur ajoutée privée) et le timing du traitement sont influencés par la dynamique démographique. Nous considérons également la structure économique de la commune. Pour ce faire, nous mesurons le degré de diversité économique en adaptant l'indicateur construit pour des analyses internationales (Duranton et Puga, 2000 ; Renski, 2011). Cet indicateur mesure la différence absolue entre la part de l'emploi local dans chaque secteur et la part de l'emploi dans ce secteur au niveau national :

$$DIV_{it} = \left(\sum_j \left| \frac{EMP_{ijt}}{EMP_{it}} - \frac{EMP_{jt}}{EMP_{FRt}} \right| \right)^{-1} \quad (5.6)$$

Avec i la commune, j le secteur défini selon la classification NAF38 et t l'année. L'indice de diversification augmente avec le niveau de diversité économique de la commune.

Nous considérons enfin la spécialisation économique en construisant pour chaque secteur (NAF 38) le ratio de l'emploi dans ce secteur dans la commune i rapporté au ratio de l'emploi de ce secteur au niveau national :

$$SPEC_{ijt} = \frac{EMP_{ijt}/EMP_{it}}{EMP_{j,FRt}/EMP_{FRt}} \quad (5.7)$$

Si un secteur est surreprésenté dans la commune i l'année t alors $SPEC_{ijt} > 1$ et inversement si un secteur est sous-représenté dans la commune.

2.3. Méthode économétrique

Comme discuté dans le Chapitre 3, nous utilisons un modèle de différence-en-différence échelonnée (« *staggered difference-in-difference* ») comme suit :

$$y_{it} = \alpha_i + \beta_t + \sum_{d \neq -1} \gamma_d \cdot 1_{t-c_i=d} + \nabla X_{it} + \epsilon_{it} \quad (5.8)$$

avec les indicateurs i et t représentant l'unité d'analyse (commune⁷⁷) et l'année, respectivement.

Trois variables expliquées construites au niveau municipal sont considérées (y_{it}): la valeur ajoutée, l'emploi et la productivité du travail. La valeur ajoutée au coût des facteurs est extraite au niveau des entreprises directement des données FARE (variable r004).⁷⁸ Les effectifs au niveau de chaque établissement sont obtenus à partir du fichier SIRET du répertoire SIRENE. Afin d'obtenir l'effectif de chaque entreprise, utile pour les analyses d'hétérogénéité menées ensuite, nous construisons l'effectif total de l'entreprise considérée comme la somme des effectifs de chacun de ses établissements. Afin de s'assurer de la validité de cette approche, nous avons croisé les données FARE sur les effectifs⁷⁹ et SIRENE pour les données d'emploi

⁷⁷ En Annexe, nous analysons l'effet au niveau des entreprises, l'unité d'analyse est dès lors l'entreprise et non plus la commune. Nous expliquons l'adaptation du cadre général au niveau entreprise en Annexe.

⁷⁸ Les données de valeur ajoutée sont déflatées à partir des déflateurs fournis par l'INSEE par secteur (NAS 129). En leur absence, l'indice des prix à la consommation par produits est utilisé. Sinon l'indice des prix général est utilisé.

⁷⁹ Une approche alternative aurait été d'extraire les données des effectifs de chaque entreprise en exploitant les données du fichier FARE qui fournit deux mesures de l'effectif : effectif salarié en personne physique (variable redi_e001) et l'effectif salarié en équivalent temps plein (variable redi_e200). Or, ces deux variables ne sont pas renseignées pour l'ensemble des entreprises avant 2015 et absentes pour l'année 2017. Cette absence est préjudiciable pour l'analyse qui porte sur la période 2010-2019. La raison évoquée par l'INSEE tient au passage progressif de la Déclaration annuelle des données sociales (DASD) à la Déclaration sociale nominative (DSN).

disponibles dans les deux bases. Bien que l'INSEE recommande d'être prudent dans l'utilisation des données d'effectifs issues de SIRENE, les données sont corrélées à plus de 95% entre les deux bases nous assurant que les données SIRENE sont utilisables en l'état.

La mesure du traitement ($1_{t-c_i=d}$) est obtenue à partir des données de couverture issues de l'Agence du Numérique et de l'Arcep. Deux niveaux de débit sont considérés :

- Accès au très haut débit (30 Mbps et plus) ;
- Accès à la fibre (plus de 100 Mbps).

En première analyse, nous considérons qu'une commune est traitée si elle a un accès au THD ou à la fibre, quel que soit le niveau de couverture de la commune. Les paramètres estimés γ_d mesurent l'effet du traitement d périodes après l'avoir reçu, c_i correspondant à la date d'accès au traitement de la commune, c'est à dire sa cohorte (pour les unités de contrôle, $c_i = +\infty$). Les coefficients γ_d représentent l'évolution de l'effet du traitement. Si l'effet du traitement est dynamique, les coefficients devraient s'éloigner de 0 graduellement pour $d > 0$.

Le modèle est augmenté avec l'inclusion de variable de contrôle au niveau de l'unité d'analyse variant dans le temps (X_{it}). Afin de contrôler pour les variables inobservables au niveau de l'unité d'analyse, nous incluons des effets fixes individuels (α_i). Nous incluons également des effets fixes annuels (β_t) afin de prendre en compte la conjoncture macroéconomique pouvant influencer les performances économiques.

Il existe une littérature assez fournie sur l'estimation des modèles de différence-en-différence échelonnée en particulier lorsque le traitement est hétérogène et dynamique dans le temps (Baker et al., 2022 ; Chapitre 3 de ce rapport). En particulier, les estimateurs simples de double effets-fixes produisent des estimations biaisées. Nous adoptons par conséquent la méthode de Sun et Abraham (2020) qui permet de contourner les problèmes exposés dans cette littérature. La méthode de Sun et Abraham (2020) permet également de tester l'hypothèse de tendances parallèles avant le traitement. En particulier on s'attend à ce que ces coefficients soient non statistiquement différents de 0 pour $d < 0$, c'est à dire pour les périodes antérieures au traitement, afin que l'hypothèse de tendances parallèles entre les contrôles et les traités soit valide. Comme nous allons le voir dans la suite de ce chapitre, ce point est très important pour notre analyse statistique.

L'INSEE reconnaît que tant que ces évolutions ne sont pas finalisées, un certain nombre de traitements ne sont temporairement plus effectués au sein du système d'élaboration des statistiques d'emploi. Il est peu probable que les données harmonisées soient disponibles à court-terme. Lors de notre dernière collecte des données FARE en juin 2021, les problèmes n'étaient toujours pas réglés. Une autre raison expliquant notre choix de recourir au répertoire SIRET est qu'il est impossible d'obtenir les effectifs par établissement dans les données FARE.

Nous décrivons en détail les approches méthodologiques pour les deux analyses dans la suite du document.

3. Résultats

3.1. Estimation de l'effet global du THD et de la fibre

Les figures 5.1 et 5.2 présentent les principaux résultats des estimations à partir d'un modèle de différence-en-différence échelonnée (cf. Equation 5.8). La figure 5.1 présente l'impact de l'arrivée du THD sur la valeur ajoutée (Panel A), les effectifs (Panel B) et la productivité du travail (Panel C). Nous présentons l'effet avant le traitement (de t-5 à t-2) et après le traitement (de t à t+5) dans les Figures 5.1 et 5.2.⁸⁰ L'année de référence (omise) est l'année juste avant le traitement (t-1). Il est attendu que les paramètres estimés associés aux années avant le traitement soient non significatifs afin de valider l'hypothèse de tendances parallèles (ou au moins pour ne pas rejeter cette hypothèse). L'effet du traitement dynamique est obtenu à partir des coefficients associés aux années du traitement et post-traitement.

La figure 5.1 montre que l'accès au THD a eu un effet positif sur la valeur ajoutée, un effet possible sur les effectifs mais aucun effet sur la productivité du travail.

Le Panel A présente les résultats du traitement THD sur la valeur ajoutée communale. Tout d'abord, les coefficients estimés pour les périodes prétraitement (t-5 à t-2) mettent en évidence une absence de différence entre les communes traitées et de contrôle, nous rassurant quant au respect de l'hypothèse de tendance parallèle. Nous observons que le traitement a eu un effet positif statistiquement significatif après un an. Cet effet n'a cessé de s'amplifier au cours du temps. Ainsi trois ans après le traitement, la valeur ajoutée générée par le secteur marchand a été augmentée de 7% pour les communes traitées (IC 5% : +2.7% ; +11.1%)⁸¹. Après cinq années, la valeur ajoutée s'est accrue de près de 18% (IC 5% : +6.9% ; +29.9%). L'effet de l'accès au THD a eu un effet non seulement cumulatif mais également important.

⁸⁰ Le modèle est estimé pour l'ensemble des années avant traitement (de t-8 à t-2) et post traitement (t à t+5). Nous ne reportons que les coefficients pour les pré-traitement à partir de t-5. Les résultats complets en moindres carrés ordinaires et selon la méthode de Sun et Abraham sont disponibles en Annexe dans le tableau 5.A.2. En l'absence de précision dans la suite du chapitre, les estimations sont effectuées selon la méthode de Sun et Abraham (2020).

⁸¹ Nous reportons entre parenthèses l'intervalle de confiance à 5% des effets économiques du traitement. Les effets marginaux en pourcentage sont établis à partir de la moyenne de la valeur ajoutée des communes traitées (cf. Tableau 5.A.1).

L'analyse sur l'emploi (Panel B) met en évidence des résultats positifs bien que plus tardifs à se matérialiser et estimés de manière moins précise. Nous confirmons l'absence de différences significatives entre les deux groupes à la fois pour la période prétraitement et les premières années après le traitement (de t à $t+3$). Nous observons cependant un effet statistiquement et économiquement significatif quatre années après le traitement. Cet effet positif est confirmé et amplifié pour la cinquième année. D'un point de vue économique les effets sont loin d'être négligeables. L'emploi dans le secteur privé marchand dans les communes traitées s'est accru de 4.3% (IC : +2.9% ; +5.9%) et de 8.0% (IC : +5.0% ; +11%) 4 et 5 ans après le traitement respectivement par rapport à l'emploi moyen.

Enfin, comme l'illustre le Panel C nous ne trouvons pas d'effet pour la productivité du travail. En effet, l'arrivée du THD ne génère pas d'évolution de la productivité du travail dans les communes traitées.

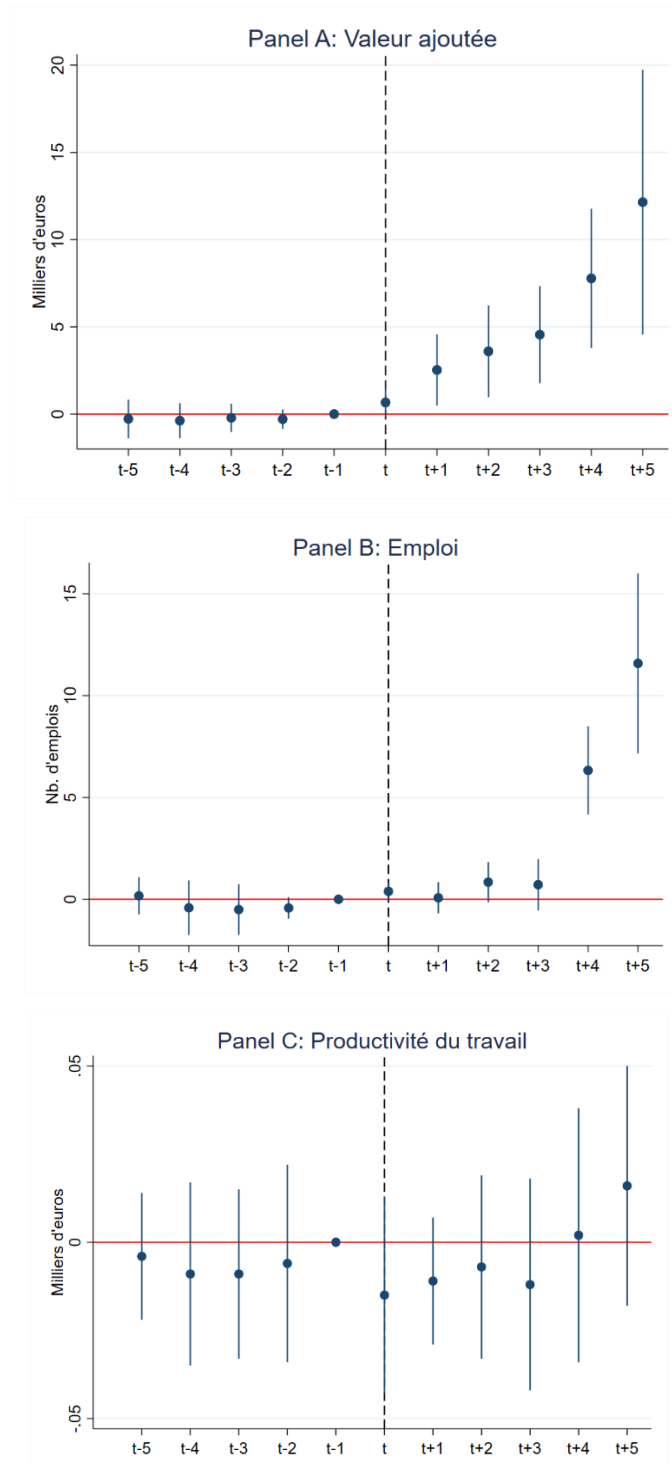
La figure 5.2 présente les mêmes estimations en considérant comme traitement l'accès à la fibre. Les résultats sont assez proches avec un effet positif (mais accentué) pour la valeur ajoutée, un effet présent mais moins précisément estimé pour l'emploi et toujours une absence d'effet pour la productivité.

Dans le détail, nous observons que l'accès à la fibre a eu un effet fort sur la création de richesse dans les communes et, comme dans le cas du THD, l'effet s'amplifie dans le temps. Les effets économiques sont également importants et proches de ceux estimés pour le THD en termes relatifs⁸². Ainsi, trois ans après le traitement, les communes traitées ont une valeur ajoutée supérieure de plus de 8.3% (IC : 0.0% ; 16.8%) par rapport à ce qui aurait prévalu en l'absence de fibre. Néanmoins, il convient de mettre un bémol à cette analyse. Les estimations sont moins précises que pour le THD (notamment $t+5$). Cela est sans doute dû au faible nombre de communes ayant atteint cette phase (le coefficient associé à $t+5$ ne peut être estimé que pour les communes traitées en 2014 soit seulement 96 communes et celui associé à $t+4$ seulement pour 465 communes soit 1.5% de l'échantillon).

Le reste de la figure 5.2 ne nous permet pas de mettre en évidence un effet de la fibre sur la productivité du travail (Panel C) et semble suggérer un effet sur l'emploi même si ce dernier est estimé avec peu de précisions et apparaît uniquement en $t+4$.

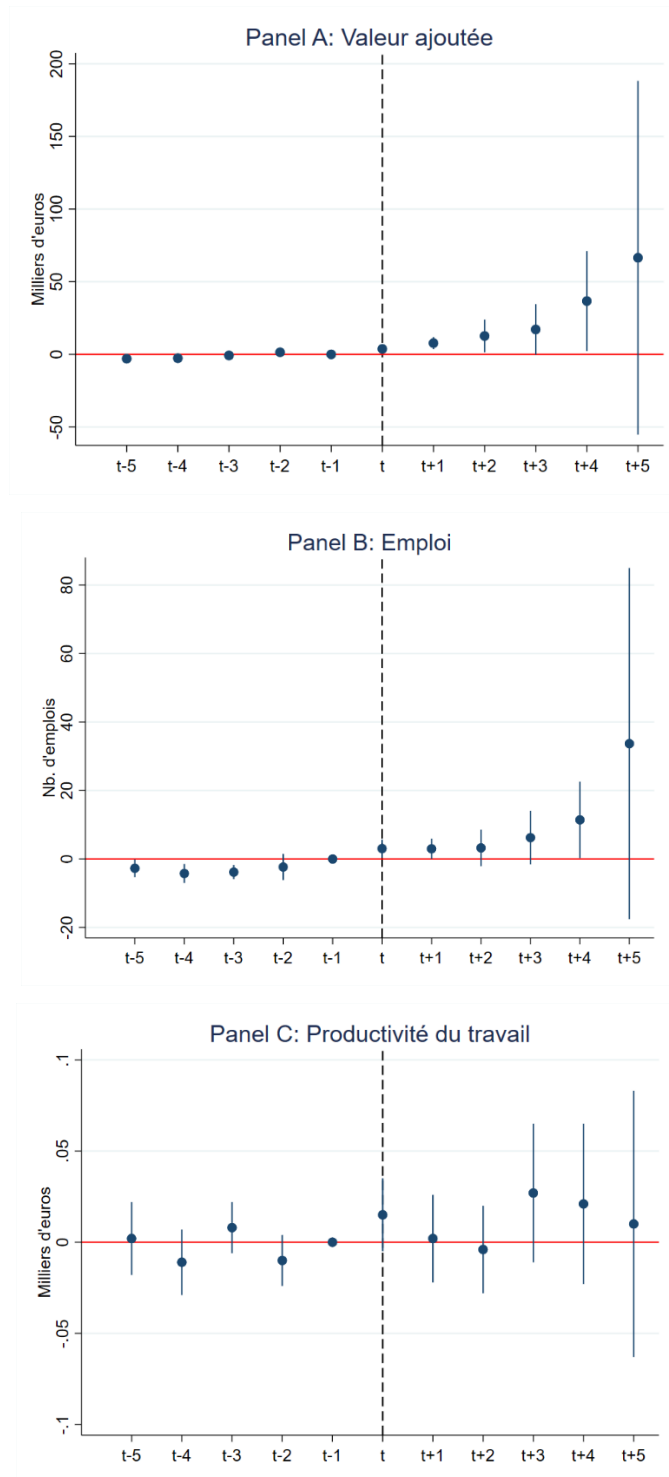
⁸² Les différences en termes de coefficients tiennent à la différence de taille entre les communes traitées par le THD et par la fibre (cf. Tableau A.5.1 en Annexe).

Figure 5.1 : Effet du THD (30 Mbps ou plus) sur la valeur ajoutée, l'emploi et la productivité du secteur marchand au niveau communal



Note : La figure reporte pour chaque estimation les coefficients et les intervalles de confiance à 5%. Les effets sont estimés en fonction de l'année précédant l'arrivée du THD (t-1). Les résultats complets sont disponibles en Annexe (tableau 5.A.2).
 Lecture : Trois ans après la mise en place du THD (t+3), la valeur ajoutée des communes ayant bénéficié de l'entrée du THD est supérieure de 5 000 euros par rapport aux communes n'ayant pas bénéficié de ce traitement.

Figure 5.2 : Effet de la fibre (100 Mbps ou plus) sur la valeur ajoutée, l'emploi et la productivité du secteur marchand au niveau communal



Note : La figure reporte pour chaque estimation les coefficients et les intervalles de confiance à 5%. Les résultats complets sont disponibles en Annexe (tableau 5.A.2).

3.2. Tests de robustesse

Avant d'étudier les hétérogénéités selon les caractéristiques des entreprises et des communes, nous évaluons la robustesse de nos résultats de base. Nous faisons ainsi varier le seuil de traitement. Jusqu'alors nous considérons qu'une commune est traitée dès l'arrivée du THD ou de la fibre sans tenir compte du pourcentage de locaux couverts. Nous reportons dans le tableau 5.3, les résultats en utilisant des seuils différents pour construire le traitement : 25%, 50% et 75% (par mesure de comparabilité nous reportons les résultats dans notre analyse de base). Les effets n'ont pas l'air de s'intensifier à mesure que la couverture (pourcentage de locaux couverts) s'améliore. Les résultats avec un niveau de couverture d'un quart des locaux couverts sont concordants avec notre analyse de base. En revanche, les effets tendent à s'estomper si nous prenons des seuils plus élevés. Deux explications peuvent justifier cette absence de résultat à mesure que le seuil est plus contraignant. D'une part, cela peut s'expliquer par la contraction du groupe de traitement qui fond à mesure que la couverture retenue pour définir le traitement augmente. D'autre part, cette conclusion peut s'expliquer par le fait que les établissements industriels et commerciaux sont sans doute parmi les premiers couverts au niveau d'une commune. Ces entreprises bénéficient sans doute rapidement de l'accès même lorsque le niveau de couverture reste limité. En revanche, accroître la couverture bénéficie surtout aux ménages.

Dans une analyse non reportée dans le document, nous prenons en compte le fait que la valeur ajoutée n'est pas disponible pour toutes les entreprises (et donc établissements). Nous restreignons notre échantillon aux communes pour lesquelles nous avons des informations sur la valeur ajoutée pour les établissements couvrant au moins 90% des emplois dans la commune. Nos résultats sont très similaires à nos analyses de base.

Tableau 5.3 : Analyse de sensibilité au seuil retenu

	Panel A: Effet sur la valeur ajoutée										Cohortes		
	t-5	t-4	t-3	t-2	t	t+1	t+2	t+3	t+4	t+5	Traités	Jamais	Avant 2013
A1: Très haut débit													
Baseline	-0.274	-0.375	-0.211	-0.288	0.669*	2.532**	3.598***	4.552***	7.779***	12.147***	15392	10844	4629
Seuil=25%	-0.627	-0.638	-0.398	-0.793	0.969	2.780**	3.573**	4.630**	8.002***	14.982**	12396	15504	2965
Seuil=50%	-0.148	0.411	0.230	-0.018	0.169	0.749	1.235	1.314	4.937	6.531	9083	20177	1605
Seuil=75%	-2.021	0.343	0.173	-0.573	0.505	1.849	-0.694	-1.403	-3.427	-4.142	5713	24554	598
A2: Fibre													
Baseline	-2.960*	-2.611	-0.738	1.458	3.778***	7.765***	12.670**	17.157**	36.699**	66.490	5160	25429	276
Seuil=25%	-3.344*	-1.488	-0.846	1.463	2.263*	6.381**	12.582*	24.213**	45.264**	65.344	3926	26720	219
Seuil=50%	-1.948	-0.307	-0.256	0.657	0.890	5.774**	12.861*	17.541*	31.129*	-1.463	3567	27116	182
Seuil=75%	-3.059	-0.655	-1.117	-1.089	0.767	8.464	8.080	13.354	4.111	-2.382	3130	27628	107
	Panel B: Effet sur l'emploi										Cohortes		
	t-5	t-4	t-3	t-2	t	t+1	t+2	t+3	t+4	t+5	Traités	Jamais	Avant 2013
B1: Très haut débit													
Baseline	0.169	-0.418	-0.505	-0.425	0.389*	0.074	0.843*	0.716	6.328***	11.582***	15392	10844	4629
Seuil=25%	-0.591	-1.089	-0.858	-0.742*	-0.422	-0.771	-0.962	-0.959	3.294**	7.676**	12396	15504	2965
Seuil=50%	0.457	0.000	-0.532	-0.238	-1.217	-2.271	-1.162	-2.041	0.822	1.595	9083	20177	1605
Seuil=75%	0.776	0.552	-0.180	0.386	-1.302	-1.582	0.037	-0.842	-1.920	0.636	5713	24554	598
B2: Fibre													
Baseline	-2.714**	-4.223***	-3.833***	-2.339*	3.038**	2.979**	3.240	6.246	11.425**	33.697	5160	25429	276
Seuil=25%	-3.611**	-2.815*	-1.878	-3.162***	0.532	-0.391	0.607	6.698*	15.017**	41.591*	3926	26720	219
Seuil=50%	-1.354	-0.721	-2.333*	-1.292	-0.466	-1.817	3.666*	9.555***	13.826**	-1.693	3567	27116	182
Seuil=75%	-0.789	-0.082	-1.349	-0.918	-0.858	2.173	5.545**	3.948	2.559	-3.025	3130	27628	107

Le tableau reporte les coefficients estimés pour les variables t-5 à t+4 (le modèle inclut aussi les variables t-8, t-7 et t-6 mais les résultats ne sont pas reportés). La spécification du modèle est identique au modèle de base et la méthode d'estimation retenue est celle de Sun et Abraham. *, ** et *** signale la significativité statistique au seuil de 10%, 5% et 1%, respectivement. Les variables dépendantes sont la valeur ajoutée du secteur productif dans la commune (Panel A) et le nombre d'emplois du secteur productif (Panel B).

Lecture : Cinq années après l'arrivée du THD (traitement défini au seuil de 25%), la valeur ajoutée a augmenté en moyenne de 14 982 euros dans les communes traitées.

3.3. Analyse par type d'entreprises

Précisions méthodologiques

Nous explorons à présent plus en détails l'effet du THD et de la fibre sur l'activité économique locale par catégorie d'entreprises. En effet, toutes les entreprises peuvent ne pas bénéficier également de l'accès au THD ou à la fibre. Nous considérons plusieurs caractéristiques des entreprises incluant leur présence initiale dans la commune, leur taille, leur secteur d'activité et le fait qu'elles soient des entreprises mono- ou multi-établissements⁸³.

Afin de pouvoir prendre en compte chaque dimension dans une analyse à un niveau agrégé, nous reconstruisons nos variables d'intérêt. Par exemple, lorsque nous souhaitons prendre en compte l'effet sur les petites entreprises (inférieures à 20 salariés), nous construisons pour chaque commune la valeur ajoutée et l'emploi uniquement pour ces petites entreprises. Nos variables d'intérêt deviennent donc :

$$EMP_{i(k)t} = \sum_{e \in \{i,k\}} EFF_{et} \text{ et } VA_{i(k)t} = \sum_{e \in \{i,k\}} VA_{et}$$

Avec k le critère considéré (taille, secteur, mono-établissement, multi-établissements).

Entreprises déjà présentes vs. Arrivée nette d'entreprises

Dans un premier temps, nous analysons si l'effet observé s'explique par une évolution des entreprises présentes sur la commune avant l'entrée du THD ou de la fibre ou par une arrivée nette d'entreprises plus productives suite à l'arrivée de l'internet très haut débit. Pour ce faire, classons les entreprises en deux groupes. Le premier groupe inclut les entreprises présentes en 2013 (c'est-à-dire avant l'arrivée du traitement dans toutes les communes) et toujours actives à la fin de la période. Le second groupe inclut les autres entreprises (soit celles entrées après 2013 ou celles qui ont quitté la commune avant 2019). Dans les tableaux 5.4 et 5.5, nous estimons le modèle sur ces deux groupes, respectivement pour le THD et pour la fibre. Il ressort de l'analyse que l'effet de l'arrivée de l'internet THD ou de la fibre sur la valeur ajoutée s'explique en priorité par une amélioration des performances des entreprises déjà existantes dans les communes traitées⁸⁴. Les évidences sur les entreprises non continuellement présentes dans la commune sont moins claires. Les résultats économétriques sur l'emploi sont moins limpides. Les deux groupes d'entreprises semblent avoir bénéficié de l'arrivée du THD alors que la fibre

⁸³ Nous limitons notre analyse à l'effet sur la valeur ajoutée et les effectifs (faute d'effet moyen sur la productivité).

⁸⁴ Ce résultat doit néanmoins être pris avec prudence en raison du non-respect de l'hypothèse de tendance parallèle.

a eu un effet plus fort sur les entreprises déjà présentes en termes économiques mais cet effet n'est jamais statistiquement significatif.

Tableau 5.4 : Effet du THD selon la présence initiale ou non des entreprises

	Valeur ajoutée		Emploi	
	Cont. (1)	Entrée nette (2)	Cont. (3)	Entrée nette (4)
t-8	-0.994*** (0.352)	-0.296 (0.192)	-0.510 (0.343)	-0.114 (0.316)
t-7	-0.515 (0.327)	-0.280 (0.188)	0.220 (0.377)	-0.727** (0.310)
t-6	-0.788*** (0.271)	-0.013 (0.161)	-0.192 (0.320)	0.165 (0.300)
t-5	-1.034* (0.543)	0.759*** (0.202)	-2.444*** (0.459)	2.614*** (0.329)
t-4	-2.421*** (0.473)	2.047*** (0.260)	-5.585*** (0.722)	5.166*** (0.758)
t-3	-1.402*** (0.391)	1.191*** (0.208)	-3.072*** (0.658)	2.566*** (0.253)
t-2	-0.503* (0.277)	0.215* (0.120)	-0.944*** (0.249)	0.519*** (0.173)
t	0.980** (0.391)	-0.310*** (0.118)	0.442** (0.176)	-0.054 (0.180)
t+1	2.225** (1.009)	0.307 (0.197)	-0.911*** (0.346)	0.985*** (0.272)
t+2	3.031*** (1.275)	0.567** (0.272)	-0.906** (0.372)	1.741*** (0.366)
t+3	3.848*** (1.409)	0.704** (0.328)	1.262** (0.509)	1.978*** (0.464)
t+4	6.970*** (2.059)	0.808* (0.480)	2.387*** (0.854)	3.941*** (0.724)
t+5	12.243*** (3.851)	-0.097 (0.870)	4.556*** (1.682)	7.023*** (1.431)
Moy(Y)	65.39	18.98	112.1	35.5
Obs.	258001	258001	258001	258001
# Communes	26327	26327	26327	26327
R ² (ajusté)	0.91	0.83	0.98	0.88

Le modèle estime l'effet de l'entrée de la fibre sur la valeur ajoutée communale (colonnes 1 et 2) et les effectifs totaux par communes (colonnes 3 et 4). Le modèle inclut des effets fixes communes et années. La méthode d'estimation est celle de Sun et Abraham. Chaque estimation inclut comme variables de contrôle la population totale de la commune, le degré de diversification et la spécialisation par secteurs (voir le texte pour un détail sur la construction de ces variables). Les résidus sont clustérisés au niveau communal. *, **, et *** signale la significativité à 10, 5 et 1% respectivement.

Tableau 5.5 : Effet de la fibre selon la présence initiale ou non des entreprises

	Valeur ajoutée		Emploi	
	Cont. (1)	Entrée nette (2)	Cont. (3)	Entrée nette (4)
t-8	-6.666*** (1.245)	0.596 (0.619)	-4.523*** (1.126)	2.499** (0.932)
t-7	-3.924*** (1.302)	1.351 (1.682)	0.754 (1.064)	-1.727 (1.191)
t-6	-2.854* (1.610)	-2.226*** (0.720)	2.109 (1.226)	-4.929*** (1.224)
t-5	-1.035 (1.641)	-1.925*** (0.650)	1.548 (1.361)	-4.262*** (1.191)
t-4	-1.357 (1.748)	-1.253 (0.789)	-0.213 (1.664)	-4.020*** (1.285)
t-3	1.035 (1.452)	-1.773** (0.563)	2.945** (1.359)	-5.284*** (1.077)
t-2	2.653* (1.919)	-1.195** (0.450)	2.226** (0.914)	-6.097*** (0.876)
t	6.654*** (1.178)	-2.876*** (0.719)	7.527*** (1.261)	-4.489*** (0.985)
t+1	4.972*** (1.655)	2.794** (1.250)	-0.111 (1.049)	3.091*** (1.141)
t+2	9.128* (4.862)	3.542** (1.773)	0.955 (2.048)	4.195** (1.689)
t+3	16.002* (8.741)	1.155 (2.201)	1.449 (2.579)	4.793 (2.967)
t+4	34.054** (16.929)	2.644 (5.288)	6.942 (4.960)	4.482 (4.882)
t+5	61.440 (48.732)	5.049 (26.302)	56.841 (38.469)	-23.144 (25.907)
Obs. #	301196	301196	301196	301196
Communes	30680	30680	30680	30680
R ² (ajusté)	0.96	0.87	0.98	0.56

modèle estime l'effet de l'entrée de la fibre sur la valeur ajoutée communale (colonnes 1 et 2) et les effectifs totaux par communes (colonnes 3 et 4). Le modèle inclut des effets fixes communes et années. La méthode d'estimation est celle de Sun et Abraham. Chaque estimation inclut comme variables de contrôle la population totale de la commune, le degré de diversification et la spécialisation par secteurs (voir le texte pour un détail sur la construction de ces variables). Les résidus sont clustérisés au niveau communal. *, **, et *** signale la significativité à 10, 5 et 1% respectivement.

Distinction selon la taille des entreprises

Dans le tableau 5.6, nous présentons l'effet du THD et de la fibre sur la valeur ajoutée et les effectifs par taille d'entreprises. Nous considérons quatre classes d'entreprises : petites entreprises (19 salariés et moins), entreprises moyennes (entre 20 et 49 salariés), entreprises de taille intermédiaire (entre 50 et 249 salariés) et grandes entreprises (plus de 250 salariés). Afin de simplifier la lecture nous ne reportons que les valeurs estimées entre t-5 et t+5.^{85,86}

Les résultats montrent que toutes les entreprises, à l'exception des petites entreprises, ont bénéficié du déploiement du THD en termes de création de richesse (Panel A1). En revanche, l'effet sur l'emploi a joué à plein sur l'ensemble des entreprises, petites entreprises incluses (Panel B1). Ces dernières ont d'ailleurs bénéficié rapidement de l'arrivée du THD (après deux années seulement) alors qu'il a fallu attendre un peu plus longtemps pour voir un effet se matérialiser pour que les entreprises de plus grande taille créent des emplois dans les communes connectées. Néanmoins, les grandes entreprises ont davantage profité de l'arrivée du THD. L'effet économique est proche pour les PME et ETI (15% de la valeur moyenne). Cet effet est le double pour les grandes entreprises⁸⁷.

En revanche, la fibre semble avoir bénéficié principalement aux PME et ETI mais pas aux grandes entreprises. L'analyse de l'évolution de la richesse créée par le secteur marchand selon la taille des entreprises (Panel A2) met en évidence un effet de l'arrivée de la fibre pour les PME et ETI même si cet effet perd sa significativité statistique en t+5 (sans doute dû au faible nombre de communes traitées jusqu'à t+5). En revanche, bien qu'un effet semble apparaître pour la valeur ajoutée par les grandes entreprises, ce dernier n'est pas statistiquement significatif et son amplitude est faible (d'autant que la valeur ajoutée par les grandes entreprises est beaucoup plus grande que pour les PME et ETI). Ces résultats sont convergents avec les analyses sur la création d'emplois dans chaque commune due aux différents types d'entreprises (Panel B2).

⁸⁵ Les estimations complètes, disponibles au besoin, incluent les périodes allant de t-8 à t-6.

⁸⁶ A nouveau, certains résultats sont à interpréter avec précaution car l'hypothèse des tendances parallèles n'est pas systématiquement respectée.

⁸⁷ Ces effets relatifs sont calculés à partir de la valeur moyenne de la valeur ajoutée pour chaque groupe.

Tableau 5.6 : Effet du THD et de la fibre selon la taille des entreprises

Panel A: Effet sur la valeur ajoutée										
	t-5	t-4	t-3	t-2	t	t+1	t+2	t+3	t+4	t+5
A1: Très haut débit										
Petites entreprises	-0.201**	-0.112	-0.222**	-0.125**	-0.055	-0.058	0.051	0.235	0.244	-0.551
Moyennes entreprises	0.001	-0.019	0.069	0.039	-0.013	0.044	0.137*	0.228**	0.608***	1.009***
Entreprises de taille intermédiaire	-0.075	-0.006	0.139	-0.054	0.041	0.181*	0.401***	0.479***	0.795***	0.401
Grandes entreprises	-0.350*	-0.196	-0.197	-0.107	0.393	0.323	0.532**	0.659*	1.151**	2.891***
A2: Fibre										
Petites entreprises	-1.062***	-0.997***	-0.696***	0.395	0.727*	1.487***	2.446***	1.915*	5.541	33.419
Moyennes entreprises	-0.385**	-0.316**	-0.263*	0.035	0.057	0.121	0.539	1.554***	1.271	3.142
Entreprises de taille intermédiaire	-0.468*	-0.481*	0.258	0.157	0.670***	1.782***	2.259***	4.234***	6.972**	10.053
Grandes entreprises	-0.333	.218	0.132	0.812*	0.947**	0.549	0.007	3.049	5.331	6.777
Panel B: Effet sur l'emploi										
	t-5	t-4	t-3	t-2	t	t+1	t+2	t+3	t+4	t+5
B1: Très haut débit										
Petites entreprises	0.142*	0.232***	0.369***	-0.107**	0.086	0.392***	1.277***	0.509***	2.062***	2.395***
Moyennes entreprises	0.070	0.167**	0.156**	0.089	0.064	-0.176*	-0.054	0.172	1.116***	2.641***
Entreprises de taille intermédiaire	0.176	-0.064	0.250*	0.189**	-0.037	-0.114	0.148	0.279	1.270***	1.447**
Grandes entreprises	0.256	0.110	-0.193	-0.197	0.333**	0.338	0.462	0.532	1.305**	2.912**
B2: Fibre										
Petites entreprises	-1.457***	-1.334***	0.396*	-0.935***	0.696***	1.954***	2.806***	5.242***	6.389***	16;574***
Moyennes entreprises	-0.099	-0.403*	0.389*	-0.367*	0.584***	0.753*	1.455***	2.811***	5.662***	12.572*
Entreprises de taille intermédiaire	-0.399	-0.830	0.094**	-0.343	0.771**	0.949*	2.030**	3.009**	6.577***	22.598*
Grandes entreprises	-0.204	0.565	0.656	0.449	1.439**	0.552	0.002	1.156	1.815	0.491

Le tableau reporte les coefficients estimés pour les variables t-5 à t+4 (le modèle inclut aussi les variables t-8, t-7 et t-6 mais les résultats ne sont pas reportés). La spécification du modèle est identique au modèle de base et la méthode d'estimation retenue est celle de Sun et Abraham. Les petites entreprises sont celles ayant moins de 20 salariés, les entreprises moyennes ont entre 20 et 49 salariés, les entreprises de taille intermédiaires ont entre 50 et 249 salariés et les grandes entreprises plus de 250 salariés. *, **, et *** signale la significativité à 10, 5 et 1% respectivement.

Distinction selon le secteur d'activités

Dans les deux tableaux suivants (5.7 et 5.8) nous inspectons les effets du THD et de la fibre par secteur d'activités.⁸⁸ Deux secteurs semblent avoir bénéficié de l'accès au THD en termes de création de valeur ajoutée. Il s'agit de l'hôtellerie-restauration et de l'industrie (cf. Panel A du tableau 5.7). En termes d'emplois, nous retrouvons un effet positif pour l'hôtellerie-restauration mais également pour les technologies de l'information et de la communication. Il convient de noter que l'arrivée du THD a eu un effet négatif dans le secteur de la construction (sur la valeur ajoutée et l'emploi), possiblement lié à un effet destructeur de la concurrence (si l'accès à internet augmente les débouchés des entreprises locales, elle peut également conduire les consommateurs locaux à se tourner vers des producteurs plus lointains). Ce résultat est toutefois à interpréter avec précaution, en raison de l'existence d'une tendance positive dans la période prétraitement (évolution de l'emploi plus bénéfique dans les communes traitées que les communes de contrôle), ce qui peut biaiser les résultats. Cette tendance positive peut s'expliquer par les travaux requis au niveau local pour permettre le déploiement des réseaux.

Seule l'industrie a réellement bénéficié de la fibre en termes de création de la valeur ajoutée au niveau local. En revanche, les effectifs dans l'industrie n'ont pas augmenté suite à l'entrée de la fibre. Ce résultat tend à contredire les études antérieures montrant que l'industrie bénéficie moins du très haut débit que les autres secteurs, du moins en termes de créations d'entreprises (Duvivier, 2019).

Un dernier résultat intéressant est l'absence d'impact du THD sur le commerce en zone rurale. Cet effet va à l'encontre des anticipations de nombreux acteurs locaux, qui espèrent souvent qu'une amélioration de la connectivité permettra de pallier le manque de débouchés au niveau local. Ce résultat est toutefois peu surprenant puisque, si l'arrivée du THD ouvre de nouveaux débouchés aux producteurs locaux via le commerce en ligne, elle engendre également une hausse de la concurrence puisque les consommateurs locaux ont désormais accès en ligne à des produits plus diversifiés, de meilleure qualité et/ou à moindre coût (Cumming et Johan, 2010; Freathy et Calderwood, 2016). Ainsi, sans adaptation de leur stratégie de marketing, les producteurs ruraux de biens/services échangeables risquent de ne pas bénéficier de l'arrivée du THD pourtant parfois très attendue (Lamie et al., 2011).

⁸⁸ La description de chaque secteur est fournie en Annexe (Tableau 5.A.3) avec la correspondance au niveau de la classification de l'INSEE (NAF 17 et NAF 88).

Distinction entre entreprises mono-établissement et multi-établissements

Nous distinguons ensuite les effets entre les entreprises mono-établissement et multi-établissements dans le tableau 5.9. Les deux types d'entreprises ont contribué à la création de valeur ajoutée (Panel A1) et à l'emploi (Panel B1) au niveau communal suite à l'arrivée du THD.

En revanche, les entreprises mono-établissement semblent avoir bénéficié davantage de l'entrée de la fibre que les entreprises multi-établissements tant au niveau de la création de valeur (Panel A2) que d'emplois (Panel B2).

Tableau 5.7 : Effet du THD et de la fibre sur la valeur ajoutée communale par secteur

Panel A : Très haut débit										
	t-5	t-4	t-3	t-2	t	t+1	t+2	t+3	t+4	t+5
Commerce	0.005	0.108	0.115	-0.029	-0.480	-0.154	-0.136	-0.004	-0.033	-0.845
Construction	0.116	0.159	-0.038	-0.037	-0.256***	-0.496***	-0.527***	-0.645***	-0.746***	-1.013**
Finance	-0.052	-0.039	-0.048	-0.032	-0.016	0.041	0.085***	0.066**	0.071	0.202
Hotels	0.039*	0.050**	0.040**	0.021	-0.007	0.015	0.011	0.117***	0.165**	0.051
Industrie	0.225	0.304	0.356	-0.012	0.950*	2.209**	2.499***	2.953***	4.463***	7.189***
TIC	-0.064***	-0.197***	-0.111**	-0.012	0.029	-0.238	0.332	0.090	0.149	0.324
Transport	-0.063	-0.187**	-0.121**	0.061*	0.033	0.633	0.839	0.976	1.524	3.623
Panel B : Fibre										
	t-5	t-4	t-3	t-2	t	t+1	t+2	t+3	t+4	t+5
Commerce	0.023	-0.499	-0.572	0.492	-0.170	-0.714	-1.523	1.662	0.765	-20.196
Construction	-0.133	-0.183	-0.400***	-0.209*	-0.157	0.475	0.387	0.062	-0.761	0.845
Finance	-0.147*	-0.111	-0.003	0.089	-0.031	-0.167	-0.200	-0.110	-0.370	-0.729
Hotels	-0.091*	-0.051	0.003	0.093*	-0.021	0.094	-0.063	-0.090	-0.535	-2.017
Industrie	-0.053	0.350	0.996	0.516	1.964	6.361**	7.405**	7.952*	19.625*	19.389
TIC	0.001	-0.080	0.055	0.297*	0.347**	-1.264	2.511	0.224	-0.074	-1.772
Transport	-0.178	-0.187	-0.085	0.236	0.317	0.290	0.052	-0.733	-1.051	-5.427

Le tableau reporte les coefficients estimés pour les variables t-5 à t+5 (le modèle inclut aussi les variables t-8, t-7 et t-6 mais les résultats ne sont pas reportés). La spécification du modèle est identique au modèle de base et la méthode d'estimation retenue est celle de Sun et Abraham. *, **, et *** signale la significativité à 10, 5 et 1% respectivement.

Tableau 5.8 : Effet du THD et de la fibre sur l'emploi communal par secteur

Panel A : Très haut débit										
	t-5	t-4	t-3	t-2	t	t+1	t+2	t+3	t+4	t+5
Commerce	-0.106	-0.105	0.143	0.139	0.276	-0.063	-0.159	0.158	1.056	1.317
Construction	0.493***	0.563***	0.346***	0.092*	-.310***	-0.603***	-0.595***	-0.707***	-0.140	1.203
Finance	0.004	-0.024	-0.036	-0.055*	0.001	0.042	-0.024	-0.176***	-0.097	0.011
Hotels	-0.026	-0.033	0.045	-0.043	0.208***	0.223***	0.382**	0.352***	1.279***	1.852***
Industrie	-0.199	0.378	0.169	-0.133	-0.303	-0.190	-0.314	-0.260	0.266	0.98
TIC	0.033	-0.008	-0.003	0.009	0.021	0.335	0.0443	0.140***	0.196***	0.241*
Transport	0.175*	-0.364	-0.511	0.050	0.088	-0.197	0.101	0.250	0.306	0.615
Panel B : Fibre										
	t-5	t-4	t-3	t-2	t	t+1	t+2	t+3	t+4	t+5
Commerce	0.017	-0.139	-0.306	-0.439	1.405*	-0.805	-0.823	3.319	6.431	0.414
Construction	-0.027	-0.262	-0.378*	-0.573**	0.231	0.804*	1.479**	1.637***	2.512*	16.156*
Finance	-0.161	-0.081	-0.123	-0.387***	-0.182*	0.049	-0.210	-0.516	-0.604	-1.15
Hotels	-0.165	-0.144	0.24	-0.106	0.465***	0.423*	0.389	1.198*	2.828*	1.485
Industrie	-0.109	-0.578	-1.027	-1.777**	-2.230*	-2.726	-3.927	-5.624	-10.104	-2.567
TIC	0.114	0.056	0.063	0.023	0.383	1.406	2.254	-0.156	-0.910	-10.186
Transport	-0.357	-0.479*	-0.721***	-0.012	0.165	0.593	0.391	-0.441	-0.323	-6.081

Le tableau reporte les coefficients estimés pour les variables t-5 à t+5 (le modèle inclut aussi les variables t-8, t-7 et t-6 mais les résultats ne sont pas reportés). La spécification du modèle est identique au modèle de base et la méthode d'estimation retenue est celle de Sun et Abraham. *, **, et *** signale la significativité à 10, 5 et 1% respectivement.

Tableau 5.9 : Effet du THD et de la fibre, mono-établissements et multi-établissements

Panel A: Effet sur la valeur ajoutée										
	t-5	t-4	t-3	t-2	t	t+1	t+2	t+3	t+4	t+5
A1: Très haut débit										
Mono	-0.094	-0.074	-0.307**	-0.246**	0.029	0.241	0.753***	0.907***	2.017	2.501**
Multi	-0.180	-0.300	0.096	-0.042	0.640	2.291**	2.846**	3.645***	5.761***	9.645***
A2: Fibre										
Mono	-1.593**	-1.274**	-0.599	0.329	1.696***	5.075***	6.124***	10.920***	20.358*	34.524
Multi	-1.367	-1.337	-0.139	1.128	2.082	2.690	6.546	6.222	16.340	31.065
Panel B: Effet sur l'emploi										
	t-5	t-4	t-3	t-2	t	t+1	t+2	t+3	t+4	t+5
B1: Très haut débit										
Mono	0.553**	0.437*	0.454**	-0.169	-0.029	-0.021	1.130***	0.990**	4.381***	6.595***
Multi	-0.384	-0.855	-0.960	-0.256	0.415**	0.095	-0.295	-0.275	1.946*	4.987**
B2: Fibre										
Mono	-1.841**	-2.855***	-0.780	-2.466***	1.611*	3.907***	3.519	11.443**	21.191***	7.232
Multi	-0.872	-1.368	-1.558	-1.367	1.426	-0.929	-0.278	5.197	9.776	26.464

Le tableau reporte les coefficients estimés pour les variables t-5 à t+4 (le modèle inclut aussi les variables t-8, t-7 et t-6 mais les résultats ne sont pas reportés). La spécification du modèle est identique au modèle de base et la méthode d'estimation retenue est celle de Sun et Abraham. *, **, et *** signale la significativité à 10, 5 et 1% respectivement.

3.4. Analyse selon les caractéristiques des communes

Après avoir mis en évidence des différences selon le type d'entreprises, nous étudions dans quelle mesure les caractéristiques communales conditionnent les effets observés. Pour ce faire, nous considérons deux critères en lien avec la revue de la littérature : le degré de ruralité/urbanité de la commune et le débit initial disponible avant l'arrivée du THD ou de la fibre.

Les petites communes peuvent plus ou moins bénéficier de l'arrivée de l'internet à très haut débit en fonction de leur voisinage. Une commune rurale proche d'un pôle urbain pourrait davantage tirer profit de l'arrivée du haut débit comme cela a été mis en évidence par certaines études (de Vos et al., 2020 ; Kim et Orazem, 2012 ; Liebowitz, 2002). Afin de prendre en compte ce critère nous utilisons le zonage en aires urbaines de l'INSEE et distinguons trois espaces différents comme suit⁸⁹ :

- Les grands pôles urbains (ayant plus de 10000 emplois dans le pôle) et les communes péri-urbaines en couronne de ces grands pôles ;
- Les pôles petits et moyens (ayant entre 1500 et 10000 emplois dans le pôle) et les communes rurales sous influence de ces pôles petits et moyens ;
- Les communes rurales isolées.

L'effet du THD sur l'activité économique locale est positif pour les trois groupes de communes comme l'illustre le tableau 5.10 (Panel A). Néanmoins cet effet est plus précoce pour les communes dans la couronne des grands pôles urbains et celles appartenant aux pôles petits et moyens. Il apparaît dès la première année après le traitement alors qu'il faut attendre trois ans pour percevoir un effet pour les communes rurales isolées. Le délai dans la matérialisation des effets du THD dans les communes rurales isolées peut s'expliquer par plusieurs facteurs comme une plus faible capacité d'absorption pour des raisons financières, de compétences ou d'attitude vis-à-vis des nouvelles technologies (Salemink et al., 2017).

De manière assez analogue, l'effet sur l'emploi du THD se retrouve dans la plupart des communes, à l'exception des communes isolées (Panel B1). Il convient de noter un certain décalage dans le temps entre l'effet sur la valeur ajoutée et l'effet sur l'emploi. Cela peut s'expliquer par le délai existant entre l'augmentation de l'activité (valeur ajoutée) et les

⁸⁹ La première catégorie réunit les communes étant catégorisées par l'INSEE, dans son zonage en aires urbaines (2010), comme appartenant aux catégories 111, 112 et 120 ; la seconde catégorie réunit les codes 211, 212, 221, 222 et 300 et la dernière catégorie réunit le code 400. Comparativement au chapitre sur les usages numériques, certaines catégories d'espaces ont été regroupées afin de conserver un nombre suffisant d'observations dans chaque classe (en particulier, les petits et moyens pôles et les grands pôles regroupent un nombre assez limité d'observations).

embauches qui peuvent suivre si cette hausse est structurelle. Les communes isolées pourraient dès lors bénéficier plus tard d'un effet du THD sur l'emploi.

En revanche, l'effet de la fibre apparaît surtout pour les communes localisées à proximité des grands pôles urbains et principalement sur la création de valeur. Néanmoins, il faut garder à l'esprit que peu de communes des petits et moyens pôles ont eu accès à cette technologie au cours de la période d'analyse. Dans l'ensemble, ces résultats confirment donc que le THD a tendance à favoriser les espaces plus proches des centres urbains.

Notre revue de la littérature a mis en évidence que peu d'études (Canzian et al., 2019 étant une exception) ont pris en compte le débit initial disponible pour évaluer l'effet de l'arrivée de l'internet à très haut débit. Ayant l'opportunité d'avoir des informations sur des niveaux de couverture moindres (3 Mbps et 8 Mbps), nous analysons si l'effet du THD ou de la fibre est conditionnel au débit initial dans la commune. Nous distinguons trois niveaux de couverture initial :

- Communes ayant un accès à internet inférieur à 8 Mbps (absence d'internet ou faible débit);
- Communes ayant un accès à internet compris entre 8 et 30 Mbps (accès à du bon haut débit) ;
- Communes ayant un accès à internet supérieur à 30 Mbps mais inférieur à 100 Mbps (accès à du très haut débit).

Le dernier groupe de commune n'est considéré que pour l'analyse de l'effet de la fibre. Nous construisons l'accès initial à partir des données de couverture pour 2013. En outre, de manière similaire à notre analyse d'impact, nous considérons qu'une commune a accès à un certain débit en 2013 dès lors qu'au moins un local est couvert par ce débit. Enfin nous négligeons les niveaux de couverture inférieurs ou supérieurs à 3 Mbps car très peu de communes n'avaient pas accès à un tel débit en 2013.

Les résultats, reportés au tableau 5.11 mettent en évidence que le THD a eu un effet uniquement pour les communes ayant déjà accès à un débit supérieur à 8 Mbps en 2013. En revanche les communes dont le débit était inférieur à ce seuil (ce qui concerne quand même presque 10 000 communes) ne semblent pas avoir tiré profit de l'arrivée du THD. Une explication possible est que les entreprises localisées dans ces communes ont eu besoin d'un temps d'adaptation pour bénéficier d'un fort débit internet et ont mis du temps à changer leurs habitudes. D'ailleurs, nous observons que le THD a un effet important, bien que non significatif, après cinq années d'accès au THD. Nous retrouvons un phénomène proche pour la fibre. Même si les résultats sont statistiquement moins clairs, les effets économiques de l'arrivée de la fibre

semblent plus forts pour les deux groupes de communes ayant accès à du bon haut débit (8 Mbps ou plus) ou du THD (30 Mbps ou plus). Comme l'illustre le panel B du tableau 5.11, nous ne trouvons d'effet statiquement significatifs que pour le second groupe sur l'emploi.

Tableau 5.10: Effet du THD et de la fibre selon le degré de ruralité de la commune

	Panel A: Effet sur la valeur ajoutée										Echantillon	
	t-5	t-4	t-3	t-2	t	t+1	t+2	t+3	t+4	t+5	Obs.	Communes
A1: Très haut débit												
Grands poles	-0.247	-0.071	0.032	-0.439	0.944	4.249**	6.342**	7.109**	11.332***	16.956***	119126	12104
Poles petits et moyens	0.581	0.088	-0.024	-0.104	0.236	1.124**	1.379**	2.101***	3.741***	3.946*	76767	7841
Communes isolées	-0.946**	-1.205***	-0.626**	0.076	0.370	0.409	0.394	1.717***	3.057***	6.562***	62094	6380
A2: Fibre												
Grands poles	-4.215*	-4.377	-2.297	0.631	3.624**	8.427***	16.265**	21.291**	45.057*	156.52	119126	12104
Poles petits et moyens	1.593	2.471	3.783	3.575	3.991*	6.986**	2.915	2.374	14.291**	-1.121	76767	7841
Communes isolées	0.712	0.359	0.428	1.256	0.716	-0.993	-2.602	-2.535	-5.798	1.358	62094	6380
	Panel B: Effet sur l'emploi										Echantillon	
	t-5	t-4	t-3	t-2	t	t+1	t+2	t+3	t+4	t+5	Obs.	Communes
B1: Très haut débit												
Grands poles	-0.154	-0.958	-1.120	-0.390	0.756**	0.754	2.816***	2.167*	9.671***	16.042***	119126	12104
Poles petits et moyens	0.881*	0.646	0.329	-0.569	-0.297	-0.538	-0.956	-0.525	3.604**	6.134	76767	7841
Communes isolées	0.173	-0.409	-0.085	-0.066	0.429	-0.598	-0.741	-0.501	1.121	4.767	62094	6380
B2: Fibre												
Grands poles	-3.867**	-5.405***	-3.302**	-3.498***	3.172*	3.747*	3.047	4.802	13.748*	91.289	119126	12104
Poles petits et moyens	-0.073	-2.176	-0.181	-6.189	1.159	-1.136	1.632	7.085	1.647	-6.493	76767	7841
Communes isolées	1.796	1.950	0.564	1.099	1.307	-0.460	-0.691	-1.172	-6.225	0.142	62094	6380

Le tableau reporte les coefficients estimés pour les variables t-5 à t+5 (le modèle inclut aussi les variables t-8, t-7 et t-6 mais les résultats ne sont pas reportés). La spécification du modèle est identique au modèle de base et la méthode d'estimation retenue est celle de Sun et Abraham. La définition précise des trois groupes est détaillée dans le corps du texte. *, **, et *** signale la significativité à 10, 5 et 1% respectivement.

Tableau 5.11 : Effet du THD et de la fibre selon le débit initial (en 2013)

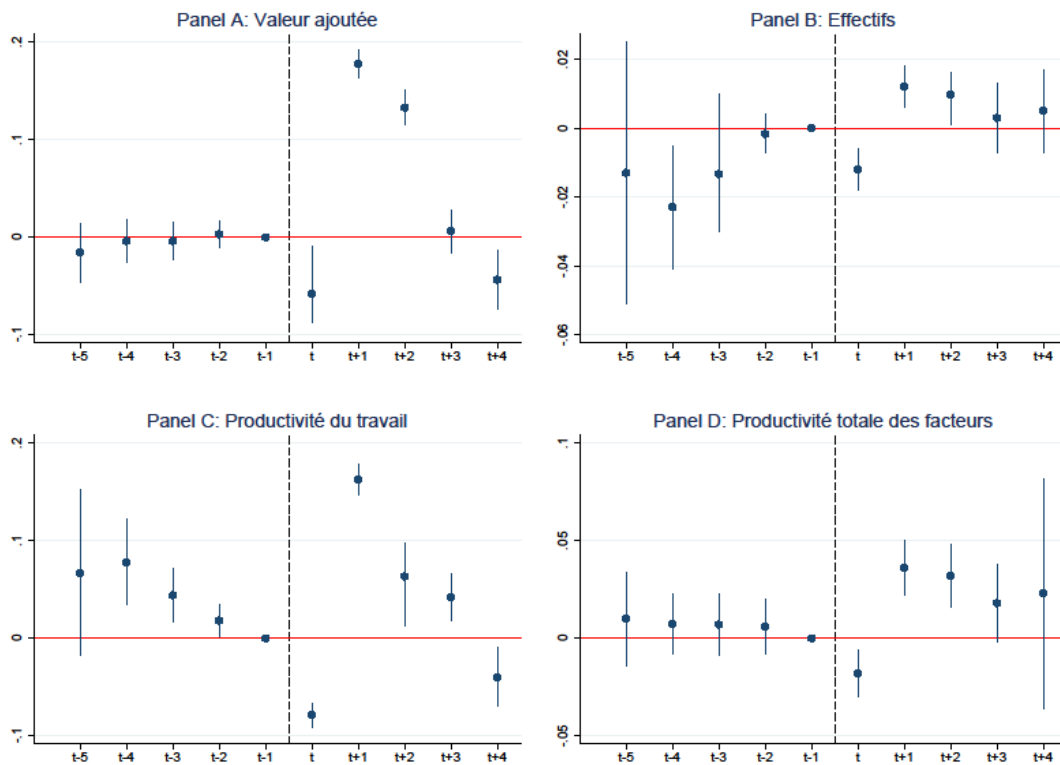
	Panel A: Effet sur la valeur ajoutée											Echantillon	
	t-5	t-4	t-3	t-2	t	t+1	t+2	t+3	t+4	t+5	Obs.	Communes	
A1: Très haut débit													
< 8Mbps	-0.834**	-0.641**	-0.376	-0.012	0.024	0.003	-0.283	-0.295	0.530	19.557	95499	9760	
> 8Mbps	0.114	-0.141	-0.062	-0.341	0.932	3.162**	4.514***	5.371***	8.391***	11.764***	161523	16335	
A2: Fibre													
< 8Mbps	-0.530	0.014	0.026	0.336	0.455	1.028*	1.364*	1.304	1.133	4.190	95499	9760	
Entre 8Mbps et 30Mbps	0.501	-0.750	0.117	0.659	1.761**	4.922**	11.746**	19.143*	34.939	-4.682	161523	16335	
> 30Mbps	-10.227*	-8.105	-3.186	3.211	10.417**	16.739**	19.927**	17.855	48.866*	278.400	43195	4353	
	Panel B: Effet sur les effectifs											Echantillon	
	t-5	t-4	t-3	t-2	t	t+1	t+2	t+3	t+4	t+5	Obs.	Communes	
B1: Très haut débit													
< 8Mbps	-0.679	-0.677	-0.263	-0.304	0.296	-0.317	0.424	0.240	1.458	18.685	95499	9760	
> 8Mbps	0.494	-0.359	-0.702	-0.517	0.519*	0.341	0.984	0.764	6.601***	11.298***	161523	16335	
B2: Fibre													
< 8Mbps	-0.791	-0.834	-1.184	-0.165	0.309	1.153	2.396**	2.966*	2.851	6.122	95499	9760	
Entre 8Mbps et 30Mbps	-1.663	-2.268	-1.903	-2.547**	1.352	1.381	5.554**	10.237**	18.496**	14.647	161523	16335	
> 30Mbps	-5.594	-10.996**	-4.832	-8.504**	7.201	5.237	-3.867	-8.065	-7.716	88.836	43195	4353	

Le tableau reporte les coefficients estimés pour les variables t-5 à t+4 (le modèle inclut aussi les variables t-8, t-7 et t-6 mais les résultats ne sont pas reportés). La spécification du modèle est identique au modèle de base et la méthode d'estimation retenue est celle de Sun et Abraham. Nous distinguons trois groupes en fonction du débit initial disponible (inférieur 8Mbps, entre 8 et 30 Mbps, supérieur à 30 Mbps). *, **, et *** signale la significativité à 10, 5 et 1% respectivement.

3.5. Analyse au niveau entreprise

Afin de conclure ce chapitre, nous étendons l'analyse en étudiant si l'arrivée du THD ou de la fibre a eu un effet sur les entreprises. L'Annexe B présente l'approche méthodologique retenue et les difficultés techniques rencontrées. Nous présentons ici uniquement les principales conclusions de cette analyse. Il convient néanmoins de noter que ces résultats doivent être pris avec de grandes précautions en raison du non-respect de l'hypothèse de tendance parallèle, qui semblent s'expliquer par un trop grand nombre d'observations (voir Annexe B pour une discussion de ce point).

Figure 5.3 : Effet du THD (30 Mbps ou plus) sur les performances des entreprises, tirage aléatoire

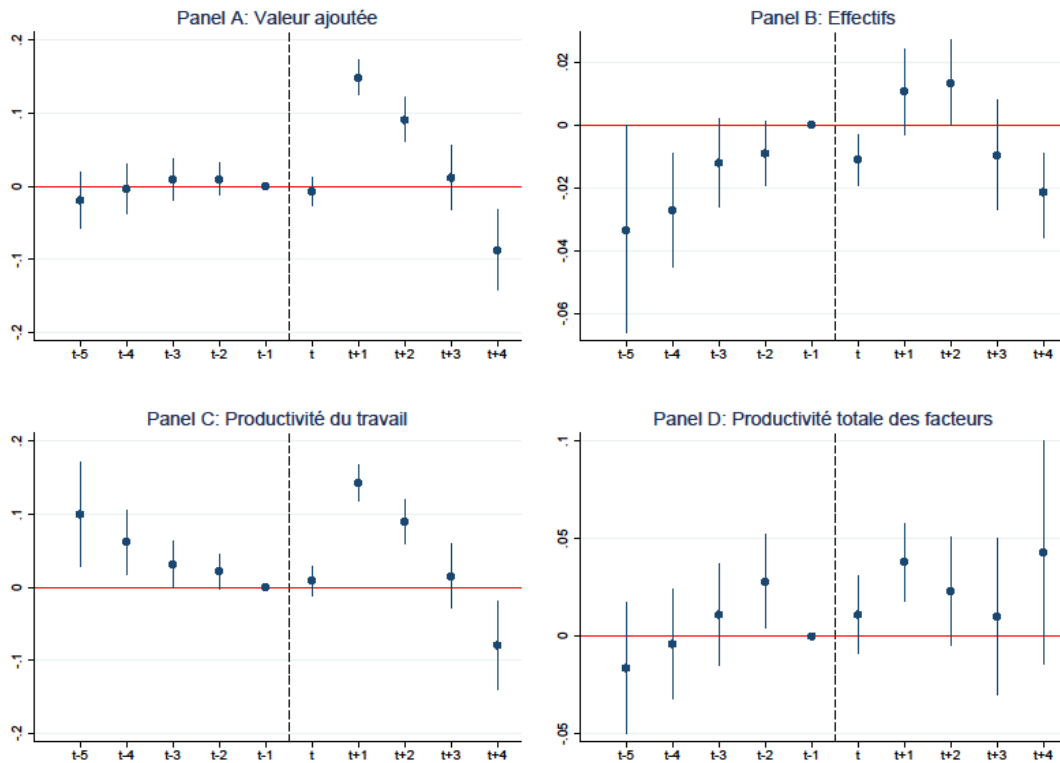


Note : La figure reporte pour chaque estimation les coefficients et les intervalles de confiance à 5%.

Les résultats extraits d'une analyse à partir d'un échantillon d'entreprises tirées aléatoirement sont présentés dans les figures 5.3 et 5.4. Cette méthode est imparfaite car le tirage aléatoire ne

permet pas d'avoir deux groupes parfaitement identiques (cf. Tableau 5.B.5 en Annexe). Néanmoins, cette méthode tend à corriger le problème d'absence de tendance parallèle. Plusieurs résultats méritent une attention particulière. L'effet du traitement est nul (pour la fibre) ou négatif (pour le THD) l'année du traitement (en t). L'effet du traitement devient très vite positif la première année pour l'ensemble des variables de performances et pour les deux traitements (THD et fibre). Ainsi l'effet du traitement au niveau entreprise n'apparaît pas uniquement sur la création de valeur mais aussi sur l'emploi et la productivité (contrairement aux évaluations communales). Néanmoins, cet effet positif se dissipe très vite. L'effet positif disparaît en général trois ans après le traitement et devient parfois négatif lors de la dernière période (Panels A et C de la Figure 5.3 ; Panels A, B, C de la Figure 5.4).

Figure 5.4 : Effet de la fibre (100 Mbps ou plus) sur les performances des entreprise, tirage aléatoire



Note : La figure reporte pour chaque estimation les coefficients et les intervalles de confiance à 5%.

4. Conclusion

L'objectif de ce chapitre est d'analyser l'effet du très haut débit et de la fibre sur les performances économiques des entreprises localisées en zone d'initiative publique. Afin de traiter cette question, nous avons étudié si la valeur ajoutée, l'emploi et la productivité (du secteur marchand) ont progressé suite au raccordement des communes au THD et à la fibre.

Nos résultats montrent que l'arrivée du THD et de la fibre ont eu des effets positifs sur la production de richesse dès la première année après le traitement et cet effet positif tend à être cumulatif. Trois ans après l'arrivée du THD, la valeur ajoutée générée par le secteur marchand a été augmentée de 7% pour les communes traitées (IC 5% : +2.7% ; +11.1%) par rapport à la situation ayant prévalu en l'absence de THD. L'arrivée de la fibre a eu des impacts économiques relativement proches avec une hausse de la valeur ajoutée du secteur productif 8.3% (IC : 0.0% ; 16.8%) trois ans après l'arrivée de cette technologie. Néanmoins, il convient de mettre un bémol à cette analyse.

Si le THD et la fibre ont stimulé la création de richesses, les résultats portant sur la création d'emplois sont plus tardifs à apparaître. L'arrivée de ces deux technologies semble avoir bénéficié à l'emploi mais les effets ont mis plus de temps à se matérialiser puisqu'il faut attendre entre 4 et 5 ans pour discerner un effet sur l'emploi productif. L'absence d'effet sur l'emploi à court-terme s'explique sans doute par un délai temporel entre la hausse d'activités et les décisions des entreprises à recruter si cette tendance se confirme.

Enfin, il n'apparaît aucun effet de l'arrivée de l'internet très haut débit ou de la fibre sur la productivité de la commune.

Dans un second temps, nous avons étudié la possibilité d'avoir des effets hétérogènes de l'arrivée du THD et de la fibre en fonction des caractéristiques des entreprises présentes dans la commune. Plusieurs résultats intéressants sont à souligner. Tout d'abord, l'effet positif observé sur la création de richesses et l'emploi est principalement dû à une hausse de l'activité des entreprises présentes avant l'arrivée d'internet et non pas par une arrivée d'entreprises plus performantes ou une sortie d'entreprises peu efficaces.

L'arrivée du THD a stimulé la création de valeur en priorité pour les grandes entreprises et les entreprises multi-établissements (même si toutes les entreprises en ont bénéficié). En revanche, l'effet sur l'emploi a été plus fort pour les PME. Les entreprises de l'industrie, de l'hôtellerie-

restauration et des TIC ont également bénéficié de l'entrée du THD. La fibre semble avoir bénéficié davantage aux PME et aux entreprises mono-établissement ainsi qu'aux entreprises industrielles.

Enfin, nous avons étudié le rôle des caractéristiques de la commune. L'analyse selon les caractéristiques des communes a mis en évidence que l'internet très haut débit et la fibre ont un effet économique rapide pour toutes les communes, à l'exception des communes isolées. Dans ce cas, seul le THD a un effet mais qui est tardif à se matérialiser. D'autre part, seules les communes ayant déjà accès à du bon haut débit ont réellement bénéficié de l'arrivée de ces nouvelles technologies.

Enfin, nous avons analysé l'effet du THD et de la fibre à un niveau plus désagrégé en s'intéressant aux entreprises. Cette analyse a posé des problèmes méthodologiques et les résultats doivent être pris avec précaution. Il apparaît de ces analyses que l'arrivée du THD et de la fibre a stimulé l'activité (valeur ajoutée, effectifs) et la productivité des entreprises. Néanmoins, cet effet est peu persistant, ce qui rentre en contradiction avec les analyses au niveau communal. Ce constat est en phase avec la comparaison des travaux existants (Section 5.1) qui a mis en évidence un effet plus important de l'accès à internet sur le développement régional plutôt que sur les entreprises prises individuellement. Une explication possible tient au fait qu'internet est un outil de coordination non seulement au sein mais aussi entre les entreprises et que les effets de débordement l'emportent sur les effets positifs purement individuels.

Annexe au chapitre 5

A.5-A / Tableaux et figures additionnels

Tableau 5.A.1 : Différence de moyenne pour les variables dépendantes

	Très haut débit (30 Mbps ou plus)			Fibre (100 Mbps ou plus)		
	Jamais	Traités	t-test	Jamais	Traités	t-test
Valeur ajoutée (<i>déviati on standard</i>)	20.27 (89.48)	66.12 (231.81)	66.71***	72.14 (264.69)	205.65 (627.87)	58.68***
Effectif (<i>déviati on standard</i>)	34.49 (142.34)	114.57 (360.26)	74.66***	126.58 (429.93)	350.72 (1001.1)	60.98***
Productivité (<i>déviati on standard</i>)	0.812 (2.355)	0.728 (0.971)	-10.13***	0.753 (1.767)	0.675 (0.537)	-5.85***
Observations	137627	97495		255000	19298	

Note : Le tableau reporte les moyennes par groupes des trois variables dépendantes pour l'analyse au niveau communal et les tests d'égalité de moyenne associés.

Tableau 5.A.2 : Effet du très haut débit (30 Mbps ou plus) et de la fibre sur les performances communales

	Effet du THD (> 30 Mbps)			Effet de la fibre (> 100 Mbps)		
	VA (1)	EMPL (2)	LP (3)	VA (4)	EMPL (5)	LP (6)
t-8	-1.239*** (0.384)	-0.625 (0.438)	0.079 (0.050)	-6.070*** (1.268)	-2.023* (1.198)	0.028** (0.014)
t-7	-0.796** (0.368)	-0.507 (0.472)	0.033 (0.022)	-2.573 (1.974)	-0.973 (1.144)	0.021 (0.014)
t-6	-0.801 (0.292)	-0.027 (0.377)	0.002 (0.011)	-5.080*** (1.541)	-2.820** (1.192)	0.012 (0.013)
t-5	-0.274 (0.560)	0.169 (0.462)	-0.004 (0.009)	-2.960* (1.575)	-2.714** (1.300)	0.002 (0.010)
t-4	-0.375 (0.505)	-0.418 (0.677)	-0.009 (0.013)	-2.611 (1.598)	-4.223*** (1.395)	-0.011 (0.009)
t-3	-0.211 (0.411)	-0.505 (0.632)	-0.009 (0.012)	-0.738 (1.370)	-3.833*** (1.038)	0.008 (0.007)
t-2	-0.288 (0.282)	-0.425 (0.265)	-0.006 (0.014)	1.458 (1.192)	-2.339* (1.92)	-0.010 (0.007)
t	0.669* (0.402)	0.389* (0.227)	-0.015 (0.014)	3.778*** (1.243)	3.038** (1.328)	0.015 (0.010)
t+1	2.532** (1.033)	0.074 (0.385)	-0.011 (0.009)	7.765*** (2.096)	2.979** (1.505)	0.002 (0.012)
t+2	3.598*** (1.328)	0.843* (0.498)	-0.007 (0.013)	12.670** (5.671)	3.240 (2.693)	-0.004 (0.012)
t+3	4.552*** (1.404)	0.716 (0.635)	-0.012 (0.015)	17.157** (8.769)	6.246 (3.954)	0.027 (0.019)
t+4	7.779*** (2.015)	6.328*** (1.093)	0.002 (0.018)	36.699** (17.388)	11.425** (5.641)	0.021 (0.022)
t+5	12.147*** (3.832)	11.582*** (2.232)	0.016 (0.017)	66.490 (61.466)	33.697 (25.905)	0.010 (0.037)
Obs.	258001	258001	217445	301196	301196	260115
# Commune	26327	26327	23839	30680	30680	28177
R ² (ajusté)	0.93	0.98	0.52	0.96	0.99	0.53

Le modèle estime l'effet de l'entrée du très haut débit sur la valeur ajoutée communale (colonnes 1 et 2) et les effectifs totaux par communes (colonnes 3 et 4). Le modèle inclut des effets fixes communes et années. La méthode d'estimation est celle de Sun et Abraham. Chaque estimation inclut comme variables de contrôle la population totale de la commune, le degré de diversification et la spécialisation par secteurs (voir le texte pour un détail sur la construction de ces variables). Les résidus sont clustérisés au niveau communal. *, **, et *** signale la significativité à 10, 5 et 1% respectivement.

Tableau 5.A.3 : Décomposition des activités par secteur

Secteur	Code (NAF 17)	Divisions (NAF 88)	Intitulé
Industrie	C1	10-12	Fabrication de denrées alimentaires, de boissons et de produits à base de tabac
	C2	19	Cokéfaction et raffinage
	C3	26-28	Fabrication d'équipements électriques, électroniques, informatiques ; fabrication de machines
	C4	29-30	Fabrication de matériels de transport
	C5	13-18,20-25, 31-33	Fabrication d'autres produits industriels
Construction	FZ	41-43	Construction
Commerce	GZ	45-47	Commerce ; réparation d'automobiles et de motocycles
Transport	HZ	49-53	Transports et entreposage
Hotels	IZ	55-56	Hébergement et restauration
Informatique	JZ	58-63	Information et communication
Finance	KZ	64-66	Activités financières et d'assurance
Exclu de l'analyse	AZ	01-03	Agriculture, sylviculture et pêche
	DE	05-09,35-39	Industries extractives, énergie, eau, gestion des déchets et dépollution
	LZ	68	Activités immobilières
	MN	69-82	Activités scientifiques et techniques ; services administratifs et de soutien
	RU	90-99	Autres activités de services
	OQ	84-88	Administration publique, enseignement, santé humaine et action sociale

Note : Secteur fait référence à la classification dans l'analyse et Code et Divisions font référence à la classification NAF 17 et NAF 88 de l'INSEE, respectivement.

A.5-B / Analyse au niveau entreprise : Précisions méthodologiques et résultats

A.5-B.1 Précisions méthodologiques

Appariement des données FARE et de couverture internet

L'analyse des performances des entreprises repose sur les données FARE, construites au niveau entreprise. L'appariement entre les deux bases ne pose aucun problème pour les entreprises mono-établissement dans lesquelles l'entreprise et l'établissement se confondent. En revanche, la procédure d'appariement est plus complexe pour les entreprises multi-établissements. Pour ces dernières, la commune considérée dans la base de données FARE est la commune du siège social de l'entreprise. Or, il est possible que l'activité économique principale soit réalisée dans d'autres établissements (localisés dans d'autres communes). Ce point est important car une entreprise multi-établissements peut bénéficier de la montée en débit d'internet (si ses établissements accèdent à un débit plus important) sans que cela soit apparent dans les données (si le débit est inchangé dans la commune du siège social). A l'opposé, l'accès à internet peut sembler progresser car la commune du siège social accède à un débit plus important, sans que les autres établissements n'en bénéficient. Le risque est donc d'imputer un « faux » traitement pour les entreprises multi-établissements.

Face à ce problème, trois choix sont possibles : (i) soit nous utilisons des observations par entreprise mais cela implique de repenser notre mesure de traitement pour les entreprises multi-établissements ; (ii) soit nous construisons des observations par établissement en utilisant une approche proche de celle développée précédemment ; (iii) soit nous excluons les entreprises multi-établissements.

Nous avons choisi de retenir la première solution, c'est-à-dire de considérer l'entreprise comme unité d'analyse tout en conservant les entreprises multi-établissements. La seconde solution (construire une analyse au niveau de l'établissement) est techniquement possible selon la même méthodologie qu'utilisée pour l'analyse communale. Le problème est que l'étude au niveau individuel implique non seulement de construire des variables dépendantes à ce niveau mais aussi des variables de contrôle. Or, ce dernier point implique des choix complexes. Par exemple, une mesure aussi simple que la taille devient difficile à construire car se pose alors la question de savoir si nous considérons la taille de l'établissement ou de l'entreprise dans son ensemble. Ce problème se retrouve pour toutes les variables mesurées au niveau de l'entreprise. Enfin, exclure les entreprises multi-établissements peut poser problème si ces entreprises bénéficient différemment

du THD et de la fibre. Or, les résultats au niveau communal tendent à mettre en évidence une différence d'effets pour les deux types d'entreprises. En outre, il est aisé d'incorporer cette analyse dans le cadre empirique retenu en excluant les entreprises multi-établissements de l'analyse (comme cela est fait plus tard dans le chapitre).

Nous incorporons dans ce rapport à la fois les entreprises mono- et multi-établissements. Pour ce faire, nous construisons une mesure de l'accès aux THD et à la fibre qui englobe ces deux situations (voir plus bas). Notre échantillon inclut toutes les entreprises ayant leur siège social en zone d'initiative publique. Pour les entreprises multi-établissements, cela implique que nous considérons des entreprises ayant des établissements qui peuvent être en zone d'initiative privée dans la mesure où le siège social est en zone d'initiative publique. De manière symétrique, des établissements situés en zone d'initiative publique sont exclus de l'analyse si l'entreprise a son siège social en zone d'initiative privée. Au-delà des avantages techniques et en termes de transparence de cette méthode⁹⁰, il n'est pas aberrant de considérer que la qualité de la connexion du siège est essentielle car ce dernier est sans doute le centre névralgique pour l'organisation interne de l'entreprise.

Construction des variables

Nous construisons des variables expliquées proches de celles utilisées dans l'analyse au niveau communal. Nous considérons pour chaque entreprise la valeur ajoutée et les effectifs de l'entreprise. Nous construisons en outre deux indicateurs de productivité. Le premier est la productivité du travail définie comme le rapport entre la valeur ajoutée et les effectifs pour chaque entreprise. Nous construisons en outre une mesure de la productivité totale des facteurs à partir de la méthode de Levinsohn et Petrin (2003).⁹¹

⁹⁰ Une approche alternative serait de définir un seuil d'activité à partir duquel on aurait considéré les entreprises comme agissant en priorité en zone d'initiative publique. Par exemple, nous aurions pu considérer que les entreprises ayant plus de la moitié de leur activité en zone publique doivent être incorporées dans l'analyse (quand bien même le siège social est en zone privée). Cette approche certes attrayante pose deux problèmes. D'une part, la définition du seuil est arbitraire (faut-il considérer les entreprises ayant au moins un établissement en zone publique, celles dont la majorité de l'activité est en zone publique, etc. ?). D'autre part, la classification à partir d'un seuil aurait induit un échantillon mouvant en raison des réallocations d'activités au sein du groupe au cours de la période étudiée. Ainsi, une entreprise aurait pu être incluse dans l'étude certaines années puis en sortir et inversement. L'approche par le siège social évite d'avoir un échantillon instable.

⁹¹ Les données de valeur ajoutée sont déflatées à partir des déflateurs par secteur (NAS 129). Les données de salaires l'ont été par l'indice annuel du coût horaire par secteurs d'activités (NAF, révisée). Les données d'investissement ont été corrigées de l'inflation en utilisant les prix à la production des biens d'investissements. Les données pour les biens intermédiaires grâce au prix à la production pour les biens intermédiaires. Toutes les données sont issues de l'INSEE.

Notre principale variable d'intérêt est l'accès au THD ou à la fibre au niveau de la commune d'implantation de l'entreprise. Afin de construire cette variable, nous suivons l'approche développée dans le chapitre sur les usages numériques. La mesure initiale de couverture (selon le débit considéré) est construite comme suit :

$$M_{i,t} = \sum_{e \in i} \theta_{e,t} C_{e \in k} \quad (5.9)$$

Avec e un établissement de l'entreprise i localisé dans la commune k à la période t . $\theta_{e,t}$ est le poids de l'établissement e dans l'activité totale de l'entreprise i (tel que défini dans l'équation 5.2). $C_{e \in k}$ est un indicateur de la couverture en THD/fibre de la commune k . L'avantage de cette mesure est qu'elle peut s'appliquer aussi aux entreprises mono-établissement pour lesquelles $\theta_{e,t} = 1$ (en raison de l'existence d'un établissement unique) ce qui implique pour ces entreprises que : $M_{i,t} = C_{i \in k}$.

Afin de faciliter l'analyse du programme, nous considérons une variable muette prenant la valeur de 1 si une entreprise est traitée, c'est-à-dire qu'elle a au moins un de ses établissements qui a accès au débit considéré. La variable de traitement est donc construite comme suit :

$$I(M)_{i,t} = \begin{cases} 1 & \text{si } M_{i,t} > 0 \\ 0, & \text{sinon} \end{cases} \quad (5.10)$$

Cette mesure binaire a l'avantage d'être moins sensible aux réallocations internes au sein d'une entreprise qui peut affecter la valeur de $M_{i,t}$ en raison d'effets de structure. En outre, elle est plus aisée à utiliser dans le cadre d'une évaluation d'impact.

Nous estimons le modèle économétrique présenté à l'équation 5.8 avec comme variables dépendantes la valeur ajoutée, l'effectif et deux mesures de productivité (productivité du travail et productivité totale des facteurs). Toutes les variables sont mesurées en logarithme. Le traitement est défini ci-dessus.

Nous ajoutons plusieurs variables de contrôle. D'une part, nous incluons les variables de contrôle utilisées dans l'analyse communale et décrites dans la section 5.2.2. Nous contrôlons également le secteur d'activités de l'entreprise.

A.5-B.2 Résultats et problèmes économétriques

Structure de l'échantillon et statistiques descriptives

Le tableau 5.B.1 présente la distribution des entreprises en fonction des cohortes étudiées. Les entreprises localisées dans des communes traitées avant 2013 sont exclues de l'analyse faute d'information sur leur couverture internet avant cette date. Les distributions des cohortes sont cohérentes avec celles du tableau 5.2. Près de 40% des entreprises sont traitées entre 2014 et 2018 pour le THD contre seulement 12% pour la fibre. La part d'entreprises non traitées est relativement importante pour la fibre (85%) alors qu'elle est autour de 20% pour le THD.

Tableau 5.B.1 : Echantillon de l'analyse au niveau entreprise par cohorte

	Très haut débit		Fibre	
	Nombre	%	Nombre	%
Déjà traitées en 2013	4,517,727	42.55	287,430	2.71
Traitées entre 2014 et 2018	4,069,206	38.32	1,272,543	11.98
- 2014	1,469,612	13.84	130,799	1.23
-2015	1,109,962	10.45	299,642	2.82
-2016	766,717	7.22	275,159	2.59
-2017	506,072	4.77	398,384	3.75
-2018	216,843	2.04	168,559	1.59
Jamais traitées	2,031,749	19.13	9,058,407	85.31

Note : Le tableau reporte le nombre d'entreprises en fonction de la cohorte : traitée avant 2013, traitée entre 2014 et 2018 et non-traitée en 2018.

L'analyse des statistiques descriptives (tableau 5.B.2) met en évidence la taille réduite de nombreuses entreprises. L'effectif moyen dépasse à peine un travailleur (sachant que le minimum est d'un en raison de la présence du propriétaire-manager). La valeur ajoutée moyenne est elle-même très faible mais avec de très fortes disparités (comme en témoigne la valeur élevée de l'écart standard). Les niveaux de productivité sont également très disparates.

Tableau 5.B.2 : Statistiques descriptives

Panel A: Très haut débit					
	Obs	Mean	Erreur std.	Min	Max
Valeur ajoutée	5,605,202	0.002	5.991	0.000	41.4
Effectifs	4,560,026	1.498	2.158	1.000	4411
Prod. travail	4,199,639	0.001	3.934	0.000	16.5
Prod. totale	3,080,175	0.015	1.923	0.000	167.7

Panel B: Fibre					
	Obs	Mean	Erreur std.	Min	Max
Valeur ajoutée	5,605,202	0.003	6.184	0.000	41.4
Effectifs	4,560,026	1.642	2.368	1.000	4411
Prod. travail	4,199,639	0.002	3.905	0.000	16.5
Prod. totale	3,080,175	0.015	1.886	0.000	167.7

Estimation de l'effet global du THD et de la fibre

Le tableau 5.B.3 présente les résultats de la méthode de différence-en-différence échelonnée au niveau des entreprises à la fois pour le THD (colonnes 1 à 4) et pour la fibre (colonnes 5 à 8).

La validité de la méthode de différence-en-différence échelonnée implique le respect de l'hypothèse de tendances parallèles entre les groupes de traitement et les groupes de contrôle. Un moyen de valider cette hypothèse, à partir de la méthode de Sun et Abraham (2020), consiste à regarder la significativité des coefficients associés aux périodes prétraitement (de t-8 à t-2 en l'espèce). Si l'absence de significativité de ces coefficients n'est pas une assurance parfaite quant au respect de cette hypothèse fondamentale, son rejet signale que cette hypothèse n'est pas respectée. Or, comme l'illustre le tableau 5.B.3, le modèle estimé rejette l'hypothèse de tendances parallèles. Ce résultat est problématique car il remet en cause la validité du modèle employé, et donc, les conclusions qui pourraient en être tirées.

Tentative d'explications du non-respect de l'hypothèse de tendances parallèles

Nous proposons dès lors de mieux comprendre pourquoi la méthode de différence-en-différence choisie ici semble mal adaptée afin de trouver une solution. Il n'y a pas a priori de raison claire expliquant le rejet de l'hypothèse de tendances parallèles, le traitement étant relativement

exogène au niveau des entreprises.⁹² La décision de l'arrivée du THD ou de la fibre a été le plus souvent prise plusieurs années avant sa mise en œuvre. Si la mise en place du THD et de la fibre peut induire des travaux perturbant l'activité des entreprises à court-terme, il est peu probable que ces perturbations soient présentes sur presque dix ans.

Dans des analyses (non reportées mais disponibles au besoin) nous avons évalué diverses sources possibles du problème. Nous avons d'abord exclu les observations extrêmes (1-99%, 5-95%) pour les variables dépendantes sans que cela n'altère nos résultats. Nous avons ensuite restreint notre échantillon aux entreprises mono-établissement. Le traitement tel que construit dans cette analyse au niveau des entreprises pourrait être la source de perturbation à l'origine de l'absence de tendances parallèles (par exemple, les entreprises ayant le plus d'établissements ont une probabilité plus forte d'être traitées en premières). Néanmoins, le fait de se restreindre aux entreprises mono-établissement, pour lesquelles le traitement est clair, ne change rien à nos conclusions avec des résultats très proches de ceux présentés au tableau 5.B.3.

⁹² Bien que le déploiement du THD et de la fibre ne soit pas exogènes au niveau des communes en termes de temporalité, l'exogénéité est sans doute plus forte au niveau des entreprises (même si les caractéristiques des communes peuvent jouer sur les performances des entreprises). Les principaux facteurs expliquant le déploiement du THD/fibre sont des caractéristiques quasi-fixes des communes (taille, géographie) qui sont prises en compte dans les analyses par les effets-fixes (effets-fixes au niveau commune dans l'analyse municipale et entreprise dans l'analyse microéconomique).

Tableau 5.B.3 : Effet du THD et de la fibre sur les performances des entreprises, échantillon total

	Accès au THD (30 Mbps ou plus)				Accès à la fibre (100 Mbps ou plus)			
	Ln(VA)	ln(EFF)	ln(PT)	ln(PTF)	Ln(VA)	ln(EFF)	ln(PT)	ln(PTF)
t-8	-0,0951*** (0,0180)	-	-	-0,0418** (0,0169)	-0,1266*** (0,0107)	-	-	-0,0085 (0,0120)
t-7	0,0582*** (-0,0078)	-	-	-0,0271*** (0,0063)	-0,1101*** (0,0058)	-	-	0,0027 (0,0047)
t-6	-0,0489*** (0,0052)	-	-	-0,0239*** (0,0038)	-0,0887*** (0,0048)	-	-	0,0070** (0,0030)
t-5	-0,0640*** (0,0029)	-0,0960** (0,0431)	0,0400*** (0,0114)	-0,0274*** (0,0028)	-0,0654*** (0,0041)	-0,0298*** (0,0042)	-0,0580*** (0,0089)	0,0040* (0,0023)
t-4	-0,0676*** (0,0029)	-0,1027*** (0,0222)	0,0305*** (0,0057)	-0,0250*** (0,0021)	-0,0354*** (0,0034)	-0,0208*** (0,0020)	-0,0400*** (0,0046)	0,0028 (0,0017)
t-3	-0,0488*** (0,0025)	-0,0716*** (0,0120)	0,0177*** (0,0033)	0,0117*** (0,0018)	-0,0162*** (0,0032)	-0,0121*** (0,0013)	-0,0270*** (0,0029)	0,0046*** (0,0016)
t-2	-0,0316*** (0,0018)	-0,0314*** (0,0060)	-0,0012 (0,0020)	-0,0063*** (0,0014)	-0,0041* (0,0023)	-0,0045*** (0,0008)	-0,0193*** (0,0019)	0,0021 (0,0013)
t	-0,0991*** (0,0023)	-0,0194** (0,0088)	-0,0778*** (0,0024)	-0,0146*** (0,0016)	-0,0329*** (0,0038)	-0,0155*** (0,0007)	-0,0973*** (0,0024)	-0,0185*** (0,0013)
t+1	0,0717*** (0,0021)	0,0544*** (0,0089)	0,0756*** (0,0022)	0,0104*** (0,0017)	0,0446*** (0,0034)	0,0068*** (0,0008)	0,0954*** (0,0018)	0,0098*** (0,0014)
t+2	0,0502*** (0,0027)	0,0876*** (0,0126)	0,0526*** (0,0029)	0,0057*** (0,0021)	0,0388*** (0,0050)	0,01018*** (0,0011)	0,1078*** (0,0025)	0,0078*** (0,0015)
t+3	0,0324*** (0,0034)	0,1051*** (0,0158)	0,0378*** (0,0035)	0,0023 (0,0025)	0,0199*** (0,0082)	0,0117*** (0,0012)	0,1114*** (0,0033)	0,0037** (0,0018)
t+4	-0,0045 (0,0044)	0,1387*** (0,0263)	0,0024 (0,0047)	0,0052 (0,0032)	0,0187 (0,0149)	0,0128*** (0,0018)	0,0911*** (0,0044)	0,0040* (0,0024)
Obs.	5038661	4123532	3781285	2631042	8918192	6844481	6325325	4891776
R ² (ajusté)	0,90	0,95	0,85	0,67	0,91	0,92	0,86	0,67

Le modèle estime l'effet de l'arrivée du THD (colonnes 1 à 4) et de la fibre (colonnes 5 à 8) sur la valeur ajoutée (colonnes 1 et 5), les effectifs totaux (colonnes 2 et 6), la productivité du travail (colonnes 3 et 7) et la productivité totale des facteurs (colonnes 4 et 8). Le modèle inclut des effets fixes communes et années ainsi que des muettes par secteurs d'activités. Chaque estimation inclut comme variables de contrôle la population totale de la commune, le degré de diversification et la spécialisation par secteurs. Les résidus sont clustérisés au niveau communal. *, **, et *** signale la significativité à 10, 5 et 1% respectivement.

La méthode la plus simple pour évaluer l'existence de tendances parallèles entre deux groupes consiste à reporter sur un graphique les évolutions moyennes de la variable dépendante avant le traitement pour les unités traitées et les unités appartenant au groupe de contrôle. Cette approche est difficilement faisable dans notre cas en raison du nombre très important d'observations (que

veut dire la moyenne ?) et du fait que le traitement n'est pas unique mais multiples avec des modifications dans les compositions des groupes (certaines entreprises initialement « contrôle » deviennent « traitées » plus tard). Néanmoins, nous avons suivi cette logique pour évaluer l'origine de l'absence de tendances parallèles. Ainsi, nous avons construit pour chaque entreprise l'évolution de chacune de nos variables dépendantes (valeur ajoutée, effectifs, productivité du travail et productivité totale des facteurs) en t-1, t-2, t-3 et t-4.⁹³ Ensuite, nous avons classé les entreprises entre groupe de traitement et groupe de contrôle pour chaque traitement sachant que le statut des entreprises peut changer en fonction des traitements (cf. tableau ci-dessous pour la classification des entreprises en fonction des cohortes et des traitements). Enfin, nous présentons la distribution des variables créées pour les périodes de t-1 à t-4 pour les entreprises de chacun des deux groupes. En l'absence de tendances parallèles, nous devrions voir des distributions différentes entre les deux groupes et, idéalement, pouvoir identifier l'origine des différences en affinant l'analyse (par sous-groupes d'entreprises par exemple).

Tableau 5.B.4 : Classement des entreprises pour tester les tendances parallèles

Année du traitement	Cohorte: Entreprises traitées						
	Avant 2013	En 2014	En 2015	En 2016	En 2017	En 2018	Après 2018
2014		T	C	C	C	C	C
2015			T	C	C	C	C
2016				T	C	C	C
2017					T	C	C
2018						T	C

Note : T se réfère au groupe de traitement, C au groupe de contrôle et l'absence de lettre indique que les entreprises sont exclues.

L'analyse graphique (non reportée mais disponible au besoin) ne met pas en évidence de différences visibles entre les deux groupes pour l'ensemble des périodes (de t-1 à t-4) et des variables considérées (valeur ajoutée, effectif, productivité). Autrement dit, il n'est pas évident d'identifier dans quelle mesure les groupes de contrôle et de traitement diffèrent dans leur comportement avant le traitement.

Plus fondamentalement, cette analyse graphique met en évidence que les techniques les plus récentes pour corriger le problème d'absence de tendances parallèles pourraient ne pas suffire dans

⁹³ Pour chaque période t-n, nous avons construit l'évolution de la variable Y entre t-n et t-n-1.

notre cas (au-delà du fait qu'elles sont le plus souvent développées pour des modèles de différence-en-différence simple). Ainsi la méthode de Rambachan et Roth (2020) consistant à extrapoler l'évolution prétraitement pour le groupe de contrôle - modulant un intervalle de confiance - ou celle de Bilinski et Hatfield (2019) qui proposent de faire des pré-tendances différentes entre les groupes de contrôle et de traitement risquent de ne pas solutionner notre problème.

Estimation sur un échantillon aléatoire

Une explication possible du rejet de l'hypothèse de tendances parallèles tient sans doute au nombre très important d'observations qui peuvent rendre statistiquement significatives des différences minimales entre les deux groupes. Afin de contourner ce problème, nous avons tiré de manière aléatoire environ 10% des entreprises et avons re-estimé le modèle sur ce sous-échantillon. Comme l'illustre le tableau 5.B.5 ci-dessous, les entreprises tirées aléatoirement ne sont pas parfaitement identiques à l'échantillon total dans leurs caractéristiques. Par exemple, l'échantillon aléatoire comporte beaucoup plus d'entreprises traitées en THD (plus de la moitié) que ce que reflète l'échantillon total. Les entreprises tirées aléatoirement sont aussi en moyenne plus petites et plus souvent des entreprises mono-établissement opérant dans la catégorie « autres secteurs ». Ces différences de caractéristiques impliquent donc de considérer les résultats avec prudence.

Les résultats présentés dans les figures 5.3 et 5.4 dans la section 3.5, ainsi que dans le tableau 5.B.6 ci-dessous soulignent que le tirage aléatoire permet de résoudre au moins partiellement le problème d'absence de tendances parallèles. Nous observons que les entreprises traitées et celles appartenant au groupe de contrôle ont des évolutions similaires avant le traitement pour les variables de valeur ajoutée (Panel A, en haut à gauche) et de productivité totale des facteurs (Panel D, en bas à droite). L'absence de différence est moins claire pour les variables d'effectifs (panel B) et de productivité du travail (Panel C).

De manière intéressante, nous retrouvons des estimations proches pour les coefficients associés aux périodes après le traitement par rapport à celles obtenues dans le tableau 5.B.3 sur l'échantillon total.

Tableau 5.B.5 : comparaison entre le groupe tiré aléatoirement et l'échantillon total

	Echantillon total	Tirage aléatoire	Reste des obs.
a) Distributions des cohortes			
<i>Très haut débit</i>			
Traité avant 2014	42.5	24.1	44.8
<i>Traité entre 2014 et 2018</i>	<i>38.3</i>	<i>51.7</i>	<i>36.7</i>
-En 2014	13.8	12.9	14.0
-En 2015	10.5	10.7	10.4
-En 2016	7.2	10.0	6.9
-En 2017	4.8	10.3	4.1
-En 2018	2.0	7.8	1.3
Jamais traité	19.1	24.1	18.5
<i>Fibre</i>			
Traité avant 2014	2.7	1.6	2.9
<i>Traité entre 2014 et 2018</i>	<i>12.0</i>	<i>11.1</i>	<i>12.0</i>
-En 2014	1.2	1.1	1.3
-En 2015	2.8	1.9	2.9
-En 2016	2.6	2.2	2.6
-En 2017	3.8	3.5	3.8
-En 2018	1.6	2.5	1.5
Jamais traité	85.3	87.3	85.1
b) Variables dépendantes			
Valeur ajoutée	0.018	0.007	0.019
Effectifs	3.811	2.137	4.020
Prod. travail	0.003	0.002	0.003
Prod. Totale	0.025	0.036	0.024
c) Autres caractéristiques			
% multi-établissement	0.068	0.038	0.072
Classe par taille			
-Petites entreprises	0.680	0.797	0.665
-Moyennes entreprises	0.315	0.195	0.330
-ETI	0.015	0.006	0.016
-Grandes entreprises	0.004	0.001	0.004
Secteurs			
-Commerce	0.197	0.190	0.198
-Construction	0.180	0.171	0.181
-Finance	0.029	0.026	0.029
-Hôtellerie-restauration	0.077	0.076	0.077
-Industrie	0.082	0.070	0.083
-TIC	0.019	0.026	0.018
-Transport	0.025	0.021	0.026
-Autre	0.392	0.420	0.388
<i>Obs.</i>	<i>10 618 380</i>	<i>1 176 396</i>	<i>9 441 984</i>

Tableau 5.B.6 : Effet du THD et de la fibre sur les performances des entreprises, échantillon aléatoire

	Accès au THD (30 Mbps ou plus)				Accès à la fibre (100 Mbps ou plus)			
	Ln(VA)	ln(EFF)	ln(PT)	ln(PTF)	Ln(VA)	ln(EFF)	ln(PT)	ln(PTF)
t-5	-0,016 (0,015)	-0,013 (0,019)	0,067 (0,043)	-0,100 (0,012)	-0,019 (0,019)	-0,034** (0,017)	0,100*** (0,036)	-0,016 (0,017)
t-4	-0,004 (0,011)	-0,023** (0,009)	0,078*** (0,022)	0,007 (0,008)	-0,003 (0,017)	-0,027*** (0,009)	0,062*** (0,022)	-0,004 (0,014)
t-3	-0,004 (0,010)	-0,020*** (0,005)	0,044*** (0,014)	0,007 (0,008)	0,009 (0,014)	-0,012* (0,007)	0,031* (0,016)	0,011 (0,013)
t-2	0,003 (0,007)	-0,002 (0,003)	0,018** (0,008)	0,006 (0,007)	0,010 (0,011)	-0,009* (0,005)	0,022* (0,012)	0,028** (0,012)
t	-0,088*** (0,006)	-0,012*** (0,003)	-0,079*** (0,006)	-0,018*** (0,006)	-0,007 (0,010)	-0,011*** (0,004)	0,009 (0,010)	0,011 (0,010)
t+1	0,178*** (0,007)	0,012*** (0,003)	0,163*** (0,008)	0,036*** (0,007)	0,149*** (0,012)	0,011** (0,007)	0,143*** (0,012)	0,038*** (0,010)
t+2	0,113*** (0,009)	0,008** (0,004)	0,098*** (0,009)	0,032*** (0,008)	0,091*** (0,015)	0,013 (0,007)	0,090*** (0,015)	0,023 (0,014)
t+3	0,0059*** (0,011)	0,003 (0,005)	0,042*** (0,012)	0,018* (0,010)	0,012 (0,022)	-0,010 (0,009)	0,015 (0,022)	0,010 (0,020)
t+4	-0,044*** (0,015)	0,005 (0,006)	-0,070*** (0,015)	0,023* (0,030)	-0,087*** (0,028)	-0,009 (0,014)	-0,079** (0,031)	0,043 (0,029)
Obs. (aléa)	576973	545310	485309	250326	807045	696330	623647	380162
R ² (ajusté)	0,88	0,85	0,81	0,65	0,89	0,87	0,82	0,66
Obs. (ens)	5038661	4123532	3781285	2631042	8918192	6844481	6325325	4891776

Le modèle estime l'effet de l'arrivée du THD (colonnes 1 à 4) et de la fibre (colonnes 5 à 8) sur la valeur ajoutée (colonnes 1 et 5), les effectifs totaux (colonnes 2 et 6), la productivité du travail (colonnes 3 et 7) et la productivité totale des facteurs (colonnes 4 et 8). L'analyse porte sur un échantillon aléatoire de 10% des entreprises. Le modèle inclut des effets fixes communes et années ainsi que des muettes par secteurs d'activités. Chaque estimation inclut comme variables de contrôle la population totale de la commune, le degré de diversification et la spécialisation par secteurs. Les résidus sont clustérisés au niveau communal. *, **, et *** signale la significativité à 10, 5 et 1% respectivement.

Chapitre 6. Performance d'innovation

1. Cadre théorique et revue de littérature

1.1. L'innovation: définitions

Pour étudier l'effet de l'accès au THD sur l'innovation, il faut d'abord définir cette notion. L'innovation est un élément central de la nouvelle théorie de la croissance (Aghion and Howitt, 1992), par l'introduction de nouveaux biens et services elle permet d'augmenter durablement la productivité par travailleur. Elle constitue la force de destruction créatrice énoncée par Schumpeter (1943), où la création de nouveaux marchés rendue possible par l'innovation en rend d'autres obsolètes et, par-là, entraîne une réorganisation de l'espace économique vers un nouvel équilibre plus productif.

Traditionnellement on distingue deux types d'innovations: les innovations de produit et de procédé. Les premières concernent la création d'un nouveau produit ou une amélioration de la qualité d'un produit existant. Les innovations de procédé portent sur le processus productif et correspondent à un moyen de produire à moindre coût, augmentant ainsi la productivité des facteurs de production.

En formalisant un peu, considérons la fonction de production suivante: $f(A, K, L)$ où la production d'une entreprise dépend de sa productivité A et des niveaux de facteurs travail L et capital K . La fonction de demande quant à elle peut être représentée par $D(p, q)$ où p est le prix et q est un vecteur d'attributs du bien produit. Alors que les innovations produits agissent sur la demande, par exemple en augmentant les attributs de qualité q du produit vendu et augmentant ainsi la demande, les innovations de procédé agissent sur la fonction de production en augmentant la productivité du capital ou du travail, par exemple ici en augmentant A . Dans les deux cas, et toutes choses égales par ailleurs, ces innovations ont pour effet d'augmenter la compétitivité des entreprises vis à vis de leurs concurrentes.

Toute innovation combine nécessairement deux éléments: a) la nouveauté, et b) le « succès ». La nouveauté va de soi car c'est l'élément constitutif de l'innovation, mais l'échelle à laquelle l'invention est nouvelle peut varier, comme détaillé ensuite. Concernant le « succès », séparons l'invention, qui est la création d'un nouveau produit ou procédé, de l'innovation. L'idée de succès

est indispensable car toute invention n'est pas une innovation: un nouveau produit peut ne pas rencontrer de demande, tout comme un nouveau processus de fabrication peut ne pas permettre de produire à moindre coût. Par cette définition, Une innovation est donc une invention qui a du succès. Ces idées seront utiles lorsque l'on décrira les mesures que l'on utilise pour mesurer l'innovation.

Revenons sur la nouveauté de l'innovation: elle peut avoir différentes échelles. Généralement, l'idée d'innovation se réfère à des inventions qui sont nouvelles au niveau international: un produit ou procédé dont aucun autre concurrent au monde ne dispose. Ce type d'innovation peut être vu comme forte. Mais se contenter de ce type d'innovation peut être restrictif. En effet, si les marchés sont segmentés alors l'introduction d'une invention existant sur un autre marché peut être considérée comme une innovation car elle offre de fait un avantage compétitif unique à l'entreprise qui l'implémente. Même au niveau de l'entreprise, implémenter une nouveauté qui n'est nouvelle que pour elle seule réduit néanmoins son écart de compétitivité avec ses concurrentes, et par-là peut être vue comme une innovation, bien qu'extrêmement locale. Il est clair néanmoins que plus l'échelle de l'innovation est locale, et moins son impact est important. Les innovations les plus importantes étant celles internationales permettant des prises autrement plus grandes. Il ne faut néanmoins pas négliger les innovations les plus locales car elles peuvent mener à des augmentations de productivité non-négligeables pour les entreprises concernées. Ces différences d'échelles de l'innovation seront importantes pour comprendre les différences entre les outils utilisés pour mesurer l'innovation.

1.2. Très haut débit et innovation: quelques mécanismes

Avec ces notions d'innovation en main, comment l'accès à l'internet en général, et au très haut débit en particulier, peut-il affecter la performance d'innovation des entreprises ?

Tout d'abord il faut prendre en compte que l'innovation n'apparaît pas ex nihilo et est le résultat d'efforts délibérés. Si l'entreprise ne cherche pas à améliorer ses produits et procédés, le THD ne peut rien y changer⁹⁴.

⁹⁴ Pour être complet, il faut souligner que même si l'entreprise n'affiche pas de volonté d'innover, le THD peut avoir un effet indirect sur sa performance d'innovation via l'attraction de main d'œuvre plus qualifiée par exemple.

Il est donc essentiel que l'entreprise fasse de la recherche et développement (R&D), et l'on considèrera qu'une entreprise fait de la R&D à partir du moment où elle cherche à améliorer ses produits ou procédés, même si elle ne dispose pas formellement d'un département de R&D. Dans cette section nous considérons uniquement le cas d'entreprises existantes faisant de la R&D.

Ici il est important de noter que, comme ce rapport porte principalement sur des zones rurales dont le débit internet maximal est potentiellement faible, l'effet de l'accès au haut débit est discuté en même temps que celui de l'accès au très haut débit. Dans la partie empirique, la question du débit sera abordée en évaluant séparément l'effet du très haut débit et de la fibre.

L'accès au THD en soi ne produit pas d'innovation mais peut affecter la productivité de la R&D dont le but est de produire de l'innovation. La littérature sur les déterminants de l'innovation est vaste (voir ex: Becheikh et al., 2006) et distingue les facteurs internes aux entreprises (comme l'âge, le type d'organisation, les ressources, etc.) et des facteurs contextuels (en particulier le secteur et les économies d'agglomération). Parmi tout cela, un élément crucial est le capital humain et les collaborations qui jouent un rôle de plus en plus important au cours du temps (ex: Jones, 2009). Par exemple, une littérature considérable en management porte sur la formation des équipes et comment la structure des collaborations influe sur la performance d'innovation (ex: Cowan and Jonard, 2004; Fang et al., 2010). Récemment, Bergé et al. (2018) démontraient un lien entre la structure du réseau de collaboration entre inventeurs et la performance d'innovation en France. L'accès au THD peut ainsi augmenter la productivité des activités de R&D, mais également modifier la technologie de collaboration en rendant possible les collaborations distantes géographiquement. Ces deux effets affectant *in fine* la performance d'innovation.

Le Tableau 6.1 liste des possibles mécanismes liant l'internet très haut débit à l'innovation. On distingue deux classes d'éléments. Les premiers concernent l'accès à des ressources, les seconds ont trait à la technologie de collaboration.

Tableau 6.1: Possibles mécanismes facilitant l'innovation grâce à l'accès au très haut débit

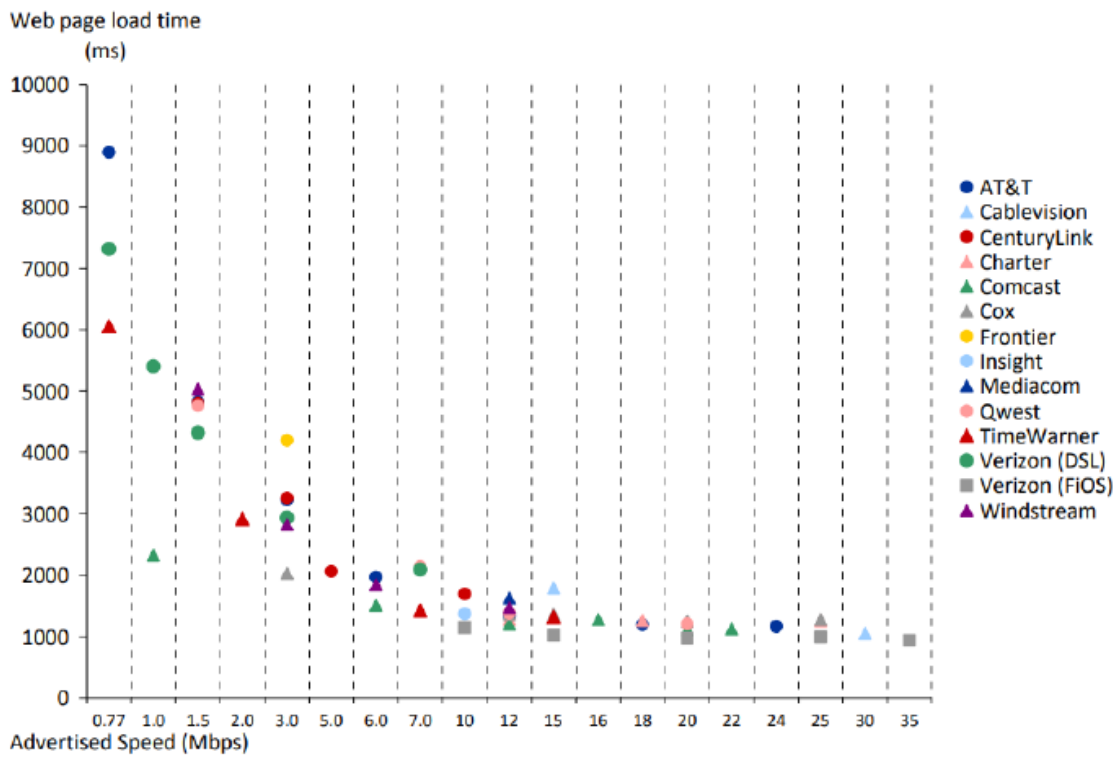
	<i>Également possible par le haut débit</i>
<i>Accès aux ressources</i>	
Recherche d'information	✓
Ressources computationnelles (<i>cloud</i>)	Partiel
Déploiement web	Partiel
<i>Technologie de collaboration</i>	
Visio-conférence	✓
Transfert de données	Partiel

Accès aux ressources

La toute première chose que permet l'accès au THD est l'affichage rapide de pages internet. Bien que cela puisse sembler trivial, les moteurs de recherche internet permettent de trouver en quelques mots-clés des informations utiles à la recherche entreprise. Si les temps de chargement sont très longs, cela limite grandement l'usage de cette technologie, réduisant ainsi la productivité. Par exemple, la Figure 6.1 montre les temps de chargements des pages internet en fonction de la vitesse de connexion (Federal Communications Commission, 2011). Pour des vitesses de connexions inférieures au mega bit par seconde, les temps de chargement peuvent dépasser les 5 secondes, réduisant ainsi substantiellement l'intérêt de la technologie. Cela est d'autant plus vrai lorsque l'entreprise est technologique et utilise des ressources bibliographiques comme les articles scientifiques ou les brevets dans leur R&D. Sur cette question du lien entre l'accès à internet et l'innovation il y a peu de résultats. Récemment Xu et al. (2019) développent un modèle lien l'accès à internet à la performance innovation via l'accès à l'information, le niveau de débit réduisant le coût d'accès à l'information. Ils effectuent ensuite une étude empirique trouvant une corrélation positive entre la production de brevets aux États-Unis et le débit d'accès à internet, au niveau des comtés. Néanmoins leurs résultats ne permettent pas d'identifier que c'est la diminution du coût d'accès à l'information qui permet l'augmentation du nombre de brevets. A noter que pour accéder à l'affichage rapide de pages internet une connexion THD n'est pas indispensable, comme le montre la Figure 6.1, une connexion haut débit rapide peut suffire. Néanmoins, les pages web intégrant de plus en plus d'images et de plug-ins de design, leur taille ne cesse d'augmenter au fil du temps, comme le montre la Figure 6.2. La taille médiane des pages web a quadruplé entre 2010 et 2020,

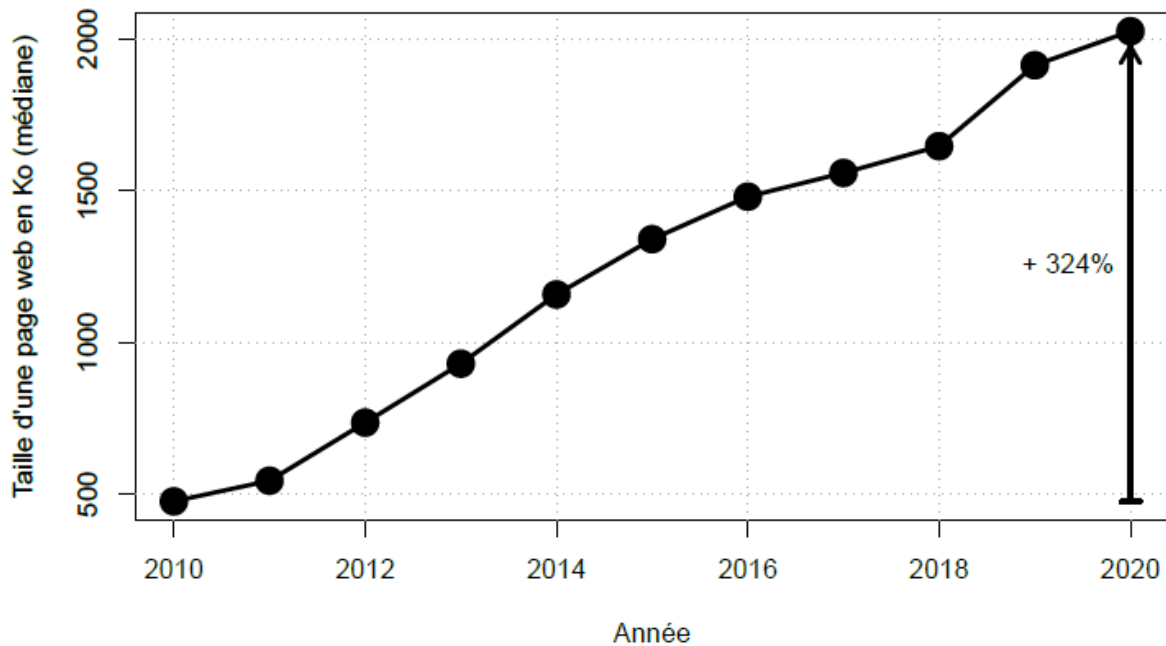
passant de moins de 500Ko à plus de 2Mo. Ce fait implique que la vitesse effective de chargement se réduit au cours du temps.

Figure 6.1: Vitesse de chargement des pages web et vitesse de connexion⁹⁵



⁹⁵ Source: Figure 10 de Federal Communications Commission (2011).

Figure 6.2: Évolution de la taille des pages web⁹⁶



Ensuite, internet permet d'accéder aux services de *cloud computing* qui ont connu une explosion ces dernières années. Créé dans les années 2000⁹⁷, ce concept consiste en la délocalisation du matériel informatique des entreprises vers des prestataires de service tiers. L'investissement en capital est uniquement porté par la société de service, les entreprises pouvant utiliser les ressources de celles-ci à la demande. Ces services de *cloud computing* ont pour effet de modifier en profondeur l'investissement en capital informatique des entreprises. Plutôt que d'investir dans du capital informatique dont il est difficile d'utiliser les ressources de façon optimale, les entreprises peuvent à la place consommer des ressources computationnelles qui s'ajustent à leur demande, limitant ainsi la problématique d'utilisation optimale du capital investi. Le secteur du *cloud computing* est en pleine expansion : ainsi, la croissance annuelle des revenus générés par ce secteur a été de 10% en moyenne entre 2014 et 2018 et tend encore à augmenter (Zykov, 2019).

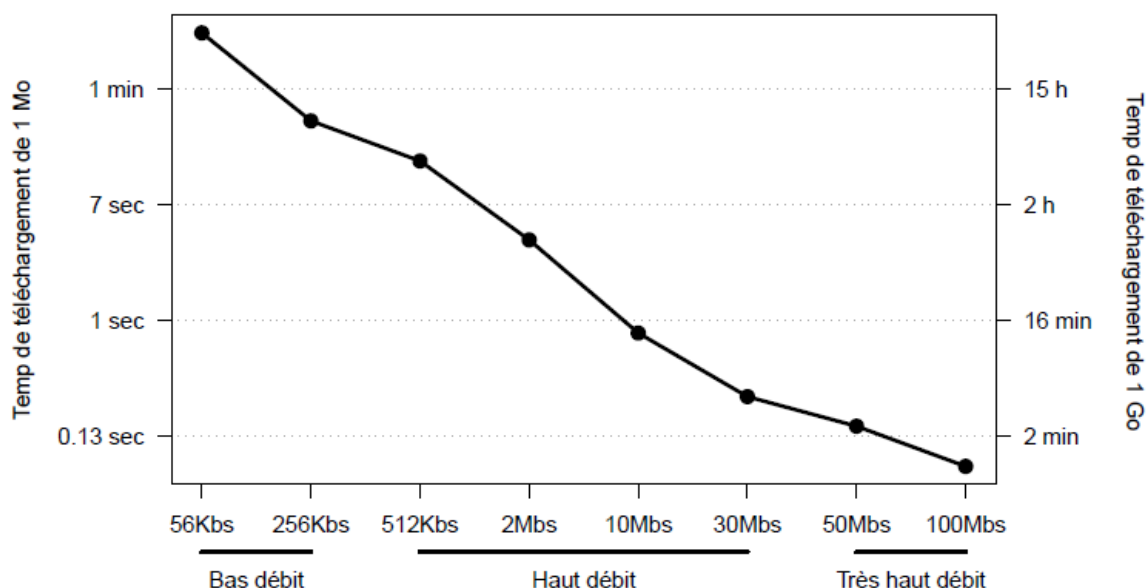
⁹⁶ Source: http archive, report on the State of the Web, <https://httparchive.org/reports/state-of-the-web>. Accès le 4/09/2020.

⁹⁷ En particulier, le précurseur Amazon Web Services a été créé en 2006.

La flexibilité permise par le *cloud computing* peut être particulièrement utile dans le cadre de la recherche et développement. L'incertitude est une caractéristique de la R&D, en particulier la recherche d'innovations radicales. Et si le développement de nouveaux produits ou procédés nécessite l'usage ponctuel de ressources informatiques importantes (comme par exemple l'usage de simulations en ingénierie), il peut être bien plus optimal de faire appel ponctuellement au *cloud computing* plutôt qu'à un investissement coûteux dont l'usage futur est incertain. Cela permet ainsi plus d'expérimentations. Une condition principale de l'accès au *cloud computing* est évidemment une connexion internet fiable. Ici la vitesse de connexion, i.e. THD versus HD, peut jouer en fonction des usages. En effet, si le transfert de données de l'entreprise vers le *cloud* (ou vice versa) est important, alors clairement le THD entraîne une augmentation de la productivité du *cloud computing* pour les entreprises. C'est moins le cas si le besoin de transfert de données est limité. Comme l'illustre la Figure 6.3, les temps de téléchargement de la frange basse de ce qui est défini comme le haut débit limite grandement le transfert fréquent de fichier.

Concernant le lien entre le *cloud computing* et l'innovation, il n'y a actuellement, à notre connaissance, pas de résultats empiriques dans la littérature, où la question étudiée porte plutôt sur les déterminants de l'adoption du *cloud computing* par les entreprises (ex : Oliveira et al., 2014).

Figure 6.3: Temps de téléchargement de fichiers en fonction du débit internet⁹⁸



Enfin, une connexion à internet permet d'accéder à des services en lignes de création de sites et ainsi d'établir facilement, sans connaissances préalables en informatique, un site web voire un site de vente en ligne. Le déploiement d'une page internet permet potentiellement l'accès à des nouveaux clients et la vente directe en ligne permet de s'affranchir des distributeurs existants. Ici une page internet en soi ne constitue pas une innovation, mais en entraînant un possible élargissement et une diversification de la demande, elle peut inciter au développement de nouveaux produits⁹⁹. Le mécanisme ici serait le « market pull », ou l'innovation tirée par la demande (DeStefano et al., 2012). L'impact de l'internet, via la création de site en ligne, sur l'innovation serait donc très indirect. Dans le cas de la création d'un site de vente en ligne avec une possible charge de connexion importante, l'internet très haut débit est nécessaire. Autrement une connexion haut débit fiable peut suffire.

⁹⁸ Source: calcul des auteurs (simple produit en croix). L'échelle de temps est logarithmique.

⁹⁹ Le commerce en ligne, rendu possible par l'accès au THD, peut également favoriser l'innovation en permettant la conquête de nouveaux marchés.

Technologie de collaboration

Un aspect grandement facilité par l'accès à internet est la collaboration distante. Comme vu précédemment, la collaboration joue un rôle important dans la production d'innovation, toute technologie réduisant les coûts et frictions à la collaboration peut ainsi faciliter l'innovation. Outre les e-mails qui peuvent être échangés même avec des connexions à très bas débit, une connexion haut-débit rend possible la visio-conférence et facilite grandement le partage de fichiers. Cela a pour conséquence de réduire les coûts de coordination au sein des l'équipe et ainsi d'augmenter leur productivité.

Ici plusieurs études existantes mettent en lien la connexion internet et la collaboration entre chercheurs. Le contexte porte à la fois sur la recherche scientifique académique mais également sur la collaboration entre industriels. Agrawal et al. (2008) étudient l'effet de l'accès de certaines universités américaines au Bitnet, un précurseur de l'internet créé au début des années 1980, sur les collaborations de leurs chercheurs. Ils se focalisent sur le domaine de l'ingénierie en examinant les collaborations dans les articles publiés dans les revues le plus prestigieuses de ce domaine. De plus, ils répartissent les universités en trois catégories en fonction de leur prestige: le premier tiers contenant les universités d'élite. Les auteurs montrent qu'une fois l'accès au Bitnet établi, les universités moins prestigieuses collaborent plus avec les universités d'élite qui leurs sont co-localisées. Ils interprètent ce résultat comme le fait qu'internet favorise une meilleure allocation de ressources, ayant un effet bénéfique sur la productivité. Ils ne montrent néanmoins pas d'effet notable sur les collaborations entre les universités d'élite. Une explication possible est que pour ces universités, les chercheurs étaient déjà connectés via un réseau antérieur: l'ARPANET, limitant ainsi le bénéfice de l'accès au Bitnet.

Ding et al. (2010) s'intéressent également à l'effet de l'adoption du Bitnet sur le comportement de collaboration dans le milieu académique. À la différence de l'étude précédente, leur analyse porte cette fois au niveau individuel contenant une riche liste de variables. En se focalisant sur le domaine de la biologie, les auteurs trouvent un effet positif de la connexion à internet sur la productivité et sur les collaborations.

Forman and Zeebroeck (2012) étudient l'effet de l'accès à internet sur la collaboration, mais cette fois dans les entreprises multi-établissements aux États-Unis. Ils utilisent les données sur les brevets pour identifier les collaborations entre chercheurs d'une même entreprise. Comme le lieu

de résidence des inventeurs doivent être indiqués sur les brevets, ils utilisent cette information pour déduire à quel établissement appartient chaque inventeur. Les auteurs concluent que la présence d'internet dans deux établissements distants augmente la probabilité que leurs chercheurs collaborent dans un brevet.

Ainsi, l'accès à internet modifie les rendements des collaborations distantes et peut augmenter ainsi la performance d'innovation. Il y a néanmoins une modération à apporter. Les collaborations à distance concernent principalement les grandes entreprises qui sont multi-établissements, ce genre de bénéfice concernera donc une minorité d'entreprise.

Recherches générales sur l'accès internet et l'innovation.

Il y a peu d'études qui analysent spécifiquement l'effet de l'accès à internet sur des mesures de l'innovation. Les seules existantes utilisent les brevets pour approximer l'innovation. L'étude de Lee et al. (2016) porte sur 40 pays de l'OCDE de 1999 à 2013 et mesure l'innovation par le nombre de brevets déposés à l'office américain de brevet (USPTO) dans le domaine des technologies de l'information et de la communication (TIC) par million d'habitants. En utilisant un panel avec effets-fixes, ils montrent une corrélation positive entre le taux d'adoption à l'internet haut débit et l'innovation.

L'étude de Xu et al. (2019), vue précédemment, porte quant à elle sur les États-Unis. Ils régressent le nombre de brevets déposés à l'USPTO dans un comté en 2015 par la proportion d'abonnés à internet dans le comté en 2013, avec différentes variables capturant le débit de l'abonnement (de 10Mbs à 100Mbs). Ils trouvent une corrélation positive entre accès à l'internet haut débit (10Mbs et plus) et l'innovation.

1.3. Mesures de l'innovation

La notion d'innovation est assez large et peut être comprise de différentes manières, il est donc important de clarifier ses termes pour préciser notre travail empirique. Ici nous nous attacherons à faire le lien entre les différentes mesures que nous utiliserons et les aspects de l'innovation qu'elles capturent. Pour ce faire, outre les divisions classiques de l'innovation entre produit et procédé, nous catégorisons l'innovation en deux éléments : a) la localisation : locale, nationale ou

internationale, et b) l'importance économique : faible ou forte. La localisation porte sur la nouveauté de l'innovation : est-elle seulement nouvelle au sein de l'entreprise, c'est à dire locale ? Ou est-elle nouvelle au niveau national ou international ? L'importance économique de l'innovation reflète quant à elle l'augmentation de la productivité de l'entreprise suite à son adoption.

Pour mesurer l'innovation nous utiliserons un ensemble de mesures, chacune capturant un certain aspect de l'innovation et ayant ses propres avantages et limites. Ces mesures sont basées sur les i) brevets et ii) l'enquête sur les pratiques d'innovation menée par l'INSEE.

Brevets

Les brevets, dont nous avons déjà parlé, sont un moyen juridique d'exclure, de façon temporaire, les concurrents d'un certain marché en rendant passible de poursuites l'imitation d'un nouveau produit ou procédé protégé par un brevet. Pour qu'une invention puisse être protégée par un brevet, une condition essentielle est la nouveauté : le nouveau produit ou procédé doit n'avoir jamais existé. Selon les critères de brevetabilité de l'INPI: « L'invention doit être nouvelle, c'est-à-dire qu'elle ne doit pas porter sur une innovation qui a déjà été rendue accessible au public, quels qu'en soient l'auteur, la date, le lieu, le moyen et la forme de cette présentation au public. »

En utilisant la classification précédente, un brevet porte donc sur une innovation internationale. Par essence ce n'est pas une innovation simplement locale, car même si le produit ou procédé est nouveau pour l'entreprise, le brevet n'est délivré que si la portée de la nouveauté est internationale.

De plus, le brevet a un coût. En France, un brevet est assez peu onéreux au niveau de l'entreprise : environ 1000€ pour un brevet protégé sur 10 ans. Mais les coûts augmentent significativement dès que l'on veut que l'invention soit protégée sur plusieurs pays, par exemple la protection au niveau de l'Union Européenne coûte plusieurs dizaines de milliers d'euros. Pour information le coût pour maintenir un brevet sur 20 ans au niveau de l'Union Européenne est d'environ 150 000€. A cela il faut ajouter le coût de juristes spécialisés servant à la fois à augmenter le taux d'acceptation de la demande de brevet et leur portée restrictive. Ainsi, les brevets ont un coût potentiellement important, il est donc raisonnable de penser que si le bénéfice du brevet est inférieur à ce coût, l'entreprise ne poursuit pas le processus de brevetage. De ce fait, les brevets ont donc tendance à être des mesures d'innovation « fortes » au niveau de l'entreprise. Il faut garder en tête néanmoins que toutes les innovations, même importantes, ne mènent pas nécessairement à

un brevet (ex: les secrets de fabrication peuvent être un mode de protection alternatif) et la validité de l'utilisation du brevet comme mesure d'innovation est plus ou moins forte entre les secteurs (ex: le lien est très fort en biotechnologie et pharmacie, moins dans le domaine des logiciels où il a plus tendance à être utilisé comme mesure stratégique).

Brevets: signification, fonctionnement et coûts.

Les brevets constituent un droit de monopole temporaire sur une invention. Toute entreprise ou individu ayant inventé un nouveau processus de fabrication ou un nouveau produit, c'est à dire une invention, peut demander à obtenir un brevet si cette invention est: a) nouvelle, b) non triviale et c) est économiquement exploitable. Si le brevet est octroyé alors il confère des droits de monopole qui peuvent être défendus en justice. Le brevet s'octroie par pays: il n'y a pas de brevet européen, mais des brevets français, allemands, états-uniens, etc. (bien qu'un brevet unitaire européen entrera en effet en 2022). L'entreprise décide donc dans quel(s) pays son invention sera protégée.

En Europe, une demande doit être effectuée dans chaque office national. Si l'on veut protéger son invention sur 25 pays, il faut donc faire 25 demandes de brevets différentes. Chaque office étant maître du processus d'examen, ce n'est pas parce qu'une invention est brevetée dans un pays qu'elle le sera dans un autre (les pratiques d'octroi pouvant être par ailleurs très différentes entre pays). Pour faciliter ce processus il existe néanmoins l'Office Européen des Brevets auquel les entreprises peuvent directement déposer leur brevet. L'OEB se charge de l'examen du brevet et, si l'OEB décide que le brevet est valide, son détenteur est alors libre de le déposer dans les 25 pays membres de l'UE que l'OEB représente. Il faudra alors payer des frais d'octroi de brevet pour chaque office, mais celui-ci est de droit accepté sans aucune examination supplémentaire.

Le dépôt d'un brevet en Europe en passant par l'OEB coûte environ 4000€. Si celui-ci est accepté, il faut alors payer des frais pour chaque office national où la protection est demandée, et ce pour chaque année de protection. Pour donner une idée, un brevet en France maintenu sur 10 ans coûte 860€, maintenu sur 20 ans cela coûte 5910€. Ces coûts sont contingents aux offices nationaux. Le coût de maintenance d'un brevet sur les 25 pays membres de l'UE associés à l'OEB est proche de 150 000€.

Nous utiliserons ici les brevets déposés à l'Office Européen des Brevets (OEB). Les avantages d'utiliser les données de brevet sont nombreux. Tout d'abord, ce sont des données publiques très riches, où les informations sont renseignées à la fois sur les entreprises et les inventeurs. Cela permet, après avoir effectué un travail d'appariement adéquat, de travailler directement au niveau établissement. De plus, ces données sont exhaustives, permettant de suivre la production totale des brevets de chaque établissement dans le temps, ce qui est un avantage certain pour notre travail empirique, autorisant ainsi l'emploi de la différence-en-différence au niveau établissement.

Ces données ont aussi des limites. Tout d'abord, bien que les brevets contiennent des informations sur les entreprises déposantes, celles-ci n'ont pas d'identifiant, ce qui fait qu'un travail d'appariement minutieux doit être effectué pour identifier quel brevet est associé à quelle entreprise. Un second problème, dont les conséquences sont plus importantes pour ce rapport, est la temporalité : bien que ces données soient publiques, le délai de publication des brevets est d'environ deux ans en moyenne. S'ensuit une période d'intégration au journal officiel puis à la base de données qui ajoute encore un délai. Les données utilisées ici sont REGPAT produit par l'OCDE, version janvier 2021. Les données dans cette base ne sont exhaustives que jusqu'à fin 2016 alors que les données de l'année 2017 sont seulement partiellement disponibles. C'est une limite importante concernant ce projet, étant donné le calendrier de déploiement de la fibre dans le cadre du Plan FTTHD.

Données de l'enquête communauté d'innovation

Le troisième ensemble de mesures mobilisées ici est issu de l'enquête communautaire sur l'innovation menée par l'INSEE. Cette enquête consiste en un questionnaire sur l'activité d'innovation fourni par les entreprises françaises.

L'avantage de cette enquête tient au grand nombre de questions qui sont posées, abordant un riche panel d'aspects de l'innovation. Elles portent sur le type d'innovation produite (ex: produit ou procédé) et leur ampleur (ex: si l'invention est nouvelle seulement pour l'entreprise ou également pour le marché). Elles portent aussi sur d'autres aspects, comme les moyens mis en place par l'entreprise pour développer l'innovation.

Bien que riches, les limites de ces données sont nombreuses. Tout d'abord, comme ce sont des données d'enquête, la qualité de l'information est tributaire des connaissances de la personne que

l'entreprise assigne au questionnaire, sa fonction influençant ses réponses. Ensuite ces données ne couvrent pas l'ensemble des entreprises: environ 20 000 entreprises sont sondées, choisies aléatoirement et stratifiées sur plusieurs critères comme le type d'activité et la tranche d'effectif. Il n'est donc pas possible d'avoir un panel d'entreprises observées de façon continue dans le temps. L'aspect aléatoire de l'échantillonnage complique également le cas de l'analyse au niveau communal. Enfin, cette enquête a lieu tous les deux ans, ce qui amène à seulement trois points d'observation sur la plage des années de déploiement de la fibre. On pourra néanmoins examiner si une tendance se dégage entre les mesures de cette enquête et l'accès au THD des entreprises, mais en gardant en tête les limites associées.

2. Analyse de la production de brevets

2.1. Construction de données

Les informations sur les brevets sont issues de la base de données REGPAT de l'OCDE, version Janvier 2021 (Maraut et al., 2008). Ces données proviennent elles-mêmes de la base Worldwide Statistical Patent Database (PATSTAT Global, Autumn 2020) de l'Office Européen des Brevets (OEB). Ces bases couvrent l'ensemble des brevets publiés à l'OEB jusqu'à fin 2020. Cela correspond à 3 698 474 brevets, allant de 1977 à 2019, tous pays confondus. Cela correspond à 249 531 brevets dont l'adresse du déposant est en France.

Les données de ces brevets incluent l'identité du déposant et des inventeurs, ainsi que leurs adresses (adresse de résidence dans le cas des inventeurs).

Nous ne pouvons pas utiliser de façon fiable l'adresse de l'entreprise déposante pour identifier le lieu de production des brevets. En effet, bien que les entreprises aient l'obligation légale de donner leur adresse, elles sont libres de ne pas révéler le lieu où a été produit le brevet. Or, beaucoup d'entreprises qui produisent des brevets sont multi-établissements et il est commun que le siège centralise les dépôts de brevet. Néanmoins, comme nous disposons de l'adresse des inventeurs, nous pouvons inférer dans quel établissement l'invention a été produite en regardant la distance entre l'adresse des inventeurs et celle des différents établissements. Cette stratégie permet ainsi de travailler non pas au niveau de l'entreprise mais au niveau de l'établissement. Cela

permettra d'analyser si l'accès au très haut débit change l'organisation des collaborations intra-entreprises et inter-établissements.

Comme nous ne disposons pas des informations sur les établissements via les données brevets, nous apparions celles-ci avec la base SIRENE établissement de l'INSEE (fichiers stock établissements et stock unités légales). Ainsi, nous disposons de diverses informations, dont l'adresse, le nom et le secteur, sur 10 442 430 entreprises et 15 495 192 établissements.

Nous décrivons maintenant le travail effectué pour apparier les données brevets aux données SIRENE. L'appariement s'effectue en deux étapes :

1. Apparier un brevet à une ou plusieurs entreprises (car il peut y avoir plusieurs déposants),
2. Sachant les entreprises appariées, associer le brevet aux établissements en utilisant leur distance géographique aux lieux de résidence des inventeurs.

Nous avons rencontré de nombreuses difficultés pour effectuer ces appariements. Tout d'abord, au sein de la base SIRENE, il a été difficile d'identifier précisément quel établissement appartient à quelle entreprise. En effet, il est commun que les entreprises appartenant à un même groupe subdivisent leurs activités entre plusieurs entreprises distinctes. C'est à dire que deux entreprises peuvent avoir le même nom et la même adresse mais des codes SIREN distincts (i.e. l'identifiant unique d'entreprise utilisé par l'INSEE). Le problème principal est donc qu'il est impossible d'associer le brevet à la bonne entreprise seulement avec le nom et l'adresse de celle-ci (seules informations dont nous disposons). Ainsi, il nous a fallu effectuer des groupements d'entreprises, c'est à dire considérer comme une seule entreprise plusieurs entités identifiées par un code SIREN. On considère ainsi l'ensemble des établissements régis par ces entreprises afin, dans la deuxième étape, d'associer le brevet à l'établissement le plus proche des inventeurs. Sur l'ensemble des SIREN, nous en avons agrégé 209 087 en 91 319 nouveaux identifiants.

Il est important de noter qu'associer le brevet à l'établissement le plus proche géographiquement des inventeurs n'est pas forcément sans poser de problème en termes de

précision. Néanmoins, la solution alternative, qui serait d'associer le brevet uniquement au siège social de l'entreprise, est bien plus problématique encore¹⁰⁰.

Par ailleurs, les adresses contenues dans les brevets ne sont pas formatées et correspondent aux adresses données par les entreprises. Cela implique qu'il y a un travail d'extraction à effectuer à partir des adresses brutes, sans compter les possibles erreurs qui y sont contenues. Le nettoyage des adresses se fait à partir d'un algorithme qui s'est amélioré au fil des nombreux essais et vérifications¹⁰¹. Concernant les données SIRENE, les données sont formatées ce qui facilite grandement la tâche. Le traitement doit toutefois amener à une adresse de format *identique* à celui des adresses nettoyées issues des données brevet. Ce travail est essentiel car l'unicité de l'adresse (et pas seulement de la commune ou de la rue) est une information essentielle pour s'assurer que deux entreprises sont bien les mêmes.

L'autre information essentielle est bien sûr le nom des entreprises. Celui-ci peut connaître des variabilités importantes, il est donc important de les corriger. Nous avons enlevé les qualificatifs d'entreprises qui peuvent apparaître à différents endroits (ex: S.A.), nous avons fusionné des acronymes qui pouvaient être écrits de différentes manières et enlevé les caractères non alphanumériques, puis fait une harmonisation à la main pour certaines entreprises.

A partir des informations harmonisées sur l'adresse et le nom des entreprises, nous avons effectué l'appariement dans l'ordre suivant :

1. Nom exact + même adresse
2. Nom proche + même adresse
3. Nom exact + même commune
4. Nom proche + même commune
5. Nom exact + même département
6. Nom proche + même département

¹⁰⁰ Dans la littérature, n'ayant pas d'information sur les établissements, il est d'usage d'associer le brevet au lieu de résidence des inventeurs plutôt qu'au siège car c'est admis que ce n'est pas le lieu de production. Comme nous disposons ici de l'information sur les établissements, notre méthode est donc bien plus précise que cette alternative.

¹⁰¹ Par exemple, notre algorithme transforme "1 ch des plattes,zone artisanale des plattes,f-69390 vourles" en "rue: 1 chemin des plattes ; commune: vourles ; cp: 69390" ou "34, avenue carnot bp 185f,63006 clermont-ferrand cedex 1" en "rue: 34 avenue carnot ; commune: clermont-ferrand ; cp: 63006", etc...

A chaque fois nous comparons le nom et l'adresse du déposant dans un brevet aux noms et adresses des entreprises (potentiellement regroupées) de la base SIRENE (i.e. une même entreprise peut avoir plusieurs variations de noms et d'adresses). Pour juger si deux noms d'entreprises sont proches, nous avons utilisé un ensemble d'algorithmes. Nous considérons que deux chaînes de caractères sont proches si : 1. si une chaîne de caractères est complètement incluse dans l'autre, avec au maximum une typo, 2. si la distance de jaro-Winkler entre les deux chaînes de caractères vaut 0.90 ou plus. Pour juger si deux adresses étaient identiques, nous avons utilisé un algorithme semblable pour estimer la similarité de deux noms de rue.

Bien sûr, un brevet apparié à une étape n'est plus à appairer dans les autres, nous sommes ainsi sûr qu'il est apparié avec le code SIREN le plus probable. Nous allons jusqu'à considérer les entreprises du même nom et même département car, dans de nombreux cas, si elles n'ont pas été appariées dans les étapes précédentes, la cause peut être simplement des erreurs de formatage.

Après avoir apparié les brevets aux entreprises, il faut y associer les établissements où il est probable que l'invention a été produite. Pour cela, nous utilisons l'adresse des inventeurs présents sur les brevets. Pour chaque inventeur sur un brevet, nous calculons la distance à tous les établissements de l'entreprise appariée. Nous assignons ensuite le brevet à l'établissement le plus proche géographiquement. Il se peut que plusieurs établissements d'une même entreprise soient assignés dans cette procédure. Il est à noter que le calcul de la distance géographique se fait au niveau de la commune et non pas au niveau de la rue. In fine, nous ne considérons donc qu'un seul établissement par commune par entreprise. Ainsi, nous assignons un brevet non pas à un établissement mais à une paire entreprise- commune qui ressemble donc à un méta-établissement qui regroupe l'ensemble des établissements d'une même entreprise dans la commune.

En utilisant la procédure décrite, nous avons pu appairer 189 947 brevets, soit 76%. Parmi ceux qui n'ont pas été appariés, 50% n'ont en fait pas d'inventeur avec une adresse française et 17% ne sont pas des entreprises mais des personnes individuelles. Environ 50% des appariement avaient les mêmes nom et adresse, suivis de 26% pour mêmes nom et commune. L'appariement avec nom proche et même département ne concerne que 5% des cas.

Une fois que les brevets sont appariés, nous construisons les différentes variables suivantes : le nombre de brevets produits, le nombre de collaborations entre établissements, et le nombre de collaborations en fonction de leur type (proches ou distantes).

Concernant la mesure d'accès au très haut débit des communes, nous utilisons les données de couverture de 2013 à 2018 de l'Observatoire du très haut débit de l'Agence du Numérique. A noter que pour les communes qui ont fusionné sur la période, nous avons calculé le taux d'accès à internet comme la moyenne (pondérée par le nombre de locaux) des taux de couverture de toutes les communes fusionnant.

A partir de là, nous construisons un panel composé de l'ensemble des établissements français de 2005 à 2018. Nous restreignons ce panel uniquement aux établissements dont la date de fermeture excède 2013, date des premières mesures de déploiement très haut débit. Nous restreignons aussi le panel aux entreprises dont la date de création est antérieure ou égale à 2015. Pour imputer le taux de déploiement pour les années 2005-2013, nous prenons simplement le taux de déploiement en 2013, il n'y aura donc aucune variation de la mesure de traitement sur cette période-là.

2.2.Stratégie empirique

Les données dont nous disposons nous permettent de faire une analyse de panel au niveau des établissements. Dans ce contexte, la méthode principale employée pour identifier l'effet du très haut débit est la méthode de Sun et Abraham (2021) sur le traitement échelonné décrite dans le chapitre 3. Dans le contexte des brevets, nous estimons :

$$y_{it} = \alpha_i + \beta_t + \sum_{d=-1} \gamma_d \cdot 1\{t - c_i = d\} + \epsilon_{it}$$

avec y_{it} une mesure de production: le nombre de brevets produits ou le nombre de collaborations (proches ou distantes). La variable $1\{t - c_i = d\}$ est une indicatrice d'accès au très haut débit (ou à la fibre) pour l'unité i à la période t , c_i étant l'année où la commune dans laquelle se situe l'établissement i a eu accès au très haut débit (ou à la fibre).

Comme les variables dépendantes sont des variables de comptage, nous utilisons également un modèle de Poisson à effets fixes pour obtenir l'estimateur de γ_d . Ce type de modèle est plus efficient pour les variables avec une dispersion élevée, comme la production de brevets (voir Silva and Tenreyno, 2006). Dans le contexte du modèle de Poisson à effets fixes, il est à noter que seules les unités ayant au moins une valeur positive de la variable dépendante sur la période seront utilisées pour l'identification.

Enfin, nous clusteriserons les erreurs au niveau de la commune étant donné que le traitement, l'accès au très haut débit (ou à la fibre), se fait à ce niveau-là.

2.3. Statistiques descriptives

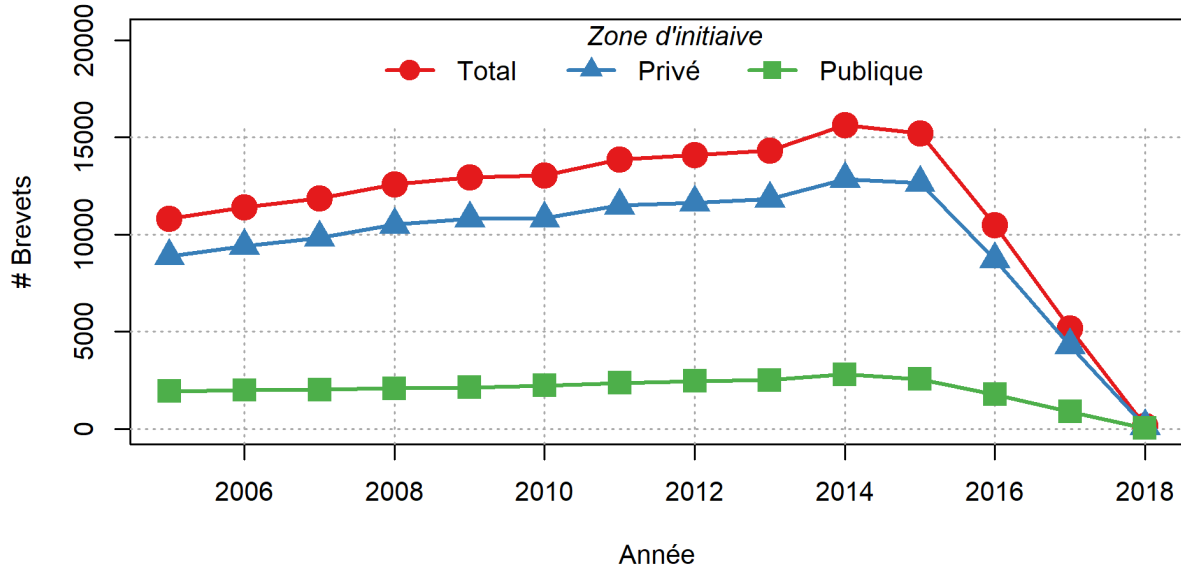
Notre panel brut comporte 94 580 045 observations, portant sur 8 341 764 établissements sur la France entière, dont 12 437 ont au moins un brevet (soit 0.15%). Si l'on se concentre sur les zones d'initiative publique, nous avons 3 013 145 établissements dont 2 997 ont au moins un brevet (soit 0.1%). Alors que la fraction d'entreprises brevetant semble très faible, il faut garder en tête que le dénominateur ici n'est pas le bon, puisque nous considérons l'ensemble des établissements alors qu'en fait une fraction seulement est concernée par la possibilité de breveter.

L'évolution de la production de brevets est représentée à la Figure 6.4. Il y a une évolution positive faible mais constante jusqu'à 2015 suivie par une chute brutale en 2016 avec quasiment plus de brevets en 2018. Comme expliqué dans la section précédente, cette baisse du nombre de brevets publiés est artificielle et intrinsèquement liée au processus de publication. En effet, cela est dû au temps de latence entre le dépôt de brevet et sa divulgation, puis ensuite sa prise en compte dans la base de données centrale. Nous pouvons donc utiliser les données jusqu'à 2015 en les considérant comme complètes. Les données des années qui suivent peuvent être utilisées en gardant en tête qu'elles sont par nature tronquées.

Nous décidons néanmoins de conserver les années 2016 et 2017 dans notre échantillon même si elles sont tronquées. L'idée étant que si la non-présence dans la base de données est suffisamment aléatoire pour affecter de façon indistincte les unités contrôle et les unités traitées de notre échantillon, alors cela implique uniquement un problème d'erreur de mesure qui biaisera

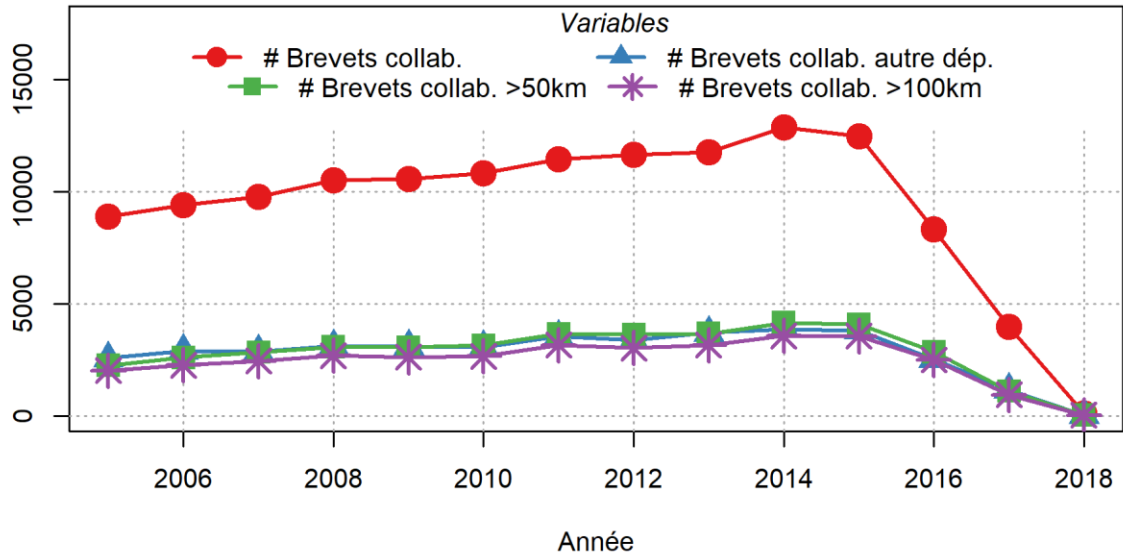
la magnitude du coefficient vers 0. Un biais de sélection positif ou négatif serait quant à lui plus problématique.

Figure 6.4: Evolution de la production de brevets par type de zone d'initiative



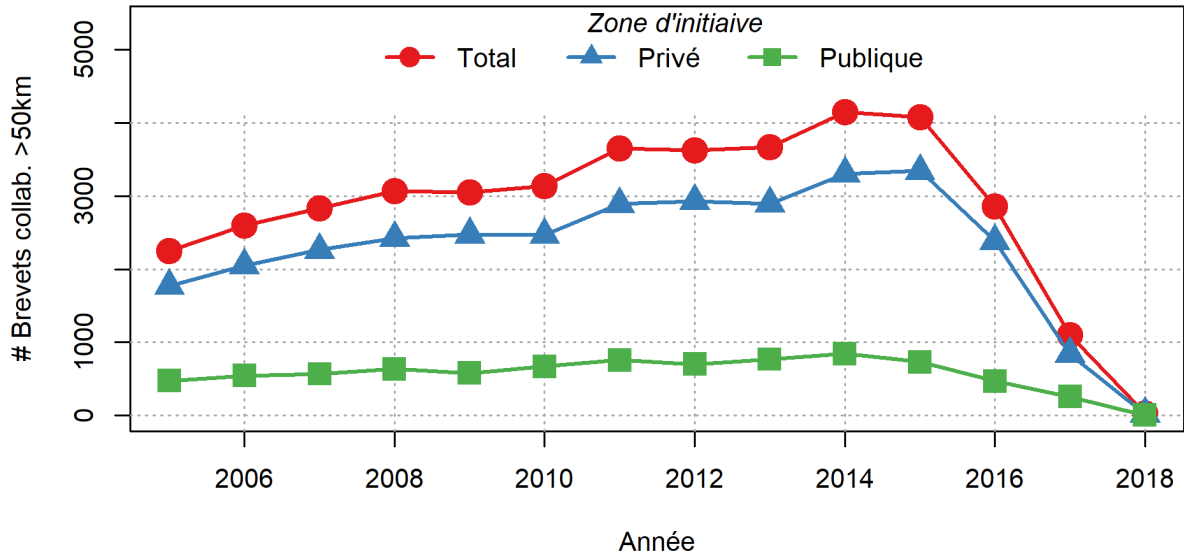
Pour évaluer l'effet de l'accès au THD sur la collaboration dans la production de brevets nous avons créé plusieurs variables. Leur évolution est décrite dans la Figure 6.5, qui dépeint l'évolution temporelle des collaborations entre établissements en fonction de leur distance. Si on ne pondère pas par la distance, on constate un nombre de collaborations très élevé. Ce nombre est néanmoins vraisemblablement issu du multi-comptage introduit par la procédure d'assignation des brevets aux établissements les plus proches des inventeurs (bien qu'il puisse contenir également des collaborations inter-entreprises). Les autres mesures de collaboration, soit avec des établissements d'autres départements ou à plus de 50 ou 100km ont toutes des valeurs proches. Par la suite, la mesure que nous retiendrons pour évaluer la collaboration dans la production de brevet est la collaboration à 50km ou plus.

Figure 6.5: Evolution de différentes mesures de collaboration dans les brevets, France entière



En retenant cette dernière mesure, la Figure 6.6 montre l'évolution des collaborations dans le temps et on peut constater qu'elle suit de près l'évolution de la production de brevets en général.

Figure 6.6: Evolution de la collaboration dans les brevets par type de zone d'initiative



Sans surprise, la distribution de la production de brevet est très inégalement répartie entre les déposants. La Figure 6.7 montre que seulement quinze entreprises produisent plus de 40% des brevets. On reconnaît à la fois des grands groupes comme Renault ou PSA, mais également de grandes institutions publiques comme le CEA ou le CNRS. La Figure 6.8 sépare la distribution dans les zones d'initiative publique et privée et on peut constater que les entreprises dominantes restent sensiblement les mêmes.

Figure 6.7: Distribution de la production de brevet par déposant, France entière

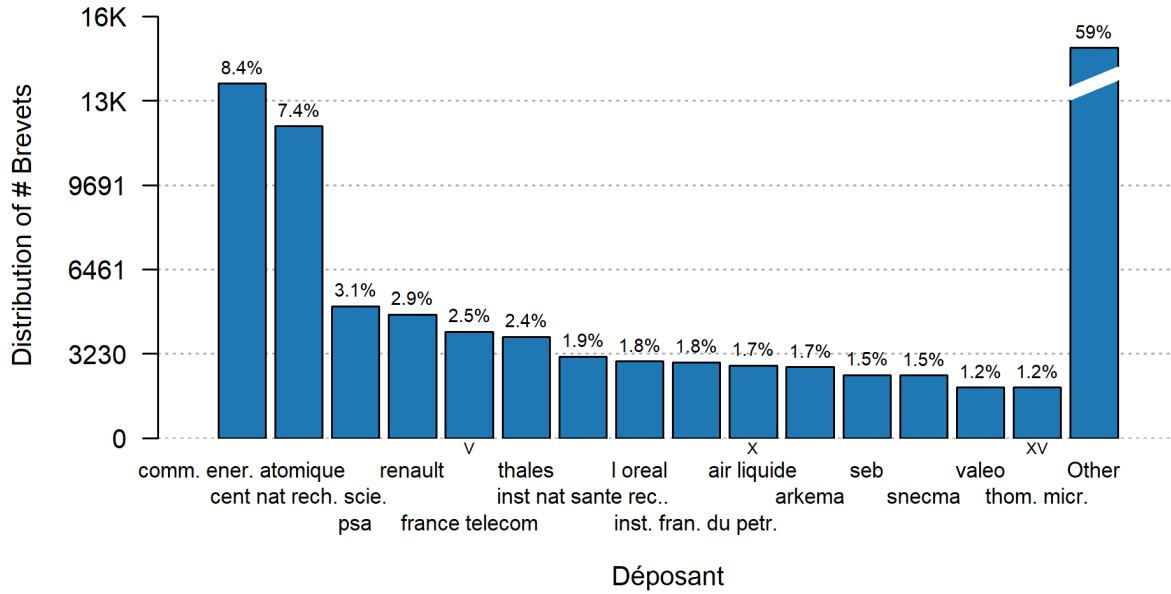
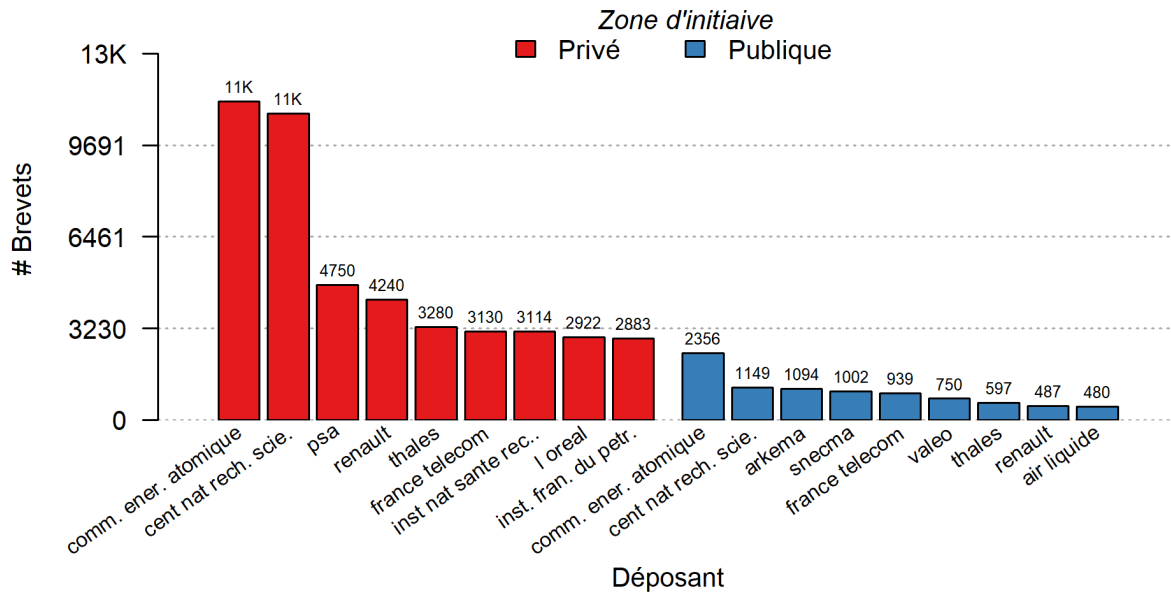
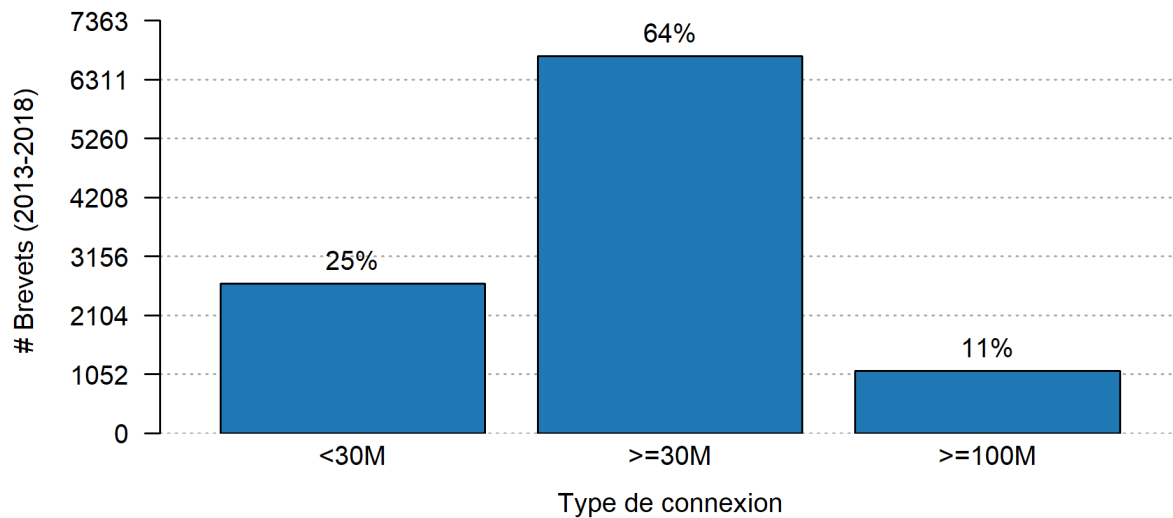


Figure 6.8: Distribution de la production de brevet par déposant, par type de zone d'initiative



Focalisons-nous maintenant sur les ZI publiques et regardons le lien entre le débit et la production. Comme le montre la Figure 6.9, la production de brevets se fait majoritairement dans les zones disposant d'un accès au très haut débit (≥ 30 M), bien que 25% soit quand même produite dans des zones avec un débit inférieur à 30M.

Figure 6.9: Distribution de la production de brevet par débit internet dans la commune

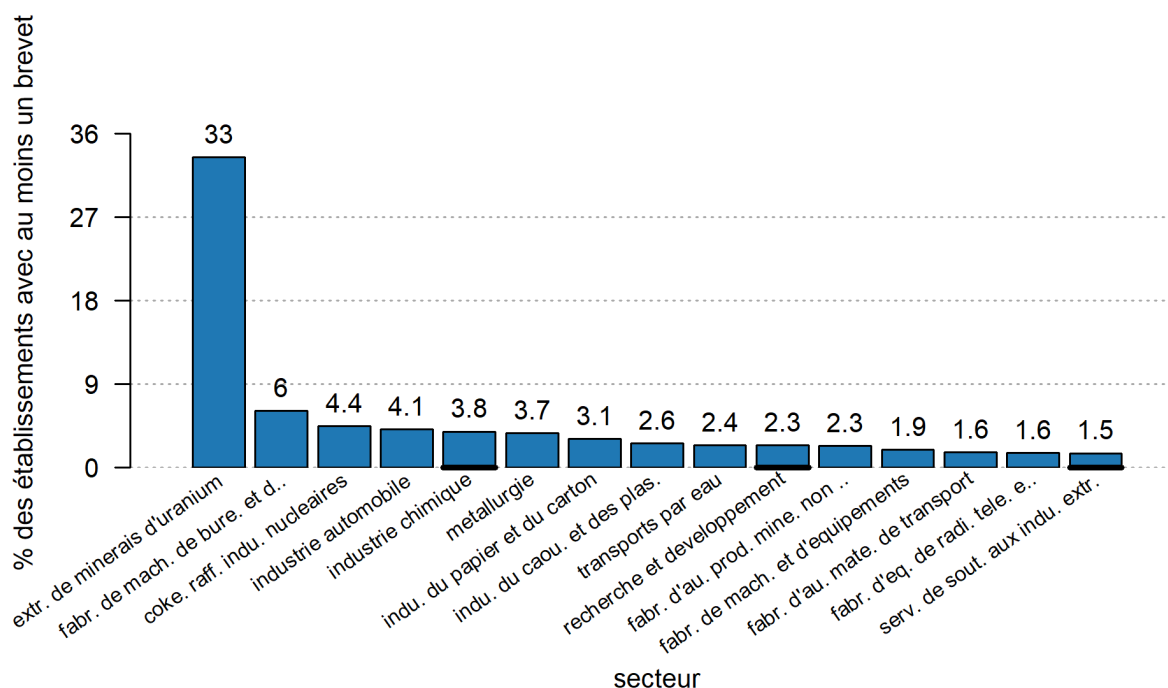


Maintenant regardons quels sont les secteurs concernés par les brevets. La Figure 6.10 représente les secteurs qui ont la plus grande proportion d'établissements avec un brevet, alors que la Figure 6.11 montre la distribution des brevets par secteur. Le secteur est ici identifié par le code NAF de l'établissement. On constate que 33% des établissements catégorisés dans l'*extraction de minerais d'uranium* ont au moins un brevet sur la période concernée. Vient ensuite la *fabrication de machines de bureau et de matériel informatique*, suivi de près par l'*industrie automobile*. Concernant la production totale de brevets: vient en première place assez logiquement le secteur

de la *recherche et développement*, avec plus de 25% des brevets déposés, suivi par les *services fournis principalement aux entreprises*¹⁰² puis par l'*industrie automobile*.

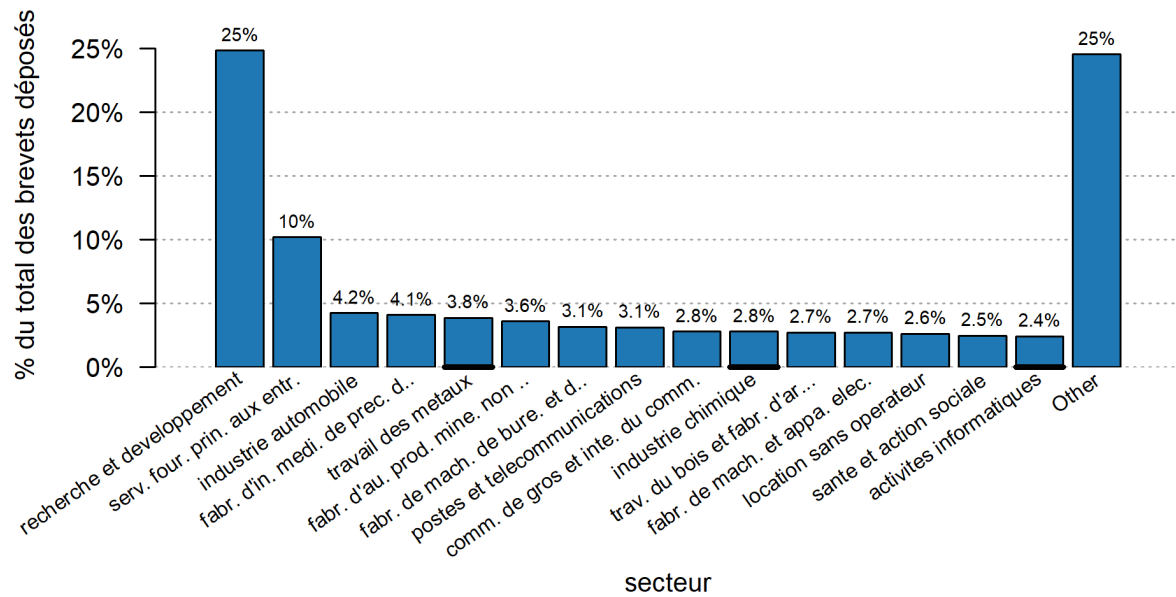
Dans la section empirique, la méthode d'estimation principale, en Poisson, ne prendra en compte que les établissements avec au moins un brevet sur la période. Par contre, lorsque nous effectuerons des estimations par OLS, il nous faudra prendre en compte également les entreprises n'ayant aucun brevet sur la période. Considérer l'ensemble des entreprises alors qu'une majorité n'est pas concernée par la problématique des brevets peut entraîner un biais vers 0 de l'estimateur. Pour éviter ce problème, nous nous concentrerons donc uniquement sur les 10 premiers secteurs en termes d'établissements avec brevet et les 10 premiers secteurs producteurs de brevets.

Figure 6.10: Distribution par secteur du pourcentage d'établissements ayant au moins un brevet, France entière



¹⁰² Cette catégorie est assez vague et on y retrouve certains établissements d'entreprises comme Michelin, Seb, Saint-Gobain, etc.

Figure 6.11: Distribution par secteur de la production totale de brevets, France entière



Pour finir, les Figures 6.12 et 6.13 représentent l'évolution de la production annuelle de brevets (pour les établissements ayant au moins un brevet sur la période) en fonction des cohortes d'accès à l'internet 30M ou 100M.

Figure 6.12: Evolution de la production annuelle de brevet pour différentes cohortes d'accès à un débit de 30M ou plus dans des ZI publiques

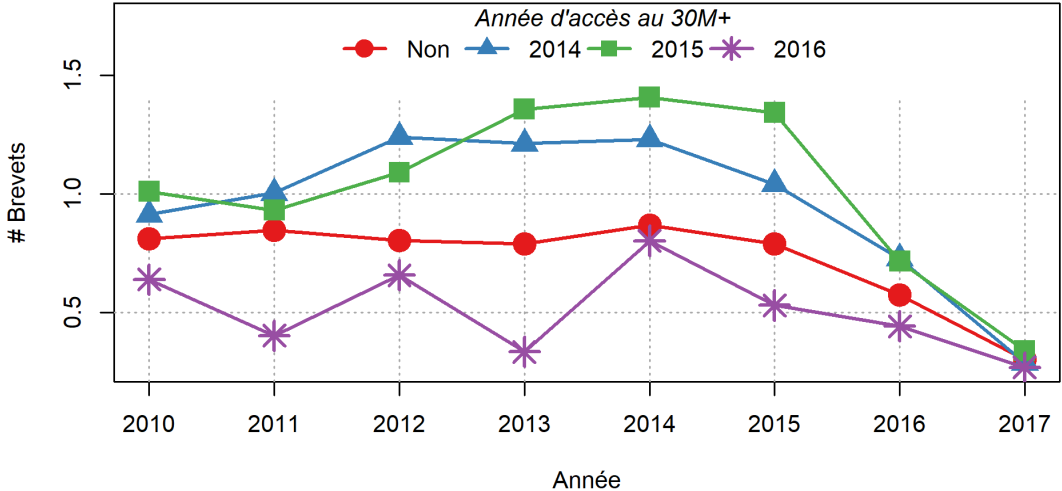
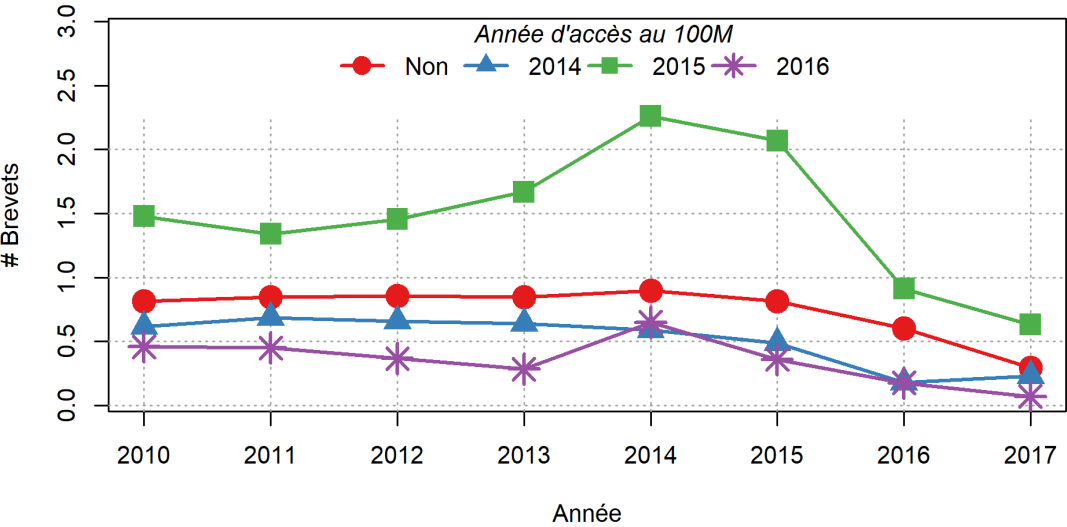


Figure 6.13: Evolution de la production annuelle de brevet pour différentes cohortes d'accès à un débit de 100M dans des ZI publiques



2.4.Résultats

Les résultats principaux sont représentés par les Figures 6.14 et 6.15. Ces figures représentent l'effet annuel du traitement, ici l'accès à un débit de 30M ou 100M, sur la production annuelle de brevet. La première chose que l'on peut constater est qu'il n'y a pas de présence de pré-tendance dans les deux cas, ce qui va dans le sens d'une bonne comparabilité entre les unités de traitement et les unités de contrôle. Concernant l'accès à 30M, il semble qu'il n'y ait pas d'effet significatif sur la production de brevets. Par contre, nous trouvons des effets négatifs et significatifs la première et deuxième année suivant l'arrivée de la fibre, avec des coefficients de -0.55 et -0.36 respectivement. Cela correspond à des baisses de 42% ($100 * (1 - \exp(-0.55))$) et 29%. La troisième année il n'y a plus d'effet significatif.

Figure 6.14: Effet de l'accès à un débit de 30M sur la production annuelle de brevet

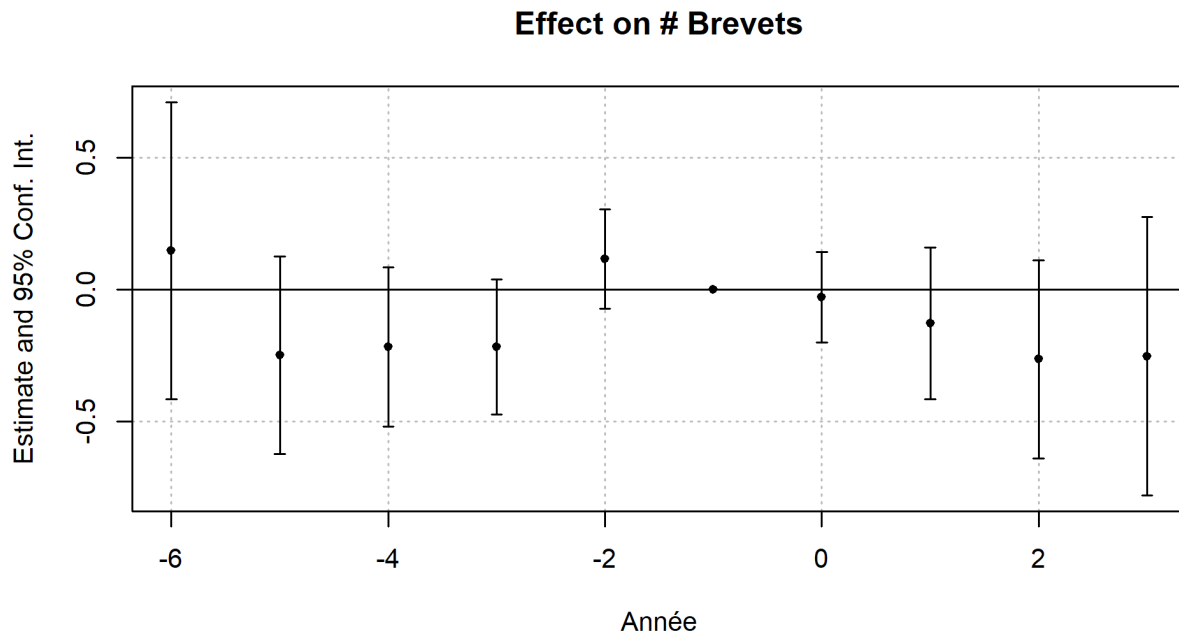
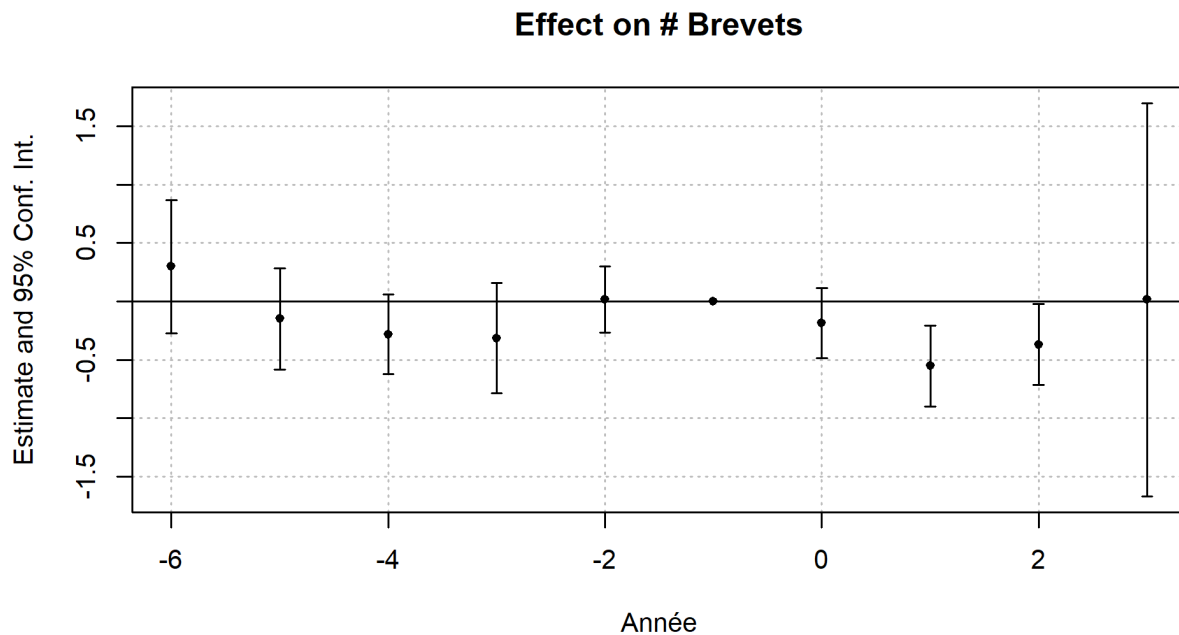


Figure 6.15: Effet de l'accès à un débit de 100M sur la production annuelle de brevet



Si on regarde l'effet agrégé sur la période après traitement, dans le Tableau 6.2, on obtient un effet total à -0.15 mais non significatif pour 30M et de -0.35 et significatif pour 100M (l'intervalle de confiance à 95% est [-0.63 ; -0.07]). Les dernières colonnes répliquent l'estimation pour 30M mais en excluant les communes qui obtiennent la fibre afin de s'assurer de ne pas mélanger l'effet des deux types d'accès. Ce faisant, l'effet se trouve encore plus réduit en magnitude et toujours non significatif.

Il est à noter que nous excluons de l'analyse les établissements qui sont traités du début à la fin de la période. Cela fait que le nombre d'observations est plus important pour l'accès à 100M que pour l'accès à 30M car il y a un grand nombre d'établissements qui sont dans des communes ayant accès à 30M sur l'ensemble de la période. Ainsi, l'échantillon utilisé pour le 100M est plus grand dû au plus grand nombre d'unités utilisées comme contrôles.

Les résultats sur les collaborations sont reportés dans le Tableau 6.3. Concernant la pré-tendance, il ne semble pas y en avoir, même si l'année 5 avant l'accès au débit de 30M a un coefficient négatif important de -0.7. Par contre, les coefficients après traitement sont toujours statistiquement non différents de zéro. Pour 100M, il n'y a pas de pré-tendance, par contre on constate une baisse des brevets collaboratifs bien plus importante suite au déploiement de la fibre.

L'effet agrégé total est ici de -0.44, plus important en magnitude que celui sur la production de brevets.

Tableau 6.2: Effet de l'accès au très haut débit sur la production de brevets.
Estimation par modèle de Poisson

Type d'accès : Model :	30M+		# Brevets 100M		≥ 30M et < 100M	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Variables</i>						
Année = -6	0.1477 (0.2877)		0.2972 (0.2909)		0.1476 (0.2878)	
Année = -5	-0.2493 (0.1910)		-0.1489 (0.2208)		-0.2521 (0.1933)	
Année = -4	-0.2186 (0.1545)		-0.2818 (0.1744)		-0.2175 (0.1550)	
Année = -3	-0.2177* (0.1308)		-0.3154 (0.2409)		-0.2225* (0.1307)	
Année = -2	0.1161 (0.0966)		0.0172 (0.1447)		0.1147 (0.0975)	
Année = 0	-0.0287 (0.0880)		-0.1859 (0.1530)		0.0010 (0.0865)	
Année = 1	-0.1287 (0.1470)		-0.5540*** (0.1761)		-0.0221 (0.1499)	
Année = 2	-0.2648 (0.1920)		-0.3698** (0.1762)		-0.2491 (0.2040)	
Année = 3	-0.2531 (0.2696)		0.0139 (0.8591)		-0.1342 (0.2838)	
Effet total		-0.1516 (0.1328)		-0.3515** (0.1431)		-0.0897 (0.1345)
<i>Fixed-effects</i>						
Etablissement	Yes	Yes	Yes	Yes	Yes	Yes
Année	Yes	Yes	Yes	Yes	Yes	Yes
<i>Fit statistics</i>						
Observations	6,330	6,330	14,687	14,687	5,983	5,983
Squared Correlation	0.90305	0.90305	0.82373	0.82373	0.90579	0.90579
Pseudo R ²	0.65621	0.65621	0.61976	0.61976	0.65654	0.65654
BIC	19,759.0	19,759.0	47,968.7	47,968.7	18,897.4	18,897.4

Clustered (Etablissement) standard-errors in parentheses
Signif. Codes : *** : 0.01, ** : 0.05, * : 0.1

Tableau 6.3: Effet de l'accès au très haut débit sur la collaboration

Type d'accès : Modèle :	# Brevets collab. >50km					
	30M+ (1)	(2)	100M (3)	(4)	≥ 30M et < 100M (5)	(6)
<i>Variables</i>						
Année = -6	-0.2075 (0.4271)		0.1918 (0.4153)		-0.2045 (0.4297)	
Année = -5	-0.7420** (0.3097)		-0.1470 (0.2133)		-0.7472** (0.3148)	
Année = -4	-0.2699 (0.2069)		-0.0517 (0.2217)		-0.2673 (0.2068)	
Année = -3	-0.1076 (0.2088)		0.0697 (0.2534)		-0.1134 (0.2079)	
Année = -2	0.1687 (0.1696)		0.1978 (0.2410)		0.1714 (0.1715)	
Année = 0	0.1249 (0.1317)		-0.3616 (0.2954)		0.1964 (0.1348)	
Année = 1	-0.0553 (0.2106)		-0.1852 (0.2745)		0.0070 (0.2136)	
Année = 2	-0.1110 (0.2973)		-0.6677** (0.3358)		0.0658 (0.3102)	
Année = 3	-0.0772 (0.4506)		-1.818** (0.8089)		-0.0214 (0.4988)	
ATT		-0.0166 (0.1951)		-0.4422* (0.2291)		0.0777 (0.1973)
<i>Fixed-effects</i>						
Etablissement	Yes	Yes	Yes	Yes	Yes	Yes
Année	Yes	Yes	Yes	Yes	Yes	Yes
<i>Fit statistics</i>						
Observations	2,417	2,417	5,882	5,882	2,265	2,265
Squared Correlation	0.74752	0.74752	0.65803	0.65803	0.75972	0.75972
Pseudo R ²	0.54470	0.54470	0.49956	0.49956	0.55140	0.55140
BIC	7,032.7	7,032.7	17,456.4	17,456.4	6,663.1	6,663.1

Clustered (Etablissement) standard-errors in parentheses
*Signif. Codes : *** : 0.01, ** : 0.05, * : 0.1*

Maintenant regardons la robustesse de ces résultats en changeant le mode d'estimation. Nous estimons ainsi le même modèle mais par les OLS. Pour cela, notre échantillon inclut l'ensemble des établissements. Comme discuté dans la section précédente, pour ne pas biaiser les résultats vers 0, nous utilisons uniquement les établissements actifs dans des secteurs qui sont les plus susceptibles de breveter. Nous avons déjà décrit ces secteurs précédemment. La base contient alors 2 714 094 observations dont 381 659 établissements. Cet échantillon diffère donc de celui d'avant

en cela que *i*) les secteurs sont sélectionnés, seulement les plus intenses en brevets sont conservés, et *ii*) tous les établissements sont inclus.

Les résultats sont illustrés par les Figures 6.16 et 6.17 et le Tableau 6.4. Ces figures sont proches de celles obtenues avec les estimations en Poisson. Il y a une absence de pré-tendance et, concernant le 100M, un effet négatif pour la première et deuxième année après le déploiement. La magnitude doit ici être interprétée en niveau et peut sembler faible, -0.0062 pour la deuxième année, mais c'est un effet moyen sur tous les établissements (357 129) et non pas seulement sur ceux qui déposent des brevets. A la différence de l'estimation en Poisson, cette fois-ci il y a des effets négatifs qui apparaissent également pour l'accès à 30M et ceux-ci suivent le même schéma que pour le 100M mais ont une magnitude plus faible. L'effet total agrégé, bien que négatif, est néanmoins non statistiquement différent de 0.

Figure 6.16: Effet de l'accès à un débit de 30M sur la production annuelle de brevet.
Estimation par OLS

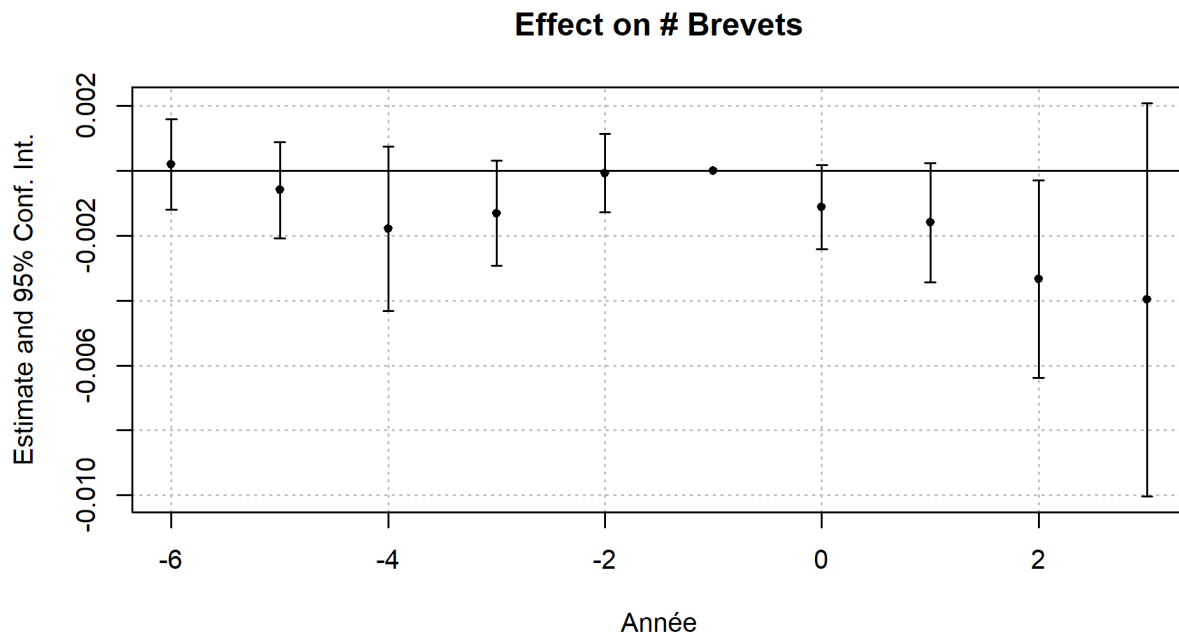


Figure 6.17: Effet de l'accès à un débit de 100M sur la production annuelle de brevet.
Estimation par OLS

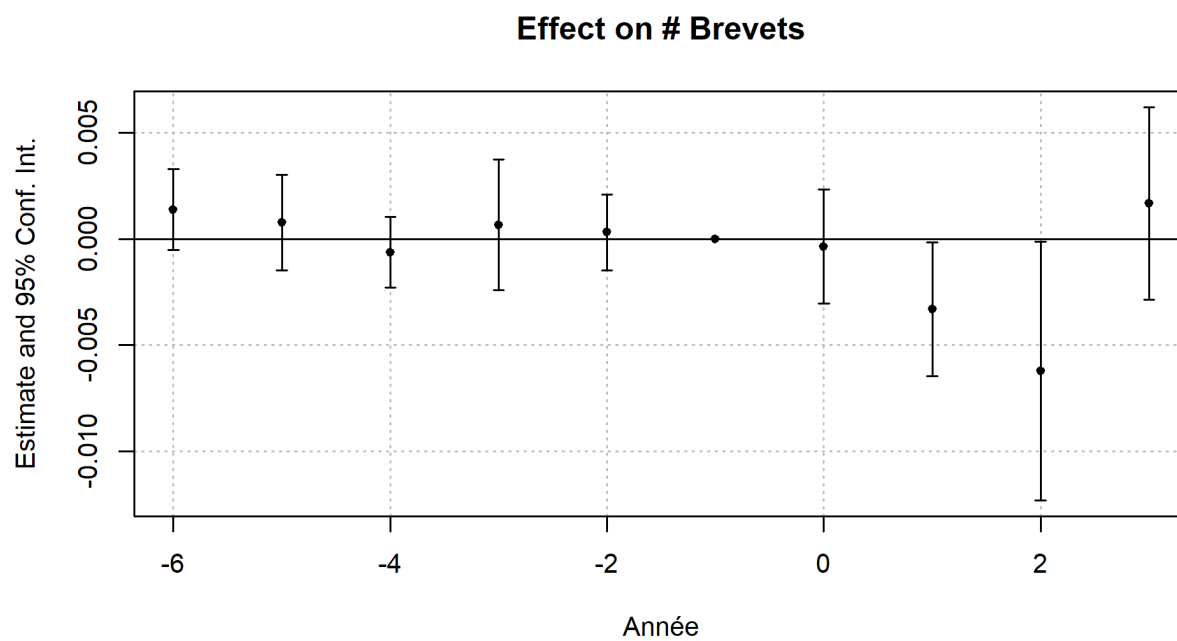


Tableau 6.4: Effet de l'accès au très haut débit sur la production de brevets et la collaboration. Estimation par OLS

Dependent Variables :	# Brevets				# Brevets collab. >50km			
Type d'accès :	30M+		100M		30M+		100M	
Model :	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Variables</i>								
Année = -6	0.0002 (0.0007)		0.0014 (0.0010)		-0.0005 (0.0004)		-0.0001 (0.0005)	
Année = -5	-0.0006 (0.0008)		0.0008 (0.0011)		-0.0012* (0.0007)		-0.0006 (0.0008)	
Année = -4	-0.0018 (0.0013)		-0.0006 (0.0008)		-0.0005 (0.0004)		-0.0003 (0.0005)	
Année = -3	-0.0013 (0.0008)		0.0007 (0.0016)		-1.77×10^{-5} (0.0004)		0.0006 (0.0010)	
Année = -2	-7.15×10^{-5} (0.0006)		0.0003 (0.0009)		-0.0001 (0.0004)		-0.0003 (0.0007)	
Année = 0	-0.0011* (0.0007)		-0.0004 (0.0014)		-0.0005 (0.0005)		-0.0013 (0.0010)	
Année = 1	-0.0016* (0.0009)		-0.0033** (0.0016)		-0.0006 (0.0005)		-0.0015 (0.0010)	
Année = 2	-0.0033** (0.0016)		-0.0062** (0.0031)		-0.0009 (0.0007)		-0.0026* (0.0016)	
Année = 3	-0.0040 (0.0031)		0.0017 (0.0023)		-9.52×10^{-5} (0.0005)		0.0006** (0.0003)	
ATT		-0.0022* (0.0011)		-0.0026 (0.0016)		-0.0006 (0.0005)		-0.0015 (0.0010)
<i>Fixed-effects</i>								
Etablissement	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Année	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>Fit statistics</i>								
Observations	1,224,365	1,224,365	2,532,690	2,532,690	1,224,365	1,224,365	2,532,690	2,532,690
R ²	0.81219	0.81219	0.78012	0.78012	0.69195	0.69195	0.65457	0.65457
Within R ²	2.65×10^{-5}	2.65×10^{-5}	2.09×10^{-5}	2.09×10^{-5}	2.82×10^{-5}	2.82×10^{-5}	2.02×10^{-5}	2.02×10^{-5}

Clustered (Etablissement) standard-errors in parentheses
 Signif. Codes : *** : 0.01, ** : 0.05, * : 0.1

Enfin, lorsque l'on regarde les collaborations (Tableau 6.4, colonnes 5 à 8), les résultats sont similaires aux résultats en Poisson, à cela près que l'effet de la troisième année après l'accès à 100M est ici positif plutôt que fortement négatif. Ces résultats laissent à penser que l'impact estimé pour cette année-là doit être pris avec précaution. En particulier, l'effet très négatif trouvé avec le modèle de Poisson, ne se basant que sur très peu d'unités traitées, n'est peut-être pas généralisable.

3. Analyse de l'enquête CIS

3.1. Présentation et construction de la base de données CIS

Les *enquêtes communautaires sur l'innovation*, devenues les *enquêtes capacité à innover et stratégie* depuis 2018, ci-nommées CIS, sont menées tous les deux ans par l'INSEE et visent à comprendre le comportement d'innovation des entreprises. Les questions principales portent sur l'introduction d'innovations avec en particulier de nombreux détails sur le type d'innovation (par exemple: est-ce l'innovation de produit concerne les services ? Est-ce que l'innovation de procédé porte sur l'organisation ou le marketing ? etc.).

Outre la production, les enquêtes abordent les moyens de l'innovation : par exemple, si les entreprises font de la R&D en interne, ou encore si elles collaborent avec d'autres entreprises ou organismes publics. De plus, sont incluses des données de cadrage sur chacune des firmes enquêtées comme le nombre d'employés, le chiffre d'affaire ou le secteur.

Dans ce chapitre, nous utilisons les enquêtes CIS 2014, 2016 et 2018 puisque les données de couverture numérique ne sont disponibles qu'à partir de 2013. Nous commençons d'abord par présenter la méthode et le périmètre de ces trois enquêtes avant d'aborder plus en détails les variables que nous utiliserons et leur cohérence entre les différentes cohortes.

Périmètre, échantillonnage et détails de l'enquête

Les enquêtes portent sur les entreprises de 10 employés ou plus. Tous les secteurs sont concernés sauf les sections AZ, OQ et RU de la NAF rev.2. Ainsi, les seuls secteurs exclus sont : agriculture, sylviculture et pêche; administration publique, enseignement, santé humaine et sociale; autres activités de service. La définition de l'entreprise est l'unité légale identifiée par son numéro de SIREN, cela signifie qu'une entreprise enquêtée peut compter de multiples établissements et que l'enquête porte donc sur la production de l'ensemble de ces établissements.

L'enquête est exhaustive pour les entreprises de 250 salariés ou plus. Pour les entreprises plus petites, l'enquête est stratifiée à la fois par secteur (au niveau A88, soit la *division*, de la NAF rev.2) et par tranche d'effectif. Suite à la stratification, les entreprises sondées sont associées à un poids de calage représentant leur probabilité inverse d'avoir été échantillonnées. Par la suite, il faudra utiliser ces poids dans les estimations et statistiques descriptives pour obtenir des statistiques représentatives de la population totale. A noter que la localisation géographique est un élément important de notre étude mais que la stratification ne tient pas

compte de la géographie.

Au total, de 10% à 15% de la population cible est échantillonnée par enquête. Par exemple, en 2014, sur une population cible de 151 294, 23 154 sont échantillonnées, menant à 18 109 questionnaires utilisables.

Les formulaires d'enquête, si remplis complètement (ce qui implique que les entreprises innovent), contiennent environ une centaine de questions. Le temps médian de réponse pour ces enquêtes, incluant le temps de recherche d'information, est de 30 minutes.

Variables et cohérence temporelle

Pour se donner une idée qualitative des types d'innovation reportées par les entreprises, nous pouvons regarder l'innovation la plus importante produite par ces entreprises sur les trois années précédant l'enquête. C'est une des seules questions sous format libre demandées par l'enquête, ce qui la rend très informative. Quelques exemples sont reportés dans le tableau 6.5 qui suit. On peut constater que les innovations rapportées sont très spécifiques aux firmes et, bien que pouvant avoir des effets positifs concrets sur la productivité, ne sont pas des innovations à portée internationale, contrairement aux brevets (à noter que cela n'est qu'une sélection, d'autres innovations plus techniques et précises ne pouvant être rapportées par souci de confidentialité).

Tableau 6.5 : Exemples d'innovations décrites par les entreprises

Question: Pouvez-vous décrire votre principale innovation au cours des trois dernières années ?

Réponses (échantillon aléatoire, enquête 2014 ou 2018)

Gestion dynamique des compétences.

Amélioration logiciel comptable.

On se sert de la chaleur produite par la salle des machines pour chauffer en partie la surface de vente.

L'innovation principale au cours des trois années concerne essentiellement la partie procédures d'entreprises, développement commercial, logistique du suivi de chantiers, de la gestion informatique. Il s'agit d'amélioration de gestion surtout.

Nouveau procédé d'extraction de matières végétales.

Mise aux normes iso 9001 versus 2015 de la société.

Création de notre site vitrine pour notre clientèle.

Sources: INSEE, enquêtes CIS 2014 et 2018.

Toutes les variables qui seront utilisées pour mesurer l'activité d'innovation dans notre analyse sont listées dans le Tableau 6.6. Ces variables principales sont de trois types.

Premièrement les mesures de **production d'innovation**, que nous décomposons par types et sous-types. C'est-à-dire si l'entreprise a produit une innovation de produit, en distinguant deux sous-types (produit ou service), et/ou une innovation de procédé (avec 5 sous-types) durant ses trois dernières années d'activité. Ces variables sont binaires et ne prennent donc pas en compte la quantité. Par exemple, une entreprise ayant développé 5 innovations produits est, par ces mesures, similaire aux entreprises n'en ayant produit qu'une seule. A noter que certaines de ces variables ne sont pas comparables entre les différentes vagues de questionnaires, nous détaillerons le traitement effectué par la suite.

Deuxièmement, nous utiliserons les mesures de l'importance des innovations produites – ces mesures reflétant en quelque sorte la **qualité de l'innovation**. Il y a deux types de mesures. La première série de mesures, portant uniquement sur les innovations de produit, concerne la portée géographique de la nouveauté de l'innovation, c'est-à-dire si l'innovation est nouvelle en France, en Europe ou dans le monde. La seconde série de mesures est plus quantitative et correspond à la part de marché associée aux produits de ces innovations, en fonction de leur portée géographique.

Troisièmement, un ensemble de mesures portant sur l'aspect **collaboratif** de la production d'innovation. Dans le cas où l'entreprise déclare une collaboration, deux caractéristiques de la collaboration sont analysées. Tout d'abord, avec quel type de partenaire l'entreprise développe son innovation. Ensuite, quelle est la dimension géographique de la collaboration : est-ce une collaboration nationale, européenne ou mondiale ?

Le Tableau A.6.1 en annexe détaille précisément la construction de toutes ces variables.

Cohérence temporelle

Concernant la cohérence temporelle de ces variables, la majorité d'entre elles se retrouve dans l'ensemble des trois questionnaires de 2014, 2016 et 2018, avec une formulation quasi-identique des questions.

Les variables relatives aux innovations de procédé (à la fois en général et par sous-types) manquent toutefois de comparabilité. En effet, dans les cohortes 2014 et 2016, les innovations marketing et organisationnelles sont des catégories à part dans l'enquête. Elles bénéficient de sections entières où de nombreuses sous-questions sont posées à leur sujet. A contrario, dans l'enquête 2018, les innovation marketing et organisationnelles sont considérées comme des innovations de procédé classiques et, à ce titre, ne bénéficient d'aucune mise en avant particulière dans le questionnaire. Ainsi, de sections entières en 2014-2016, elles passent à seulement deux lignes en 2018. Cela implique mécaniquement que les innovations marketing et organisationnelles sont moindres en 2018 comparé à 2014-2016. Ce phénomène est intéressant et montre par-là les limites de ces données d'enquête car, dans les deux cas, les idées d'innovation marketing et organisationnelle sont les mêmes mais les valeurs obtenues diffèrent en fonction de la forme des questions. Pour en tenir compte, lorsque nous analyserons les innovations de procédé, nous distinguerons ainsi les années 2014-2016 et 2018.

Concernant les mesures de la qualité d'innovation, l'enquête 2018 ne contient pas la question de la portée géographique de la nouveauté des innovations de produit. Ainsi cette information n'est disponible que pour les années 2014 et 2016. Toujours sur ces variables, la question relative à la part de chiffre d'affaire provenant de produits nouveaux pour le marché a été modifiée en 2018. Cela implique une cassure bien que l'information reportée est censée être la même.

Tableau 6.6 : Liste des variables dépendantes utilisées dans l'analyse

Variables	
Production	
► <i>Les variables qui suivent sont binaires. Elles valent 1 si au moins une innovation du type décrit a été réalisée dans les trois dernières années précédant l'enquête.</i>	
Innovation produit : tous types (<i>au moins une innovation de produit, quelle qu'elle soit</i>)	
Innovation produit : nouveau produit	Innovation produit : nouveau service
Innovation procédé : tous types (<i>au moins une innovation de procédé, quelle qu'elle soit</i>)	
Innovation procédé : fabrication	Innovation procédé : logistique
Innovation procédé : marketing	Innovation procédé : organisationnel
Innovation procédé : support	
Qualité	
► <i>Les variables qui suivent sont binaires. Elles valent 1 si au moins une innovation du type décrit a été réalisée dans les trois dernières années précédant l'enquête. Ces variables sont disponibles pour les enquêtes 2014 et 2016 uniquement.</i>	
Innovation produit : nouveau pour la France	Innovation produit : nouveau pour l'Europe
Innovation produit : nouveau pour le monde	
► <i>La variable qui suit varie entre 0 et 100.</i>	
Fraction du CA due à un produit nouveau pour le marché	
Collaboration (seulement les collaborations dans le but d'innover)	
► <i>Les variables qui suivent sont binaires. Elles valent 1 si au moins une collaboration du type décrit a été réalisée dans les trois dernières années précédant l'enquête.</i>	
Collaboration avec un partenaire pour innover (<i>quel que soit le partenaire</i>)	
Collaboration avec une autre entreprise du groupe	Collaboration avec un partenaire hors groupe
Collaboration avec les clients	Collaboration avec une entreprise privée
Collaboration avec une entreprise du secteur public	
Collaboration avec un partenaire situé en France	Collaboration avec un partenaire situé en Europe
Collaboration avec un partenaire situé hors de l'Europe	
Collaboration avec une autre entreprise du groupe, située en France	Collaboration avec une autre entreprise du groupe, située en Europe
Collaboration avec une autre entreprise du groupe, située dans le monde	

Variables	
Collaboration avec un partenaire hors groupe, situé en France	Collaboration avec un partenaire hors groupe, situé en Europe
Collaboration avec un partenaire hors groupe, situé dans le monde	

Les informations disponibles dans l'enquête CIS nous permettent donc d'analyser l'effet de l'implantation du très haut débit dans les zones rurales sur la **production** et la **qualité** d'innovation, ainsi que sur les **collaborations** associées. Avant d'effectuer l'analyse économétrique nous regardons d'abord les statistiques descriptives sur ces mesures.

3.2. Statistiques descriptives des données CIS

Nous commentons séparément les trois types de variables utilisées. Les données rapportées ne concernent que les entreprises situées dans des communes de la ZIPU.

Production d'innovation

Tout d'abord, concernant la production d'innovation : les statistiques descriptives sont reportées sur le Tableau 6.7. Ce tableau rapporte, pour chaque enquête et variable, les moyennes pour les entreprises situées dans les communes ayant accès à la fibre et celles n'ayant pas accès.

Sur les trois années, environ 20% des entreprises reportent avoir développé des innovations de produit, les nouveaux produits (env. 15% des firmes en reportent) étant plus fréquents que les nouveaux services (env. 10% des firmes en reportent).

Les innovations de procédé sont plus fréquentes que les innovations produit, avec plus de 40% des entreprises en reportant. Néanmoins ce nombre tombe à 30% en 2018, en raison de la redéfinition des variables d'innovation marketing et organisationnelles dans cette enquête comme expliqué précédemment. En regardant plus dans le détail, cette chute au niveau des innovations marketing et organisationnelles est notable : passant d'environ 25% à 12% et d'environ 30% à 20% pour chacune d'elles.

Tableau 6.7 : Statistiques descriptives sur les variables de production d'innovation

Accès fibre	2014			2016			2018		
	Non	Oui	Diff.	Non	Oui	Diff.	Non	Oui	Diff.
Innovation produit									
Tous type	0.18 (0.38)	0.24 (0.43)	0.067 [0.000]	0.18 (0.39)	0.21 (0.41)	0.027 [0.000]	0.22 (0.42)	0.26 (0.44)	0.036 [0.000]
Produit	0.13 (0.34)	0.19 (0.39)	0.060 [0.000]	0.14 (0.35)	0.16 (0.37)	0.020 [0.000]	0.18 (0.38)	0.21 (0.4)	0.030 [0.000]
Service	0.094 (0.29)	0.13 (0.33)	0.032 [0.000]	0.1 (0.3)	0.11 (0.31)	0.010 [0.011]	0.13 (0.33)	0.15 (0.36)	0.023 [0.000]
Innovation de procédé									
Tous types	0.41 (0.49)	0.47 (0.5)	0.055 [0.000]	0.44 (0.5)	0.49 (0.5)	0.045 [0.000]	0.3 (0.46)	0.33 (0.47)	0.031 [0.000]
Fabrication	0.15 (0.36)	0.16 (0.37)	0.014 [0.011]	0.16 (0.37)	0.17 (0.37)	0.007 [0.115]	0.16 (0.37)	0.18 (0.39)	0.023 [0.000]
Logistique	0.078 (0.27)	0.078 (0.27)	0.001 [0.843]	0.08 (0.27)	0.094 (0.29)	0.014 [0.000]	0.083 (0.28)	0.089 (0.29)	0.007 [0.027]
Marketing	0.2 (0.4)	0.25 (0.43)	0.047 [0.000]	0.24 (0.43)	0.26 (0.44)	0.024 [0.000]	0.11 (0.31)	0.12 (0.33)	0.014 [0.000]
Organisationnel	0.28 (0.45)	0.33 (0.47)	0.053 [0.000]	0.32 (0.47)	0.36 (0.48)	0.043 [0.000]	0.18 (0.38)	0.21 (0.41)	0.034 [0.000]
Support	0.094 (0.29)	0.11 (0.32)	0.019 [0.000]	0.11 (0.31)	0.13 (0.34)	0.026 [0.000]	0.18 (0.38)	0.21 (0.4)	0.025 [0.000]
N (pondéré)	44,799	4,519		40,395	7,967		35,965	11,517	

Notes : les données sont issues des enquêtes CIS 2014, 2016 et 2018. Les calculs sont ceux des auteurs. Toutes les entreprises font partie des ZIPU. Les variables reportées sont les moyennes et leurs différences. Les variables entre parenthèses sont les écarts-types et les valeurs entre crochets sont les p-values d'un test de différence de moyenne. Chaque entreprise est pondérée par son poids d'enquête.

Passons maintenant aux différences entre les zones ayant accès à la fibre et celles non couvertes. D'une façon générale, il est clair que les entreprises dans les communes ayant accès à la fibre reportent des mesures de production d'innovation toutes supérieures aux entreprises

sans accès fibre, et ce quelle que soit la cohorte. Par exemple, pour 2014, il y a 5% d'entreprises en plus, en point de pourcentage, déclarant des innovations de produit dans les zones fibre ; de même pour les innovations de procédé. Il semble ainsi, que d'un point de vue purement descriptif, au sein des ZIPU, les entreprises dans les communes ayant accès à la fibre sont plus innovantes que les autres.

Qualité d'innovation

Le Tableau 6.8 reporte les statistiques descriptives pour la qualité d'innovation. Plus précisément, *pour l'ensemble des entreprises ayant au moins une innovation de produit*, il reporte la portée géographique de l'innovation ainsi que la part de chiffre d'affaire (CA) dégagé par l'innovation nouvelle pour le marché.

Assez logiquement, le nombre d'innovations nouvelles pour la France (environ 55%) sont plus nombreuses que celles nouvelles pour l'Europe (environ 40%), elles-mêmes moins nombreuses que celles nouvelles au niveau mondial (environ 30%).

Le CA issu de produits nouveaux pour le marché est d'environ 12% sur 2014-2016, et d'environ 7% pour 2018. On peut constater une cassure nette avec l'enquête 2018 où la formulation de la question a été modifiée.

Maintenant en se concentrant sur les différences entre les entreprises selon leur accès à la fibre, on constate, comme pour les mesures de production, que les entreprises implantée dans des communes ayant accès à la fibre ont des mesures de qualité d'innovation systématiquement supérieures.

Par exemple, les entreprises ayant accès à la fibre ont 10% (en points de pourcentage) de plus d'innovations nouvelles au niveau mondial. C'est une différence très importante. Mais c'est également le cas pour les innovations nouvelles pour la France (env. 7% de plus) ou nouvelles pour l'Europe (env. 8%). A noter que l'avantage le plus important est au niveau mondial.

Ce schéma se retrouve aussi lorsque l'on regarde le CA dû aux produits nouveaux pour le marché : d'une différence négative mais quasi non significative en 2014, on passe à un avantage pour les entreprises ayant accès à la fibre en 2016 et 2018.

Tableau 6.8 : Statistiques descriptives sur la qualité d'innovation

Accès fibre	2014			2016			2018		
	Non	Oui	Diff.	Non	Oui	Diff.	Non	Oui	Diff.
Introduction d'un produit nouveau pour...									
... la France	0.53 (0.5)	0.58 (0.49)	0.049 [0.007]	0.55 (0.5)	0.62 (0.48)	0.078 [0.000]			
... l'Europe	0.35 (0.48)	0.44 (0.5)	0.090 [0.000]	0.35 (0.48)	0.39 (0.49)	0.040 [0.012]			
... le monde	0.25 (0.43)	0.35 (0.48)	0.108 [0.000]	0.25 (0.43)	0.33 (0.47)	0.084 [0.000]			
Fraction du CA dû à un produit nouveau pour									
... le marché	11 (21)	10 (18)	-1.298 [0.027]	13 (22)	16 (25)	3.328 [0.000]	5.6 (8.7)	8.9 (15)	3.366 [0.000]
N (pondéré)	7,941	1,106		7,402	1,672		7,981	2,971	

Notes : les données sont issues des enquêtes CIS 2014, 2016 et 2018. Les calculs sont ceux des auteurs. Toutes les entreprises font partie des ZIPU. Les variables reportées sont les moyennes et leurs différences. Les variables entre parenthèses sont les écarts-types et les valeurs entre crochets sont les p-values d'un test de différence de moyenne. Chaque entreprise est pondérée par son poids d'enquête. CA signifie chiffre d'affaire.

Collaborations

Nous décrivons à présent les statistiques pour la collaboration dans l'innovation, décrites dans le Tableau 6.9.

Tout d'abord, environ 8% des entreprises sans accès à la fibre déclarent collaborer concernant leur activité d'innovation. Ce nombre monte à environ 12% pour les entreprises ayant accès à la fibre. Au niveau temporel, on peut constater que le nombre de collaborations baisse légèrement pour les entreprises de communes ayant accès à la fibre, passant de 13% en 2014 à 11% en 2016 et 2018.

Si l'on détaille les types de collaborations, on peut constater qu'environ 5% des collaborations se font avec d'autres entreprises du groupe. Si l'on regarde les collaborations avec un partenaire qui ne fait pas partie du groupe, ce nombre passe à 9% environ. Au niveau de la dynamique temporelle, ces chiffres sont relativement stables.

Ensuite, que ce soit des collaborations avec des clients, avec d'autres entreprises privées, ou des entités publiques, de la même façon les entreprises ayant accès à la fibre collaborent plus. Néanmoins, il n'y a pas vraiment de dynamique temporelle qui se dégage.

Les entreprises ayant accès à la fibre collaborent presque deux fois plus (2%) avec des partenaires situés hors de France et hors Europe que les entreprises sans accès (1%).

Pour les autres variables, qui combinent le type de partenaire avec la zone géographique de collaboration, le schéma qui ressort est toujours le même : les entreprises ayant accès à la fibre collaborent plus que les entreprises sans accès, mais sans qu'il y ait de dynamique temporelle évidente.

Tableau 6.9 : Statistiques descriptives sur la collaboration dans les activités d'innovation

Accès fibre	2014			2016			2018		
	Non	Oui	Diff.	Non	Oui	Diff.	Non	Oui	Diff.
Collaboration									
Tous types	0.081	0.13	0.050	0.091	0.11	0.014	0.076	0.11	0.034
	(0.27)	(0.34)	[0.000]	(0.29)	(0.31)	[0.000]	(0.26)	(0.31)	[0.000]
Collaboration avec...									
...une autre entreprise du groupe	0.039	0.079	0.039	0.05	0.065	0.016	0.028	0.047	0.020
	(0.19)	(0.27)	[0.000]	(0.22)	(0.25)	[0.000]	(0.16)	(0.21)	[0.000]
...un partenaire hors groupe	0.075	0.11	0.034	0.085	0.099	0.014	0.073	0.1	0.030
	(0.26)	(0.31)	[0.000]	(0.28)	(0.3)	[0.000]	(0.26)	(0.3)	[0.000]
... des clients	0.028	0.044	0.016	0.054	0.062	0.008	0.037	0.055	0.018
	(0.16)	(0.2)	[0.000]	(0.23)	(0.24)	[0.007]	(0.19)	(0.23)	[0.000]
...une entreprise privée	0.064	0.089	0.025	0.076	0.089	0.013	0.068	0.093	0.024
	(0.24)	(0.28)	[0.000]	(0.26)	(0.28)	[0.000]	(0.25)	(0.29)	[0.000]
... une entreprise du secteur public	0.03	0.061	0.032	0.034	0.045	0.011	0.032	0.059	0.027
	(0.17)	(0.24)	[0.000]	(0.18)	(0.21)	[0.000]	(0.17)	(0.24)	[0.000]
Collaboration avec un partenaire situé...									
... en France	0.077	0.12	0.045	0.088	0.1	0.015	0.071	0.099	0.028
	(0.27)	(0.33)	[0.000]	(0.28)	(0.3)	[0.000]	(0.26)	(0.3)	[0.000]
... en Europe	0.025	0.042	0.017	0.033	0.049	0.016	0.021	0.033	0.012
	(0.16)	(0.2)	[0.000]	(0.18)	(0.22)	[0.000]	(0.14)	(0.18)	[0.000]
... dans le monde	0.011	0.02	0.009	0.016	0.021	0.005	0.0097	0.02	0.011
	(0.1)	(0.14)	[0.000]	(0.13)	(0.14)	[0.008]	(0.098)	(0.14)	[0.000]
Collaboration avec une autre entreprise du groupe située...									
... en France	0.032	0.063	0.031	0.043	0.052	0.010	0.022	0.038	0.016
	(0.18)	(0.24)	[0.000]	(0.2)	(0.22)	[0.000]	(0.15)	(0.19)	[0.000]

Accès fibre	2014			2016			2018		
	Non	Oui	Diff.	Non	Oui	Diff.	Non	Oui	Diff.
... en Europe	0.011 (0.11)	0.022 (0.15)	0.011 [0.000]	0.011 (0.1)	0.022 (0.15)	0.010 [0.000]	0.0091 (0.095)	0.013 (0.11)	0.004 [0.001]
... dans le monde	0.0052 (0.072)	0.014 (0.12)	0.009 [0.000]	0.0057 (0.075)	0.013 (0.11)	0.008 [0.000]	0.0031 (0.056)	0.0068 (0.082)	0.004 [0.000]
Collaboration avec un partenaire hors groupe situé...									
... en France	0.071 (0.26)	0.11 (0.31)	0.034 [0.000]	0.083 (0.28)	0.094 (0.29)	0.011 [0.001]	0.068 (0.25)	0.091 (0.29)	0.023 [0.000]
... en Europe	0.02 (0.14)	0.032 (0.18)	0.012 [0.000]	0.029 (0.17)	0.043 (0.2)	0.014 [0.000]	0.018 (0.13)	0.028 (0.16)	0.010 [0.000]
... dans le monde	0.0077 (0.087)	0.011 (0.1)	0.003 [0.044]	0.013 (0.11)	0.015 (0.12)	0.002 [0.262]	0.0076 (0.087)	0.018 (0.13)	0.011 [0.000]
N (pondéré)	44,799	4,519		40,395	7,967		35,965	11,517	

Notes : les données sont issues des enquêtes CIS 2014, 2016 et 2018. Les calculs sont ceux des auteurs. Toutes les entreprises font partie des ZIPU. Les variables reportées sont les moyennes et leurs différences. Les variables entre parenthèses sont les écarts-types et les valeurs entre crochets sont les p-values d'un test de différence de moyenne. Chaque entreprise est pondérée par son poids d'enquête.

Conclusion des statistiques descriptives

Les statistiques descriptives dépeignent un tableau cohérent où les entreprises des communes ayant accès à la fibre ont des mesures d'innovation, tant sur la production, la qualité ou la collaboration, qui sont systématiquement supérieures à celles des entreprises de communes n'ayant pas accès.

Néanmoins, les différences de performance d'innovation peuvent simplement résulter du fait que les entreprises des communes ayant accès à la fibre sont différentes des entreprises de communes non couvertes (par exemple, en termes de taille, de secteur, etc.). Pour isoler l'effet de l'accès à la fibre sur la performance d'innovation, nous utilisons donc des méthodes économétriques décrites dans la section suivante.

3.3.Méthode économétrique

Comme expliqué dans la section précédente, les enquêtes CIS sont menées sur des échantillons de la population d'entreprises, avec une représentativité sectorielle. De ce fait, la structure des données correspond à des coupes répétées. Autrement dit, nous ne pouvons observer l'évolution d'une même entreprise au cours temps, nous empêchant de contrôler pour les caractéristiques propres à chaque entreprise et constantes dans le temps.

Cette structure de données apporte ses challenges propres pour évaluer de façon causale l'effet de l'accès à la fibre. La méthode la plus appropriée qu'il nous est possible d'utiliser est l'appariement. Nous expliquons tout d'abord le principe général de la méthode avant de présenter son application aux données CIS.

Principe

L'estimation de l'effet causal de l'accès à la fibre sur l'innovation se fait en deux étapes : 1) un appariement, et 2) une estimation.

Principe de l'appariement

Dans la première étape nous créons deux groupes d'entreprises : les entreprises traitées et les entreprises de contrôle. Le groupe des traités est simplement composé des entreprises disposant d'un accès à la fibre, le groupe de contrôle étant composé des autres entreprises. Ensuite chaque entreprise traitée est appariée à une entreprise de contrôle en se basant sur un ensemble de variables. L'objectif étant que, pour chaque entreprise traitée, il y ait une entreprise de contrôle lui ressemblant trait pour trait vis-à-vis des variables utilisées pour l'appariement.

Il y a de nombreuses façon de réaliser un appariement, dans la pratique nous utiliserons la méthode basée sur le score de propension (voir Caliendo et Kopeinig, 2008). Le score de propension correspond à la probabilité de faire partie du groupe de traités en fonction des caractéristiques observables. Mathématiquement :

$$p(W_i) = E(T_i|W_i)$$

avec $p(W_i)$ le score de propension, W_i l'ensemble des variables d'appariement de l'observation i et T_i une indicatrice valant 1 si l'observation i est traitée. Le score de propension est estimé avec un modèle Logit de telle sorte que :

$$p(W_i) = E(T_i|W_i) = \text{Logit}(\hat{\beta}W_i),$$

avec $\hat{\beta}$ le vecteur de coefficients estimé par le modèle Logit.

Chaque observation traitée est ensuite appariée à une unité de contrôle ayant le score de propension le plus proche. Les unités de contrôle non appariées sont retirées de l'échantillon d'estimation.¹⁰³ Le résultat de cette étape est donc un nouvel échantillon, plus restreint que l'échantillon initial, et un score de propension qui sera utilisé pour pondérer les unités de contrôle.

Estimation de l'ATT

Dans la seconde étape, nous estimons l'effet du traitement sur les unités traitées par :

$$y_i = \alpha + \gamma T_i + \epsilon_i,$$

où y_i est une variable qui capture la performance d'innovation de l'entreprise i . La variable T_i vaut 1 si la commune de l'entreprise i dispose de la fibre l'année précédant l'enquête. Le coefficient d'intérêt est ici γ qui correspond à l'effet *causal* du traitement pour les unités traitées (i.e. *average treatment effect for the treated*, ou ATT). Plus généralement, nous ne sommes pas contraints par la forme linéaire, et l'ATT peut être estimé par :

$$E(y_i|T_i) = f(\alpha + \gamma T_i)$$

avec f une fonction qui lie les variables explicatives à l'espérance de la variable dépendante, cela incluant toute forme non linéaire, comme par exemple le modèle Logit, ou encore le modèle de Poisson.

A noter que les estimations sont pondérées à la fois par le poids lié à l'appariement et par les poids associés à l'enquête (cf. Section 4.3 du Chapitre 4).

Hypothèse de causalité

L'hypothèse centrale sur laquelle se base l'interprétation causale de cette méthode est que les unités de contrôle sont des contrefactuels plausibles des unités traitées. C'est-à-dire que celles-ci auraient évolué de façon similaire à l'entreprise contrôle s'il elles n'avaient pas reçu le traitement.

Comme détaillé par la suite, nous utiliserons un riche ensemble de variables d'appariement pour essayer de s'approcher au mieux de la plausibilité d'une interprétation causale, celle-ci restant néanmoins, par nature, une hypothèse.

¹⁰³ À noter que 5 unités traitées sont hors support, c'est-à-dire ne peuvent être appariées à une unité de contrôle dû à une distance de score de propension trop importante.

Opérationnalisation sur les données CIS : appariement

Nous appariions les entreprises avec un ensemble de mesures qui ont trait aux caractéristiques des entreprises et des communes, de façon similaire au Chapitre 4. L'objectif est que ces variables captent au mieux le comportement d'innovation des entreprises, de sorte qu'elles permettent de déterminer correctement des trajectoires d'innovations si aucune entreprise n'avait accès à la fibre.

Comme le traitement, c'est-à-dire l'accès à la fibre, a lieu au niveau communal, il est nécessaire de faire en sorte que les communes des unités traitées et de contrôle soient le plus similaires possible car certaines caractéristiques communales peuvent affecter les activités d'innovation. Les variables utilisées au niveau communal sont la population, le stock d'établissements, la proximité à la ville moyenne la plus proche, une indicatrice de zone montagneuse, et une variable reportant le mauvais accès au THD en 2013.

Les informations sur la population et le stock d'établissements sont issues du recensement de la population et de la base SIRENE, ces deux variables étant construites pour 2013. La proximité aux villes moyennes est une indicatrice valant 1 si la commune se trouve à moins de 20 minutes d'une ville d'au moins 20 000 habitants (en 2013). La variable de zone montagneuse est une indicatrice valant 1 si la commune se situe en zone de moyenne ou haute montagne. Enfin, la variable de mauvaise connexion au THD est une indicatrice qui vaut 1 si moins de 25% des locaux de la commune ont un accès au THD (>30 Mbps) en 2013.

Concernant les variables au niveau entreprise, nous utilisons le secteur, la taille d'entreprise, l'âge de l'entreprise, et enfin si l'entreprise est mono- ou multi-établissements.

Le secteur est bien sûr un déterminant important de l'activité d'innovation. Nous utilisons ici cinq grandes catégories : 1) la manufacture, 2) la construction, 3) le commerce, transport et hébergement, 4) le tertiaire supérieur, et 5) les autres services. Ces catégories sont assez larges afin d'éviter d'avoir un problème de support commun entre les unités de contrôle et les unités traitées. Elles sont construites à partir de la classification NAF rev. 2 issue du répertoire SIRENE.

La taille de l'entreprise est une variable catégorielle de quatre groupes : 1) 0-9 salariés, 2) 10-49 salariés, 3) 50-249 salariés, et 4) plus de 250 salariés. L'âge de l'entreprise, correspondant au nombre d'années depuis sa création, est également une variable catégorielle, de trois groupes : 1) entre 0 et 5 ans, 2) entre 6 et 9 ans, et 3) 10 ans ou plus. Enfin la variable mono-établissement est une indicatrice valant 1 si l'entreprise n'a qu'un seul établissement. Chacune

de ces variables est calculée à partir de la base SIRENE.

Enfin, pour effectuer l'appariement nous utiliserons également des indicatrices pour chaque département et chaque année.

Statistiques descriptives

Le Tableau 6.10 reporte les statistiques descriptives pour les variables d'appariement décrites ci-dessus.

L'échantillon de départ (avant appariement) est constitué de toutes les entreprises appartenant à une des trois cohortes. Cela correspond à 121 160 entreprises de contrôle et 24 003 entreprises traitées (ces nombres prennent en compte la pondération des poids de l'enquête).

Comme on peut le constater, les entreprises ayant accès à la fibre diffèrent sensiblement des entreprises sans accès.

Tout d'abord, les entreprises traitées font partie de communes bien plus grandes (environ 5000 habitants de plus en moyenne, soit le double) et sont également bien plus proches d'une ville moyenne puisque 34% sont situées à moins de 20 minutes d'une ville moyenne contrairement à 16% des unités de contrôle. Les différences concernant le stock d'établissements, l'appartenance à une zone montagneuse ou le mauvais accès au THD sont moins flagrantes bien que toutes soient statistiquement significatives.

Au niveau des caractéristiques des entreprises, la distribution de l'âge des entreprises entre les deux groupes est identique. En termes de secteurs, la proportion des entreprises dans chaque secteur est comparable entre les groupes. Néanmoins on peut constater que les entreprises traitées sont plus fréquemment dans le tertiaire supérieur (12% vs 6%) et moins souvent dans la construction (13% vs 21%).

Les différences les plus importantes concernent la taille des entreprises et leur type. Dans les zones avec accès à la fibre seules 42% des entreprises sont mono établissements alors que ce nombre atteint 75% dans les zones sans accès. Il est aussi notable que les entreprises traitées sont plus grandes que celles de contrôle (26% d'entreprises avec plus de 50 salariés contre 13%).

Tableau 6.10 : Différences entre unités de contrôle et traitées, avant et après appariement

Accès fibre	Avant appariement			Après appariement		
	Non	Oui	Diff.	Non	Oui	Diff.
Variables au niveau commune						
Population	5,069 (32,474)	9,746 (46,397)	4,676.43 [0.000]	9,518 (77,647)	8,459 (48,768)	-1,058.8 [0.104]
Stock établissements	1.8 (10)	1.4 (1.3)	-0.375 [0.000]	1.4 (1.4)	1.4 (1.4)	0.031 [0.028]
Proximité ville moyenne	0.16 (0.37)	0.34 (0.47)	0.175 [0.000]	0.27 (0.44)	0.3 (0.46)	0.026 [0.000]
Zone montagneuse	0.14 (0.35)	0.11 (0.31)	-0.039 [0.000]	0.11 (0.31)	0.11 (0.31)	0.001 [0.800]
Mauvais accès THD	0.64 (0.48)	0.54 (0.5)	-0.099 [0.000]	0.56 (0.5)	0.56 (0.5)	0.004 [0.466]
Variables au niveau entreprise						
Mono-établissement	0.75 (0.43)	0.42 (0.49)	-0.334 [0.000]	0.41 (0.49)	0.43 (0.5)	0.026 [0.000]
Taille d'entreprise						
...0-9 salariés	0.069 (0.25)	0.045 (0.21)	-0.024 [0.000]	0.042 (0.2)	0.046 (0.21)	0.003 [0.110]
...10-49 salariés	0.79 (0.41)	0.69 (0.46)	-0.103 [0.000]	0.72 (0.45)	0.71 (0.45)	-0.008 [0.087]
...50-249 salariés	0.12 (0.33)	0.2 (0.4)	0.078 [0.000]	0.2 (0.4)	0.19 (0.4)	-0.001 [0.847]
...>= 250	0.014 (0.12)	0.063 (0.24)	0.049 [0.000]	0.044 (0.21)	0.049 (0.22)	0.005 [0.013]
Age d'entreprise						
...0-5 ans	0.086 (0.28)	0.086 (0.28)	-0.001 [0.789]	0.087 (0.28)	0.084 (0.28)	-0.002 [0.411]
...6-9 ans	0.091 (0.29)	0.092 (0.29)	0.000 [0.942]	0.094 (0.29)	0.092 (0.29)	-0.002 [0.506]

Accès fibre	Avant appariement			Après appariement		
	Non	Oui	Diff.	Non	Oui	Diff.
...10 ans et +	0.82 (0.38)	0.82 (0.38)	0.000 [0.888]	0.82 (0.38)	0.82 (0.38)	0.004 [0.270]
Secteur						
...Manufacture	0.32 (0.47)	0.29 (0.45)	-0.025 [0.000]	0.31 (0.46)	0.3 (0.46)	-0.013 [0.005]
...Construction	0.21 (0.41)	0.13 (0.34)	-0.073 [0.000]	0.13 (0.34)	0.14 (0.35)	0.011 [0.002]
...Commerce, trans., héberg.	0.38 (0.48)	0.41 (0.49)	0.030 [0.000]	0.42 (0.49)	0.41 (0.49)	-0.007 [0.159]
...Tertiaire sup.	0.059 (0.24)	0.12 (0.32)	0.057 [0.000]	0.1 (0.31)	0.11 (0.31)	0.002 [0.526]
...Autres services	0.04 (0.2)	0.051 (0.22)	0.011 [0.000]	0.042 (0.2)	0.049 (0.22)	0.007 [0.001]
N (pondéré)	121,160	24,003		20,529	19,503	

Notes : les entreprises sont issues des enquêtes CIS 2014, 2016 et 2018. Les variables sont issues du répertoire SIRENE et du recensement de population 2013. Les calculs sont ceux des auteurs. Toutes les entreprises font partie des ZIPU. Les variables reportées sont les moyennes et leurs différences. Les variables entre parenthèses sont les écarts-types et les valeurs entre crochets sont les p-values d'un test de différence de moyenne. Chaque entreprise est pondérée par son poids d'enquête. Les trois cohortes sont regroupées.

Comme nous venons de le voir, avant appariement, les entreprises traitées et de contrôle sont très différentes. Nous procédons donc maintenant à l'appariement par le score de propension afin de constituer un échantillon d'entreprises de contrôle et traitées comparables.

Appariement par le score de propension

Comme vu précédemment, le score de propension peut être estimé par un modèle Logit où la variable dépendante est l'indicatrice de traitement, ici l'accès à la fibre, et les variables explicatives sont l'ensemble des variables d'appariement décrites précédemment.

Nous effectuons l'appariement avec le logiciel R MatchIt (Ho, Imai et King, 2011). Les résultats du modèle Logit sont décrits dans le Tableau 6.11, qui rapporte à la fois le coefficient de l'estimation et l'Odd Ratio (ou rapport de côtes). On remarque qu'un doublement de la

population augmente la probabilité d'avoir accès à la fibre de 29% par rapport à la celle de ne pas y avoir accès ($1.259 = 1.395^{\ln(2)}$). Aussi, être une firme mono-établissement réduit le rapport de chances de près de 86%.

Tableau 6.11 : Estimation du score de propension pour l'échantillon principal

	Logit	
	Coef.	Odd Ratio
Variables au niveau commune		
Population (ln)	0.3330*** (0.0707)	1.395*** (0.0987)
Stock établissements (ln)	0.1821 (0.1927)	1.200 (0.2312)
Proximité ville moyenne	0.4699** (0.1802)	1.600** (0.2884)
Montagne	-0.3910* (0.1575)	0.6764* (0.1066)
Mauvais accès THD	-0.2353 (0.1416)	0.7904 (0.1119)
Variables au niveau entreprise		
Mono-établissement	-1.985*** (0.1361)	0.1374*** (0.0187)
<i>Taille d'entreprise (réf : ≥250 salariés)</i>		
...0-9 salariés	-1.152*** (0.2227)	0.3162*** (0.0704)
...10-49 salariés	-1.030*** (0.1394)	0.3569*** (0.0498)
...50-249 salariés	-0.6884*** (0.1215)	0.5024*** (0.0610)

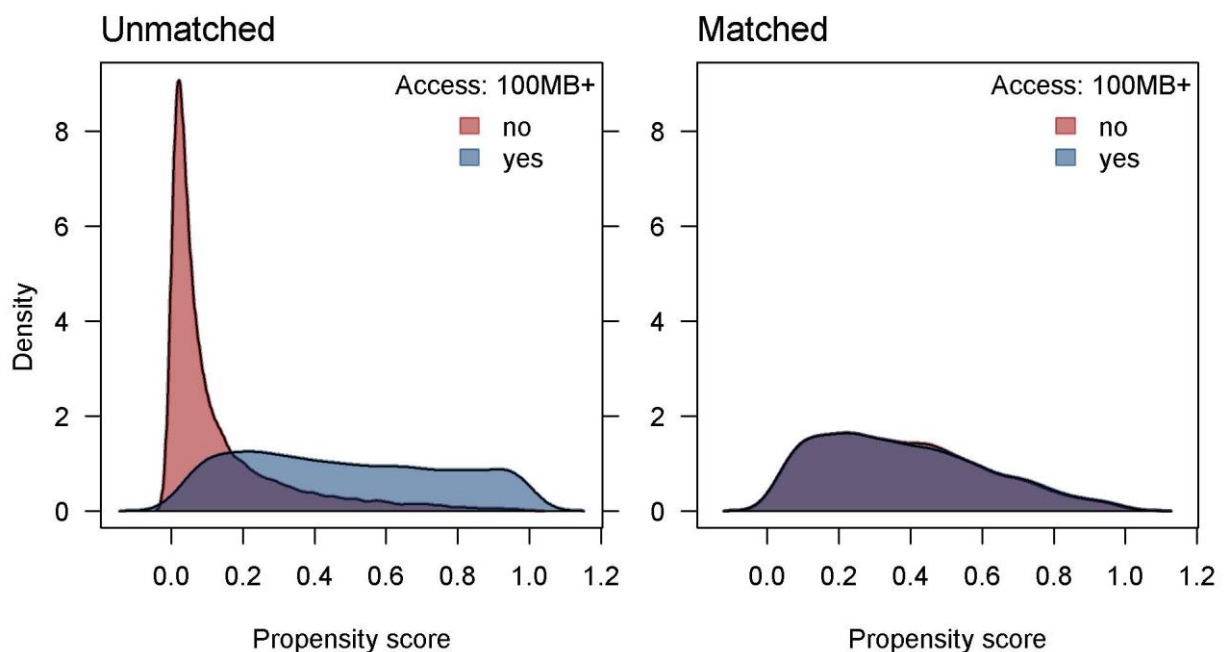
	Logit	
	Coef.	Odd Ratio
<i>Age d'entreprise (réf : 10+ ans)</i>		
...0-5 ans	0.2055 (0.1158)	1.228 (0.1422)
...6-9 ans	0.1105 (0.0943)	1.117 (0.1053)
<i>Secteur (réf : Autres services)</i>		
...Manufacture	-0.5272* (0.2083)	0.5903* (0.1230)
...Construction	-0.3367 (0.2038)	0.7141 (0.1456)
...Commerce, trans., héberg.	-0.0944 (0.1902)	0.9099 (0.1731)
...Tertiaire sup.	0.2759 (0.2044)	1.318 (0.2694)
Poids de l'enquête	-0.0361*** (0.0108)	0.9646*** (0.0104)
Effets-fixes département	Oui	Oui
Effets-fixes années	Oui	Oui
N	15,294	15,294
Pseudo R2	0.33707	0.33707
% correctement classées	0.84386	0.84386

*Notes : les entreprises sont issues des enquêtes CIS 2014, 2016 et 2018. Les variables sont issues du répertoire SIRENE et du recensement de population 2013. Les calculs sont ceux des auteurs. Toutes les entreprises font partie des ZIPU. La variable dépendante est l'indicatrice d'accès à la fibre. Les erreurs-standards sont clusterisées au niveau département et reportées entre parenthèses. Contrairement aux statistiques précédentes, les entreprises ne sont pas pondérées par leur poids d'enquête, d'où le nombre plus faible d'observations. Les trois cohortes sont regroupées. Les deux colonnes reportent la même estimation, la colonne Odd Ratio, égale au coefficient exponentialisé, ayant été ajoutée pour faciliter l'interprétation des coefficients. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.*

Ce modèle nous permet d'obtenir le score de propension, c'est-à-dire la probabilité d'accès à la fibre basée sur les caractéristiques observables et telle qu'estimée par le modèle Logit.

A partir de ce score de propension, chaque unité traitée est alors appariée à l'unité de contrôle qui a le score de propension le plus proche. Il y a de nombreux détails liés à l'implémentation de l'appariement en tant que tel, en particulier nous pouvons décider : est-ce qu'une unité de contrôle peut être appariée à plusieurs unités traitées ? est-ce que passé une certaine distance dans les valeurs de propension, les appariements sont considérés comme invalides ? Dans notre cas, il se trouve que les résultats sont peu modifiés par la variation de ces différents paramètres. Dans la pratique nous n'utilisons pas de remplacement (i.e. une unité de contrôle ne peut être appariée qu'une fois) avec considérons qu'un appariement est invalide quand la distance au score de propension dépasse 5%.

Figure 6.18 : Distribution des scores de propension avant et après appariement



Le résultat de l'appariement est représenté dans la Figure 6.18 qui reporte la distribution des scores de propension pour les unités traitées et de contrôle avant et après l'appariement. On

peut constater que les deux distributions, bien que très différentes au début, sont quasiment équivalentes après appariement.

En revenant sur les caractéristiques utilisées pour appairier les entreprises, le Tableau 6.10 reporte les moyennes et différences de moyennes entre les groupes de contrôle et de traités après appariement. On constate en effet que dorénavant les deux groupes sont statistiquement équivalents (i.e. affichent des moyennes non statistiquement différentes) sur quasiment l'ensemble de ces mesures. Les quelques différences subsistantes ont une magnitude très faible (par exemple que les unités traitées sont en moyenne plus proches des villes moyennes de 2 points de pourcentage).

3.4. Résultats principaux

A partir de l'échantillon apparié décrit précédemment nous allons effectuer les estimations pour obtenir l'effet causal de l'accès à la fibre sur l'activité d'innovation. Ces estimations sont de la forme :

$$E(y_i|T_i) = f(\alpha + \gamma T_i)$$

avec f une fonction faisant le lien entre la forme fonctionnelle et l'espérance conditionnelle de y , la variable dépendante. Par exemple, dans le cas d'une estimation linéaire, f est l'identité ($f(x) = x$), dans le cas d'une estimation par modèle Logit, on a $f(x) = \text{Logit}(x)$, et dans le cas d'une estimation par modèle de Poisson $f(x) = \exp(x)$. L'interprétation du coefficient d'intérêt, γ , diffère en fonction de la forme fonctionnelle utilisée. Dans le cas d'une estimation linéaire, l'effet causal est capturé par une variation en niveau : $E(Y(1) - Y(0)) = \hat{\gamma}$, où $E(Y(1) - Y(0))$ représente l'espérance de la différence des variables dépendantes avec et sans traitement. Dans le cas d'une estimation par modèle de Poisson, l'effet causal est capturé par une variation relative : $E(Y(1)/Y(0)) = \exp(\hat{\gamma})$. Dans le cas d'un modèle Logit, l'interprétation est bien plus complexe car $\exp(\hat{\gamma})$ correspond à l'effet sur le rapport de côte, ou Odd Ratio, qui est en soi difficilement interprétable.

La liste des variables dépendantes de cette analyse est présentée dans le Tableau 6.6. La totalité de celles-ci sont des variables binaires à l'exception des variables relatives à la part de profit issue de nouveaux produits.

Dans une première étape, nous utiliserons les modèles linéaires et Poisson pour toutes les variables et le modèle Logit pour les variables binaires. Dans la suite de l'analyse, comme les

résultats sont sensiblement identiques entre méthodes, nous emploierons uniquement un modèle linéaire qui permet une grande facilité d'interprétation des résultats. Pour effectuer les estimations, nous utiliserons le logiciel R fixest (Bergé, 2018).

Résultats pour la production d'innovation

Les résultats pour la production d'innovations sont reportés dans les tableaux 6.12 et 6.13. Le Tableau 6.12 reporte les estimations pour l'ensemble de la période (2014-2018) alors que le Tableau 6.13 distingue les cohortes 2014-2016 d'une part, et 2018 d'autre part, afin de limiter les problèmes de comparabilité des mesures d'innovation de procédé détaillées précédemment. Pour chacun de ces échantillons, le travail d'appariement est répliqué (les résultats de l'appariement ne sont pas présentés par souci d'espace mais sont disponibles sur demande). Comme un grand nombre d'estimations a été réalisé¹⁰⁴, nous présentons les résultats de manière synthétique en ne reportant que le coefficient d'intérêt dans les tableaux 6.12 et 6.13.

De façon flagrante, il est à noter que l'effet de l'accès à fibre est quasi nul pour l'ensemble de ces mesures. Que ce soit en niveau (régression linéaire), ou en relatif (Poisson, Logit), les coefficients obtenus reflétant l'effet causal du traitement ne sont jamais statistiquement différents de 0 et, de plus, la magnitude de ces coefficients est de l'ordre du point de pourcentage, soit très faible.

Ces résultats valent également lorsque les cohortes sont séparées selon la vague de l'enquête.

Ainsi, pour les innovations de produit comme pour les innovations de procédé, il ne semble pas y avoir d'effet causal positif de l'accès à la fibre.

¹⁰⁴ 9 variables dépendantes x 3 estimateurs pour le Tableau 6.12, et 6 variables dépendantes x 2 estimateurs x 2 échantillons pour le Tableau 6.13.

Tableau 6.12 : Estimations principales pour les variables de production d'innovation

	\bar{y}	OLS	logit	Poisson
Innovation produit				
Tous type	0.322	-0.0014 (0.0155)	-0.0063 (0.0677)	-0.0063 (0.0681)
Produit	0.271	-0.0012 (0.0144)	-0.0065 (0.0806)	-0.0066 (0.0811)
Service	0.154	0.0084 (0.0125)	0.0709 (0.1092)	0.0685 (0.1020)
Innovation de procédé				
Tous types	0.484	0.0194 (0.0160)	0.0507 (0.0426)	0.0494 (0.0406)
Fabrication	0.241	0.0050 (0.0125)	0.0314 (0.0794)	0.0309 (0.0770)
Logistique	0.124	-0.0019 (0.0111)	-0.0204 (0.1191)	-0.0206 (0.1216)
Marketing	0.239	0.0160 (0.0136)	0.0898 (0.0802)	0.0860 (0.0736)
Organisationnel	0.333	0.0209 (0.0148)	0.0806 (0.0589)	0.0775 (0.0545)
Support	0.179	0.0062 (0.0107)	0.0417 (0.0735)	0.0409 (0.0706)
N (pondéré)		40,032	40,032	40,032

Notes : les entreprises sont issues des enquêtes CIS 2014, 2016 et 2018. Les calculs sont ceux des auteurs. L'échantillon utilisé consiste en un appariement des unités traitées et de contrôle. Toutes les entreprises font partie des ZIPU. Chaque ligne correspond à une estimation où le nom de la variable dépendante est spécifié dans la colonne de gauche. La deuxième colonne contient la moyenne de la variable dépendante. Pour chaque estimation, le coefficient associé à l'indicatrice de traitement est reporté (γ). Les erreurs-standards sont clusterisées au niveau département et reportées entre parenthèses. Les entreprises sont pondérées par les poids de l'enquête et de l'appariement. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Tableau 6.13 : Estimation principale sur les productions d'innovations de procédé, en séparant par cohortes

	2014-2016			2018		
	\bar{y}	OLS	Poisson	\bar{y}	OLS	Poisson
Innovation de procédé						
Tous types	0.55	0.0226 (0.0224)	0.0483 (0.0477)	0.379	0.0125 (0.0228)	0.0408 (0.0747)
Fabrication	0.25	0.0076 (0.0167)	0.0465 (0.1019)	0.225	0.0020 (0.0187)	0.0122 (0.1169)
Logistique	0.131	-0.0041 (0.0142)	-0.0438 (0.1512)	0.113	0.0006 (0.0163)	0.0062 (0.1823)
Marketing	0.302	0.0195 (0.0166)	0.0774 (0.0664)	0.137	0.0093 (0.0156)	0.0830 (0.1398)
Organisationnel	0.407	0.0164 (0.0228)	0.0473 (0.0658)	0.217	0.0227 (0.0181)	0.1252 (0.1010)
Support	0.153	0.0028 (0.0147)	0.0230 (0.1199)	0.22	0.0113 (0.0199)	0.0612 (0.1084)
N (pondéré)		21,329	21,329		18,703	18,703

Notes : les entreprises sont issues des enquêtes CIS 2014, 2016 et 2018. Les calculs sont ceux des auteurs. Deux échantillons sont utilisés : un pour la période 2014-2016 et un pour la période 2018. Dans chaque cas, l'appariement entre les traités et contrôles est répété. Toutes les entreprises font partie des ZIPU. Chaque ligne correspond à une estimation où le nom de la variable dépendante est spécifié dans la colonne de gauche. La deuxième colonne contient la moyenne de la variable dépendante. Pour chaque estimation, le coefficient associé à l'indicatrice de traitement est reporté (γ). Les erreurs-standards sont clusterisées au niveau département et reportées entre parenthèses. Les entreprises sont pondérées par les poids de l'enquête et de l'appariement.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Résultats pour la qualité d'innovation

Passons maintenant à la qualité d'innovation, mesurée par la portée géographique de l'innovation (disponible pour 2014-2016) et la part des nouveaux produits dans le chiffre d'affaire (disponible pour 2014-2018).

Le Tableau 6.14 reporte l'ensemble des estimations. De la même façon que pour les estimations concernant la production d'innovation, l'ensemble des coefficients mesurant l'impact de l'accès à la fibre sont non statistiquement différents de 0.

En termes de magnitude, dans la régression linéaire le coefficient associé à l'effet sur l'introduction d'un nouveau produit pour le monde est de 0,1% pour une moyenne de 9,5%. Ainsi, bien que le coefficient soit positif, même s'il était statistiquement significatif, sa magnitude serait très faible.

Au vu de ces résultats, l'accès à la fibre ne semble pas avoir d'influence sur la qualité de l'innovation produite, ni de façon positive, ni de façon négative.

Tableau 6.14 : Estimations principales pour la qualité d'innovation.

	\bar{y}	OLS	logit	Poisson
Introduction d'un produit nouveau pour...				
... la France	0.185	0.0102 (0.0150)	0.1102 (0.1588)	0.1045 (0.1431)
	N (pondéré)	20,467	20,467	20,467
... l'Europe	0.13	-0.0030 (0.0101)	0.0604 (0.1644)	0.0586 (0.1550)
	N (pondéré)	19,918	19,918	19,918
... le monde	0.0946	0.0014 (0.0088)	0.0580 (0.1854)	0.0564 (0.1753)
	N (pondéré)	19,557	19,557	19,557
Fraction du CA dû à un produit nouveau pour				
... le marché	3.04	0.4205 (0.3468)		0.1790 (0.1472)
	N (pondéré)	40,032		40,032

Notes : les entreprises sont issues des enquêtes CIS 2014, 2016 et 2018. Les calculs sont ceux des auteurs. L'échantillon sur les introductions de produits porte sur 2014-2016, alors l'échantillon de la dernière estimation porte sur toutes les périodes. L'appariement entre traités et contrôles est répété pour chaque échantillon. Les différences d'échantillon sont dues à des données manquantes. Toutes les entreprises font partie des ZIPU. Chaque ligne correspond à une estimation où le nom de la variable dépendante est spécifié dans la colonne de gauche. La deuxième colonne contient la moyenne de la variable dépendante. Pour chaque estimation, le coefficient associé à l'indicatrice de traitement est reporté (γ). Les erreurs-standards sont clusterisées au niveau département et reportées entre parenthèses. Les entreprises sont pondérées par les poids de l'enquête et de l'appariement.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Résultats pour la collaboration

Comme détaillé dans la revue de littérature, le développement de l'accès à la fibre est supposé avoir un impact positif sur la collaboration, via la facilitation des échanges à distance, des échanges de fichiers, du travail collaboratif en ligne, etc.

Le Tableau Tableau 6.15 illustre toutefois, comme dans les résultats précédents, l'absence d'effet significatif de l'accès à la fibre. Quelle que soit les mesures utilisées (la collaboration en général, la collaboration en distinguant le type de partenaire ou la portée géographique) aucun coefficient n'est statistiquement différent de 0.

De plus la magnitude des coefficients est faible. Ainsi, sur la mesure de collaboration de tout type, l'effet linéaire estimé est de 0,4% pour une moyenne de 17,7%.

Ainsi, il semble que l'accès à la fibre n'ait pas d'effet non plus sur la collaboration dans l'activité d'innovation.

Tableau 6.15 : Estimations principales pour les collaborations relatives à l'innovation

	\bar{y}	OLS	logit	Poisson
Collaboration				
Tous types	0.177	0.0044 (0.0182)	0.0412 (0.1754)	0.0403 (0.1684)
Collaboration avec...				
...une autre entreprise du groupe	0.0947	-0.0024 (0.0114)	-0.0517 (0.2406)	-0.0531 (0.2537)
...un partenaire hors groupe	0.17	0.0019 (0.0164)	0.0187 (0.1647)	0.0186 (0.1617)
... des clients	0.0941	0.0054 (0.0125)	0.1056 (0.2580)	0.1004 (0.2334)
...une entreprise privée	0.16	-0.0072 (0.0165)	-0.0754 (0.1660)	-0.0784 (0.1795)
... une entreprise du secteur public	0.106	0.0039 (0.0099)	0.0712 (0.1906)	0.0688 (0.1780)

	\bar{y}	OLS	logit	Poisson
Collaboration avec un partenaire situé...				
... en France	0.165	-0.0017 (0.0172)	-0.0170 (0.1731)	-0.0171 (0.1761)
... en Europe	0.0749	-0.0078 (0.0094)	-0.2110 (0.2165)	-0.2370 (0.2744)
... dans le monde	0.0432	-7.34e-5 (0.0060)	-0.0040 (0.3274)	-0.0040 (0.3288)
Collaboration avec une autre entreprise du groupe située...				
... en France	0.0725	0.0003 (0.0111)	0.0084 (0.3118)	0.0084 (0.3092)
... en Europe	0.0372	-0.0058 (0.0051)	-0.3523 (0.2237)	-0.4343 (0.3453)
... dans le monde	0.0222	0.0001 (0.0023)	0.0197 (0.4393)	0.0195 (0.4308)
Collaboration avec un partenaire hors groupe situé...				
... en France	0.158	-0.0066 (0.0156)	-0.0701 (0.1603)	-0.0726 (0.1723)
... en Europe	0.0677	-0.0078 (0.0089)	-0.2285 (0.2160)	-0.2594 (0.2800)
... dans le monde	0.0366	0.0013 (0.0059)	0.0889 (0.4164)	0.0852 (0.3824)
N (pondéré)		16,125	16,125	16,125

Notes : les entreprises sont issues des enquêtes CIS 2014, 2016 et 2018. Les calculs sont ceux des auteurs. L'échantillon utilisé consiste en un appariement des unités traitées et contrôles. Toutes les entreprises font partie des ZIPU. Chaque ligne correspond à une estimation où le nom de la variable dépendante est spécifié dans la colonne de gauche. La deuxième colonne contient la moyenne de la variable dépendante. Pour chaque estimation, le coefficient associé à l'indicatrice de traitement est reporté (γ). Les erreurs-standards sont clusterisées au niveau département et reportées entre parenthèses. Les entreprises sont pondérées par les poids de l'enquête et de l'appariement. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Extension des résultats principaux

Nous procédons à une série de tests supplémentaires pour nous assurer de la robustesse des résultats obtenus.

Dans un premier temps nous ajoutons les variables utilisées pour l'appariement (variables listées au Tableau 6.11) comme variables de contrôle. Si, normalement, les différences entre les groupes de contrôle et de traités sont gommées par l'appariement, il peut néanmoins rester une hétérogénéité résiduelle vis-à-vis de ces variables affectant la performance d'innovation. Pour s'assurer que ce n'est pas le cas, nous intégrons les variables d'appariement comme contrôles.

Ensuite, nous restreignons le groupe de contrôle aux entreprises situées dans des communes dotées d'une connexion d'au moins 30Mbps. Cela permet d'exclure du groupe de comparaison les communes avec un niveau de connectivité internet très faible et dont les contextes économiques sont souvent particuliers (ex : très petites communes en perte de vitesse). L'effet de la fibre est alors obtenu en comparant les entreprises traitées aux entreprises situées dans des communes ayant déjà un bon accès internet. Ainsi, si l'accès à la fibre a un effet, la magnitude de cet effet devrait être moindre dans ce cas-ci.

Enfin, nous restreignons le groupe traité aux communes ayant un taux de couverture strictement supérieur à 1% afin d'éviter d'avoir dans le groupe traité des entreprises qui *in fine* sont situées dans des communes disposant d'un très faible accès à la fibre, et donc, qui n'en bénéficient vraisemblablement pas. Cette estimation revient en quelque sorte à retirer de possibles faux-positifs (c'est-à-dire des entreprises considérées à tort comme ayant accès à la fibre). Par contre, il est possible que l'on crée des faux-négatifs (risque d'exclure des entreprises ayant accès à la fibre dans des communes où le taux de couverture est inférieur à 1%).

L'ensemble de ces nouveaux résultats sont reportés dans le Tableau 6.16. Dans un souci d'économie d'espace, nous ne reportons les résultats que pour les variables principales.

Globalement, ces nouvelles estimations ne changent que très peu les résultats des estimations initiales. Tous les résultats sont quasiment nuls en magnitude et non statistiquement significatifs¹⁰⁵.

¹⁰⁵ Le seul résultat statistiquement significatif est un effet négatif de l'accès à la fibre sur l'introduction d'innovations nouvelles au niveau européen (lorsque le groupe de contrôle est constitué uniquement d'entreprises situées dans des communes ayant accès au THD ou quand les entreprises de communes ayant peu accès à la fibre sont retirées). Néanmoins les résultats ne sont pas concordants en fonction des méthodes employées. Alors qu'un effet négatif de -3% en niveau est révélé par le modèle linéaire, le coefficient de l'estimation par un modèle de Poisson n'est pas statistiquement significatif, ce résultat est donc peu robuste.

Tableau 6.16 : Résultats principaux en ajoutant des variables de contrôle et en modifiant l'échantillon

	+ Contrôles		G. Contrôle > 30M		G. Traitement > 1%	
	OLS	Poisson	OLS	Poisson	OLS	Poisson
Production						
<i>Innovation de produit</i>						
Tous types	0.0003 (0.0139)	0.0063 (0.0593)	-0.0084 (0.0163)	-0.0363 (0.0703)	-0.0033 (0.0157)	-0.0147 (0.0693)
<i>Innovation de procédé</i>						
Tous types	0.0208 (0.0150)	0.0512 (0.0375)	0.0218 (0.0157)	0.0558 (0.0399)	0.0190 (0.0161)	0.0484 (0.0408)
N (pondéré)	40,032	40,032	36,522	36,522	38,283	38,283
Qualité						
<i>Introduction d'un produit nouveau pour...</i>						
... la France	0.0140 (0.0121)	0.1527 (0.1265)	-0.0092 (0.0150)	0.0591 (0.1407)	-0.0219 (0.0154)	0.1262 (0.1445)
N (pondéré)	20,467	20,467	18,605	18,605	19,638	19,638
... l'Europe	-0.0004 (0.0084)	0.1432 (0.1488)	-0.033** (0.0114)	-0.0072 (0.1583)	-0.031** (0.0116)	0.0478 (0.1665)
N (pondéré)	19,918	19,918	18,100	18,100	19,089	19,089
... le monde	0.0025 (0.0073)	0.1200 (0.1670)	-0.0016 (0.0090)	0.0306 (0.1844)	-0.0127 (0.0089)	0.0009 (0.1853)
N (pondéré)	19,557	19,557	17,770	17,770	18,742	18,742
<i>Fraction du CA dû à un produit nouveau pour</i>						
... le marché	0.4549 (0.3421)	0.2078 (0.1450)	0.3329 (0.3871)	0.1390 (0.1628)	0.4348 (0.3669)	0.1846 (0.1550)
	40,032	40,032	36,522	36,522	38,283	38,283
Collaboration						
Tous types	-0.0017 (0.0093)	-0.0153 (0.0801)	-0.0045 (0.0104)	-0.0413 (0.0965)	-0.0050 (0.0100)	-0.0476 (0.0943)

	+ Contrôles		G. Contrôle > 30M		G. Traitement > 1%	
	OLS	Poisson	OLS	Poisson	OLS	Poisson
<i>Collaboration avec un partenaire situé...</i>						
... en France	-0.0049 (0.0096)	-0.0483 (0.0880)	-0.0063 (0.0104)	-0.0624 (0.1028)	-0.0064 (0.0101)	-0.0646 (0.1009)
... en Europe	-7.21e-5 (0.0050)	0.0690 (0.1349)	-0.0029 (0.0057)	-0.0763 (0.1470)	-0.0013 (0.0049)	-0.0352 (0.1360)
... dans le monde	0.0022 (0.0032)	0.1795 (0.1994)	0.0020 (0.0035)	0.1181 (0.2096)	0.0013 (0.0034)	0.0818 (0.2134)
N (pondéré)	40,032	40,032	36,522	36,522	38,283	38,283

Notes : les entreprises sont issues des enquêtes CIS 2014, 2016 et 2018. Les calculs sont ceux des auteurs. L'échantillon utilisé consiste en un appariement des unités traitées et contrôles. Toutes les entreprises font partie des ZIPU. Chaque ligne correspond à une estimation où le nom de la variable dépendante est spécifié dans la colonne de gauche. Pour chaque estimation, le coefficient associé à l'indicatrice de traitement est reporté (γ). Les erreurs-standards sont clusterisées au niveau département et reportées entre parenthèses. Les entreprises sont pondérées par les poids de l'enquête et de l'appariement. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

3.5. Hétérogénéité des résultats en fonction des caractéristiques des entreprises

Après avoir examiné l'effet moyen de l'accès à la fibre, nous analysons maintenant son impact en fonction de sous-ensembles d'entreprises. En effet, bien que nous n'ayons pu mettre en évidence d'effet général du traitement sur l'innovation, peut-être les bénéfices engendrés par l'accès à la fibre sont-ils spécifiques à certains types d'entreprises seulement.

Le Tableau 6.17 reporte les résultats des estimations selon a) la structure organisationnelle de l'entreprise : mono- ou multi-établissements, et b) la taille de l'entreprise : quatre groupes de taille allant de moins de 10 employés à plus de 250.

Comme pour les résultats principaux, nous ne trouvons aucun effet significatif de l'accès à la fibre sur les variables d'innovation. Ceci est vrai à l'exception des innovations de procédé.

En effet, nous trouvons que les entreprises multi-établissements produisent plus d'innovation de procédé suite au déploiement de la fibre, avec un effet statistiquement significatif de +6%. C'est-à-dire que les entreprises multi-établissements qui ont accès à la fibre ont en moyenne une probabilité de développer des innovations de procédé supérieure de 6% à celles qui n'y ont pas accès. Ces résultats sont robustes à l'introduction de variables de contrôle,

ou à la modification du groupe de contrôle ou du groupe traité, comme dans le Tableau 6.16.¹⁰⁶ Lorsque nous utilisons un modèle de Poisson, nous trouvons également une augmentation statistiquement significative de 13% d'innovation de procédé en comparaison aux entreprises sans accès à la fibre.

Tableau 6.17 : Effet de l'accès à la fibre sur l'innovation. Hétérogénéité en fonction du type et de la taille de entreprises

	Mono- ou multi-établissements		Taille d'entreprise (# salariés)			
	Mono	Multi	0-9	10-49	50-249	≥ 250
Production						
<i>Innovation de produit</i>						
Tous types	0.0082 (0.0179)	0.0249 (0.0158)	0.0553 (0.0608)	0.0155 (0.0149)	-0.0049 (0.0360)	-0.0193 (0.0342)
\bar{y}	0.24	0.362	0.128	0.199	0.347	0.577
<i>Innovation de procédé</i>						
Tous types	-0.0225 (0.0281)	0.0598** (0.0217)	0.0518 (0.1110)	0.0386 (0.0218)	0.0061 (0.0327)	0.0016 (0.0417)
\bar{y}	0.394	0.527	0.255	0.365	0.562	0.702
N (pondéré)	13,604	22,816	1,151	27,051	7,580	1,414
Qualité						
<i>Introduction d'un produit nouveau pour...</i>						
... la France	0.0173 (0.0316)	-0.0218 (0.0155)	0.1076 (0.1294)	0.0236 (0.0146)	-0.0517 (0.0381)	-0.0451 (0.0474)
\bar{y}	0.133	0.221	0.0635	0.0891	0.179	0.404
N (pondéré)	5,930	13,407	672	13,847	4,110	916
... le monde	0.0397* (0.0154)	0.0028 (0.0106)	0.0202 (0.1022)	0.0079 (0.0068)	-0.0097 (0.0230)	0.0282 (0.0559)
\bar{y}	0.0582	0.115	0.0161	0.0336	0.0789	0.243
N (pondéré)	5,643	12,736	665	13,348	3,861	800

¹⁰⁶ Ces résultats ne sont pas reportés mais sont disponibles sur demande.

	Mono- ou multi- établissements		Taille d'entreprise (# salariés)			
	Mono	Multi	0-9	10-49	50-249	≥ 250
Collaboration						
Tous types	0.0170 (0.0146)	-0.0060 (0.0108)	-0.0007 (0.0310)	0.0037 (0.0093)	0.0250 (0.0316)	-0.0479 (0.0294)
\bar{y}	0.112	0.217	0.0213	0.077	0.231	0.408
<i>Collaboration avec un partenaire situé...</i>						
... en France	0.0116 (0.0141)	-0.0112 (0.0109)	-0.0007 (0.0310)	0.0008 (0.0089)	0.0169 (0.0300)	-0.0436 (0.0311)
\bar{y}	0.104	0.206	0.0213	0.0733	0.216	0.389
... hors France	0.0084 (0.0082)	-0.0060 (0.0070)	-- --	0.0030 (0.0048)	-0.0134 (0.0193)	-0.0511 (0.0319)
\bar{y}	0.0491	0.113	0	0.0235	0.0934	0.265
N (pondéré)	5,643	12,736	665	13,348	3,861	800

*Notes : les entreprises sont issues des enquêtes CIS 2014, 2016 et 2018. Les calculs sont ceux des auteurs. L'échantillon utilisé dans chaque colonne consiste en un appariement des unités traitées et contrôles pour le sous-ensemble d'entreprises considérées. Toutes les entreprises font partie des ZIPU. Chaque ligne correspond à une estimation où le nom de la variable dépendante est spécifié dans la colonne de gauche. Chaque estimation est une régression linéaire, le coefficient associé à l'indicatrice de traitement est reporté (γ). Les erreurs-standards sont clusterisées au niveau département et reportées entre parenthèses. Les entreprises sont pondérées par les poids de l'enquête et de l'appariement. Chaque estimation contient les variables utilisées pour l'appariement comme contrôle (à l'exception de la variable sur laquelle l'échantillon est restreint). * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.*

Regardons maintenant plus en détail quels types d'innovations de procédé sont concernées. Tout d'abord rappelons-nous la différence de questionnaires entre les enquêtes 2014-2016 et celle de 2018 concernant les questions d'innovations procédé. Pour éviter tout problème de comparabilité, nous divisons l'échantillon en deux groupes : un pour les années 2014-2016 et un second groupe pour l'année 2018. A chaque fois nous ré-effectuons un appariement des groupes de traités et de contrôles et effectuons une estimation par type d'innovation de procédé. Les résultats pour 2014-2016, non reportés, montrent que la fibre n'a aucun effet sur les innovations de procédé : aucun coefficient de l'effet du traitement n'est significatif.

Par contraste, nous trouvons des effets significatifs et importants pour l'année 2018, dont

les résultats sont reportés dans le Tableau 6.18. Ce dernier reporte à la fois les estimations en OLS et en Poisson, et ce pour quatre spécifications différentes (avec ou sans variables de contrôle, et en faisant varier le groupe des contrôles ou des traités). La première chose à constater est que les effets du traitement qui sont significatifs le sont pour toutes les spécifications, et à la fois en niveau (OLS) et en relatif (Poisson).

Pour les entreprises multi-établissements, l'accès à la fibre augmente les innovations : 1) de marketing, 2) organisationnelles, et 3) de support.¹⁰⁷ Le niveau de l'augmentation est similaire entre ces trois mesures : environ +20% en relatif.

Les innovations marketing correspondent à une large gamme d'éléments : elles peuvent porter sur la tarification, la promotion, ou encore la relation client, etc. Ici l'accès à la fibre a pu améliorer, ou pousser à améliorer certains de ces aspects en facilitant la présence en ligne.

Les entreprises multi-établissements sont les entreprises avec potentiellement le plus de besoins en termes organisationnels étant donné la nécessaire coordination entre les différents établissements. On peut supposer que l'accès à la fibre a facilité l'usage des nouveaux outils de communication et de collaboration à distance, améliorant ainsi l'organisation entre établissements.

Enfin, les innovations de support correspondent en général à une amélioration des systèmes informatiques via le déploiement ou l'achat de logiciels. Par ailleurs ceux-ci peuvent être en lien avec les innovations de marketing et d'organisation s'ils facilitent celles-ci. Ici, l'accès à la fibre semble avoir été complémentaire avec l'achat et le déploiement de nouveaux logiciels.

Deux questions intéressantes se posent : a) Pourquoi est-ce que l'effet positif sur les innovations de procédé n'apparaît que pour les entreprises multi-établissements ? b) Pourquoi l'effet positif de l'accès à la fibre n'est visible que pour l'enquête 2018 ?

A la première question, on peut raisonnablement supposer que les entreprises multi-établissements (EME) sont celles qui ont à la fois les plus gros besoins en termes d'organisation, et celles qui ont les capacités les plus importantes de mise en place des innovations de procédé. En effet, par le besoin de coordination entre de multiples établissements, les EME ont une utilité marginale à l'implémentation d'innovations organisationnelles supérieure ; de plus, ce sont aussi celles qui ont les plus grandes capacités d'infrastructure et de mobilisation des moyens financiers. Ces simples mécanismes impliqueraient que ce sont celles qui à la fois bénéficient

¹⁰⁷ Les activités de support incluent l'amélioration des systèmes informatiques (de traitement de l'information, de communication, de comptabilité, etc.).

le plus d'innovation de procédé et les mettraient en place le plus rapidement suite à l'accès à la fibre. L'accès à la fibre serait ici très complémentaire aux innovation de procédé pour les EME.

Sur la seconde question, le fait que l'effet ne soit visible que pour 2018 peut vouloir dire qu'il y a besoin d'un délai raisonnable entre l'implantation de la fibre et l'apparition d'effets positifs sur l'innovation.

Ces résultats font échos aux conclusions du chapitre 4, selon lequel les effets de la fibre sur l'usage d'outils numériques ne sont visibles qu'au bout de trois ans en moyenne et sont essentiellement limités aux entreprises multi-établissements.

Tableau 6.18 : Effet de la fibre sur l'innovation de procédé des EME pour l'enquête 2018

	Principal		+ Contrôles		G. Contrôle > 30M		G. Traitement > 1%	
	OLS	Poisson	OLS	Poisson	OLS	Poisson	OLS	Poisson
Innovation de procédé								
Tous types	0.0570*	0.1322*	0.0598***	0.1392**	0.0564*	0.1308*	0.0549*	0.1277*
$\bar{y} = 0.53$	(0.0228)	(0.0526)	(0.0217)	(0.0502)	(0.0230)	(0.0533)	(0.0232)	(0.0534)
Fabrication	0.0128	0.0734	0.0103	0.0548	0.0120	0.0686	0.0153	0.0870
$\bar{y} = 0.26$	(0.0158)	(0.0900)	(0.0150)	(0.0858)	(0.0159)	(0.0909)	(0.0160)	(0.0906)
Logistique	0.0209	0.2192	0.0204	0.1983	0.0261*	0.2821*	0.0246	0.2534*
$\bar{y} = 0.13$	(0.0121)	(0.1270)	(0.0118)	(0.1269)	(0.0115)	(0.1245)	(0.0127)	(0.1289)
Marketing	0.0442*	0.2031*	0.0444**	0.2048**	0.0394*	0.1793*	0.0452*	0.2074*
$\bar{y} = 0.27$	(0.0171)	(0.0795)	(0.0166)	(0.0774)	(0.0169)	(0.0779)	(0.0177)	(0.0815)
Organisationnel	0.0622**	0.2084**	0.0637**	0.2081**	0.0628**	0.2108**	0.0604**	0.2028**
$\bar{y} = 0.37$	(0.0209)	(0.0686)	(0.0200)	(0.0654)	(0.0213)	(0.0705)	(0.0214)	(0.0698)
Support	0.0338*	0.1974*	0.0339*	0.1904*	0.0356*	0.2089*	0.0344*	0.2005*
$\bar{y} = 0.20$	(0.0154)	(0.0893)	(0.0153)	(0.0904)	(0.0158)	(0.0929)	(0.0159)	(0.0915)
N (pondéré)	22,816	22,816	22,816	22,742	20,812	20,812	21,860	21,860

Notes : les entreprises sont issues de l'enquête CIS 2018. Les calculs sont ceux des auteurs. L'échantillon utilisé consiste en un appariement des unités traitées et contrôles pour 2018. Toutes les entreprises font partie des ZIPU. Chaque ligne correspond à une estimation où le nom de la variable dépendante est spécifié dans la colonne de gauche. Pour chaque estimation, le coefficient associé à l'indicateur de traitement est reporté (γ). Les erreurs-standards sont clusterisées au niveau département et reportées entre parenthèses. Les entreprises sont pondérées par les poids de l'enquête et de l'appariement. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Nous poursuivons l'étude d'hétérogénéité en groupant les entreprises par secteur et par âge.

En cohérence avec le chapitre 4, nous créons trois grands secteurs : i) manufacture et construction, ii) commerce, transport et hébergement, et iii) services. Ces trois catégories reflètent des types importants d'entreprises avec des profils d'innovation différents. Les entreprises avec le plus d'innovation traditionnelles appartiennent au secteur manufacturier, qui contient également les entreprises les plus importantes en termes de taille, et en termes de personnes affectées à la recherche et développement. La deuxième catégorie des commerces, transports et hébergement est celle dont on attend le moins d'innovations, bien qu'elle puisse être présente. Pour finir, le secteur des services peut être innovant mais est constitué d'entreprises de taille moindre.

Nous séparons également les entreprises en trois groupes d'âge : moins de 5 années d'existence, de 6 à 9 ans, et plus de 10 ans. Nous nous attendons à ce que les effets de l'accès à la fibre soient différenciés en fonction de l'âge de l'entreprise car l'âge peut affecter la capacité de l'entreprise à mettre en place des nouveaux procédés, et est corrélé avec les moyens financiers de l'entreprise.

Les résultats sont identiques pour l'ensemble des sous-échantillons : l'accès à la fibre ne semble affecter aucune de ces variables capturant l'activité d'innovation (Tableau 6.19).

Tableau 6.19 : Effets de l'accès à la fibre sur l'innovation. Hétérogénéité en fonction du secteur et de l'âge des entreprises

	Secteur			Age d'entreprise		
	Manuf.- Constr.	Comm., trans., hébergement.	Services	0-5	6-9	10+
Production						
<i>Innovation de produit</i>						
Tous types	0.0054 (0.0174)	-0.0344 (0.0226)	0.0620 (0.0486)	-0.0110 (0.0439)	-0.0111 (0.0434)	0.0076 (0.0155)
\bar{y}	0.436	0.168	0.304	0.236	0.288	0.336
<i>Innovation de procédé</i>						
Tous types	0.0023 (0.0224)	0.0340 (0.0284)	0.0213 (0.0506)	-0.0198 (0.0706)	0.0754 (0.0624)	0.0246 (0.0167)
\bar{y}	0.553	0.386	0.498	0.464	0.457	0.498
N (pondéré)	17,654	15,762	5,059	2,629	3,023	32,712
Qualité						
<i>Introduction d'un produit nouveau pour...</i>						
... la France	-0.0118 (0.0225)	0.0118 (0.0111)	0.0557 (0.0599)	0.0473 (0.0645)	-0.1063 (0.0858)	0.0132 (0.0152)
\bar{y}	0.284	0.0587	0.188	0.137	0.15	0.204
N (pondéré)	8,653	8,767	2,923	1,629	1,573	17,017
... le monde	-0.0008 (0.0158)	0.0098 (0.0059)	-0.0241 (0.0331)	0.0009 (0.0521)	-0.0696 (0.0393)	-0.0022 (0.0096)
\bar{y}	0.162	0.0191	0.0828	0.0759	0.0476	0.102
N (pondéré)	8,073	8,609	2,747	1,563	1,510	16,183
Collaboration						
Tous types	-0.0047 (0.0136)	0.0078 (0.0109)	0.0270 (0.0299)	0.0703 (0.0445)	-0.0208 (0.0260)	-0.0065 (0.0098)
\bar{y}	0.26	0.0706	0.158	0.148	0.116	0.203

	Secteur			Age d'entreprise		
	Manuf.- Constr.	Comm., trans., hébergement.	Services	0-5	6-9	10+
<i>Collaboration avec un partenaire situé...</i>						
... en France	-0.0053 (0.0127)	0.0025 (0.0106)	0.0166 (0.0290)	0.0772 (0.0426)	-0.0265 (0.0245)	-0.0091 (0.0091)
\bar{y}	0.246	0.065	0.154	0.136	0.106	0.192
... hors France	0.0012 (0.0098)	0.0067 (0.0045)	0.0203 (0.0156)	0.0029 (0.0244)	0.0118 (0.0145)	-0.0022 (0.0059)
\bar{y}	0.151	0.0239	0.0641	0.076	0.0728	0.103
N (pondéré)	8,073	8,609	2,747	1,563	1,510	16,183

*Notes : les entreprises sont issues des enquêtes CIS 2014, 2016 et 2018. Les calculs sont ceux des auteurs. L'échantillon utilisé dans chaque colonne consiste en un appariement des unités traitées et contrôles pour le sous-ensemble d'entreprises considérées. Toutes les entreprises font partie des ZIPU. Chaque ligne correspond à une estimation où le nom de la variable dépendante est spécifié dans la colonne de gauche. Chaque estimation est une régression linéaire, le coefficient associé à l'indicatrice de traitement est reporté (γ). Les erreurs-standards sont clusterisées au niveau département et reportées entre parenthèses. Les entreprises sont pondérées par les poids de l'enquête et de l'appariement. Chaque estimation contient les variables utilisées pour l'appariement comme contrôle (à l'exception de la variable sur laquelle l'échantillon est restreint). * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.*

4. Conclusion

Ce chapitre a considéré un ensemble de mesures basées sur les brevets et l'enquête CIS afin de capturer des aspects différents de l'innovation.

Analyses des données de brevets

Les résultats dépeignent un impact plutôt négatif à court terme sur la production de brevet, surtout pour l'accès à 100M. L'effet est également légèrement négatif sur la collaboration. Ces résultats sont toutefois à interpréter avec prudence puisque les données de brevets ne sont exhaustives que jusqu'en 2016, et donc, l'analyse ne capte que les effets de très court terme des premiers déploiements de fibre. Or, les chapitres sur les usages et la performance ont montré que certains effets économiques n'étaient identifiables que deux à trois années après le déploiement. L'analyse gagnerait donc à être répliquée dans plusieurs années lorsque davantage de données sur les brevets seront disponibles.

Analyses de l'enquête CIS

Le tableau dressé par l'ensemble de nos résultats est très clair : à l'exception des entreprises multi-établissements dont nous reparlerons ensuite, il semble que l'accès à la fibre n'a pas influencé la production d'innovation des entreprises, ni même leurs schémas collaboratifs, du moins lorsque l'on considère les mesures de l'enquête CIS. Celles-ci ont toutefois des caractéristiques particulières qui peuvent éventuellement expliquer l'absence d'effet.

Premièrement, des erreurs de mesure dans l'auto-évaluation des entreprises sont possibles. Comme nous avons pu le voir (Tableau 6.5), ce qui peut être considéré comme une innovation varie grandement entre entreprises. On peut donc considérer qu'il y a un bruit important entre la perception subjective de la personne répondant au questionnaire et une mesure objective de l'innovation, alors que seule cette dernière est intéressante pour comprendre la performance d'innovation. Une autre illustration de ce bruit est, par exemple, la reformulation de la question concernant les innovations marketing et organisationnelles qui a entraîné un changement important sur les valeurs reportées, alors que la mesure est censée être la même. Cette erreur de mesure systématique, qui n'est vraisemblablement pas corrélée à l'accès à la fibre, a pour conséquence de réduire la précision des coefficients estimés (Millimet et Parmeter, 2022).

Deuxièmement, les mesures de l'enquête CIS rapportent un effet extensif et non pas intensif. C'est-à-dire que pour chaque entreprise nous savons si elle a effectué *au moins une* innovation mais nous ne savons pas combien. Il se peut que la fibre ait eu un effet positif sur la marge intensive mais pas sur la marge extensive. Dans ce cas, les données dont nous disposons nous rendent incapables d'identifier l'effet de la fibre.

Enfin, et comme discuté dans les sections précédentes, l'innovation se fait sur un temps long. Ainsi, les effets à très court terme, pour l'année 2014 sont logiquement peu probables (le déploiement de la fibre étant encore très récent).

Malgré ces éléments qui réduisent la probabilité d'observer un effet positif de l'accès à la fibre, notre étude montre que les entreprises multi-établissements augmentent leur taux d'innovation de procédé suite à l'accès à la fibre. Nous mettons en évidence un effet positif pour les innovations organisationnelles, marketing et de support. En cohérence avec l'idée qu'il faut du temps pour que les innovations prennent, l'effet observé est apparent uniquement sur l'enquête 2018, mais de façon très importante.

Annexe au chapitre 6 : construction des variables

Tableau A.6.1 : Définition des variables construites à partir des enquêtes CIS

Variables	Définition
Production : innovation produit	
	Binaire : 0 ou 1
Nouveau produit	<p><i>Version 2014-2016.</i> Au cours des trois années 2014 à 2016, votre entreprise a-t-elle introduit : des biens nouveaux ou améliorés de façon significative (à l'exclusion de la simple revente en l'état de nouveaux biens achetés à d'autres entreprises et des modifications exclusivement esthétiques ou de simple conditionnement) ?</p> <p><i>Version 2018.</i> Au cours des trois années 2016 à 2018, votre entreprise a-t-elle introduit : un bien nouveau ou amélioré ?</p>
	Binaire : 0 ou 1
Nouveau service	<p><i>Version 2014-2016.</i> Au cours des trois années 2014 à 2016, votre entreprise a-t-elle introduit : des prestations de services nouvelles ou améliorées de façon significative ?</p> <p><i>Version 2018.</i> Au cours des trois années 2016 à 2018, votre entreprise a-t-elle introduit : un service nouveau ou amélioré* ?</p>
Toute sorte	Binaire : 0 ou 1. Vaut 1 si nouveau produit ou nouveau service.
Production : innovation de procédé	
	Binaire : 0 ou 1
Fabrication	<p><i>Version 2014-2016.</i> Au cours des trois années 2014 à 2016, votre entreprise a-t-elle introduit des nouveautés ou des améliorations significatives concernant : vos procédés de fabrication ou de production de biens ou de prestations de services ?</p> <p><i>Version 2018.</i> Au cours des trois années 2016 à 2018, votre entreprise a-t-elle introduit un procédé nouveau ou significativement amélioré relatif aux méthodes dans ces différents domaines ? Production de biens ou de services (y compris les méthodes de développement)</p>
	Binaire : 0 ou 1
Logistique	<p><i>Version 2014-2016.</i> Au cours des trois années 2014 à 2016, votre entreprise a-t-elle introduit des nouveautés ou des améliorations significatives concernant : vos méthodes de logistique, de fourniture ou de distribution de matières premières, biens ou prestations de services ?</p> <p><i>Version 2018.</i> Au cours des trois années 2016 à 2018, votre entreprise a-t-elle introduit un procédé nouveau ou significativement amélioré relatif aux méthodes dans ces différents domaines ? Production de biens ou de services (y compris les méthodes de développement)</p>

Variables	Définition
Marketing	<p>Binaire : 0 ou 1</p> <p><i>Version 2014-2016</i> Oui à une des quatre questions suivantes.</p> <p>Au cours des trois années 2014 à 2016, votre entreprise a-t-elle introduit les innovations de marketing suivantes :</p> <ul style="list-style-type: none"> • modifications significatives du design ou de l'emballage d'un bien ou service ? • utilisation de nouvelles techniques ou de nouveaux médias pour la promotion des biens ou des services ? • nouvelles méthodes (ou modifications significatives des méthodes) de vente ou de distribution ? • nouvelles stratégies de tarification de vos produits (biens ou services) ? <p><i>Version 2018.</i> Au cours des trois années 2016 à 2018, votre entreprise a-t-elle introduit un procédé nouveau ou significativement amélioré relatif aux méthodes dans ces différents domaines ? Marketing, promotion, emballage, tarification, placement de produit, service après-vente</p>
Organisationnel	<p>Binaire 0 ou 1.</p> <p><i>Version 2014-2016 : Oui à une des trois questions suivantes.</i></p> <p>Au cours des trois années 2014 à 2016, votre entreprise a-t-elle introduit les innovations d'organisation suivantes :</p> <ul style="list-style-type: none"> • de nouveaux modes de fonctionnement dans l'organisation des procédures ? • de nouvelles méthodes d'organisation du travail et de prise de décision ? • de nouvelles méthodes d'organisation des relations externes avec d'autres entreprises ou organismes ? <p><i>Version 2018 : Oui à une des deux questions suivantes.</i></p> <p>Au cours des trois années 2016 à 2018, votre entreprise a-t-elle introduit un procédé nouveau ou significativement amélioré relatif aux méthodes dans ces différents domaines ?</p> <p>Organisation des procédures, relations externes Organisation du travail, processus de décision, management des ressources humaines</p>
Support	<p>Binaire : 0 ou 1</p> <p><i>Version 2014-2016.</i> Au cours des trois années 2014 à 2016, votre entreprise a-t-elle introduit des nouveautés ou des améliorations significatives concernant : vos activités de soutien ou de support, comme les activités de maintenance, d'achat, de comptabilité ou informatiques ?</p> <p><i>Version 2018 : Oui à une des deux questions suivantes.</i></p> <p>Au cours des trois années 2016 à 2018, votre entreprise a-t-elle introduit un procédé nouveau ou significativement amélioré relatif aux méthodes dans ces différents domaines ?</p> <ul style="list-style-type: none"> • Traitement de l'information et communication • Comptabilité, opérations administratives, progiciel de Gestion Intégré (PGI – ERP)
Tous types	<p>Binaire : 0 ou 1. Vaut 1 si au moins un des sous-types d'innovation de procédé vaut 1.</p>
Qualité : Innovation produit	
Produit nouveau pour la France	<p>Binaire : 0 ou 1.</p> <p><i>Uniquement pour 2014-2016.</i> Au cours des trois années 2014 à 2016, l'une de vos innovations de produits était-elle une nouveauté sur : le marché français ?</p>

Variables	Définition
Produit nouveau pour l'Europe	<p>Binaire : 0 ou 1.</p> <p><i>Uniquement pour 2014-2016.</i> Au cours des trois années 2014 à 2016, l'une de vos innovations de produits était-elle une nouveauté sur : le marché européen ?</p>
Produit nouveau pour le monde	<p>Binaire : 0 ou 1.</p> <p><i>Uniquement pour 2014-2016.</i> Au cours des trois années 2014 à 2016, l'une de vos innovations de produits était-elle une nouveauté sur : le marché mondial ?</p>
Fraction du CA due à un produit nouveau pour le marché	<p>Variable entre 0 et 100.</p> <p><i>Version 2014-2016.</i> Estimez la part de votre chiffre d'affaires en 2016 relative à : des produits nouveaux ou améliorés de façon significative, introduits au cours des années 2014 à 2016, nouveaux pour l'un de vos marchés</p> <p><i>Version 2018.</i> Pouvez-vous estimer la part de votre chiffre d'affaires en 2018 relative aux biens et services nouveaux ou améliorés ? Pouvez-vous décomposer cette part selon que ces biens ou services nouveaux ou améliorés étaient : nouveaux comparés aux produits déjà offerts par vos concurrents ?</p>
Collaboration : types principaux	
Collaboration avec une autre entreprise du groupe	<p>Binaire : 0 ou 1.</p> <p><i>Version 2014-2016.</i> Au cours des trois années 2014 à 2016, votre entreprise a-t-elle coopéré avec d'autres entreprises ou organismes pour ses activités d'innovation ? Si oui, avec quel(s) partenaire(s) et dans quel(s) pays ?</p> <ul style="list-style-type: none"> • Autres entreprises de votre groupe / réseau d'enseigne <p><i>Version 2018.</i> Au cours des trois années 2016 à 2018, avec quel type de partenaires votre entreprise a-t-elle coopéré pour ses activités d'innovation ou de R&D ?</p> <ul style="list-style-type: none"> • Entreprises appartenant à votre groupe
Collaboration avec une entreprise du secteur public	<p>Binaire : 0 ou 1.</p> <p><i>Version 2014-2016.</i> Au cours des trois années 2014 à 2016, votre entreprise a-t-elle coopéré avec d'autres entreprises ou organismes pour ses activités d'innovation ? Si oui, avec quel(s) partenaire(s) et dans quel(s) pays ?</p> <ul style="list-style-type: none"> • Universités ou établissements d'enseignement supérieur • Organismes gouvernementaux ou publics de recherche <p><i>Version 2018.</i> Au cours des trois années 2016 à 2018, avec quel type de partenaires votre entreprise a-t-elle coopéré pour ses activités d'innovation ou de R&D ?</p> <ul style="list-style-type: none"> • Universités, grandes écoles, établissements d'enseignement supérieur • Organismes du service public (administration, école, hôpital, agence publique)
Collaboration avec des clients	<p>Binaire 0 ou 1.</p> <p><i>Version 2014-2016.</i> Au cours des trois années 2014 à 2016, votre entreprise a-t-elle coopéré avec d'autres entreprises ou organismes pour ses activités d'innovation ? Si oui, avec quel(s) partenaire(s) et dans quel(s) pays ?</p> <ul style="list-style-type: none"> • Clients ou consommateurs <p><i>Version 2018.</i> Au cours des trois années 2016 à 2018, avec quel type de partenaires votre entreprise a-t-elle coopéré pour ses activités d'innovation ou de R&D ?</p> <ul style="list-style-type: none"> • Clients ou consommateurs

Variables	Définition
	Binaire 0 ou 1. <i>Version 2014-2016.</i> Au cours des trois années 2014 à 2016, votre entreprise a-t-elle coopéré avec d'autres entreprises ou organismes pour ses activités d'innovation ? Si oui, avec quel(s) partenaire(s) et dans quel(s) pays ?
Collaboration avec une entreprise privée	<ul style="list-style-type: none"> • Fournisseurs d'équipements, matériels, composants, logiciels • Concurrents ou autres entreprises de votre secteur d'activité • Consultants, laboratoires commerciaux ou privés <i>Version 2018.</i> Au cours des trois années 2016 à 2018, avec quel type de partenaires votre entreprise a-t-elle coopéré pour ses activités d'innovation ou de R&D ? <ul style="list-style-type: none"> • Entreprises privées n'appartenant pas à votre groupe
Collaboration avec un partenaire hors groupe	Binaire : 0 ou 1. Vaut 1 si une des variables « collaboration avec ... » a) une entreprise du secteur public, b) des clients, ou c) une entreprise du secteur privé, vaut 1.
Collaboration avec un partenaire pour innover (<i>quel que soit le partenaire</i>)	Binaire : 0 ou 1. Vaut 1 si une des variables de sous-types de collaboration vaut 1.
Collaboration : géographie	
	Binaire : 0 ou 1. <i>Version 2014-2016.</i> Au moins un oui pour les variables de sous-types décrites précédemment pour les zones géographiques suivantes : <ul style="list-style-type: none"> • Même région que vous • France, en dehors de votre région <i>Version 2018.</i> Au moins un oui pour les variables de sous-types décrites précédemment pour les zones géographiques suivantes : En France
Collaboration avec un partenaire situé en France	<i>Version 2014-2016.</i> Au moins un oui pour les variables de sous-types décrites précédemment pour les zones géographiques suivantes : <ul style="list-style-type: none"> • autres pays d'Europe <i>Version 2018.</i> Au moins un oui pour les variables de sous-types décrites précédemment pour les zones géographiques suivantes : Dans les pays de l'Union européenne ou de l'AELE (EFTA)
Collaboration avec un partenaire situé en Europe	<i>Version 2014-2016.</i> Au moins un oui pour les variables de sous-types décrites précédemment pour les zones géographiques suivantes : <ul style="list-style-type: none"> • États-Unis • Chine ou Inde • Autres pays <i>Version 2018.</i> Au moins un oui pour les variables de sous-types décrites précédemment pour les zones géographiques suivantes : Dans les autres pays
Collaboration avec un partenaire situé hors de l'Europe	
Collaboration : type et géographie	
Collaboration avec une autre entreprise du groupe, située en France	Toutes ces variables sont binaires 0 ou 1.
Collaboration avec une autre entreprise du groupe, située en Europe	Elles valent 1 si l'entreprise répond oui concernant une collaboration pour le type donné et la géographie donnée. Le type et la géographie sont décrits précédemment.

Variables	Définition
	Collaboration avec une autre entreprise du groupe, située dans le monde
	Collaboration avec un partenaire hors groupe, situé en France
	Collaboration avec un partenaire hors groupe, situé en Europe
	Collaboration avec un partenaire hors groupe, situé dans le monde

Notes: Les définitions correspondent au verbatim des questions posées dans les enquêtes. Les questions des enquêtes 2014 et 2016 sont similaires. Ainsi, pour les version 2014-2016, ce sont les verbatims des questions de l'enquête 2016 qui sont reportés.

Références

- Acemoglu, D., Dell, M., 2010. Productivity differences between and within countries. *American Economic Journal: Macroeconomics* 2(1), 169-188.
- Aghion, P., Howitt, P., March 1992. A model of growth through creative destruction. *Econometrica* 60(2), 323–351.
- Agrawal, A., Kapur, D., McHale, J., 2008. How do spatial and social proximity influence knowledge flows ? evidence from patent data. *Journal of urban economics* 64(2), 258–269.
- Akerman, A., Gaarder, I., Mogstad, M., 2015. The Skill Complementarity of Broadband Internet. *Q. J. Econ.* 130, 1781–1824.
- Alshamaila, Y., Papagiannidis, S., Li, F., 2013. Cloud computing adoption by SMEs in the north east of England: A multi-perspective framework. *Journal of Enterprise Information Management* 26, 250–275.
- Aronica, M., Bonfanti, R.C., Piacentino, D., 2021. Social media adoption in Italian firms. Opportunities and challenges for lagging regions. *Papers in Regional Science*. DOI: 10.1111/pirs.12606
- Austin, P. C., Jembere, N., & Chiu, M., 2018. Propensity score matching and complex surveys. *Statistical methods in medical research*, 27(4), pp. 1240-1257.
- Bai, Y., 2017. The faster, the better? The impact of internet speed on employment. *Inf. Econ. Policy* 40, 21–25.
- Baker, A. C., Larcker, D. F., & Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates?. *Journal of Financial Economics*, 144(2), 370-395.
- Battisti, G., Stoneman, P., 2003. Inter- and intra-firm effects in the diffusion of new process technology. *Research Policy* 32, 1641–1655.
- Battisti, G., Stoneman, P., 2005. The intra-firm diffusion of new process technologies. *International Journal of Industrial Organization* 23, 1–22.
- Bayo-Moriones, A., Lera-López, F., 2007. A firm-level analysis of determinants of ICT adoption in Spain. *Technovation* 27, 352–366.
- Becheikh, N., Landry, R., Amara, N., 2006. Lessons from innovation empirical studies in the manufacturing sector : A systematic review of the literature from 1993–2003. *Technovation* 26(5-6), 644–664.
- Ben-Aoun Peltier, L., Ben Youssef, A., *Does Internet Speed Matter? Impact of Internet Speed on E-Applications Adoption by Firms in Luxembourg*. Manuscript non publié.
- Bergé L (2018). “Efficient estimation of maximum likelihood models with multiple fixed-effects: the R package FENmlm.” CREA Discussion Papers.
- Bergé, L. R., Carayol, N., Roux, P., 2018. How do inventor networks affect urban invention? *Regional Science and Urban Economics* 71, 137–162.
- Bertschek, I., Briglauer, W., Hüschelrath, K., Kauf, B., Niebel, T., 2015. The economic impacts of broadband internet: A survey. *Review of Network Economics*, 14(4), 201-227.
- Bertschek, I., Cerquera, D., Klein, G.J., 2013. More bits – more bucks? Measuring the impact of broadband internet on firm performance. *Inf. Econ. Policy* 25, 190–203.
- Bilinski, A., Hatfield, L.A., 2019. *Nothing to see here? Non-inferiority approaches to parallel trends and other model assumptions*. ArXiv 1805.03273v5, disponible à : <https://arxiv.org/pdf/1805.03273.pdf>

- Blinder, A.S., 1973. Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources* 8, 436–455.
- Bloom, N., Sadun, R., Van Reenen, J., 2012. Americans Do IT Better: US Multinationals and the Productivity Miracle. *Am. Econ. Rev.* 102, 167–201.
- Bocquet, R., Brossard, O., 2007. The variety of ICT adopters in the intra-firm diffusion process: Theoretical arguments and empirical evidence. *Structural Change and Economic Dynamics* 18, 409–437.
- Briglauer, W., Dürr, N.S., Falck, O., Hüschelrath, K., 2019. Does state aid for broadband deployment in rural areas close the digital and economic divide? *Information Economics and Policy* 46, 68–85.
- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1), 31-72.
- Callaway, B., & Sant’Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200-230.
- Canzian, G., Poy, S., Schüller, S., 2019. Broadband upgrade and firm performance in rural areas: Quasi-experimental evidence. *Regional Science and Urban Economics* 77, 87–103.
- Chatzoglou, P.D., Vraimaki, E., Diamantidis, A., Sarigiannidis, L., 2010. Computer acceptance in Greek SMEs. *Journal of Small Business and Enterprise Development*.
- Ciapanna, E., Colonna, F., 2019. Is your Broadband really broad? Internet Speed, Labour Demand and Productivity Outcomes: Evidence from Italian Firms 39.
- Ciccone, A., Hall, R.E., 1996. Productivity and the density of economic activity. *The American Economic Review* 86, 54–70.
- Coad, A., Daunfeldt, S.-O., Hölzl, W., Johansson, D., Nightingale, P., 2014. High-growth firms: introduction to the special section. *Ind. Corp. Change* 23, 91–112.
- Colombo, M.G., Croce, A., Grilli, L., 2013. ICT services and small businesses’ productivity gains: An analysis of the adoption of broadband Internet technology. *Inf. Econ. Policy* 25, 171–189.
- Cowan, R., Jonard, N., 2004. Network structure and the diffusion of knowledge. *Journal of Economic Dynamics and Control* 28(8), 1557–1575.
- Cumming, D., Johan, S., 2010. The differential impact of the internet on spurring regional entrepreneurship. *Entrepreneurship Theory and Practice*, 34(5), 857-883.
- de Chaisemartin, C., d’Haultfoeuille, X., 2020. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9), 2964-96.
- de Vos, D., Lindgren, U., van Ham, M., Meijers, E., 2020. Does broadband internet allow cities to ‘borrow size’? Evidence from the Swedish labour market. *Regional Studies*, 54(9), 1175-1186.
- DeStefano, T., Backer, K.D., Moussiégt, L., 2017. *Determinants of digital technology use by companies*. OECD Policy Papers No. 40.
- DeStefano, G., Gambardella, A., Verona, G., 2012. Technology push and demand pull perspectives in innovation studies : Current findings and future research directions. *Research policy* 41(8), 1283–1295.
- DeStefano, T., Kneller, R., Timmis, J., 2018. Broadband infrastructure, ICT use and firm performance: Evidence for UK firms. *Journal of Economic Behavior & Organization* 155, 110–139.

- DeStefano, T., Kneller, R., Timmis, J., 2020. *Cloud Computing and Firm Growth*, SSRN Scholarly Paper No. ID 3618829. Social Science Research Network, Rochester, NY.
- Ding, W. W., Levin, S. G., Stephan, P. E., Winkler, A. E., 2010. The impact of information technology on academic scientists' productivity and collaboration patterns. *Management Science* 56(9), 1439–1461.
- Draca, M., Sadun, R., Reenen, J.V., 2009. *Productivity and ICTs: A review of the evidence*, in: The Oxford Handbook of Information and Communication Technologies. Oxford University Press, Oxford, UK, pp. 100–147.
- DuGoff, E. H., Schuler, M., & Stuart, E. A., 2014. Generalizing observational study results: applying propensity score methods to complex surveys. *Health services research*, 49(1), pp. 284-303.
- Duranton, G., Puga, D., 2000. Diversity and specialization in cities: why, where and when does it matter? *Urban Studies* 37(3), 533-555.
- Duvivier, C., 2019. Broadband and Firm Location: Some Answers to Relevant Policy and Research Issues using Meta-analysis. *Canadian Journal of Regional Science*, 42(1), pp. 24-45.
- Duvivier, C. Bussière, C. 2022. The Contingent Nature of Broadband as an Engine for Business Startups in Rural Areas. *Journal of Regional Science*. 1-29. <https://doi.org/10.1111/jors.12605>
- Duvivier, C., Cazou, E., Truchet-Aznar, S., Brunelle, C., Dubé, J., 2021. When, where, and for what industries does broadband foster establishment births? *Papers in Regional Science*, 100(6), pp. 1377-1401.
- Duvivier, C., Truchet, S., Mauhé, N., Mbarek, M., 2018. Déploiement du très haut débit et création d'entreprises dans les zones rurales: Une évaluation du programme Auvergne Très Haut Débit. *Economie et Prévision*, 214, pp. 97-139.
- Elder, T., & Jepsen, C., 2014. Are Catholic primary schools more effective than public primary schools?. *Journal of Urban Economics*, 80, pp. 28-38.
- Ezell, S.J., Atkinson, R.D., Castro, D., Ou, G., 2009. *The Need for Speed: The Importance of Next-Generation Broadband Networks*, SSRN Scholarly Paper No. ID 1354032. Social Science Research Network, Rochester, NY.
- Fabling, R., Grimes, A., 2016. *Picking up Speed: Does Ultrafast Broadband Increase Firm Productivity?* SSRN Scholarly Paper No. ID 2884330. Social Science Research Network, Rochester, NY.
- Fang, C., Lee, J., Schilling, M. A., 2010. Balancing exploration and exploitation through structural design : The isolation of subgroups and organizational learning. *Organization Science* 21(3), 625–642.
- Faquet, R., Malardé, V., 2020. Numérisation des entreprises françaises. Direction générale du Trésor, *Trésor-Eco*, n°271 1–8.
- Federal Communications Commission, 2011. *Measuring broadband America. A report on consumer wireline broadband performance in the U.S.* FCC's Office of Engineering and Technology and Consumer and Governmental Affairs Bureau.
- Ford, G.S., 2018. Is faster better? Quantifying the relationship between broadband speed and economic growth. *Telecommun. Policy*, 42, 766–777.
- Forman, C., Goldfarb, A., Greenstein, S., 2005a. Geographic location and the diffusion of Internet technology. *Electronic Commerce Research and Applications* 4, 1–13.
- Forman, C., Goldfarb, A., Greenstein, S., 2005b. How did location affect adoption of the commercial

- Internet? Global village vs. urban leadership. *Journal of Urban Economics* 58, 389–420.
- Forman, C., Zeebroeck, N. v., 2012. From wires to partners: How the internet has fostered R&D collaborations within firms. *Management Science* 58(8), 1549–1568.
- France Stratégie, 2020. *Déploiement du très haut débit et Plan France très haut débit, Évaluation socioéconomique - Rapport d'étape*. 80 pages.
- Freathy, P., & Calderwood, E. (2016). Coping with change: the implications of e-commerce adoption for island consumers. *Regional Studies*, 50(5), 894-908.
- Galliano, D., Lethiais, V., Soulié, N., 2008. Faible densité des espaces et usages des TIC par les entreprises : besoin d'information ou de coordination ? *Revue d'économie industrielle* 41–64.
- Galliano, D., Roux, P., 2006. Les inégalités spatiales dans l'usage des TIC : Le cas des firmes industrielles françaises. *Revue Economique* 57, 1449-1475.
- Galliano, D., Roux, P., 2008. Organisational motives and spatial effects in Internet adoption and intensity of use : evidence from French industrial firms. *The Annals of Regional Science* 42, 425–448.
- Giotopoulos, I., Kontolaimou, A., Korra, E., Tsakanikas, A., 2017. What drives ICT adoption by SMEs? Evidence from a large-scale survey in Greece. *Journal of Business Research* 81, 60–69.
- Grimes, A., Ren, C., Stevens, P., 2012. The need for speed: impacts of internet connectivity on firm productivity. *J. Product. Anal.* 37, 187–201.
- Haller, S.A., Lyons, S., 2019. Effects of broadband availability on total factor productivity in service sector firms: Evidence from Ireland. *Telecommun. Policy* 43, 11–22.
- Haller, S.A., Siedschlag, I., 2011. Determinants of ICT adoption: evidence from firm-level data. *Applied Economics* 43, 3775–3788.
- Hasbi, M. (2020). Impact of very high-speed broadband on company creation and entrepreneurship: Empirical Evidence. *Telecommunications Policy*, 44(3), 101873.
- Hirano, K., Imbens, G.W., 2004. The propensity score with continuous treatment. In: Gelman, A., Meng, X.L. (Eds.), *Applied Bayesian Modelling and Causal Inference from Missing Data Perspectives*. Wiley, pp. 73-84.
- Ho DE, Imai K, King G, Stuart EA (2011). “MatchIt: Nonparametric Preprocessing for Parametric Causal Inference.” *Journal of Statistical Software*, 42(8), 1–28.
- Hsieh, C-T., Klenow, P. J., 2009. Misallocation and manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4), 1403-1448.
- Ivus, O., Boland, M., 2015. The employment and wage impact of broadband deployment in Canada. *Canadian Journal of Economics* 48(5), 1803-1830.
- Jones, B. F., 2009. The burden of knowledge and the ‘death of the renaissance man’s: is innovation getting harder? *Review of Economic Studies* 76(1), 283–317.
- Kandilov, I. T., Renkow, M., 2010. Infrastructure investment and rural economic development: an evaluation of USDA's broadband loan program. *Growth and Change*, 41(2), 165-191.
- Kim, Y., Orazem, P.F., 2017. Broadband Internet and New Firm Location Decisions in Rural Areas. *American Journal of Agricultural Economics* 99, 1–18.
- Kolko, J., 2012. Broadband and local growth. *Journal of Urban Economics* 71(1), 100-113.
- Komarek, T. M., 2018. Crime and natural resource booms: Evidence from unconventional natural gas production. *The Annals of Regional Science*, 61(1), 113-137.

- Lamie, R.D., Barkley, D.L., & Markley D.M. (2011). Positive Examples and Lessons Learned from Rural Small Business Adoption of E-Commerce Strategies. *Journal of Extension* 49 (6), pp. 8.
- Lee, S., Nam, Y., Lee, S., Son, H., 2016. Determinants of ICT innovations: A cross-country empirical study. *Technological Forecasting and Social Change*, 110, 71–77.
- Lenis, D., Nguyen, T. Q., Dong, N., & Stuart, E. A., 2019. It's all about balance: propensity score matching in the context of complex survey data. *Biostatistics*, 20(1), pp. 147-163.
- Levinsohn, J., Petrin, A., 2003. Estimating production function using inputs to control for unobservables. *The review of Economic Studies*, 70(2), 317-341.
- Liebowitz, S.J., 2002. *Re-thinking the Network Economy: The True Forces That Drive the Digital Marketplace*. AMACOM Books, New York, NY.
- Mack, E. A., 2014. Broadband and knowledge intensive firm clusters: Essential link or auxiliary connection?. *Papers in Regional Science*, 93(1), 3-29.
- Mack, E. A., 2015. Variations in the broadband-business connection across the urban hierarchy. *Growth and Change*, 46(3), 400-423.
- Mack, E. A., Anselin, L., Grubestic, T. H., 2011. The importance of broadband provision to knowledge intensive firm location. *Regional Science Policy & Practice*, 3(1), 17-35.
- Mack, E.A., Rey, S.J., 2014. An econometric approach for evaluating the linkages between broadband and knowledge intensive firms. *Telecommunications Policy* 38(1), 105-118.
- Manlove, J., Whitacre, B., 2019. Understanding the trend to mobile-only internet connections: A decomposition analysis. *Telecommunications Policy*, 43(1), 76-87.
- Maraut, S., Dernis, H., Webb, C., Spiezia, V., Guellec, D., 2008. *The OECD REGPAT database: a presentation*. Documents de travail de l'OCDE sur la science, la technologie et l'industrie, n° 2008/02, Éditions OCDE, Paris, <https://doi.org/10.1787/241437144144>.
- Michaels, G., Natraj, A., Van Reenen, J., 2013. Has ICT Polarized Skill Demand? Evidence from Eleven Countries over Twenty-Five Years. *Rev. Econ. Stat.* 96, 60–77.
- Millimet, D. L., & Parmeter, C. F. (2022). Accounting for skewed or one-sided measurement error in the dependent variable. *Political Analysis*, 30(1), 66-88.
- Nicoletti, G., von Rueden, C., Andrews, D., 2020. Digital technology diffusion: A matter of capabilities, incentives or both? *European Economic Review* 128.
- Oaxaca, R., 1973. Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review* 14, 693–709.
- Oliveira, R., Moura, K., Viana, J., Tigre, R., Sampaio, B., 2015. Commute duration and health: Empirical evidence from Brazil. *Transportation Research Part A: Policy and Practice*, 80, 62-75.
- Oliveira, T., Thomas, M., Espadanal, M., 2014. Assessing the determinants of cloud computing adoption: An analysis of the manufacturing and services sectors. *Information & Management* 51(5), 497–510.
- Rambachan, A., Roth, J., 2020. *An honest approach to parallel trends*. Mimeo, Harvard University, disponible à: https://scholar.harvard.edu/jroth/files/honestparalleltrends_main.pdf
- Renski, H., 2011. External economies of localization, urbanization and industrial diversity and new firm survival. *Papers in Regional Science* 90(3), 473-502.

- Ridgeway, G., Kovalchik, S. A., Griffin, B. A., & Kabeto, M. U. 2015. Propensity score analysis with survey weighted data. *Journal of causal inference*, 3(2), pp. 237-249.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.
- Ruiz-Rodríguez, F., Lucendo-Monedero, A.L., González-Relaño, R., 2018. Measurement and characterisation of the Digital Divide of Spanish regions at enterprise level. A comparative analysis with the European context. *Telecommunications Policy* 42, 187–211.
- Rupasingha, A., Marré, A. W., 2020. Moving to the hinterlands: agglomeration, search costs and urban to rural business migration. *Journal of Economic Geography* 20(1), 123-153.
- Salemink, K., Strijker, D., Bosworth, G., 2017. Rural development in the digital age: A systematic literature review on unequal ICT availability, adoption, and use in rural areas. *Journal of Rural Studies* 54, 360–371.
- Schumpeter, J. A., 1943. *Capitalism, socialism and democracy*. Reprint 2010. Routledge.
- Silva, J. S., & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, 88(4), 641-658.
- Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2), 175-199.
- Thong, J.Y.L., 2001. Resource constraints and information systems implementation in Singaporean small businesses. *Omega* 29, 143–156.
- Thonipara, A., Sternberg, R.G., Proeger, T., Haefner, L., 2020. *Assessing the “digital divide” and its regional determinants: Evidence from a web-scraping analysis*. IFH Working Paper No. 25/2020.
- Townsend, L., Sathiaselan, A., Fairhurst, G., Wallace, C., 2013. Enhanced broadband access as a solution to the social and economic problems of the rural digital divide. *Local Economy* 28, 580–595.
- Whitacre, B., Gallardo, R., Stover, S., 2014. Does rural broadband impact jobs and income? Evidence from spatial and first-differenced regressions. *The Annals of Regional Science* 53, 649–670.
- Xu, X., Watts, A., Reed, M., 2019. Does access to internet promote innovation? A look at the US broadband industry. *Growth and Change* 50(4), 1423–1440.
- Yu, L., Orazem, P. F., Jolly, R. W., 2011. Why do rural firms live longer?. *American Journal of Agricultural Economics*, 93(3), 673-692.
- Zanutto, E. L., 2006. A comparison of propensity score and linear regression analysis of complex survey data. *Journal of data Science*, 4(1), pp. 67-91.
- Zykov, S. V., 2019. Market analysis on cloud computing 2020. *American Journal of Computer Science and Information Technology* 7(3).